Likelihood Ratio Based Tests for Markov Regime Switching*

Zhongjun Qu[†] Fan Zhuo[‡]

Boston University Boston University

May 24, 2020

Abstract

Markov regime switching models are very common in economics and finance. Despite persisting interest in them, the asymptotic distributions of likelihood ratio based tests for detecting regime switching remain unknown. This study examines such tests and establishes their asymptotic distributions in the context of nonlinear models, allowing multiple parameters to be affected by regime switching. The analysis addresses three difficulties: (i) some nuisance parameters are unidentified under the null hypothesis; (ii) the null hypothesis yields a local optimum; and (iii) the conditional regime probabilities follow stochastic processes that can only be represented recursively. Addressing these issues permits substantial power gains in empirically relevant settings. This study also presents the following results: (1) a characterization of the conditional regime probabilities and their derivatives with respect to the model's parameters; (2) a high order approximation to the log likelihood ratio; (3) a refinement of the asymptotic distribution; and (4) a unified algorithm to simulate the critical values. For models that are linear under the null hypothesis, the elements needed for the algorithm can all be computed analytically. Furthermore, the above results explain why some bootstrap procedures can be inconsistent, and why standard information criteria can be sensitive to the hypothesis and the model structure. When applied to US quarterly real GDP growth rate data, the methods detect relatively strong evidence favoring the regime switching specification. Lastly, we apply the methods in the context of dynamic stochastic equilibrium models, and obtain similar results as the GDP case.

Keywords: Hypothesis testing, likelihood ratio, Markov switching, nonlinearity.
JEL codes: C12, C22, E32.

*We are grateful to the editor Aureo de Paula and four anonymous referees for their comments and suggestions that substantially improved the paper. We thank James Hamilton, Hiroaki Kaido, Frank Kleibergen, Pierre Perron, Douglas Steigerwald, and the seminar participants at Amsterdam, Brown, BU Statistics, Cemfi, UCSD, Barcelona GSE 2017 Summer Forum, the 2016 ES Winter Meeting, the NBER-NSF Time Series Conference, 2014 JSM, the 23rd SNDE, and the 11th World Congress of the Econometric Society for their suggestions and discussions, as well as Carrasco, Hu, and Ploberger for making their code available.

[†]Department of Economics, Boston University, 270 Bay State Road, Boston, MA, 02215 (qu@bu.edu).

[‡]Department of Economics, Boston University, 270 Bay State Road, Boston, MA, 02215 (zhuo@bu.edu).

1 Introduction

Markov regime switching models are widely used in economics and finance. Hamilton (1989) is a seminal contribution, which provides not only a framework for describing economic recessions, but also a general algorithm for filtering, smoothing, and estimation. Surveys of this voluminous body of literature are available in Hamilton (2008, 2016) and Ang and Timmermann (2012).

Three approaches have been considered for detecting regime switching. The first approach involves treating the issue as testing for parameter homogeneity against heterogeneity. To detect heterogeneity, Neyman and Scott (1966) studied the $C(\alpha)$ test, while Chesher (1984) derived a score test and showed that it is related to the information matrix test of White (1982). Carrasco, Hu, and Ploberger (2014) further developed this approach by allowing the parameters to follow flexible weakly dependent processes. They analyzed a class of tests, showing that they are asymptotically locally optimal against a specific alternative characterized in their study. The above tests have two features. First, they require estimation under the null hypothesis only. Second, they are designed to detect general parameter heterogeneity, not Markov regime switching specifically. Their power can be substantially lower than that achievable if the parameters follow a finite state Markov chain.

The second approach, proposed by Hamilton (1996), involves conducting generic tests for the hypothesis that a K-regime model (e.g., $K = 1$) adequately describes the data. If the hypothesis holds, then the score function should have mean zero, and follow a martingale difference sequence. Otherwise, the model should be enriched to allow for additional features, possibly by introducing an additional regime. Hamilton (1996) demonstrated the implementation of such tests as a by-product of calculating the smoothed regime probability, making them widely applicable. However, it remains desirable to have testing procedures designed to detect Markov switching alternatives.

The third approach proceeds under the (quasi) likelihood ratio principle, where the (quasi) likelihood functions are constructed assuming a single regime under the null hypothesis, and two regimes under the alternative hypothesis. The analysis faces three challenges. (i) The transition probabilities (p and q) are unidentified under the null hypothesis. This is known as the Davies' (1977) problem. In other words, some parameters are identified only under the alternative hypothesis, and consequently, they are not consistently estimable under the null hypothesis. (ii) The null hypothesis yields a local optimum (cf., Hamilton, 1990); that is, the score function is equal to zero when evaluated at the restricted MLE. Because the resulting unrestricted MLE converges slower than $T^{-1/2}$, the second order Taylor approximation is insufficient for studying the asymptotic properties of the likelihood ratio. (iii) The conditional regime probability (the probability of occupying a particular regime at time t, given the observed information up to $t-1$) follows a stochastic process that can only be represented recursively. The first two difficulties are also present when testing for mixtures. The simultaneous occurrence of all three difficulties challenges the study of the likelihood ratio in the current context. For example, when studying a high order expansion of the likelihood ratio, it is necessary to consider high order derivatives of the conditional regime probability with respect to the model's parameters. Thus far, their statistical properties have remained elusive. Consequently, the asymptotic distribution of the log likelihood ratio remains unknown.

Important progress is documented in Hansen (1992), Garcia (1998), Cho and White (2007), and Carter and Steigerwald (2012). Hansen (1992) explained why difficulties (i) and (ii) cause the conventional approximation to the likelihood ratio to break down. Furthermore, he treated the likelihood function as a stochastic process, and obtained a bound for its asymptotic distribution. His result provides a platform for conducting conservative inference. Garcia (1998) suggested an approximation to the log likelihood ratio that would follow if the score had a positive variance at the restricted MLE. Results in this current study will show that this distribution is, in general, different from the actual limiting distribution. Recently, Cho and White (2007) made significant progress. They suggested a *quasi* likelihood ratio (QLR) test against a two-component mixture alternative, a model in which the regime arrives independently of its past values. The difficulty (iii) is avoided because the conditional regime probability is reduced to a constant, which can be treated as an unknown parameter. Later, Carter and Steigerwald (2012) discussed a consistency issue related to the QLR test. Our study uses several important techniques in Cho and White (2007). Simultaneously, it goes beyond their framework to directly confront Markov switching alternatives. As we will show, the power gains from doing so can be quite substantial.

Specifically, this study considers a family of likelihood ratio based tests and establishes their asymptotic distributions in the context of nonlinear models, allowing multiple parameters to be affected by regime switching. The analysis under the null hypothesis, presented in Sections 4 and 5, takes five steps. Step 1 provides a characterization of the conditional regime probability and its derivatives with respect to the model's parameters. When evaluated under the null hypothesis, this probability reduces to a constant, and its derivatives can all be represented as linear first order difference equations with lagged coefficients $p+q-1$. Because $0<p,q<1$, the equations are stable and amenable to the application of uniform laws of large numbers and functional central limit theorems. This novel characterization is a critical step that makes the subsequent analysis feasible. Step 2 examines a fourth order Taylor approximation of the likelihood ratio for fixed (p,q). This step builds on Cho and White (2007), but goes beyond to account for the effect on the likelihood ratio of the time variation in the conditional regime probability. Step 3 derives an approximation of the likelihood ratio, as an empirical process over $\{(p,q)\colon \epsilon \leq p,q \leq 1-\epsilon \text{ and } p+q \geq 1+\epsilon\}$, where ϵ is a small positive constant. The empirical process perspective follows several existing studies, including Hansen (1992), Garcia (1998), Cho and White (2007), and Carrasco, Hu, and Ploberger (2014). Step 4 provides a finite sample refinement, motivated by the observation that while the limiting distribution in Step 3 is adequate for a broad class of models, it can lead to over-rejections when a singularity (specified later) is present. This problem is addressed by examining a sixth order expansion of the likelihood ratio along the line $p+q=1$, and an eighth order expansion at $p=q=0.5$. The leading terms are then incorporated into the asymptotic distribution to safeguard against their effects. The resulting refined distribution delivers reliable approximations in all of our experiments. Finally, Step 5 presents an algorithm that simulates the refined approximation. For linear models, the elements of this algorithm can all be computed analytically.

The null asymptotic distribution has several unusual features. First, the nuisance parameters, though constrained not to switch, can affect the limiting distribution. Second, the properties of the regressors (e.g., whether they are strictly or weakly exogenous) also affect the distribution. Third, this distribution depends on which parameter is allowed to switch. These features imply that some bootstrap procedures may be inconsistent and that the standard information criteria, such as the BIC, can be sensitive to the hypothesis and the model structure; see Section 7 for details.

Next, we study the likelihood ratio under the alternative hypothesis. The results explain the potential local power difference between the likelihood ratio test and the tests of Cho and White (2007) and Carrasco, Hu, and Ploberger (2014) in an empirically important setting.

We conduct simulations using a data generating process (DGP) considered in Cho and White (2007). The results show that the power difference can indeed be quite large when the regimes are persistent, a situation that is common in practice. We also apply the testing procedure to US quarterly real GDP growth rate data for the period 1960:I-2014:IV and to a range of subsamples. The results consistently favor the regime switching specification. In addition, we apply the methods in the context of dynamic stochastic equilibrium models, and obtain results similar to those of the GDP case. To the best of our knowledge, this is the first time such consistent evidence for regime switching in the US business cycle dynamics has been documented using hypothesis testing.

Empirical studies have estimated regime switching models on a wide range of time series, including exchange rates, output growth, interest rates, debt-output ratios, bond prices, equity returns, and consumption and dividend processes (Hamilton, 2008). Regime switching has also been incorporated into DSGE models; see Schorfheide (2005), Liu, Waggoner, and Zha (2011), Bianchi

(2013), and Lindé, Smets, and Wouters (2016). However, because of the lack of methods with good power properties, the regime switching hypothesis is rarely formally tested from a frequentist perspective. The methods proposed here can potentially narrow this gap in the literature.

This study contributes to the literature on hypothesis testing when some regularity conditions fail to hold. Some related studies are as follows. Davies (1987), King and Shively (1993), Andrews and Ploberger (1994, 1995), and Hansen (1996) considered tests when a nuisance parameter is unidentified under the null hypothesis. Hartigan (1985), Lindsay (1995), Liu and Shao (2003), and Kasahara and Shimotsu (2012, 2015) tackled the issues of zero score and/or unidentified nuisance parameters in the context of mixture models. Rotnitzky, Cox, Bottai, and Robins (2000) developed a theory for deriving the asymptotic distribution of the likelihood ratio statistic when the rank of the information matrix is one less than full. Our study is the first in the hypothesis testing literature to simultaneously address the difficulties (i) to (iii) described earlier. We conjecture that the techniques and results will have implications for hypothesis testing in other contexts that involve hidden Markov structures.

This paper is structured as follows. Section 2 presents the model and hypotheses. Section 3 introduces the test statistics. Section 4 examines the log likelihood ratio for fixed p and q, while Section 5 presents the limiting distribution, a finite sample refinement, and an algorithm for simulating the critical values. Section 6 examines a boundary issue and the local power properties. Section 7 discusses some implications of the theory for bootstrapping and information criteria. Section 8 examines the finite sample properties. Section 9 considers an application to US real GDP growth rate data and several other applications in the context of DSGE models. Section 10 concludes the paper, and the online appendix contains all proofs. Readers interested in the empirical applications can first read Sections 2, 3, and 9 and then return to Sections 4 and 5.

The following notation is used throughout. $||x||$ is the Euclidean norm of a vector x. $||X||$ is the vector induced norm of a matrix X. $x^{\otimes k}$ and $X^{\otimes k}$ denote the k-fold Kronecker product of x and X, respectively. $\text{vec}(A)$ is the vectorization of an array A. For example, for a three dimensional array A, with n elements along each dimension, $\text{vec}(A)$ returns an n^3-vector with the $(i + (j-1)n + (k-1)n^2)$-th element equal to $A(i,j,k)$. $\mathbf{1}_{\{\cdot\}}$ is the indicator function. For a real valued function $f(\theta)$ of $\theta \in R^p$, $\nabla_\theta f(\theta_0)$ denotes a p-by-1 vector of partial derivatives evaluated at θ_0, $\nabla_{\theta'} f(\theta_0)$ is its transpose, and $\nabla_{\theta_{j_1}} \nabla_{\theta_{j_2}} \cdots \nabla_{\theta_{j_k}} f(\theta_0)$ is the k-th order partial derivative of $f(\theta)$ with respect to the $j_1, j_2, ..., j_k$-th element of θ at θ_0. The symbols "$\Rightarrow$", "$\rightarrow^d$", and "$\rightarrow^p$" denote weak convergence under the Skorohod topology, convergence in distribution, and convergence in probability, and $O_p(\cdot)$ and $o_p(\cdot)$ is the usual notation for the orders of stochastic magnitude.

2 Model, likelihood functions, and hypotheses

The model is as follows. Let $\{(y_t, x_t')\}$ be a sequence of random vectors, where y_t is a scalar, and x_t is a finite dimensional vector. Let s_t be a latent binary variable, the value of which determines the regime at time t. Define the information set at time $t-1$ as $\Omega_{t-1} = \sigma\text{-field}\{..., x_{t-1}', y_{t-2}, x_t', y_{t-1}\}$. Let $f(\cdot|\Omega_{t-1}; \beta, \delta)$ denote the conditional density of y_t given Ω_{t-1}, and assume that it satisfies

$$y_t|(\Omega_{t-1}, s_t) \backsim \begin{cases} f(\cdot|\Omega_{t-1}; \beta, \delta_1), \text{ if } s_t = 1, \\ f(\cdot|\Omega_{t-1}; \beta, \delta_2), \text{ if } s_t = 2, \end{cases} \quad (t = 1, ..., T), \tag{1}$$

where δ_1, δ_2, and β are unknown parameters. Henceforth, we abbreviate the densities on the right hand side of (1) as $f_t(\beta, \delta_1)$ and $f_t(\beta, \delta_2)$, respectively. The regimes are assumed to be Markovian, with $P(s_t = 1|\Omega_{t-1}, s_{t-1} = 1, s_{t-2}, ...) = P(s_t = 1|s_{t-1} = 1) = p$ and $P(s_t = 2|\Omega_{t-1}, s_{t-1} = 2, s_{t-2}, ...) = P(s_t = 2|s_{t-1} = 2) = q$. The stationary probability for $s_t = 1$ is thus given by

$$\xi_* \equiv \xi_*(p, q) = \frac{1-q}{2-p-q}. \tag{2}$$

Under (1), the log likelihood function, evaluated at $0 < p, q < 1$, is

$$\begin{aligned} & \mathcal{L}^A(p, q, \beta, \delta_1, \delta_2) \qquad (3) \\ = & \textstyle\sum_{t=1}^T \log \left\{ f_t(\beta, \delta_1)\xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2) + f_t(\beta, \delta_2)(1 - \xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2)) \right\}, \end{aligned}$$

where $\xi_{t|v}(\cdot)$ denotes the conditional probability of $s_t = 1$ given Ω_v; that is,

$$\xi_{t|v}(p, q, \beta, \delta_1, \delta_2) = P(s_t = 1|\Omega_v; p, q, \beta, \delta_1, \delta_2) \quad (t = 1, ..., T), \tag{4}$$

which satisfies

$$\xi_{t|t}(p, q, \beta, \delta_1, \delta_2) = \frac{f_t(\beta, \delta_1)\,\xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2)}{f_t(\beta, \delta_1)\,\xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2) + f_t(\beta, \delta_2)\,(1 - \xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2))}, \tag{5}$$

$$\xi_{t+1|t}(p, q, \beta, \delta_1, \delta_2) = p\xi_{t|t}(p, q, \beta, \delta_1, \delta_2) + (1-q)(1 - \xi_{t|t}(p, q, \beta, \delta_1, \delta_2)). \tag{6}$$

Without loss of generality, we set $\xi_{1|0} = \xi_*$ throughout the paper. When $\delta_1 = \delta_2 = \delta$, the log likelihood (3) reduces to

$$\mathcal{L}^N(\beta, \delta) = \textstyle\sum_{t=1}^T \log f_t(\beta, \delta). \tag{7}$$

Note that (2) and (3)-(6) are implied by the derivations in Sections 2 and 4 of Hamilton (1989).

In this paper, we study tests based on (7) and (3). The null and alternative hypotheses are

$$H_0 \;:\; \delta_1 = \delta_2 = \delta_* \text{ for some unknown } \delta_*;$$

$$H_1 \;:\; (\delta_1, \delta_2) = (\delta_1^*, \delta_2^*) \text{ for some unknown } \delta_1^* \neq \delta_2^* \text{ and } (p, q) \in (0, 1) \times (0, 1).$$

To proceed, we impose the following assumptions on the DGP and the parameter space. Let n_β and n_δ denote the dimensions of β and δ, respectively.

Assumption 1 *(i) The random vector* (x_t', y_t) *is strictly stationary, ergodic, and absolutely regular, with mixing coefficients* b_τ *satisfying* $b_\tau \leq c\rho^\tau$ *for some* $c > 0$ *and* $\rho \in [0, 1)$*. (ii) Under the null hypothesis,* y_t *is generated by* $f(\cdot|\Omega_{t-1}; \beta_*, \delta_*)$*, where* β_* *and* δ_* *are interior points of* $\Theta \subset \mathbb{R}^{n_\beta}$ *and* $\Delta \subset \mathbb{R}^{n_\delta}$*, respectively, and* Θ *and* Δ *are compact.*

Assumption 1(i) is the same as Assumption A.1(i) in Cho and White (2007). It allows regime switching in x_t under the null hypothesis. Assumption 1(ii) specifies the true parameter values, where the interior point condition ensures that the expansions considered later are well-defined.

Assumption 2 *Under the null hypothesis: (i)* (β_*, δ_*) *uniquely solves* $\max_{(\beta,\delta)\in\Theta\times\Delta} E[\mathcal{L}^N(\beta, \delta)]$*; and (ii) for any* $0 < p, q < 1$*,* $(\beta_*, \delta_*, \delta_*)$ *uniquely solves* $\max_{(\beta,\delta_1,\delta_2)\in\Theta\times\Delta\times\Delta} E[\mathcal{L}^A(p, q, \beta, \delta_1, \delta_2)]$*.*

Part (i) of the assumption implies that (β, δ) is globally identified at (β_*, δ_*) under the null hypothesis. Part (ii) rules out observational equivalence, that is, no two-regime specification with some $\delta_1 \neq \delta_2$ is observationally equivalent to the single-regime specification with $\delta_1 = \delta_2 = \delta_*$.

Assumption 3 *Under the null hypothesis: (i)* $T^{-1}[\mathcal{L}^N(\beta, \delta) - E\mathcal{L}^N(\beta, \delta)] = o_p(1)$ *holds uniformly over* $(\beta, \delta) \in \Theta \times \Delta$*, and* $T^{-1}\sum_{t=1}^T \nabla_{(\beta',\delta')'} \log f_t(\beta, \delta)\nabla_{(\beta',\delta')} \log f_t(\beta, \delta)$ *is positive definite in an open neighborhood of* (β_*, δ_*) *with probability tending to one; (ii) for any* $0 < p, q < 1$*,* $T^{-1}[\mathcal{L}^A(p, q, \beta, \delta_1, \delta_2) - E\mathcal{L}^A(p, q, \beta, \delta_1, \delta_2)] = o_p(1)$ *holds uniformly over* $(\beta, \delta_1, \delta_2) \in \Theta \times \Delta \times \Delta$*.*

Assumption 3 requires (7) and (3) to satisfy the uniform law of large numbers, while allowing (3) to have multiple local maxima. Under the null hypothesis and Assumptions 2 and 3, the maximizers of (7) and (3) for $0 < p, q < 1$ converge in probability to (β_*, δ_*) and $(\beta_*, \delta_*, \delta_*)$, respectively.

Below, we introduce a model that we will use throughout the paper to illustrate the theory.

An illustrative model. A prominent application of regime switching in macroeconomics is to linear models with normal errors:

$$y_t = z_t'\alpha + w_t'\gamma_1\mathbf{1}_{\{s_t=1\}} + w_t'\gamma_2\mathbf{1}_{\{s_t=2\}} + u_t \text{ with } u_t \sim i.i.d.N(0, \sigma^2). \tag{8}$$

This model encompasses finite order AR and ADL models as special cases. In relation to (1), we have $\Omega_{t-1} = \sigma\text{-field}\{..., z_{t-1}', w_{t-1}', y_{t-2}, z_t', w_t', y_{t-1}\}$ and $x_t' = (z_t', w_t')$. We now use this model to illustrate Assumptions 1-3. For Assumption 1, the absolute regularity of (x_t', y_t) is satisfied if x_t follows a stationary VARMA(P,Q) process $\sum_{j=0}^P B_j x_{t-j} = \sum_{j=0}^Q A_j\varepsilon_{t-j}$, where ε_t is an i.i.d. random vector with mean zero and density that is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^{\dim(\varepsilon_t)}$; see Mokkadem (1988). Other processes that are absolutely regular with a geometric rate of decay, as reviewed in Chen (2013), include those generated by threshold autoregressive models, functional coefficient autoregressive models, and GARCH and stochastic volatility models. For Assumption 2, its part (i) is satisfied if Ex_tx_t' has full rank, and its part (ii) is satisfied if the single-regime specification and the two-regime specification with $\delta_1 \neq \delta_2$ are not observationally equivalent. Finally, Assumption 3 requires that $T^{-1}\sum_{t=1}^{T} x_tx_t'$ is positive definite in large samples and that the uniform laws of large numbers hold. Because $\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2)$ is bounded between zero and one, they hold under Assumption 1 and mild conditions on the moments of y_t and x_t. ■

3 The test statistic

We first introduce a family of test statistics based on the likelihood ratio, and then examine empirically relevant values of the transition probabilities p and q. The second issue is important, not only for making the tests empirically relevant, but also for the technical analysis presented later.

Let $\widetilde{\beta}$ and $\widetilde{\delta}$ denote the restricted MLE, i.e., $(\widetilde{\beta},\widetilde{\delta}) = \arg\max_{\beta,\delta}\mathcal{L}^N(\beta,\delta)$. The log likelihood ratio, evaluated at some $0 < p, q < 1$, is equal to

$$LR(p,q) = 2\left[\max_{\beta,\delta_1,\delta_2}\mathcal{L}^A(p,q,\beta,\delta_1,\delta_2) - \mathcal{L}^N(\widetilde{\beta},\widetilde{\delta})\right]. \tag{9}$$

This leads to the following test statistic:

$$\text{SupLR}(\Lambda_\epsilon) = \operatorname*{Sup}_{(p,q)\in\Lambda_\epsilon} LR\,(p,q)\,,$$

where Λ_ϵ is a compact set, specified below, and the supremum is taken to obtain the strongest evidence against the null hypothesis. The supremum can be replaced by other operators. For example, following Andrews and Ploberger (1994), one can consider $\text{ExpLR}(\Lambda_\epsilon) = \int_{\Lambda_\epsilon} LR\,(p,q)\,dJ(p,q)$, with $J(p,q)$ being a weight function. Such considerations lead to a family of tests based on $LR(p,q)$ and Λ_ϵ. We focus on $\text{SupLR}(\Lambda_\epsilon)$; the results extend immediately to other tests including $\text{ExpLR}(\Lambda_\epsilon)$.

Hamilton (2008) reviews twelve articles that apply regime switching models in a wide range of contexts. Of these, ten articles consider two-regime specifications with constant transition probabilities, related to exchange rates (Jeanne and Masson, 2000), output growth (Hamilton, 1989; Chauvet and Hamilton, 2006), interest rates (Hamilton, 1988, 2005; Ang and Bekaert, 2002b), debt-output ratio (Davig, 2004), bond prices (Dai, Singleton, and Yang, 2007), equity returns (Ang and Bekaert, 2002a), and consumption and dividend processes (Garcia, Luger, and Renault, 2003). Eighteen sets of estimates are reported, where the values of the transition probabilities range between 0.855 and 0.998 for the more persistent regime, and between 0.740 and 0.997 for the other regime. These estimates have two features: (i) $p+q$ is substantially above one; and (ii) p (or q) is close to one. Motivated by these two features, we suggest specifying Λ_ϵ as

$$\Lambda_\epsilon = \{(p,q) : p+q \geq 1+\epsilon \text{ and } \epsilon \leq p,q \leq 1-\epsilon \text{ with } \epsilon > 0\}. \tag{10}$$

In our Monte Carlo experiments, we experiment with small values of ϵ, such that the features (i) and (ii) are consistent with the analysis. The results suggest that $\epsilon = 0.02$ is a reasonable choice.

The set in (10) can be generalized to allow for different trimming proportions, leading to $\{(p,q) : p+q \geq 1+\epsilon_1 \text{ and } \epsilon_2 \leq p,q \leq 1-\epsilon_3 \text{ with } \epsilon_1,\epsilon_2,\epsilon_3 > 0\}$. It can also be modified to incorporate additional information. For example, if p and q are known to be higher than 0.5, we can use

$$\{(p,q) : 0.5+\epsilon \leq p,q \leq 1-\epsilon \text{ with } \epsilon > 0\}. \tag{11}$$

In this paper, we focus on (10). The results hold for the latter two specifications of Λ_ϵ, provided that the Λ_ϵ in the limiting distribution is modified accordingly. As a limitation, (10) and (11) exclude some of the estimates reported above when they are very close to one. To fully resolve this issue, we would need to allow p or q to approach one as $T \to \infty$, and study the asymptotic distribution of the likelihood ratio. We leave this topic for future research.

4 Log likelihood ratio under given p and q

The time-varying conditional regime probability $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ represents the key difference between the Markov switching model and the mixture model. In this section, we first study this probability and its partial derivatives with respect to β,δ_1, and δ_2 (Subsection 4.1), and then apply the result to develop an asymptotic expansion of the concentrated log likelihood under the null hypothesis (Subsection 4.2). All results presented in this section are uniform with respect to $(p,q) \in [\epsilon, 1-\epsilon] \times [\epsilon, 1-\epsilon]$, where ϵ can be any constant satisfying $0 < \epsilon < 1/2$.

4.1 The conditional regime probability

The following two observations are key to our analysis. First, (5) and (6) can be combined to produce a first order difference equation that relates $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ to $\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2)$:

$$\begin{aligned} & \xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2) \\ = \; & p + (p+q-1)\frac{f_t(\beta,\delta_2)(\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2)-1)}{f_t(\beta,\delta_1)\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2) + f_t(\beta,\delta_2)(1-\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2))}. \end{aligned} \tag{12}$$

From this representation, it is clear that the partial derivatives of $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ with respect to β,δ_1, and δ_2 all follow first order difference equations. Second, although these difference equations are nonlinear when δ_1 and δ_2 are unrestricted, they simplify greatly if $\delta_1 = \delta_2$. Because the asymptotic expansions here are around the restricted MLE, focusing on $\delta_1 = \delta_2$ is sufficient.

The next lemma characterizes $\xi_{t+1|t}(p, q, \beta, \delta_1, \delta_2)$ and its derivatives evaluated at $(\beta, \delta_1, \delta_2) = (\beta, \delta, \delta)$, with β and δ denoting generic values in the parameter space. Define

$$\theta = (\beta', \delta_1', \delta_2')', \tag{13}$$

$I_0 = \{1, ..., n_\beta\}$, $I_1 = \{n_\beta + 1, ..., n_\beta + n_\delta\}$, and $I_2 = \{n_\beta + n_\delta + 1, ..., n_\beta + 2n_\delta\}$, where I_0, I_1, and I_2 are index sets that refer to the elements of β, δ_1, and δ_2, respectively, which are needed for Lemma 1. Define $\bar{\xi}_{t+1|t} = \xi_{t+1|t}(p, q, \beta, \delta, \delta)$ and $\bar{f}_t = f_t(\beta, \delta)$. Let $\nabla_{\theta_{j_1}} ... \nabla_{\theta_{j_k}} \bar{\xi}_{t|t-1}$, $\nabla_{\theta_{j_1}} ... \nabla_{\theta_{j_k}} \bar{f}_{1t}$, and $\nabla_{\theta_{j_1}} ... \nabla_{\theta_{j_k}} \bar{f}_{2t}$ denote the k-th order partial derivatives of $\xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2)$, $f_t(\beta, \delta_1)$, and $f_t(\beta, \delta_2)$ with respect to the j_1-th,..., j_k-th elements of θ, evaluated at generic β and $\delta_1 = \delta_2 = \delta$.

Lemma 1 *Let $\rho = p + q - 1$ and $r = \rho\xi_*(1 - \xi_*)$, with ξ_* defined in (2). Then, for any $t \geq 1$:*

1. $\bar{\xi}_{t+1|t} = \xi_*$.
2. $\nabla_{\theta_j} \bar{\xi}_{t+1|t} = \rho \nabla_{\theta_j} \bar{\xi}_{t|t-1} + \bar{\mathcal{E}}_{j,t}$, *where $\bar{\mathcal{E}}_{j,t}$ is equal to zero if $j \in I_0$, $r \nabla_{\theta_j} \log \bar{f}_{1t}$ if $j \in I_1$, and $-r \nabla_{\theta_j} \log \bar{f}_{2t}$ if $j \in I_2$.*
3. $\nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t+1|t} = \rho \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1} + \bar{\mathcal{E}}_{jk,t}$, *where $\bar{\mathcal{E}}_{jk,t}$ is equal to (Let (I_a, I_b) denote the case $j \in I_a$ and $k \in I_b$; $a, b = 0, 1, 2$.)*

$$\begin{aligned}
&(I_0, I_0) : 0 \\
&(I_0, I_1) : -\frac{r\nabla_{\theta_j} \bar{f}_{2t}}{\bar{f}_t} \frac{\nabla_{\theta_k} \bar{f}_{1t}}{\bar{f}_t} + \frac{r\nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{1t}}{\bar{f}_t} \\
&(I_0, I_2) : \frac{r\nabla_{\theta_j} \bar{f}_{2t}}{\bar{f}_t} \frac{\nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t} - \frac{r\nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t} \\
&(I_1, I_1) : \frac{\rho(1-2\xi_*)\nabla_{\theta_j} \bar{\xi}_{t|t-1} \nabla_{\theta_k} \bar{f}_{1t}}{\bar{f}_t} + \frac{\rho(1-2\xi_*)\nabla_{\theta_k} \bar{\xi}_{t|t-1} \nabla_{\theta_j} \bar{f}_{1t}}{\bar{f}_t} + \frac{r\nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{1t}}{\bar{f}_t} - \frac{2r\xi_* \nabla_{\theta_j} \bar{f}_{1t}}{\bar{f}_t} \frac{\nabla_{\theta_k} \bar{f}_{1t}}{\bar{f}_t} \\
&(I_1, I_2) : \frac{\rho(2\xi_*-1)\nabla_{\theta_j} \bar{\xi}_{t|t-1} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t} - \frac{\rho(2\xi_*-1)\nabla_{\theta_k} \bar{\xi}_{t|t-1} \nabla_{\theta_j} \bar{f}_{1t}}{\bar{f}_t} + \frac{r(2\xi_*-1)\nabla_{\theta_j} \bar{f}_{1t}}{\bar{f}_t} \frac{\nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t} \\
&(I_2, I_2) : \frac{\rho(2\xi_*-1)\nabla_{\theta_j} \bar{\xi}_{t|t-1} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t} + \frac{\rho(2\xi_*-1)\nabla_{\theta_k} \xi_{t|t-1} \nabla_{\theta_j} \bar{f}_{2t}}{\bar{f}_t} - \frac{r\nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t} - \frac{2r(\xi_*-1)\nabla_{\theta_j} \bar{f}_{2t}}{\bar{f}_t} \frac{\nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t}
\end{aligned}.$$

4. $\nabla_{\theta_j} \nabla_{\theta_k} \nabla_{\theta_l} \bar{\xi}_{t+1|t} = \rho \nabla_{\theta_j} \nabla_{\theta_k} \nabla_{\theta_l} \bar{\xi}_{t|t-1} + \bar{\mathcal{E}}_{jkl,t}$, *where the expressions of $\bar{\mathcal{E}}_{jkl,t}$, with $j, k, l \in \{I_a, I_b, I_c\}$ and $a, b, c = 0, 1, 2$, are given in the appendix.*

Remark 1 *Lemma 1 holds in finite samples. When $\delta_1 = \delta_2$, the conditional regime probability reduces to a constant that is simply its stationary probability, whereas its derivatives up to the third order all follow linear first order difference equations. Because the lagged coefficients equal $p + q - 1$ with $0 < p, q < 1$, these difference equations are all stable, and therefore are amenable to the application of uniform laws of large numbers and functional central limit theorems. Lemma 1 is the key result that makes our subsequent analysis feasible.*

4.2 Concentrated log likelihood and its expansion

To obtain the limiting distribution of the log likelihood ratio (9), a standard approach is to expand $\mathcal{L}^A(p,q,\beta,\delta_1,\delta_2)$ around the restricted MLE, $(\widetilde{\beta},\widetilde{\delta},\widetilde{\delta})$, and study the supremum of this expansion over the parameter space. Unfortunately, this approach is infeasible for the current problem because $\mathcal{L}^A(p,q,\beta,\delta_1,\delta_2)$ can have multiple local optima. In particular, when computing the supremum, the first order condition with respect to $(\beta,\delta_1,\delta_2)$ can produce multiple zeros, which makes subsequent analysis difficult. Cho and White (2007) encountered a similar problem, and proceeded by working with the concentrated likelihood. We follow their insightful strategy. This allows us to break the analysis into two steps. In the first step, we quantify the relationship between (β,δ_1) and δ_2 using the first order conditions that define the concentrated likelihood (Lemma A.3). This removes β and δ_1 from the subsequent analysis, reducing the dimension of the problem by half or more. This step uses the property that, once δ_2 and (p,q) are fixed, the likelihood has a unique maximum. In the second step, we expand the concentrated likelihood around $\delta_2=\widetilde{\delta}$ (Lemma 2), and obtain an approximation to $LR(p,q)$. Because the conditional regime probability is time varying, the analysis here is substantially more challenging than that in Cho and White (2007).

Let $\hat{\beta}(\delta_2)$ and $\hat{\delta}_1(\delta_2)$ denote the maximizer of the log likelihood for a generic $\delta_2\in\Delta$ (the dependence of $\hat{\beta}$ and $\hat{\delta}_1$ on p and q is suppressed to shorten the expressions):

$$(\hat{\beta}(\delta_2),\hat{\delta}_1(\delta_2))=\arg\max_{\beta,\delta_1}\mathcal{L}^A(p,q,\beta,\delta_1,\delta_2). \tag{14}$$

Let $\mathcal{L}(p,q,\delta_2)$ be the concentrated log likelihood: $\mathcal{L}(p,q,\delta_2)=\mathcal{L}^A(p,q,\hat{\beta}(\delta_2),\hat{\delta}_1(\delta_2),\delta_2)$. Because (9) satisfies $\max_{\beta,\delta_1,\delta_2}\mathcal{L}^A(p,q,\beta,\delta_1,\delta_2)=\max_{\delta_2}\mathcal{L}(p,q,\delta_2)$ and $\mathcal{L}^N(\widetilde{\beta},\widetilde{\delta})=\mathcal{L}(p,q,\widetilde{\delta})$, we have

$$LR(p,q)=2\max_{\delta_2}\left[\mathcal{L}(p,q,\delta_2)-\mathcal{L}(p,q,\widetilde{\delta})\right]. \tag{15}$$

For any $k\geq 1$, let $\mathcal{L}^{(k)}_{i_1\ldots i_k}(p,q,\delta_2)$ $(i_1,\ldots,i_k\in\{1,\ldots,n_\delta\})$ denote the k-th order derivative of $\mathcal{L}(p,q,\delta_2)$ with respect to the i_1-th, ..., i_k-th elements of δ_2. Let d_j $(j\in\{1,\ldots,n_\delta\})$ denote the j-th element of $(\delta_2-\widetilde{\delta})$. A fourth order Taylor expansion of $\mathcal{L}(p,q,\delta_2)$ around $\widetilde{\delta}$ is

$$\begin{aligned}\mathcal{L}(p,q,\delta_2)-\mathcal{L}(p,q,\widetilde{\delta}) &= \sum_{j=1}^{n_\delta}\mathcal{L}^{(1)}_j(p,q,\widetilde{\delta})d_j+\frac{1}{2!}\sum_{j=1}^{n_\delta}\sum_{k=1}^{n_\delta}\mathcal{L}^{(2)}_{jk}(p,q,\widetilde{\delta})d_jd_k \\ &\quad+\frac{1}{3!}\sum_{j=1}^{n_\delta}\sum_{k=1}^{n_\delta}\sum_{l=1}^{n_\delta}\mathcal{L}^{(3)}_{jkl}(p,q,\widetilde{\delta})d_jd_kd_l \\ &\quad+\frac{1}{4!}\sum_{j=1}^{n_\delta}\sum_{k=1}^{n_\delta}\sum_{l=1}^{n_\delta}\sum_{m=1}^{n_\delta}\mathcal{L}^{(4)}_{jklm}(p,q,\bar{\delta})d_jd_kd_ld_m,\end{aligned} \tag{16}$$

where $\bar{\delta} = \widetilde{\delta} + c(\delta_2 - \widetilde{\delta})$ for some $c \in (0,1)$ by Taylor's theorem in several variables.

In a standard testing problem, a second order Taylor expansion around the restricted MLE is sufficient to establish the asymptotic distribution of the likelihood ratio. This is because the score vector and the Hessian matrix are $O_p(T^{1/2})$ and $O_p(T)$, respectively, and the unrestricted MLE converges at rate $T^{-1/2}$. In the current problem (see Lemmas 2 and A.6), the score vector in (16) is identically zero, and the Hessian matrix and the fourth order derivatives are $O_p(T^{1/2})$ and $O_p(T)$, respectively, while the unrestricted MLE converges at rate $T^{-1/4}$ over Λ_ϵ. The third order derivatives are $O_p(T^{1/2})$; therefore, their effect on the expansion is dominated by that of the second and fourth order derivatives. Consequently, the second and fourth order derivatives play the roles of the first and second order derivatives in the standard problem, and a fourth order Taylor expansion is needed to derive the distribution of the log likelihood ratio under the null hypothesis.

Assumption 4 *There exists an open neighborhood of* (β_*, δ_*)*, denoted by* $B(\beta_*, \delta_*)$*, and a sequence of positive, strictly stationary, and ergodic variables* $\{\upsilon_t\}$ *satisfying* $E\upsilon_t^{1+c} < L < \infty$ *for some* $c > 0$*, such that* $\sup_{(\beta,\delta_1)\in B(\beta_*,\delta_*)} |[\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_k}} f_t(\beta, \delta_1)]/f_t(\beta,\delta_1)|^{\alpha(k)/k} < \upsilon_t$ *for all* $i_1, ..., i_k \in \{1, ..., n_\beta + n_\delta\}$*, where* $1 \le k \le 5$*,* $\alpha(k) = 6$ *if* $k = 1, 2, 3$*, and* $\alpha(k) = 5$ *if* $k = 4, 5$*.*

This assumption is slightly stronger than Assumption A5 (iii) in Cho and White (2007). There, instead of $\alpha(k)/k$, the values are $4, 2, 2$, and 1 for $k = 1, 2, 3$, and 4, respectively.

Assumption 5 *There exists* $\eta > 0$ *such that* $\sup_{p,q\in[\epsilon,1-\epsilon]} \sup_{|\delta-\widetilde{\delta}|<\eta} T^{-1}|\mathcal{L}^{(5)}_{jklmn}(p,q,\delta)| = O_p(1)$ *for all* $j,k,l,m,n \in \{1,...,n_\delta\}$*, where* ϵ *is an arbitrarily small constant satisfying* $0 < \epsilon < 1/2$*.*

In a standard problem, we need $\mathcal{L}^{(2)}_{jk}(p,q,\delta)$ to be continuous in δ, or $\mathcal{L}^{(3)}_{jkl}(p,q,\delta)$ to be $O_p(T)$ around $\widetilde{\delta}$, to ensure that a quadratic expansion is an adequate approximation of the log likelihood ratio. Here, because the fourth order derivatives play the role of the usual second order derivatives, we need to impose Assumption 5 on the fifth order derivatives.

Let $\widetilde{\xi}_{t+1|t}$ and $\widetilde{f}_t$ denote $\xi_{t+1|t}(p,q,\widetilde{\beta},\widetilde{\delta},\widetilde{\delta})$ and $f_t(\widetilde{\beta},\widetilde{\delta})$. Let $\nabla_{\delta_{1i_1}} ... \nabla_{\delta_{1i_k}} \widetilde{\xi}_{t|t-1}$ and $\nabla_{\delta_{1i_1}} ... \nabla_{\delta_{1i_k}} \widetilde{f}_{1t}$ denote the k-th order derivatives of $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ and $f_t(\beta,\delta_1)$ with respect to the i_1-th, ..., i_k-th elements of δ_1, evaluated at $(\widetilde{\beta},\widetilde{\delta},\widetilde{\delta})$. Define

$$\begin{aligned}
\widetilde{U}_{jk,t} &= \frac{1}{\widetilde{f}_t}\left\{\left(\frac{1-\xi_*}{\xi_*}\right)\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t} + \frac{1}{\xi_*^2}\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t} + \frac{1}{\xi_*^2}\nabla_{\delta_{1j}}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}\right\}, \quad (17) \\
\widetilde{D}_{jk,t} &= \frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{U}_{jk,t}, \quad \widetilde{I}_t = \frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}, \\
\widetilde{V}_{jklm} &= T^{-1}\sum_{t=1}^{T}\widetilde{U}_{jk,t}\widetilde{U}_{lm,t}, \quad \widetilde{D}_{lm} = T^{-1}\sum_{t=1}^{T}\widetilde{D}_{lm,t}, \quad \widetilde{I} = T^{-1}\sum_{t=1}^{T}\widetilde{I}_t.
\end{aligned}$$

Lemma 2 *Under the null hypothesis and Assumptions 1-5, for any* $j,k,l,m \in \{1,...,n_\delta\}$*, we have* $\mathcal{L}_j^{(1)}(p,q,\widetilde{\delta})=0$, $T^{-1/2}\mathcal{L}_{jk}^{(2)}(p,q,\widetilde{\delta})=T^{-1/2}\sum_{t=1}^T \widetilde{U}_{jk,t}+o_p(1)$, $T^{-3/4}\mathcal{L}_{jkl}^{(3)}(p,q,\widetilde{\delta})=O_p\left(T^{-1/4}\right)$*, and* $T^{-1}\mathcal{L}_{jklm}^{(4)}(p,q,\widetilde{\delta})=-\{\widetilde{V}_{jklm}-\widetilde{D}'_{jk}\widetilde{I}^{-1}\widetilde{D}_{lm}+\widetilde{V}_{jmkl}-\widetilde{D}'_{jm}\widetilde{I}^{-1}\widetilde{D}_{kl}+\widetilde{V}_{jlkm}-\widetilde{D}'_{jl}\widetilde{I}^{-1}\widetilde{D}_{km}\}+o_p(1)$.

This lemma characterizes how various terms in the expansion in (16) affect the limiting distribution of the likelihood ratio. The score vector is identically zero; thus, it has no effect whatsoever. The Hessian matrix is nonzero. However, because it converges to a random matrix after division by $T^{1/2}$, its effect is very different to that in the standard situation. The third order derivatives do not affect the limiting distribution, but we will need to study them when establishing a finite sample refinement. Finally, the fourth order derivatives are $O_p(T)$. Their leading terms provide consistent estimates for the limiting variances and covariances of $T^{-1/2}\sum_{t=1}^T \widetilde{U}_{jk,t}$, for $j,k \in \{1,...,n_\delta\}$.

Remark 2 *Of the three components of* $\widetilde{U}_{jk,t}$*, the first component* $((1-\xi_*)/\xi_*)\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t}/\widetilde{f}_t$ *is familiar in the mixture literature as a measure of dispersion. For example, it is in the test of Cho and White (2007) against mixtures. The remaining two components are new, and are the result of the Markov switching structure. They can be rewritten as* $((1-\xi_*)/\xi_*)\sum_{s=1}^{t-1}\rho^s\nabla_{\delta_{1j}}\log\widetilde{f}_{1(t-s)}\nabla_{\delta_{1k}}\log\widetilde{f}_{1t}$ *and* $((1-\xi_*)/\xi_*)\sum_{s=1}^{t-1}\rho^s\nabla_{\delta_{1k}}\log\widetilde{f}_{1(t-s)}\nabla_{\delta_{1j}}\log\widetilde{f}_{1t}$*, respectively, and as such, they measure the serial dependence caused by the regime switching. Their relative magnitudes increase with* $|\rho|$*. This suggests that the power difference between testing against Markov switching alternatives and mixture alternatives can be substantial when the regimes are persistent, that is, when* ρ *is close to one.*

The illustrative model (cont'd). Suppose that in model (8), only the regression coefficients are allowed to switch. Then, $\widetilde{U}_{jk,t}$ and $\widetilde{D}_{jk,t}$ in Lemma 2 are equal to

$$\left(\tfrac{1-\xi_*}{\xi_*}\right)\left\{\tfrac{w_{jt}w_{kt}}{\widetilde{\sigma}^2}\left(\tfrac{\widetilde{u}_t^2}{\widetilde{\sigma}^2}-1\right)+\sum_{s=1}^{t-1}\rho^s\left(\tfrac{w_{j(t-s)}\widetilde{u}_{t-s}}{\widetilde{\sigma}^2}\right)\left(\tfrac{w_{kt}\widetilde{u}_t}{\widetilde{\sigma}^2}\right)+\sum_{s=1}^{t-1}\rho^s\left(\tfrac{w_{k(t-s)}\widetilde{u}_{t-s}}{\widetilde{\sigma}^2}\right)\left(\tfrac{w_{jt}\widetilde{u}_t}{\widetilde{\sigma}^2}\right)\right\} \tag{18}$$

and

$$\left[\begin{array}{ccc}\frac{z_t'\widetilde{u}_t}{\widetilde{\sigma}^2} & \frac{1}{2\widetilde{\sigma}^2}\left(\frac{\widetilde{u}_t^2}{\widetilde{\sigma}^2}-1\right) & \frac{w_t'\widetilde{u}_t}{\widetilde{\sigma}^2}\end{array}\right]'\widetilde{U}_{jk,t}, \tag{19}$$

respectively, where $\widetilde{u}_t$ are the OLS residuals and $\widetilde{\sigma}^2=T^{-1}\sum_{t=1}^T\widetilde{u}_t^2$. Thus, $\widetilde{U}_{jk,t}$ and $\widetilde{D}_{jk,t}$ depend on the regressors and the OLS residuals only, and the covariance function of $T^{-1/2}\mathcal{L}_{jk}^{(2)}(p,q,\widetilde{\delta})$ is consistently estimable. This feature is useful for computing the critical values for the test. ■

5 Asymptotic approximations

This section presents three sets of results: (1) the limiting distribution of SupLR(Λ_ϵ); (2) a finite sample refinement to further improve the aforementioned limiting distribution for an important special case; and (3) a unified algorithm to compute the critical values.

5.1 Limiting distribution of $\mathrm{SupLR}(\Lambda_\epsilon)$

Let $T^{-1/2}\mathcal{L}^{(2)}(p,q,\widetilde{\delta})$ denote an n_δ-by-n_δ matrix whose (j,k)-th element is equal to $T^{-1/2}\mathcal{L}^{(2)}_{jk}(p,q,\widetilde{\delta})$. Under Assumptions 1-5, $T^{-1/2}\operatorname{vec}\mathcal{L}^{(2)}(p,q,\widetilde{\delta})$ converges weakly to a vector of Gaussian processes over $\epsilon \le p,q \le 1-\epsilon$ (Lemma A.5). Below, we express the limiting distribution of $\mathrm{SupLR}(\Lambda_\epsilon)$ in terms of this Gaussian process. For any $0 < p_r, q_r, p_s, q_s < 1$ and $j,k,l,m \in \{1,2,...,n_\delta\}$, define

$$\omega_{jklm}\left(p_r,q_r;p_s,q_s\right) = V_{jklm}\left(p_r,q_r;p_s,q_s\right) - D'_{jk}(p_r,q_r)I^{-1}D_{lm}(p_s,q_s), \tag{20}$$

with $V_{jklm}\left(p_r,q_r;p_s,q_s\right) = E\left[U_{jk,t}\left(p_r,q_r\right)U_{lm,t}\left(p_s,q_s\right)\right], D_{jk}(p_r,q_r) = ED_{jk,t}(p_r,q_r)$, and $I = EI_t$, where $U_{jk,t}\left(p_r,q_r\right)$, $D_{jk,t}(p_r,q_r)$ and I_t equal $\widetilde{U}_{jk,t}$, $\widetilde{D}_{jk,t}$, and $\widetilde{I}_t$ in (17), respectively, but are evaluated at $(p_r,q_r,\beta_*,\delta_*)$ instead of $(p_r,q_r,\widetilde{\beta},\widetilde{\delta})$. Let $\Omega(p_r,q_r;p_s,q_s)$ be an n_δ^2-by-n_δ^2 matrix, with the $(j+(k-1)n_\delta, l+(m-1)n_\delta)$-th element equal to $\omega_{jklm}\left(p_r,q_r;p_s,q_s\right)$. Let $\Omega(p,q) = \Omega(p,q;p,q)$.

Assumption 6 $\min_{x\in R^{n_\delta}, \|x^{\otimes 2}\|=1} \left(x^{\otimes 2}\right)'\Omega(p,q)\left(x^{\otimes 2}\right) > L$ *for some* $L > 0$ *and all* $(p,q)\in\Lambda_\epsilon$.

This assumption is not restrictive because Λ_ϵ is bounded away from $p+q=1$. In the appendix (pp. A26-A29), we illustrate this assumption in two ways. First, we consider some cases for which $\Omega(p,q)$ can be computed analytically. In addition to showing that this assumption holds, the results also reveal that $\Omega(p_r,q_r;p_s,q_s)$ is affected by the following: (i) the model's dynamic properties (e.g., whether the regressors are weakly or strictly exogenous); (ii) which parameters are allowed to switch (e.g., the regressions coefficients or the variance of the errors); and (iii) whether nuisance parameters are present in the model. Next, we consider $\Omega(p,q)$ in the context of the model given in (8), and explain intuitively why this assumption is expected to hold in general cases.

Proposition 1 *Under the null hypothesis and Assumptions 1-6, for* Λ_ϵ *given by (10), we have*

$$\mathrm{SupLR}(\Lambda_\epsilon) \Longrightarrow \sup_{(p,q)\in\Lambda_\epsilon}\sup_{\eta\in R^{n_\delta}} \mathcal{W}^{(2)}(p,q,\eta) \tag{21}$$

where $\mathcal{W}^{(2)}(p,q,\eta) = \left(\eta^{\otimes 2}\right)'\operatorname{vec}G\left(p,q\right) - (1/4)\left(\eta^{\otimes 2}\right)'\Omega(p,q)\left(\eta^{\otimes 2}\right)$, *and* $\operatorname{vec}G\left(p,q\right)$ *is an* n_δ^2*-vector of zero-mean continuous Gaussian processes such that* $E[\operatorname{vec}G\left(p_r,q_r\right)\operatorname{vec}G\left(p_s,q_s\right)'] = \Omega(p_r,q_r;p_s,q_s)$ *for any* $(p_r,q_r),(p_s,q_s)\in\Lambda_\epsilon$.

In Proposition 1 and its proof, the weak convergence is in the space of continuous functions defined on a compact set. If $n_\delta = 1$, the optimization over η can be solved analytically, leading to $\mathrm{SupLR}(\Lambda_\epsilon) \Rightarrow \max[0, \sup_{(p,q)\in\Lambda_\epsilon} G\left(p,q\right)/\sqrt{\Omega(p,q)}]^2$. If $n_\delta > 1$, it needs to be solved numerically. Because $\mathcal{W}^{(2)}(p,q,\eta)$ is a quadratic function of $\eta^{\otimes 2}$ with η unrestricted, the computation is standard.

The illustrative model (cont'd). We use the following special case of (8) to illustrate the limiting distribution in (21), and compare it with the finite sample distribution:

$$y_t = \mu + \alpha y_{t-1} + u_t, \tag{22}$$

where $u_t \sim i.i.d.\ N(0,\sigma^2)$; $(\mu,\alpha,\sigma^2) = (0, 0.5, 1)$; $T = 250$; and Λ_ϵ is given by (11) with $\epsilon = 0.05$.

The results for testing α and μ are reported in Figures 1(a) and 1(b), respectively. The distributions in panel (a) are substantially different from those in panel (b), which confirms that the distribution of $\text{SupLR}(\Lambda_\epsilon)$ is not invariant to which parameter is allowed to switch. Meanwhile, although the curves are close to each other in panel (a), a gap is observed in (b). In the latter case, applying the asymptotic approximation will lead to an over-rejection of the null hypothesis.

The gap is due to the structure of $T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_{jk,t}$. When testing μ, we have

$$T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_{jk,t} = \frac{1}{\widetilde{\sigma}^2}\left(\frac{1-\xi_*}{\xi_*}\right)\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\widetilde{u}_t^2}{\widetilde{\sigma}^2}-1\right) + \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\rho^s\frac{\widetilde{u}_{t-s}}{\widetilde{\sigma}}\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)\right\}, \tag{23}$$

where the first summation is equal to zero because $\widetilde{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\widetilde{u}_t^2$, and the second is small when $p+q$ is close to one. Thus, although asymptotically, $T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_{jk,t}$ is the leading term in the likelihood ratio expansion over Λ_ϵ, in finite samples, its value can be too small to dominate the omitted higher order terms such as $T^{-3/4}\mathcal{L}_{jkl}^{(3)}(p,q,\widetilde{\delta})$ when $p+q$ is close to one. This omission produces the gap in panel (b). The situation is different when testing for switching in α, where

$$T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_{jk,t} = \frac{1}{\widetilde{\sigma}^2}\left(\frac{1-\xi_*}{\xi_*}\right)\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\widetilde{u}_t^2}{\widetilde{\sigma}^2}-1\right)y_{t-1}^2 + \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\rho^s\frac{y_{t-s-1}\widetilde{u}_{t-s}}{\widetilde{\sigma}}\frac{y_{t-1}\widetilde{u}_t}{\widetilde{\sigma}}\right)\right\}.$$

Because the first term in the braces has a positive variance, this complication does not arise. ■

5.2 A refinement

This subsection presents a refined approximation to the distribution of $\text{SupLR}(\Lambda_\epsilon)$ under the null hypothesis. We begin with the following assumption, which, in practice, is used to determine whether a refinement is needed for a particular testing problem.

Assumption 7 *The following linear relationship holds for some $i_1, i_2 \in \{1,...,n_\delta\}$ and all t:*

$$\nabla_{\delta_{1i_1}}\nabla_{\delta_{1i_2}}\widetilde{f}_{1t} = \alpha_{i_1i_2}^{(1)\prime}\nabla_\beta\widetilde{f}_{1t} + \alpha_{i_1i_2}^{(2)\prime}\nabla_{\delta_1}\widetilde{f}_{1t}, \tag{24}$$

where $\alpha_{i_1i_2}^{(1)}$ and $\alpha_{i_1i_2}^{(2)}$ are n_β- and n_δ-dimensional known vectors of constants, respectively.

If (24) holds, then certain elements of the second order derivatives of the log likelihood cancel out because of the linear dependencies (see, e.g., (23)), implying that a refinement is needed.

For the model in (8), checking (24) can lead to three outcomes, depending on which parameter is allowed to switch. If only the intercept γ is allowed to switch, then (24) is satisfied with $\nabla_\gamma \nabla_\gamma \widetilde{f}_{1t} = 2\nabla_{\sigma^2} \widetilde{f}_{1t}$. If the intercept is not allowed to switch, then (24) is violated for all $i_1, i_2 \in \{1, ..., n_\delta\}$. If the intercept and some other parameters are allowed to switch, then (24) is satisfied when δ_{1i_1} and δ_{1i_2} both represent the intercept, but not otherwise. In the first and third cases, Assumption 7 holds, and a refinement to (21) is needed. In the second case, no complication arises, and thus no refinement is needed. For other models, checking (24) is expected to remain simple, because it requires computing only the first and second order derivatives of the density under the null hypothesis. For example, it is simple to verify that (24) is satisfied when testing γ in the generalized linear model $y_t = g(\gamma + z_t' \alpha + u_t)$, where $g(\cdot)$ is a smooth invertible function and $u_t \sim N(0, \sigma^2)$, and that it is satisfied when testing the intercepts in Gaussian vector autoregressions. The relevant computational steps for the above three models are presented in the appendix (pp. A37-A40), which illustrates how to check this assumption in various situations.

The next assumption strengthens Assumption 4. It is similar to A.5(iv) in Cho and White (2007). The subsequent analysis makes heavy use of their results in Section 2.3.2.

Assumption 8 *There exists an open neighborhood of (β_*, δ_*), $B(\beta_*, \delta_*)$, and a sequence of positive, strictly stationary, and ergodic random variables $\{\upsilon_t\}$, satisfying $E\upsilon_t^{1+c} < \infty$ for some $c > 0$, such that the supremums of the following quantities over $B(\beta_*, \delta_*)$ are bounded from above by υ_t: $\left|\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_k}} f_t(\beta, \delta_1) / f_t(\beta, \delta_1)\right|^4$, $\left|\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_m}} f_t(\beta, \delta_1) / f_t(\beta, \delta_1)\right|^2$, $\left|\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_8}} f_t(\beta, \delta_1) / f_t(\beta, \delta_1)\right|$, $\left|\nabla_{\theta_{j_1}} \nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_7}} f_t(\beta, \delta_1) / f_t(\beta, \delta_1)\right|$, and $\left|\nabla_{\theta_{j_1}} \nabla_{\theta_{j_2}} \nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_6}} f_t(\beta, \delta_1) / f_t(\beta, \delta_1)\right|$, where $k = 1, 2, 3, 4$, $m = 5, 6, 7$, $i_1, ..., i_7 \in \{1, ..., n_\beta + n_\delta\}$, and $j_1, j_2 \in \{1, ..., n_\beta\}$.*

Because the goal of the refinement is to adequately account for the effects of the higher order terms when (p, q) is close to $p + q = 1$, in Lemma 3 below, we study a sixth order Taylor expansion of the log likelihood ratio along $p + q = 1$ and an eighth order expansion at $p = q = 1/2$. To approximate the third and sixth order derivatives of the concentrated log likelihood, define

$$\widetilde{s}_{jkl,t}(p, q) = \frac{(1-p)(p-q)}{(1-q)^2} \frac{\nabla_{\delta_{1j}} \nabla_{\delta_{1k}} \nabla_{\delta_{1l}} \widetilde{f}_{1t}}{\widetilde{f}_t}, \tag{25}$$

and let $G^{(3)}_{jkl}(p, q)$ be a zero-mean continuous Gaussian process, satisfying

$$\begin{aligned} & \omega^{(3)}_{jklmnu}(p_r, q_r; p_s, q_s) \\ = \;& \operatorname{Cov}(G^{(3)}_{jkl}(p_r, q_r), G^{(3)}_{mnu}(p_s, q_s)) \\ = \;& E\left[s_{jkl,t}(p_r, q_r) s_{mnu,t}(p_s, q_s)\right] - E\left[\frac{\nabla_{(\beta', \delta_1')} f_{1t}}{f_t} s_{jkl,t}(p_r, q_r)\right] I^{-1} \left[\frac{\nabla_{(\beta', \delta_1')'} f_{1t}}{f_t} s_{mnu,t}(p_s, q_s)\right], \end{aligned}$$

where $s_{jkl,t}(p,q)$ is equal to $\widetilde{s}_{jkl,t}(p,q)$, but is evaluated at the true parameter values. Let $\omega^{(3)}_{jklmnu}(p,q) = \omega^{(3)}_{jklmnu}(p,q;p,q)$. To approximate the fourth and eighth order derivatives, define

$$\begin{aligned}\widetilde{k}_{jklm,t}(p,q) &= \frac{1-p}{2-p-q}\left(1+\left(\frac{1-p}{1-q}\right)^3\right)\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t} \\ &\quad+\left(\frac{1-p}{1-q}\right)^2\sum_{(i_1,i_2,i_3,i_4)\in S}\frac{1}{\widetilde{f}_t}\Bigg\{-\nabla_{\delta_{1i_1}}\nabla_{\delta_{1i_2}}\nabla_{\beta'}\widetilde{f}_{1t}\alpha^{(1)}_{i_3i_4}-\left(\frac{1-p}{1-q}\right)\nabla_{\delta_{1i_1}}\nabla_{\delta_{1i_2}}\nabla_{\delta_1'}\widetilde{f}_{1t}\alpha^{(2)}_{i_3i_4} \\ &\quad+\frac{1}{2}\alpha^{(1)\prime}_{i_1i_2}\nabla_{\beta}\nabla_{\beta'}\widetilde{f}_{1t}\alpha^{(1)}_{i_3i_4}+\alpha^{(1)\prime}_{i_1i_2}\nabla_{\beta}\nabla_{\delta_1'}\widetilde{f}_{1t}\alpha^{(2)}_{i_3i_4}\Bigg\},\end{aligned} \tag{26}$$

and let $G^{(4)}_{i_1i_2i_3i_4}(p,q)$ denote a zero-mean continuous Gaussian process, satisfying

$$\begin{aligned}\omega^{(4)}_{i_1i_2\ldots i_8}(p_r,q_r;p_s,q_s) &= \text{Cov}\left(G^{(4)}_{i_1i_2i_3i_4}(p_r,q_r),G^{(4)}_{i_5i_6i_7i_8}(p_s,q_s)\right) \\ &= E\left[k_{i_1i_2i_3i_4,t}(p_r,q_r)\,k_{i_5i_6i_7i_8,t}(p_s,q_s)\right] \\ &\quad -E\left[\frac{\nabla_{(\beta',\delta_1')}f_{1t}}{f_t}k_{i_1i_2i_3i_4,t}(p_r,q_r)\right]I^{-1}\left[\frac{\nabla_{(\beta',\delta_1')'}f_{1t}}{f_t}k_{i_5i_6i_7i_8,t}(p_s,q_s)\right],\end{aligned}$$

where $S=\{jklm, jlkm, jmkl, kljm, kmjl, lmjk\}$, and $k_{i_1i_2i_3i_4,t}(p,q)$ is equal to $\widetilde{k}_{i_1i_2i_3i_4,t}(p,q)$, but is evaluated at the true parameter values. Let $\omega^{(4)}_{i_1i_2\ldots i_8}(p,q)=\omega^{(4)}_{i_1i_2\ldots i_8}(p,q;p,q)$.

Lemma 3 *Under the null hypothesis and Assumptions 1-8 with (24) satisfied for all* $i_1,i_2\in\{1,\ldots,n_\delta\}$ *:*

1. *The following results hold uniformly over* $\{(p,q):\epsilon\le p,q\le 1-\epsilon, p+q=1\}$: $T^{-1/2}\mathcal{L}^{(3)}_{jkl}(p,q,\widetilde{\delta}) = T^{-1/2}\sum_{t=1}^T\widetilde{s}_{jkl,t}(p,q)+o_p(1)\Rightarrow G^{(3)}_{jkl}(p,q)$; $T^{-1/2}\mathcal{L}^{(4)}_{jklm}(p,q,\widetilde{\delta})=O_p(1)$; $T^{-1/2}\mathcal{L}^{(5)}_{jklmn}(p,q,\widetilde{\delta})=O_p(1)$; $T^{-1}\mathcal{L}^{(6)}_{jklmnr}(p,q,\widetilde{\delta})=-\sum_{(i_1,i_2,\ldots,i_6)\in IND_1}\omega^{(3)}_{i_1i_2\ldots i_6}(p,q)+o_p(1)$, *where* $IND_1=\{jklmnr$, $jkmlnr, jknlmr, jkrlmn, jlmknr, jlnkmr, jlrkmn, jmnklr, jmrkln, jnrklm\}$.

2. *The following results hold at* $p=q=1/2$: $T^{-1/2}\mathcal{L}^{(3)}_{jkl}(p,q,\widetilde{\delta})=o_p(1)$; $T^{-1/2}\mathcal{L}^{(4)}_{jklm}(p,q,\widetilde{\delta}) = T^{-1/2}\sum_{t=1}^T\widetilde{k}_{jklm,t}(p,q)+o_p(1)\Rightarrow G^{(4)}_{jklm}(p,q)$; $T^{-1/2}\mathcal{L}^{(k)}_{i_1i_2\ldots i_k}(p,q,\widetilde{\delta})=O_p(1)$, *where* $i_1,\ldots,i_k\in\{1,\ldots,n_\delta\}$ *for k=5,6 and 7;* $T^{-1}\mathcal{L}^{(8)}_{jklmnrsu}(p,q,\widetilde{\delta})=-\sum_{(i_1,i_2,\ldots,i_8)\in IND_2}\omega^{(4)}_{i_1i_2\ldots i_8}(p,q)+o_p(1)$, *where the elements of* IND_2 *are defined as follows:* $i_1=j$, *each triplet* (i_2,i_3,i_4) *corresponds to one of the 35 outcomes of picking three elements from* $\{k,l,m,n,r,s,u\}$ *(the ordering does not matter), and* i_5,i_6,i_7, *and* i_8 *correspond to the remaining elements.*

When $p+q=1$ and $p\neq 1/2$, $T^{-1/2}\sum_{t=1}^T\widetilde{s}_{jkl,t}(p,q)$ serves as the leading term of the Taylor expansion. As a result, a sixth order expansion is needed to approximate the likelihood ratio. When $p=q=1/2$, $T^{-1/2}\sum_{t=1}^T\widetilde{k}_{jklm,t}(p,q)$ becomes the leading term, and an eighth order expansion is needed. Lemma 3 assumes that (24) is satisfied for all $i_1, i_2 \in \{1, ..., n_\delta\}$. If this relationship holds for a subset of derivatives, then we set $\alpha^{(1)}_{i_1 i_2} = 0$ and $\alpha^{(2)}_{i_1 i_2} = 0$ for the cases that do not satisfy (24).

Proposition 1 and Lemma 3 lead to three different approximations to $LR(p,q)$. The approximation in Proposition 1 is expected to perform well when (p,q) is not close to the boundary $p+q=1$. The approximation implied by Lemma 2.1, which is based on the limit of $T^{-1/2}\sum_{t=1}^{T}\widetilde{s}_{jkl,t}(p,q)$, is expected to perform well when $p+q=1$ but $p \neq 1/2$, as well as when (p,q) is local to such a point because of the continuity of $LR(p,q)$ with respect to p and q. Finally, the approximation implied by Lemma 2.2, which is based on the limit of $T^{-1/2}\sum_{t=1}^{T}\widetilde{k}_{jklm,t}(p,q)$, is expected to perform well when (p,q) is equal to, or is local to $(1/2,1/2)$. These three approximations complement each other. By merging their leading terms in a proper way, according to how they appear in the Taylor expansion, it is potentially possible to obtain an approximation that performs well over a wide range of transition probabilities. This is the intuition behind our refined approximation. We present this approximation below, and examine it further in Subsection 6.1.

Let $G^{(3)}(p,q)$ be an n_δ^3- dimensional vector, with the $(j+(k-1)n_\delta+(l-1)n_\delta^2)$-th element given by $G^{(3)}_{jkl}(p,q)$, and $\Omega^{(3)}(p,q)$ an n_δ^3-by-n_δ^3 matrix, with the $(j+(k-1)n_\delta+(l-1)n_\delta^2, m+(n-1)n_\delta+(r-1)n_\delta^2)$-th element given by $\omega^{(3)}_{jklmnr}(p,q)$. Define

$$\mathcal{W}^{(3)}(p,q,\eta) = T^{-1/4}\frac{1}{3}\left(\eta^{\otimes 3}\right)' \operatorname{vec} G^{(3)}(p,q) - T^{-1/2}\frac{1}{36}\left(\eta^{\otimes 3}\right)' \Omega^{(3)}(p,q)\left(\eta^{\otimes 3}\right).$$

Similarly, let $G^{(4)}(p,q)$ be an n_δ^4-dimensional vector, with the $(j+(k-1)n_\delta+(l-1)n_\delta^2+(m-1)n_\delta^3)$-th element equal to $G^{(4)}_{jklm}(p,q)$, and $\Omega^{(4)}(p,q)$ an n_δ^4-by-n_δ^4 matrix, with the $(j+(k-1)n_\delta+(l-1)n_\delta^2+(m-1)n_\delta^3, n+(r-1)n_\delta+(s-1)n_\delta^2+(u-1)n_\delta^3)$-th element equal to $\omega^{(4)}_{jklmnrsu}(p,q)$. Define

$$\mathcal{W}^{(4)}(p,q,\eta) = T^{-1/2}\frac{1}{12}\left(\eta^{\otimes 4}\right)' \operatorname{vec} G^{(4)}(p,q) - T^{-1}\frac{1}{576}\left(\eta^{\otimes 4}\right)' \Omega^{(4)}(p,q)\left(\eta^{\otimes 4}\right).$$

We propose using $S_\infty(\Lambda_\epsilon)$ as the refined approximation to $\text{SupLR}(\Lambda_\epsilon)$, where

$$S_\infty(\Lambda_\epsilon) \equiv \sup_{(p,q)\in\Lambda_\epsilon}\sup_{\eta\in R^{n_\delta}}\left\{\mathcal{W}^{(2)}(p,q,\eta)+\mathcal{W}^{(3)}(p,q,\eta)+\mathcal{W}^{(4)}(p,q,\eta)\right\}. \tag{27}$$

Corollary 1 *Under the null hypothesis and Assumptions 1-8 with Λ_ϵ equal to (10), we have* $\Pr(\text{SupLR}(\Lambda_\epsilon)\le s) - \Pr(S_\infty(\Lambda_\epsilon)\le s) \to 0$ *for any $s \in R$.*

The illustrative model (cont'd). For the model in (22), when testing μ, (25) and (26) equal

$$\frac{(1-p)(p-q)}{(1-q)^2}\frac{1}{\widetilde{\sigma}^3}\left\{\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^3 - 3\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right\} \quad \text{and} \quad \left[\frac{1-p}{2-p-q}\left(1+\left(\frac{1-p}{1-q}\right)^3\right) - 3\left(\frac{1-p}{1-q}\right)^2\right]\frac{1}{\widetilde{\sigma}^4}\left\{\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^4 - 6\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^2 + 3\right\},$$

respectively. The refined approximation is reported in Figure 1(b). The improvement over the original approximation is substantial. When testing α, by Assumption 7, no refinement is needed.■

5.3 Critical values

This section presents an algorithm to produce the critical values of $S_{\infty}(\Lambda_{\epsilon})$ in (27). The idea is to sample from the estimated distribution of the vector Gaussian process

$$[\text{vec}\, G\,(p,q)'\,, \text{vec}\, G^{(3)}\,(p,q)'\,, \text{vec}\, G^{(4)}\,(p,q)'], \tag{28}$$

and then, for each draw, solve the maximization problem (27). Similar algorithms are considered in Hansen (1992) and Garcia (1998). The main steps are as follows.

STEP 1 (Estimate the parameters). Estimate the model after imposing the null hypothesis to obtain $\widetilde{\beta}$ and $\widetilde{\delta}$. Create an equidistant grid over Λ_{ϵ}, and denote the grid points by $\{p_i, q_i\}_{i=1}^{n}$.

STEP 2 (Estimate the covariance function of the Gaussian process). Compute $\widetilde{U}_{jk,t}(p_i, q_i)$, $\widetilde{s}_{jkl,t}(p_i, q_i)$, and $\widetilde{k}_{jklm,t}(p_i, q_i)$ using (17), (25), and (26), respectively, where $j, k, l, m \in \{1, ..., n_{\delta}\}$ and $i \in \{1, ..., n\}$. For each i, store their values in a vector as

$$\widetilde{\mathcal{G}}_t\,(p_i, q_i) = \begin{bmatrix} \widetilde{U}_t^{(2)}(p_i, q_i) \\ \widetilde{U}_t^{(3)}(p_i, q_i) \\ \widetilde{U}_t^{(4)}(p_i, q_i) \end{bmatrix}, \tag{29}$$

where $\widetilde{U}_t^{(2)}(p, q)$ is an n_{δ}^2-vector, with the $(j + (k-1)n_{\delta})$-th element equal to $\widetilde{U}_{jk,t}(p, q)$, $\widetilde{U}_t^{(3)}(p, q)$ is an n_{δ}^3-vector, with the $(j + (k-1)n_{\delta} + (l-1)n_{\delta}^2)$-th element equal to $\widetilde{s}_{jkl,t}(p, q)$, and $\widetilde{U}_t^{(4)}(p, q)$ is an n_{δ}^4-vector, with the $(j + (k-1)n_{\delta} + (l-1)n_{\delta}^2 + (m-1)n_{\delta}^3)$-th element equal to $\widetilde{k}_{jklm,t}(p, q)$. For each $i, s \in \{1, ..., n\}$, compute

$$\begin{aligned} \widetilde{\Omega}\,(p_i, q_i; p_s, q_s) \quad &= \quad T^{-1}\sum_{t=1}^{T} \widetilde{\mathcal{G}}_t\,(p_i, q_i)\,\widetilde{\mathcal{G}}_t\,(p_s, q_s)' \\ &\quad - \left\{T^{-1}\sum_{t=1}^{T} \widetilde{\mathcal{G}}_t\,(p_i, q_i)\,\frac{\nabla_{(\beta', \delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\right\}\widetilde{I}^{-1}\left\{T^{-1}\sum_{t=1}^{T} \widetilde{\mathcal{G}}_t\,(p_s, q_s)\,\frac{\nabla_{(\beta', \delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\right\}'. \end{aligned} \tag{30}$$

Let $\widetilde{\Omega}(p_i, q_i)$, $\widetilde{\Omega}^{(3)}(p_i, q_i)$, and $\widetilde{\Omega}^{(4)}(p_i, q_i)$ denote the three consecutive diagonal blocks of $\widetilde{\Omega}\,(p_i, q_i; p_i, q_i)$ of dimensions n_{δ}^2, n_{δ}^3, and n_{δ}^4, respectively.

STEP 3 (Sampling). Generate an n-by-1 zero-mean normal random vector, with covariance equal to (30), and repeat B times. Save the values as $[\text{vec}\, G_b\,(p_i, q_i)'\,, \text{vec}\, G_b^{(3)}\,(p_i, q_i)'\,, \text{vec}\, G_b^{(4)}\,(p_i, q_i)']$, where $i = 1, ..., n$, and $b = 1, ..., B$.

STEP 4 (Optimization). For each $b \in \{1, ..., B\}$, solve

$$S_b(\Lambda_{\epsilon}) \equiv \sup_{\{p_i, q_i\}_{i=1}^{n}} \sup_{\eta \in R^{n_{\delta}}} \sum_{j=2}^{4} \mathcal{W}_b^{(j)}(p_i, q_i, \eta), \tag{31}$$

where

$$\begin{aligned}
\mathcal{W}_b^{(2)}(p_i,q_i,\eta) &= \left(\eta^{\otimes 2}\right)' \operatorname{vec} G_b\left(p_i,q_i\right) - \frac{1}{4}\left(\eta^{\otimes 2}\right)' \widetilde{\Omega}(p_i,q_i)\left(\eta^{\otimes 2}\right), \qquad (32)\\
\mathcal{W}_b^{(3)}(p_i,q_i,\eta) &= T^{-1/4}\frac{1}{3}\left(\eta^{\otimes 3}\right)' \operatorname{vec} G_b^{(3)}(p_i,q_i) - T^{-1/2}\frac{1}{36}\left(\eta^{\otimes 3}\right)' \widetilde{\Omega}^{(3)}(p_i,q_i)\left(\eta^{\otimes 3}\right),\\
\mathcal{W}_b^{(4)}(p_i,q_i,\eta) &= T^{-1/2}\frac{1}{12}\left(\eta^{\otimes 4}\right)' \operatorname{vec} G_b^{(4)}(p_i,q_i) - T^{-1}\frac{1}{576}\left(\eta^{\otimes 4}\right)' \widetilde{\Omega}^{(4)}(p_i,q_i)\left(\eta^{\otimes 4}\right).
\end{aligned}$$

Sort the values of $S_b(\Lambda_\epsilon)$ (b=1, ..., B) to obtain the desired critical value.

The illustrative model (cont'd). When testing the intercept of (22), we have

$$\begin{aligned}
&\widetilde{U}_t^{(2)}(p,q) = \tfrac{2}{\widetilde{\sigma}^2}\left(\tfrac{1-p}{1-q}\right)\textstyle\sum_{s=1}^{t-1}(p+q-1)^s \tfrac{\widetilde{u}_{t-s}\widetilde{u}_t}{\widetilde{\sigma}^2},\ \widetilde{U}_t^{(3)}(p,q) = \tfrac{(1-p)(p-q)}{(1-q)^2}\tfrac{1}{\widetilde{\sigma}^3}\left\{\left(\tfrac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^3 - 3\tfrac{\widetilde{u}_t}{\widetilde{\sigma}}\right\},\\
&\widetilde{U}_t^{(4)}(p,q) = \left[\tfrac{1-p}{2-p-q}\left(1+\left(\tfrac{1-p}{1-q}\right)^3\right) - 3\left(\tfrac{1-p}{1-q}\right)^2\right]\tfrac{1}{\widetilde{\sigma}^4}\left\{\left(\tfrac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^4 - 6\left(\tfrac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^2 + 3\right\}, \qquad (33)\\
&\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t} = \left[\begin{array}{ccc}\frac{y_{t-1}\widetilde{u}_t}{\widetilde{\sigma}^2} & \frac{1}{2\widetilde{\sigma}^2}\left(\frac{\widetilde{u}_t^2}{\widetilde{\sigma}^2}-1\right) & \frac{\widetilde{u}_t}{\widetilde{\sigma}^2}\end{array}\right].
\end{aligned}$$

The covariance function in (30) follows from (33) and (29). When testing the same hypothesis in AR(p) or ADL(p,q) models, the covariance function can be computed in the same way, except that y_{t-1} in (33) needs to be replaced by $(y_{t-1},...,y_{t-p})$ and $(y_{t-1},...,y_{t-p},x_{t-1},...,x_{t-q})$, respectively.■

6 The boundary issue and local power properties

In this section, we study two issues. First, we study the likelihood ratio under the null hypothesis when the transition probabilities are close to the boundary:

$$\{(p,q) : p+q=1 \text{ and } \epsilon \le p,q \le 1-\epsilon \text{ for some } 0<\epsilon<0.5\}. \qquad (34)$$

The results further justify the refined approximation presented and implemented in Subsections 5.2-5.3. Next, we study the likelihood ratio under the alternative hypothesis. The results explain the potential local power difference between the likelihood ratio test and the tests of Cho and White (2007) and Carrasco, Hu, and Ploberger (2014) in an empirically important setting.

6.1 The boundary issue

We allow the transition probabilities to depend on the sample size, such that they approach (34) as $T\to\infty$. The analysis below is based on an eighth order expansion of $\mathcal{L}(p_T,q_T,\delta_2)$ around $\widetilde{\delta}$:

$$\begin{aligned}
\mathcal{L}(p_T,q_T,\delta_2) - \mathcal{L}(p_T,q_T,\widetilde{\delta}) &= \frac{1}{k!}\sum_{k=2}^{7}\sum_{i_1,\ldots,i_k=1}^{n_\delta}\mathcal{L}_{i_1\ldots i_k}^{(k)}(p_T,q_T,\widetilde{\delta})d_{i_1}\ldots d_{i_k} \qquad (35)\\
&\quad + \frac{1}{8!}\sum_{i_1,\ldots,i_8=1}^{n_\delta}\mathcal{L}_{i_1\ldots i_8}^{(8)}(p_T,q_T,\bar{\delta})d_{i_1}\ldots d_{i_8},
\end{aligned}$$

where d_j is the j-th element of $(\delta_2 - \widetilde{\delta})$ and $\bar{\delta} = \widetilde{\delta} + c(\delta_2 - \widetilde{\delta})$ for some $c \in (0,1)$. We assume (24) is satisfied for all, not just some $i_1, i_2 \in \{1, ..., n_\delta\}$. Relaxing this assumption is left for future work.

Assumption 9 *(a)* $\sup_{(\beta,\delta_1)\in B(\beta_*,\delta_*)} |\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_k}} f_t(\beta,\delta_1) / f_t(\beta,\delta_1)|^{\alpha(k)/k} < \upsilon_t$ *holds for* $k = 1, 2, 5, 8$, *where* $i_1, ..., i_8 \in \{1, ..., n_\beta + n_\delta\}$, $\alpha(k) = 12$ *if* $k \in \{1, 2, 5\}$, $\alpha(8) = 8$, *and* $B(\beta_*, \delta_*)$ *and* υ_t *are defined in Assumption 8; (b)* $E(\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_k}} f_t(\beta_*, \delta_*) \nabla_\theta f_t(\beta_*, \delta_*) / f_t(\beta_*, \delta_*)^2 |\Omega_{t-1}) = 0$ *holds for* $k = 3, 4$, *where* $i_1, ..., i_4 \in \{n_\beta + 1, ..., n_\beta + n_\delta\}$, *and* β_* *and* δ_* *denote the true parameter values.*

Assumption 9(a) strengthens Assumptions 4 and 8, enabling us to use the CLT and LLN to study (35). Assumption 9(b) requires that the score and $\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_k}} f_t(\beta_*, \delta_*)/f_t(\beta_*, \delta_*)$ are uncorrelated for $k = 3, 4$, which holds in the model in (8), regardless of which parameters are allowed to switch.

We first study the refined approximation under the following two drifting sequences of (p_T, q_T):

$$\begin{aligned} \text{SEQ1} &: \quad p_T = p + c_1 T^{-a_1} \text{ and } q_T = q + c_2 T^{-a_2} \text{ for some } (p,q) \neq (0.5, 0.5), \\ \text{SEQ2} &: \quad p_T = 0.5 + c_1 T^{-a_1} \text{ and } q_T = 0.5 + c_2 T^{-a_2}, \end{aligned} \tag{36}$$

where (p, q) denotes a point in (34), $a_1, a_2 \geq 0$, and $c_1, c_2 \neq 0$. Let $a = \min(a_1, a_2)$. We assume $c_1 \neq -c_2$ when $a_1 = a_2$; otherwise, ρ_T is zero for any a. Define

$$\begin{aligned} \widetilde{\mathcal{W}}^{(2)}(p_T, q_T, \eta) &= (\eta^{\otimes 2})' \left(T^{-1/2} \sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T) \right) - \frac{1}{4} (\eta^{\otimes 2})' \widetilde{\Omega}(p_T, q_T) (\eta^{\otimes 2}), \qquad (37) \\ \widetilde{\mathcal{W}}^{(3)}(p_T, q_T, \eta) &= T^{-1/4} \frac{1}{3} (\eta^{\otimes 3})' \left(T^{-1/2} \sum_{t=1}^{T} \widetilde{U}_t^{(3)}(p_T, q_T) \right) - T^{-1/2} \frac{1}{36} (\eta^{\otimes 3})' \widetilde{\Omega}^{(3)}(p_T, q_T) (\eta^{\otimes 3}), \\ \widetilde{\mathcal{W}}^{(4)}(p_T, q_T, \eta) &= T^{-1/2} \frac{1}{12} (\eta^{\otimes 4})' \left(T^{-1/2} \sum_{t=1}^{T} \widetilde{U}_t^{(4)}(p_T, q_T) \right) - T^{-1} \frac{1}{576} (\eta^{\otimes 4})' \widetilde{\Omega}^{(4)}(p_T, q_T) (\eta^{\otimes 4}). \end{aligned}$$

Let $S_b(p_T, q_T)$ denote the output from STEP 4 of the algorithm when $(p_i, q_i) = (p_T, q_T)$, that is, $S_b(p_T, q_T) = \sup_{\eta \in R^{n_\delta}} \sum_{j=2}^{4} \mathcal{W}_b^{(j)}(p_T, q_T, \eta)$; see (31)-(32).

Proposition 2 *Suppose that the null hypothesis and Assumptions 1-9 hold, with (24) satisfied for all* $i_1, i_2 \in \{1, ..., n_\delta\}$. *Then:*

1. *Under* SEQ1, *if* $\min\limits_{||x^{\otimes 2}||=1} \frac{(x^{\otimes 2})' \widetilde{\Omega}(p_T, q_T)(x^{\otimes 2})}{\rho_T^2} > L$ *and* $\min\limits_{||x^{\otimes 3}||=1} \frac{(x^{\otimes 3})' \widetilde{\Omega}^{(3)}(p_T, q_T)(x^{\otimes 3})}{(p_T - q_T)^2} > L$ *for some* $L > 0$ *in probability, and* $\sup_{||\delta - \widetilde{\delta}|| < c} T^{-1} |\mathcal{L}^{(7)}_{i_1 ... i_7}(p_T, q_T, \delta)| = O_p(1)$ *for some* $c > 0$ *and any* $i_1, ..., i_7 \in \{1, ..., n_\delta\}$, *then* $\Pr(LR(p_T, q_T) \leq x) - \Pr(\widetilde{S}(p_T, q_T) \leq x) \to 0$ *for* $x \in R$, *where*

$$\widetilde{S}(p_T, q_T) = \begin{cases} \sup_{\eta \in R^{n_\delta}} \widetilde{\mathcal{W}}^{(2)}(p_T, q_T, \eta) & \text{if } a < 1/6 \\ \sup_{\eta \in R^{n_\delta}} \{\widetilde{\mathcal{W}}^{(2)}(p_T, q_T, \eta) + \widetilde{\mathcal{W}}^{(3)}(p_T, q_T, \eta)\} & \text{if } a = 1/6 \\ \sup_{\eta \in R^{n_\delta}} \widetilde{\mathcal{W}}^{(3)}(p_T, q_T, \eta) & \text{if } a > 1/6 \end{cases}.$$

2. *Under* SEQ2, *if* $\min_{||x^{\otimes 2}||=1} \frac{(x^{\otimes 2})'\widetilde{\Omega}(p_T,q_T)(x^{\otimes 2})}{\rho_T^2} > L$ *and* $\min_{||x^{\otimes 4}||=1}(x^{\otimes 4})'\widetilde{\Omega}^{(4)}(p_T,q_T)(x^{\otimes 4}) > L$ *for some* $L > 0$ *in probability, and* $\sup_{||\delta-\widetilde{\delta}||<c} T^{-1}|\mathcal{L}^{(9)}_{i_1...i_9}(p_T,q_T,\delta)|=O_p(1)$ *for some* $c > 0$ *and any* $i_1,...,i_9 \in \{1,...,n_\delta\}$, *then* $\Pr(LR(p_T,q_T) \leq x) - \Pr(\widetilde{S}(p_T,q_T) \leq x) \to 0$ *for* $x \in R$, *where*

$$\widetilde{S}(p_T,q_T) = \begin{cases} \sup_{\eta \in R^{n_\delta}} \widetilde{\mathcal{W}}^{(2)}(p_T,q_T,\eta) & \text{if } a < 1/4 \\ \sup_{\eta \in R^{n_\delta}} \{\widetilde{\mathcal{W}}^{(2)}(p_T,q_T,\eta) + \widetilde{\mathcal{W}}^{(4)}(p_T,q_T,\eta)\} & \text{if } a = 1/4 \\ \sup_{\eta \in R^{n_\delta}} \widetilde{\mathcal{W}}^{(4)}(p_T,q_T,\eta) & \text{if } a > 1/4 \end{cases}.$$

3. *The distribution of* $S_b(p_T,q_T)$ *is a weakly consistent estimator of the limiting distribution of* $LR(p_T,q_T)$ *under* SEQ1 *and* SEQ2 *if the above conditions in 2.1 and 2.2 are satisfied.*

The first two results of Proposition 2 reveal how the distribution of $LR(p_T,q_T)$ differs between SEQ1 and SEQ2, and how it changes with a. The third result shows that the refined approximation is a consistent estimator of the limiting distribution of $LR(p_T,q_T)$ under SEQ1 and SEQ2, for any $a \geq 0$. The consistency holds because the refined approximation encompasses all terms that potentially matter for the limiting distribution of $LR(p_T,q_T)$ under SEQ1 and SEQ2. For example, although $\mathcal{W}_b^{(4)}(p_T,q_T,\eta)$ is asymptotically negligible when (p_T,q_T) follows SEQ1, including it ensures consistency under SEQ2. In fact, if any $\mathcal{W}_b^{(k)}(p_T,q_T,\eta)$ $(k = 2,3,4)$ is removed, the resulting approximation will become inconsistent under SEQ1 or SEQ2 for some $a > 0$. The assumptions on $\mathcal{L}^{(k)}_{i_1...i_k}(p_T,q_T,\delta)$ $(k = 7,9)$ serve a similar purpose as Assumption 5. The condition on $(x^{\otimes 2})'\widetilde{\Omega}(p_T,q_T)(x^{\otimes 2})/\rho_T^2$ is illustrated using a linear model in Lemma A.21 of the appendix.

The next result allows (p_T,q_T) to follow sequences that are more general than (36).

Corollary 2 *Assume* $\rho_T \to 0$ *as* $T \to \infty$. *Suppose that the conditions in Proposition 2 hold, and that* $E(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_3}} f_t(\beta_*,\delta_*)\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}} f_t(\beta_*,\delta_*)/f_t(\beta_*,\delta_*)^2|\Omega_{t-1}) = 0$ *for* $i_1,...,i_4 \in \{n_\beta+1,...,n_\beta+n_\delta\}$. *Then,* $\Pr(LR(p_T,q_T) \leq x) - \Pr(\sup_{\eta \in R^{n_\delta}} \sum_{j=2}^4 \widetilde{\mathcal{W}}^{(j)}(p_T,q_T,\eta) \leq x) \to 0$ *for any* $x \in R$.

Corollary 2 implies that the refined approximation is valid under general drifting sequences. Because (p_T,q_T) may not converge to any point, we do not obtain a consistency result, as in Proposition 2.3. The assumption on $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}} f_t(\beta_*,\delta_*)$ $(k = 3,4)$ ensures that $\mathcal{L}^{(7)}_{i_1,...,i_7}(p_T,q_T,\widetilde{\delta})= O_p(T^{-1/2}+|\rho_T|+|p_T-q_T|^2)$. This holds in the model in (8) if the variance is not allowed to switch.

In the above analysis, we have allowed the transition probabilities of a Markov switching model to approach those of a mixture model. The results show that the distribution of $LR(p_T,q_T)$ is path dependent and that the number of terms needed for a consistent approximation to it varies accordingly. Because the refined approximation encompasses all terms that are potentially nonnegligible, it provides an adequate approximation over a wide range of transition probabilities.

6.2 Local power properties

We first derive a refined approximation to the distribution of $LR(p,q)$ under a simple DGP. Then, we apply it to study the power properties of the tests. By focusing on a simple DGP, we lose some generality, but we gain analytical results that clearly show the differences between the tests.

The DGP is

$$y_t = \mu_* + A_T 1_{\{s_t=2\}} + e_t, \tag{38}$$

where $e_t \sim i.i.d.N(0,\sigma_*^2)$; $A_T = c_* T^{-1/4}$; μ_*, σ_*^2, and c_* are independent of T; and the transition probabilities p_* and q_* are fixed and belong to Λ_ϵ. The local alternatives are $O(T^{-1/4})$ (proved in Lemma A.29 in the appendix), as in Carrasco, Hu, and Ploberger (2014).

We consider two estimated models in order to reflect the empirical literature and to quantify the effect of the model specification on the testing power:

$$\begin{aligned} &(\text{Static Model}) && y_t = \delta_1 1_{\{s_t=1\}} + \delta_2 1_{\{s_t=2\}} + u_t, \\ &(\text{Dynamic Model}) && y_t = \delta_1 1_{\{s_t=1\}} + \delta_2 1_{\{s_t=2\}} + \alpha y_{t-1} + u_t, \end{aligned} \tag{39}$$

where $u_t \sim N(0,\sigma^2)$; and $\delta_1, \delta_2, \alpha$, and σ^2 are unknown parameters. The static model is considered in Cecchetti, Lam, and Mark (1990), Chauvet and Hamilton (2006), and Hamilton (2016); and the dynamic model is considered in Davig (2004), Hansen (1992), and Cho and White (2007). To simplify the analysis and focus on the main issue, we assume (p_*, q_*) is known.

To present the results, let $\rho_* = p_* + q_* - 1$, and define the following noncentrality parameters:

$$\begin{aligned} A_{2s}(p_*,q_*) &= \frac{2c_*^2}{\sigma_*^4}\left(\frac{1-p_*}{1-q_*}\right)\sum_{j=1}^{\infty}\rho_*^j \operatorname{Cov}\left(1_{\{s_t=2\}}, 1_{\{s_{t-j}=2\}}\right), \\ A_{2d}(p_*,q_*) &= \frac{2c_*^2}{\sigma_*^4}\left(\frac{1-p_*}{1-q_*}\right)\sum_{j=2}^{\infty}\rho_*^j \operatorname{Cov}\left(1_{\{s_t=2\}}, 1_{\{s_{t-j}=2\}}\right), \\ A_3(p_*,q_*) &= \frac{c_*^3}{3\sigma_*^6}\frac{(1-p_*)(p_*-q_*)}{(1-q_*)^2} E\left(1_{\{s_t=2\}} - E1_{\{s_t=2\}}\right)^3, \\ A_4(p_*,q_*) &= \frac{c_*^4}{12\sigma_*^8}\left(\frac{1-p_*}{2-p_*-q_*}\left(1+\left(\frac{1-p_*}{1-q_*}\right)^3\right) - 3\left(\frac{1-p_*}{1-q_*}\right)^2\right) \\ &\quad \times \left\{E\left(1_{\{s_t=2\}} - E1_{\{s_t=2\}}\right)^4 - 3E\left(1_{\{s_t=2\}} - E1_{\{s_t=2\}}\right)^2\right\}. \end{aligned} \tag{40}$$

The above quantities depend only on (p_*, q_*) and c_*/σ_*^2. Their analytical expressions are given in Lemma A.31. Let $\sup_{\eta\in R}\{\sum_{j=2}^4 \mathcal{W}^{(j)}(p_*,q_*,\eta)\}$ denote the refined approximation to $LR(p_*,q_*)$ under the null hypothesis, i.e., with $c_* = 0$; see (27) and its simulated version in (32).

Proposition 3 *Under DGP (38) and Assumption 5,* $\Pr(LR(p_*, q_*) \leq s) - \Pr(\mathcal{S}(p_*, q_*) \leq s) \to 0$ *for any* $s \in R$*, where* $\mathcal{S}(p_*, q_*)$ *equals*

$$\begin{aligned} &\sup_{\eta \in R}\{\textstyle\sum_{j=2}^{4} \mathcal{W}^{(j)}(p_*, q_*, \eta) + A_{2s}(p_*, q_*)\eta^2 + T^{-1/2}A_3(p_*, q_*)\eta^3 + T^{-1}A_4(p_*, q_*)\eta^4\}, \\ &\sup_{\eta \in R}\{\textstyle\sum_{j=2}^{4} \mathcal{W}^{(j)}(p_*, q_*, \eta) + A_{2d}(p_*, q_*)\eta^2 + T^{-1/2}A_3(p_*, q_*)\eta^3 + T^{-1}A_4(p_*, q_*)\eta^4\}, \end{aligned} \tag{41}$$

in the static and dynamic cases, respectively.

Proposition 3 shows that, under (38)-(39), regime switching affects the local power of $LR(p_*, q_*)$ through three channels: (i) serial dependence, measured by $A_{2s}(p_*, q_*)$ or $A_{2d}(p_*, q_*)$; (ii) asymmetry, measured by $T^{-1/2}A_3(p_*, q_*)$; and (iii) tail behavior, measured by $T^{-1}A_4(p_*, q_*)$. Although (ii) and (iii) appear as high order terms in the refined approximation, their values can exceed (i) in finite samples. For example, $T^{-1/2}A_3(p_*, q_*)$ is greater than $A_{2s}(p_*, q_*)$ and $A_{2d}(p_*, q_*)$ when $(T, c_*, \sigma_*, p_*, q_*) = (500, 6, 1, 0.9, 0.2)$, and it remains greater than $A_{2d}(p_*, q_*)$ after q_* increases to 0.4. Therefore, neglecting (ii) or (iii) can lead to a poor, even misleading representation of the power properties of the likelihood ratio test. Proposition 3 also shows that, in the dynamic case, the first order dependence caused by the regime switching no longer helps the power. This result explains why detecting regime switching in the mean is more difficult when allowing for linear dynamics.

The QLR test of Cho and White (2007) captures the mixture properties, but not the serial correlation caused by the regime switching. When the mixing probability π is set to $\xi_* = (1 - q_*)/(2 - p_* - q_*)$, a refined approximation to its distribution is given by

$$\sup_{\eta \in R}\left\{\mathcal{W}^{(3)}(p_*, q_*, \eta) + \mathcal{W}^{(4)}(p_*, q_*, \eta) + T^{-1/2}A_3(p_*, q_*)\eta^3 + T^{-1}A_4(p_*, q_*)\eta^4\right\}. \tag{42}$$

The $TS(\rho_*)$ test of Carrasco, Hu, and Ploberger (2014) captures the serial correlation, but not the mixture properties. From their analysis, it follows that $2TS(\rho_*)$ converges to

$$\sup_{\eta \in R}\left\{\mathcal{W}^{(2)}(p_*, q_*, \eta) + A_{2s}(p_*, q_*)\eta^2\right\} \text{ and } \sup_{\eta \in R}\left\{\mathcal{W}^{(2)}(p_*, q_*, \eta) + A_{2d}(p_*, q_*)\eta^2\right\} \tag{43}$$

in the static and dynamic cases, respectively. Therefore, under (38)-(39), (i) does not contribute to the local power of the QLR test, and (ii) and (iii) do not affect the $TS(\rho_*)$ test.

We now use simulations to quantify these differences. We set $(T, \sigma_*, p_*) = (500, 1, 0.9)$ and $(q_*, c_*) = (0.2, 7.3)$,$(0.4, 6.4)$,$(0.6, 4.7)$,$(0.8, 3.0)$, where the values of c_* are chosen such that the power of $LR(p_*, q_*)$ is about 50% in the static case at the 5% nominal level. For $c_* = 7.3$, the change between the regimes is 1.54, while for $c_* = 3.0$, it is 0.63. As the standard deviation is 1.0, these values imply that the change needs to be big for the test to have good power and that the power increases with the persistence of the regimes. Table 1 reports the rejection frequencies computed using (41), (42), and (43). The distributions that produce these values are reported in Figures S2-S9, along with the finite sample distributions to reflect the adequacy of the approximations.

In the static case (Panel (a) of Table 1), the power of $LR(p_*, q_*)$ is substantially higher than that of $QLR(\pi_*)$ when $q_* = 0.6, 0.8$, and substantially higher than that of $TS(\rho_*)$ when $q_* = 0.2, 0.4$, where the differences reach 47.36% and 44.90%, respectively. This confirms that serial dependence is important for power when the regimes are persistent, and that the asymmetry and tail behavior are important otherwise. The dynamic case (Panel (b)) shows a similar pattern, where the differences reach 31.31% and 48.71%, respectively. Figures S2-S9 show that, overall, the approximations are close to their finite sample distributions. The approximation improves further when T is increased to 1000; see Figures S10-S17.

Proposition 3 implies that the local power difference between $LR(p_*, q_*)$ and $TS(\rho_*)$ should decrease as the sample size increases. To examine this further, we repeat the above analysis using $T = 1000, 5000, 20000$, and 50000. We focus on $(q_*, c_*) = (0.2, 7.3)$, because in this case, the power difference is the largest. For the static model, the power differences at the 5% level computed using (41) and (43) for the four sample sizes are 35.58%, 17.75%, 7.35%, and 3.12%, respectively. Although the value decreases, the rate is slow, and it remains significant for sample sizes that are enormous from an empirical perspective. The pattern is similar for the dynamic model, with the rate of the decrease being even slower, where the corresponding values are 41.60%, 21.90%, 12.24%, and 7.02%, respectively. The results also confirm that the refined approximations accurately represent their finite sample distributions in all cases; see Figures S18-S21. Therefore, the asymptotic of the LR test kicks in only in very large samples, and for many applications, the finite sample correction term $T^{-1/2}A_3(p_*, q_*)\eta^3 + T^{-1}A_4(p_*, q_*)\eta^4$ is necessary.

In summary, we have examined how regime switching affects the local power of the likelihood ratio test by altering the serial dependence, symmetry, and tail behavior of a time series. The results show that these three channels are all potentially important for power. The tests of Cho and White (2007) and Carrasco, Hu, and Ploberger (2014) turn off some channels and, as a result, their power can be lower than that achievable.

7 Implications for bootstrap procedures and information criteria

The results in the previous sections can be used to evaluate the consistency, or the lack thereof, of various bootstrap procedures. We illustrate some important aspects using the linear model in (8).

Bootstrap procedures. We begin with the important special case where the model specifies a stationary AR(p) process with normal errors. A standard parametric bootstrap procedure is as follows. (1) Estimate the model under the null hypothesis. (2) Sample from a normal distribution with mean zero and variance equal to the sample variance of the residuals. Use the sampled values and the estimated coefficients to generate a new AR(p) series. (3) Compute the test using this series. (4) Repeat steps (1)-(3). This procedure preserves the normality of the errors and the autoregressive structure. The covariance function in the bootstrap world is thus in agreement with that in Proposition 1. As a result, the procedure is asymptotically valid.

Next, we consider the more general situation where a second variable is present among the regressors; for example, an autoregressive distributed lags (ADL) model. Because this model does not specify the joint distribution of the dependent variable and the regressors, the bootstrap procedure described above is no longer applicable. Two alternative approaches deserve consideration.

The first approach involves keeping the regressors fixed at their original values when generating the data; that is, we use the fixed regressor bootstrap. This procedure is asymptotically valid in the context of testing for structural breaks (Hansen, 2000). However, in the current context, it is, in general, inconsistent. In contrast to the original DGP, the regressors are strictly, but not weakly exogenous in the bootstrap world. As a result, the covariance function in the bootstrap world differs from that in Proposition 1. We now illustrate the potential severity of the size distortion using the setting in (22) with $T = 250$. The finite sample distribution and the bootstrap distribution for testing regime switching in the intercept are reported in Figure S22. Using the critical values from the fixed regressor bootstrap, the rejection rates at the 10% and 5% levels are 21.8% and 10.0%, respectively. The overrejection does not decrease when the sample size is increased to 500.

The second approach involves specifying the joint distribution of the data. For example, if we have an ADL model with normal errors, we specify a full model that is a Gaussian vector autoregression and apply the parametric bootstrap to the augmented model. This bootstrap procedure is consistent if it reproduces the covariance function in Proposition 1 asymptotically.

Information criteria. The asymptotic results imply that the performance of conventional information criteria, such as the BIC, can be sensitive to the structure of the model and to the choice of which parameters are allowed to switch. This is because the distribution of the likelihood ratio depends on which parameter is allowed to switch, whereas the penalty term in the BIC depends only on the dimension of the model and the sample size. We illustrate this sensitivity using the model in (22) for two cases. In the first case, we apply the BIC to determine whether there is regime switching in the intercept. The second case is the same as the first, except that the slope parameter is allowed to switch instead. In the simulated data, no regime switching is present; $\mu = 0, \alpha = 0.5$, and $\sigma^2 = 1$. The set Λ_ϵ is specified as (11) with $\epsilon = 0.05$. The sample size is 250. Of the 5000 realizations, the BIC falsely classifies 12.5% in the first case, while only 2.4% in the second case. Because the penalty terms in the Akaike information criterion and the Hannan–Quinn information criterion have the same structure, they are expected to exhibit the same sensitivity.

8 Monte Carlo

We examine the finite sample properties of the SupLR(Λ_ϵ) test, and compare these properties with those of the tests of Cho and White (2007) and Carrasco, Hu and Ploberger (2014). The DGP is

$$y_t = \mu_1 \cdot 1_{\{s_t=1\}} + \mu_2 \cdot 1_{\{s_t=2\}} + \alpha y_{t-1} + e_t \text{ with } e_t \sim i.i.d.\ N(0, \sigma^2), \tag{44}$$

where $P(s_t = 1|s_{t-1} = 1) = p$, $P(s_t = 2|s_{t-1} = 2) = q$, $\alpha = 0.5$, and $\sigma^2 = 1$. This DGP is considered in Cho and White (2007), and it is a sensible approximation to the postwar U.S. quarterly real GDP growth series, as shown in Section 9. Throughout this section, Λ_ϵ is given by (10) with $\epsilon = 0.05, 0.02$. For the supTS of Carrasco, Hu, and Ploberger (2014), the supremum is taken over $\rho \in [0.05, 0.90]$ or $\rho \in [0.02, 0.96]$. Because $\rho = p + q - 1$, these two sets for ρ are consistent with $\Lambda_{0.05}$ and $\Lambda_{0.02}$, specified above. The resulting tests are denoted by supTS_1 and supTS_2, respectively. The rejection frequencies are based on 5000 replications.

The results under the null hypothesis are reported in Table 2. The rejection frequencies of SupLR(Λ_ϵ) are overall close to the nominal levels, although some mild over-rejections do exist. When $T = 200$, the rejection rates at the 5% and 10% levels are 6.78% and 14.36% for $\epsilon = 0.05$, and 6.86% and 14.58% for $\epsilon = 0.02$. Similar rejection rates are observed when $T = 500$. The results confirm that the QLR and supTS tests exhibit excellent size properties.

For power properties, following Cho and White (2007), we let $\mu_1 = -\mu_2$ with $\mu_2 = 0.2, 0.6, 1.0$. Motivated by the estimates discussed in Section 3, we set (p, q) to $(0.70, 0.70)$, $(0.70, 0.90)$, and $(0.90, 0.90)$. The rejection rates at the 5% nominal level are reported in Table 3.

Because the alternatives are not mixtures, the power of the SupLR(Λ_ϵ) test is higher than that of the QLR. For example, when $(p, q) = (0.7, 0.7)$, the rejection rates of the SupLR($\Lambda_{0.05}$) test are 18.38% and 96.38% for $\mu_2 = 0.6$ and 1.0, and the corresponding values of the QLR are 9.46% and 68.83%. When $(p, q) = (0.9, 0.9)$, the values become 59.26% and 100% for SupLR($\Lambda_{0.05}$), and 7.06% and 7.30% for the QLR. Therefore, although the QLR test can be valuable for detecting mixtures, the SupLR(Λ_ϵ) test can offer substantial power gains when the regimes are dependent.

The power of SupLR(Λ_ϵ) is substantially higher than that of the supTS test. In addition to the explanation in Subsection 6.2, the power difference also arises from the following channel. Note that a key element of supTS is $\mu_{2,t}(\rho) = (1/(2\widetilde{\sigma}^4)) \sum_{s<t} \rho^{t-s} \widetilde{e}_t \widetilde{e}_s$, which measures the serial correlation in the residuals ($\widetilde{e}_t$) computed under the null hypothesis. When the parameters are estimated under the null hypothesis, the regime switching is removed from the model and forced into the residuals, which causes $\widetilde{e}_t$ to be *positively* serially correlated because $p + q - 1 > 0$. At the same time, the autoregressive coefficient $\widetilde{\alpha}$ is upward biased, and the bias is stronger when the data are more persistent. The bias in $\widetilde{\alpha}$ leads to overdifferencing the data and, consequently, making $\widetilde{e}_t$ *negatively* serially correlated. In finite samples, these two opposite effects can offset each other, making the value of $\mu_{2,t}(\rho)$ insensitive to the departure from the null hypothesis. (A similar phenomenon is studied in Perron (1990, 1991) in a structural change context.) This finding is consistent with the simulation results in Carrasco, Hu, and Ploberger (2014, Table II), which show that the test can have good power properties when a lagged dependent variable is not allowed in the model.

Next, we examine the situation where the DGP is a mixture model with $(p, q) = (0.5, 0.5)$. The results (the last five rows of Table 3) show that the power of the QLR test is higher than that of the SupLR(Λ_ϵ) test. However, the maximum difference is only 8.86% for the cases considered.

9 Applications

We first study the quarterly US real GDP growth rate, and then consider additional applications in the context of dynamic stochastic equilibrium models. The model in (44) is used for the analysis. The samples contain quarterly observations for the period 1960:I–2014:IV, unless it is stated otherwise. All tests are evaluated at the 5% nominal level. The relevant p-values are also reported.

9.1 US GDP growth

Following the influential work of Hamilton (1989), a large body of literature has modeled US real output growth as a regime switching process. Here, we apply the SupLR(Λ_ϵ) test to assess the empirical evidence for this specification.

Testing results. When applied to the full sample, the SupLR($\Lambda_{0.02}$) test is equal to 8.75, and the p-value is 0.033 (Table 4), implying that the null hypothesis is rejected at the 5% level. To exclude the influence of the Great Recession, we consider the subsample 1960:I–2006:IV. Now, the test is equal to 8.57, and the p-value is 0.035, implying the same conclusion. To take the analysis further, we use 1960:I–1980:I as the first subsample and incorporate observations quarter by quarter. This leads to 140 subsamples of increasing sizes; see Figure 2(a), in which the null hypothesis is rejected in 102 subsamples. Therefore, there is consistent evidence favoring the regime switching specification. The results of the QLR and supTS tests over the same subsamples are shown in Figures 2(b)-(c). The null hypothesis is rejected only after the Great Recession is included. Therefore, the empirical evidence for regime switching is substantially weaker when viewed through these two tests.

Recession probability. Figures 3(a) and 3(b) display the smoothed recession probabilities when the model in (44) is applied to 1960:I–2006:IV and 1960:I–2014:IV, respectively. The NBER business cycle indicators are also included, with the shaded areas corresponding to recessions. The results show that the model is informative. For the subsample, the recession probabilities implied by the model closely track the NBER's recession indicators. For the full sample, they are broadly similar, with the main difference being that the model assigns low recession probabilities to the relatively shallow recessions of 1969:IV-1970:IV and 2001:I-2001:IV. This difference arises because the estimated mean growth rate in recessions decreases from -0.18 to -0.67 when the Great Recession is included. Therefore, it reflects the unusual nature of the Great Recession. The parameter estimates are reported in Table 4.

Robustness check. In practice, the order of the autoregression under the null hypothesis is often determined by some information criterion. To reflect this practice, we estimate the lag order using the BIC for each subsample, and repeat the analysis. The minimum and maximum orders are set to 1 and 4, respectively. The null hypothesis is rejected for 92 of the 140 subsamples. Therefore, the evidence of regime switching remains considerable. We also repeat the analysis using reverse recursive subsamples. We let 1994:IV–2014:IV be the first subsample, and then incorporate additional observations backward quarter by quarter. The lag order is determined using the BIC for each subsample. The null hypothesis is rejected in 120 of the 140 subsamples. Finally, we exclude the Great Recession (i.e., we let 1986:IV-2006:IV be the first subsample), and then incorporate additional observations backward quarter by quarter. The null hypothesis is rejected in 47 out of the 108 subsamples. Therefore, although the evidence is weaker in this case, it remains considerable and fairly consistent across the subsamples.

Calibrated simulations. We generate data from the model in (44), using the empirical estimates in Table 4. The sample sizes are set to those implied by the subsample (1960:I–2006:IV) and the full sample (1960:I–2014:IV), respectively. Under the null hypothesis, in the subsample case, the rejection rates at the 5% level are 6.86% for $\text{SupLR}(\Lambda_{0.05})$ and 7.16% for $\text{SupLR}(\Lambda_{0.02})$, respectively, while in the full sample case, they are 6.90% and 7.14%. These values are consistent with those in Table 2. Under the alternative hypothesis, at the 5% level, the rejection rates of SupLR($\Lambda_{0.02}$) are 66% and 65% in the two cases, respectively. In contrast, the rejection rates of the QLR test are 14% and 25%, and that of the supTS$_2$ test are 24% and 10%. The power differences are substantial.

We also compute the p-value of the full-sample SupLR($\Lambda_{0.02}$) test using the parametric bootstrap in Section 7. The resulting value is 0.041, slightly above the value 0.033 reported in Table 4.

9.2 Other applications

Here, we consider three sets of applications in the context of dynamic stochastic equilibrium models. A description of the relevant data sets can be found in the footnote of Table 5.

Hours worked and capital utilization. Regime switching in real output growth has testable implications when viewed through the lens of medium scale DSGE models. In such models, the real output is usually modeled as a Cobb-Douglas function of capital stock K_t, capital utilization U_t, hours worked H_t, and some exogenous productivity process Z_t, that is, $Y_t = {Z_t}^{1-\alpha}(U_t K_{t-1})^{\alpha} H_t^{\beta}$; see Smets and Wouters (2007) and Schmitt-Grohé and Uribe (2012). In terms of growth rates (i.e., $y_t = \log(Y_t/Y_{t-1})$), this implies $y_t = \alpha\,(u_t + k_{t-1}) + \beta h_t + (1-\alpha)z_t$. Because of the linearity, at least one endogenous variable among u_t, k_{t-1}, and h_t must show regime switching. Otherwise, we arrive at a contradiction, implying that the production function is misspecified for the data under consideration, or the regime switching conclusion with regard to y_t should be revisited.

Motivated by this observation, we examine the regime switching hypothesis for u_t and h_t. The k_t series is not considered, because it is difficult to measure, and its official data are available only at the annual frequency. The SupLR($\Lambda_{0.02}$) test rejects the null hypothesis in both cases (see Table 5). The smoothed recession probabilities, see Figures S23-S24, are consistent with those in Figure 3 based on the GDP series. Overall, the results show internal consistency, and they support using the above production function as a building block of DSGE models.

Unemployment. The aggregate unemployment is a focal point of business cycle analysis. The empirical findings obtained thus far, particularly those related to hours worked, suggest that the dynamics of this variable may also exhibit regime switching. Table 5 shows that the null hypothesis is indeed rejected by the SupLR($\Lambda_{0.02}$) test. Furthermore, the estimated parameter values reveal an important asymmetry: the unemployment rate rises sharply at 1.03% per quarter in recessions, while it decreases slowly at only 0.10% per quarter in expansions. The asymmetry becomes invisible under the linear model in (44). Interestingly, the recession probabilities computed based on the change in the unemployment rate correlate closely with the NBER's indicators (Figure S25). Therefore, although the level of unemployment is a lagging indicator of the business cycle, the change of this variable should be viewed as a coincident indicator. This appears to be a new finding.

Consumption. Some dynamic stochastic equilibrium models have modeled the mean growth rate of aggregate consumption as a regime switching process in order to reproduce the observed properties of asset returns, including those related to the riskless rate and the equity premium; see Cecchetti, Lam, and Mark (1990), Kandel and Stambaugh (1991), and Ju and Miao (2012). However, partly because of the technical difficulty, formal statistical testing of this regime switching hypothesis is rare, particularly after allowing for linear dynamics. After applying the $\mathrm{SupLR}(\Lambda_{0.02})$ test, we find that the null hypothesis of no regime switching is rejected for the nondurable consumption expenditure series; see row D of Table 5. Also, the estimated recession probabilities (Figure S26) are consistent with those based on the other series. These findings provide formal statistical support for allowing for regime switching in aggregate consumption in these models.

Finally, we apply the supTS_2 and QLR tests to the same four time series. The SupTS_2 does not reject any null hypothesis. Its values in the four cases are (critical values in parentheses): 1.28 (2.65), 2.22 (2.64), 2.07 (2.64), and 2.02 (2.34). The QLR test rejects the null hypothesis only for the capacity utilization and the unemployment series. Its values are: 4.84 (6.18), 19.00 (6.34), 24.11 (6.28), and 6.06 (6.11). Therefore, similarly to the GDP case, the evidence of regime switching is weaker when viewed through these two tests.

10 Conclusion

We have examined a family of likelihood ratio based tests for detecting Markov regime switching. In addition to obtaining the limiting distribution under the null hypothesis and a finite sample refinement, thus resolving a long standing problem in the literature, we provide a unified algorithm for simulating the relevant critical values. Working with a simple DGP, we show analytically why these tests can be more powerful than some other tests that are based on alternative testing principles. When applied to the US real GDP growth data and four other time series in the context of dynamic stochastic equilibrium models, the proposed methods detect consistent evidence favoring the regime switching specification. We conjecture that the techniques and results presented here can have implications for hypothesis testing in other contexts, such as testing for Markov switching in state space models and multivariate regressions. Such investigations are currently in progress.

References

Andrews, D.W.K. (1993): " Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821-856.

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

—— (1995): "Admissibility of the Likelihood Ratio Test When a Nuisance Parameter is Present Only Under the Alternative," *The Annals of Statistics*, 23, 1609-1629.

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

—— (2002b): "Regime Switches in Interest Rates," *Journal of Business and Economic Statistics*, 20, 163-182.

Ang, A. and A. Timmermann: (2012): "Regime Changes and Financial Markets," *Annual Reviews of Financial Economics,* 4, 313–337.

Bai, J. and P. Perron (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47-78.

Bianchi, F. (2013): "Regime Switches, Agents' Beliefs, and Post-World War II U.S. Macroeconomic Dynamics," *Review of Economic Studies* 80, 463-490.

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

Carter, A.V. and D.G. Steigerwald (2012): "Testing for Regime Switching: A Comment," *Econometrica,* 80, 1809-1812.

Cecchetti, S.G., P. Lam, and N.C. Mark (1990): "Mean Reversion in Equilibrium Asset Prices," *American Economic Review*, 80, 398-418.

Chauvet, M. and J. Hamilton (2006): "Dating Business Cycle Turning Points," In *Nonlinear Time Series Analysis of Business Cycles*, (Milas, C., P. Rothman, and D. van Dijk, eds.), Elsevier.

Chen, X. (2013): "Penalized Sieve Estimation and Inference of Semi-nonparametric Dynamic Models:A Selective Review," In *Advances in Economics and Econometrics*, Cambridge University Press.

Chesher, A. (1984): "Testing for Neglected Heterogeneity," *Econometrica*, 22, 865–872.

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

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

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

—— (1987): "Hypothesis Testing when a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, 74, 33–43.

Davig, T. (2004): "Regime-Switching Debt and Taxation," *Journal of Monetary Economics*, 51, 837-859.

Garcia, R. (1998): "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models," *International Economic Review*, 39, 763–788.

Garcia, R., R. Luger and E. Renault (2003): "Empirical Assessment of an Intertemporal Option Pricing Model with Latent Variables," *Journal of Econometrics*, 116, 49-83.

Hamilton, J. (1988): "Rational-Expectations Econometric Analysis of Changes in Regime : An Investigation of the Term Structure of Interest Rates," *Journal of Economic Dynamics and Control*, 12, 385-423.

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

—— (1990): "Analysis of Time Series Subject to Changes in Regime," *Journal of Econometrics*, 45, 39-70.

—— (1996): "Specification Testing in Markov-Switching Time-Series Models," *Journal of Econometrics* 70, 127-157.

—— (2005): "What's Real About the Business Cycle?" *Federal Reserve Bank of St. Louis Review*, 435-452.

—— (2008): "Regime-Switching Models," In *New Palgrave Dictionary of Economics*, edited by S. Durlauf and L. Blume. London: Palgrave Macmillan.

—— (2016): "Macroeconomic Regimes and Regime Shifts," *Handbook of Macroeconomics*, Vol 2, 163–201.

Hansen, B. (1992): "The Likelihood Ratio Test under Non-Standard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, S61–S82.

—— (1996): "Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413-430.

—— (2000): "Testing for Structural Change in Conditional Models," *Journal of Econometrics,* 97, 93-115.

Hartigan, J.A. (1985): "A Failure of Likelihood Asymptotics for Normal Mixtures," In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer*, 2, 807-810.

Jeanne, O. and P. Masson (2000): "Currency Crises, Sunspots and Markov-Switching Regimes," *Journal of International Economics*, 50, 327-350.

Ju, N. and J. Miao (2012): " Ambiguity, Learning, and Asset Returns," *Econometrica*, 80, 559-591.

Kandel, S. and R.F. Stambaugh (1991): "Asset Returns and Intertemporal Preferences," *Journal of Monetary Economics*, 27, 39-71.

Kasahara, H. and K. Shimotsu (2012): "Testing the Number of Components in Finite Mixture Models," *Working paper*, University of British Columbia.

—— (2015): "Testing the Number of Components in Normal Mixture Regression Models," *Journal of the American Statistical Association* 110, 1632-1645.

King, M.L. and T.S. Shively (1993): "Locally Optimal Testing When a Nuisance Parameter is Present Only Under the Alternative," *Review of Economics and Statistics*, 75, 1-7.

Lindsay, B.G. (1995): *Mixture Models: Theory, Geometry and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics.

Lindé, J., F. Smets and R. Wouters. (2016): "Challenges for Central Banks' Macro Models," *Sveriges Riksbank Research Paper Series, no. 323.*

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

Liu, Z., D.F. Waggoner, and T. Zha (2011): "Sources of Macroeconomic Fluctuations: A Regime-Switching DSGE Approach", *Quantitative Economics*, 2, 251–301.

Mokkadem, A. (1988): "Mixing Properties of ARMA Processes," *Stochastic Processes and their Applications*, 29, 309-315.

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

Perron, P. (1990): "Testing for a Unit Root in a Time Series with a Changing Mean," *Journal of Business and Economic Statistics,* 8, 153-162.

—— (1991): "A Test for Changes in a Polynomial Trend Function For a Dynamic Time Series," Working paper, Boston University.

Rotnitzky, A., D.R. Cox, M. Bottai, and J. Robins (2000): "Likelihood-based Inference with Singular Information Matrix," *Bernoulli*, 6, 243-284.

Schmitt-Grohé, S. and M. Uribe (2012): "What's News in Business Cycles," *Econometrica*, 80, 2733-2764.

Schorfheide, F. (2005): "Learning and Monetary Policy Shifts," *Review of Economic Dynamics*, 8, 392–419.

Smets, F. and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586-606.

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

Table 1: Local power under a simple DGP

(a) Static model				
(p_*, q_*, c_*)	$(0.9, 0.2, 7.3)$	$(0.9, 0.4, 6.4)$	$(0.9, 0.6, 4.7)$	$(0.9, 0.8, 3.0)$
LR	55.14	53.00	53.53	53.15
QLR	53.69	37.86	11.93	5.79
TS	10.24	35.47	53.13	53.64
(b) Dynamic model				
(p_*, q_*, c_*)	$(0.9, 0.2, 7.3)$	$(0.9, 0.4, 6.4)$	$(0.9, 0.6, 4.7)$	$(0.9, 0.8, 3.0)$
LR	54.44	41.42	25.54	36.90
QLR	54.57	38.94	12.23	5.59
TS	5.73	13.01	27.01	36.26

Note. LR: the likelihood ratio test; QLR: Cho and White's (2007) test; TS: Carrasco, Hu, and Ploberger's (2014) test. The transition probabilities are set to (p_*, q_*) for all three tests to ensure a fair comparison. For the LR test, the refined approximation under the null hypothesis is obtained using the algorithm in Subsection 5.3. The refined approximation under the alternative hypothesis is computed in the same way, except that the noncentrality parameters $A_{2s}(p_*, q_*)$ (or $A_{2d}(p_*, q_*)$), $T^{-1/4}A_3(p_*, q_*)$, and $T^{-1/2}A_4(p_*, q_*)$ are added to the Gaussian random vectors $G_b\left(p_*, q_*\right), G_b^{(3)}\left(p_*, q_*\right)', G_b^{(4)}\left(p_*, q_*\right)$ before the optimization. The sample size is 500. The critical values and rejection frequencies are based on 10,000 realizations.

Table 2: Rejection frequencies under the null hypothesis

	Nominal size	2.50	5.00	7.50	10.00
T=200	SupLR($\Lambda_{0.05}$)	3.42	6.78	10.44	14.36
	SupLR($\Lambda_{0.02}$)	3.50	6.86	10.62	14.58
	QLR	2.43	5.30	7.50	10.00
	supTS$_1$	2.86	5.06	7.58	10.14
	supTS$_2$	2.84	5.06	7.62	10.06
T=500	SupLR($\Lambda_{0.05}$)	3.50	7.04	10.36	14.04
	SupLR($\Lambda_{0.02}$)	3.46	7.08	10.46	14.58
	QLR	2.33	5.43	7.53	10.20
	supTS$_1$	3.04	5.80	8.06	10.86
	supTS$_2$	3.02	5.72	8.12	10.70

Note. The values corresponding to the QLR test are taken from Table II in Cho and White (2007). The values corresponding to the supTS tests are obtained using the accompanying code of Carrasco, Hu and Ploberger (2014). The number of replications: 5000. In each replication, the critical values of the SupLR test are computed using the algorithm in Subsection 5.3 with 199 realizations.

Table 3: Rejection frequencies under the alternative hypothesis

(p, q)		$\mu_2 = 0.20$	$\mu_2 = 0.60$	$\mu_2 = 1.00$
(0.70,0.70)	SupLR($\Lambda_{0.05}$)	6.38	18.38	96.38
	SupLR($\Lambda_{0.02}$)	6.60	17.48	95.62
	QLR	6.16	9.46	68.83
	supTS$_1$	5.70	10.22	32.98
	supTS$_2$	5.50	10.10	32.70
(0.70,0.90)	SupLR($\Lambda_{0.05}$)	6.80	36.58	99.72
	SupLR($\Lambda_{0.02}$)	7.28	35.16	99.72
	QLR	6.14	13.40	60.56
	supTS$_1$	4.84	5.54	18.72
	supTS$_2$	4.82	5.46	18.46
(0.90,0.90)	SupLR($\Lambda_{0.05}$)	8.56	59.26	100.00
	SupLR($\Lambda_{0.02}$)	8.96	57.54	100.00
	QLR	5.76	7.06	7.30
	supTS$_1$	6.62	12.12	4.84
	supTS$_2$	6.56	12.16	4.84
(0.50,0.50)	SupLR($\Lambda_{0.05}$)	6.84	10.54	76.14
	SupLR($\Lambda_{0.02}$)	7.16	10.20	76.78
	QLR	6.03	11.33	85.10
	supTS$_1$	4.90	5.52	5.68
	supTS$_2$	4.88	5.46	5.50

Note. The values corresponding to the QLR test are taken from Table III in Cho and White (2007). Note that there the values in the rows of 0.1 and 0.9 in their table should be exchanged. The values related to the supTS tests are obtained using the accompanying code of Carrasco, Hu and Ploberger (2014). The number of replications: 5000. In each replication, the critical values of the SupLR test are computed using the algorithm in Subsection 5.3 with 199 realizations. Nominal level: 5%. Sample size: 500.

Table 4: Results for the GDP growth series

Tests	$\text{SupLR}(\Lambda_{0.02})$		5% critical value		p-value	
1960:I–2014:IV	8.75		7.62		0.033	
1960:I–2006:IV	8.57		7.61		0.035	
Estimates under H_0	μ		α		σ^2	
1960:I–2014:IV	0.51		0.33		0.64	
1960:I–2006:IV	0.60		0.28		0.65	
Estimates under H_1	μ_1	μ_2	α	σ^2	p	q
1960:I–2014:IV	−0.54	0.75	0.19	0.49	0.66	0.96
1960:I–2006:IV	−0.16	0.97	0.09	0.48	0.77	0.94

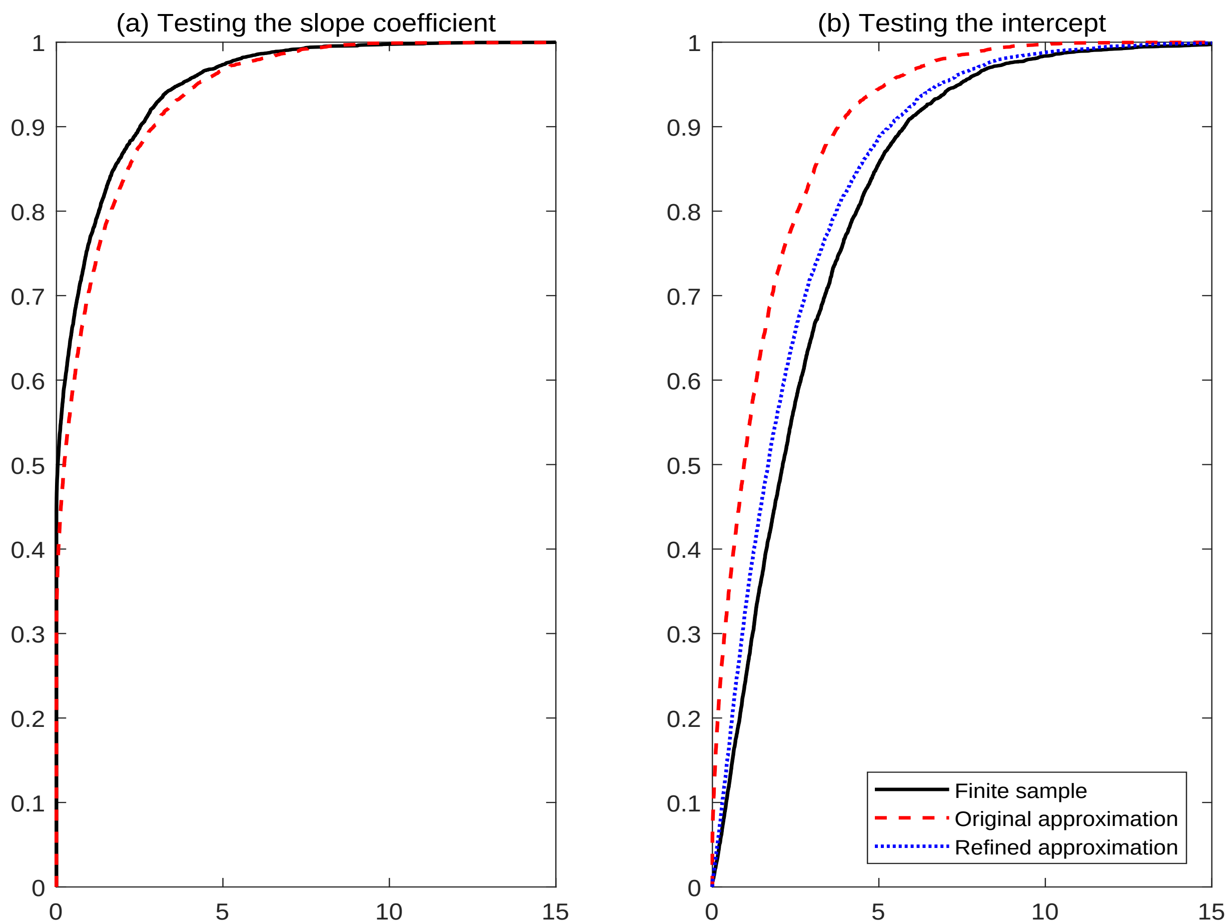
Note. The data series is GDPC1, retrieved from the St. Louis Fed website. The critical values of the SupLR test are obtained using the algorithm in Subsection 5.3 with 5000 realizations.

Table 5: Results for the other applications

Tests	$\text{SupLR}(\Lambda_{0.02})$		5% critical value		p-value	
Series A	12.86		7.80		0.009	
Series B	29.95		7.15		0.000	
Series C	32.12		7.96		0.000	
Series D	10.08		7.56		0.022	
Estimates under H_0	μ		α		σ^2	
Series A	0.12		0.63		0.40	
Series B	-0.01		0.58		1.54	
Series C	0.00		0.65		0.07	
Series D	0.27		0.20		0.50	
Estimates under H_1	μ_1	μ_2	α	σ^2	p	q
Series A	-1.31	0.21	0.52	0.32	0.51	0.98
Series B	-3.66	0.13	0.45	1.05	0.43	0.98
Series C	0.51	-0.06	0.44	0.04	0.68	0.96
Series D	-0.55	0.44	0.03	0.41	0.80	0.97

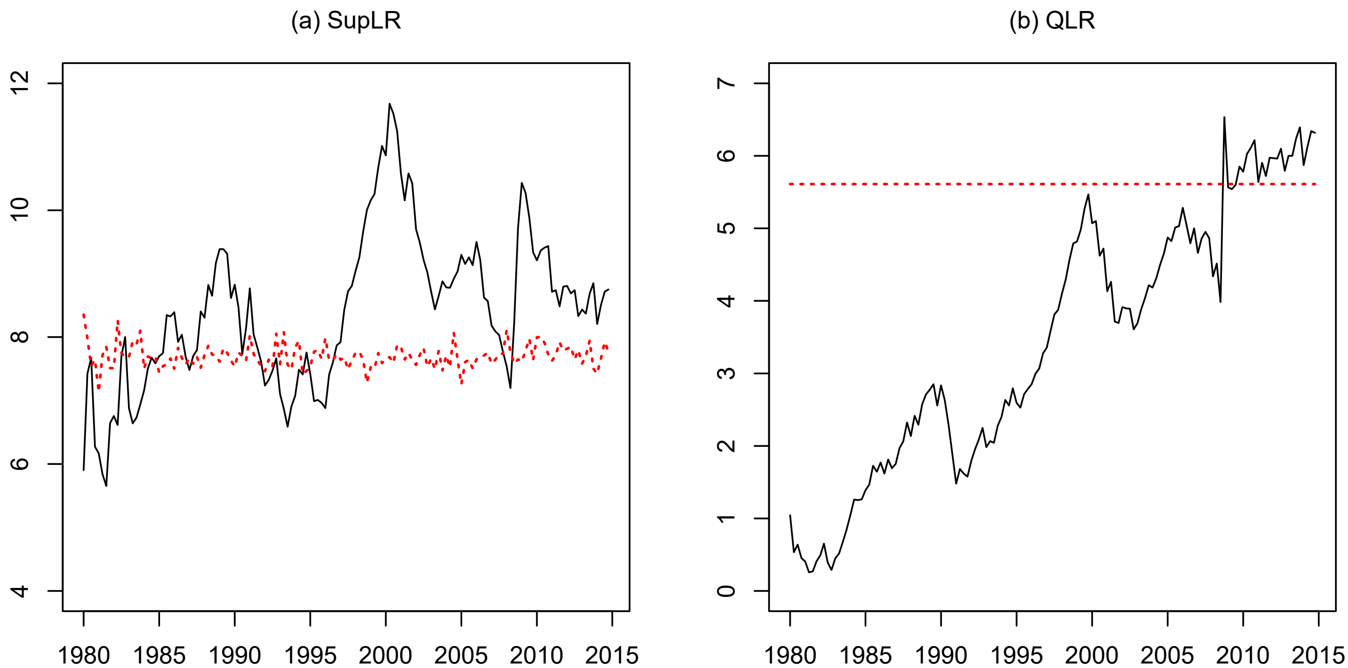
Note. Series A: hours worked (HOANBS), percent change, 1960:I-2014:IV. Series B: capacity utilization (TCU), percent change, 1967:IV-2014:IV. This sample is shorter because of data availability. Series C: unemployment (UNRATE), change, percent, 1960:I-2014:IV. Series D: Consumption (A796RX0Q048SBEA), percent change, 1960:I-2014:IV. All data series are at the quarterly frequency and are retrieved from the St. Louis Fed website. The critical values of the SupLR test are obtained using the algorithm in Subsection 5.3 with 5000 realizations.

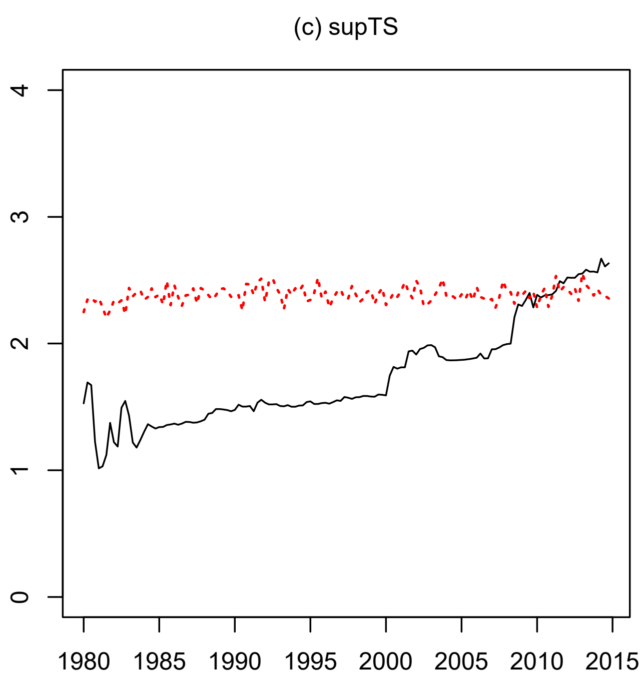
FIGURE 1. Distributions in an AR(1) model



Note. The figure displays the finite sample distributions of the SupLR(Λ_ϵ) test and their approximations for detecting regime switching in an AR(1) model: $y_t = \mu + \alpha y_{t-1} + u_t$ with $u_t \sim i.i.d.N(0, \sigma^2)$, where $\mu = 0, \alpha = 0.5, \sigma^2 = 1$, and $T = 250$. The original approximation and the refined approximation are given in Proposition 1 and Corollary 1, respectively. All results are based on 5000 replications.

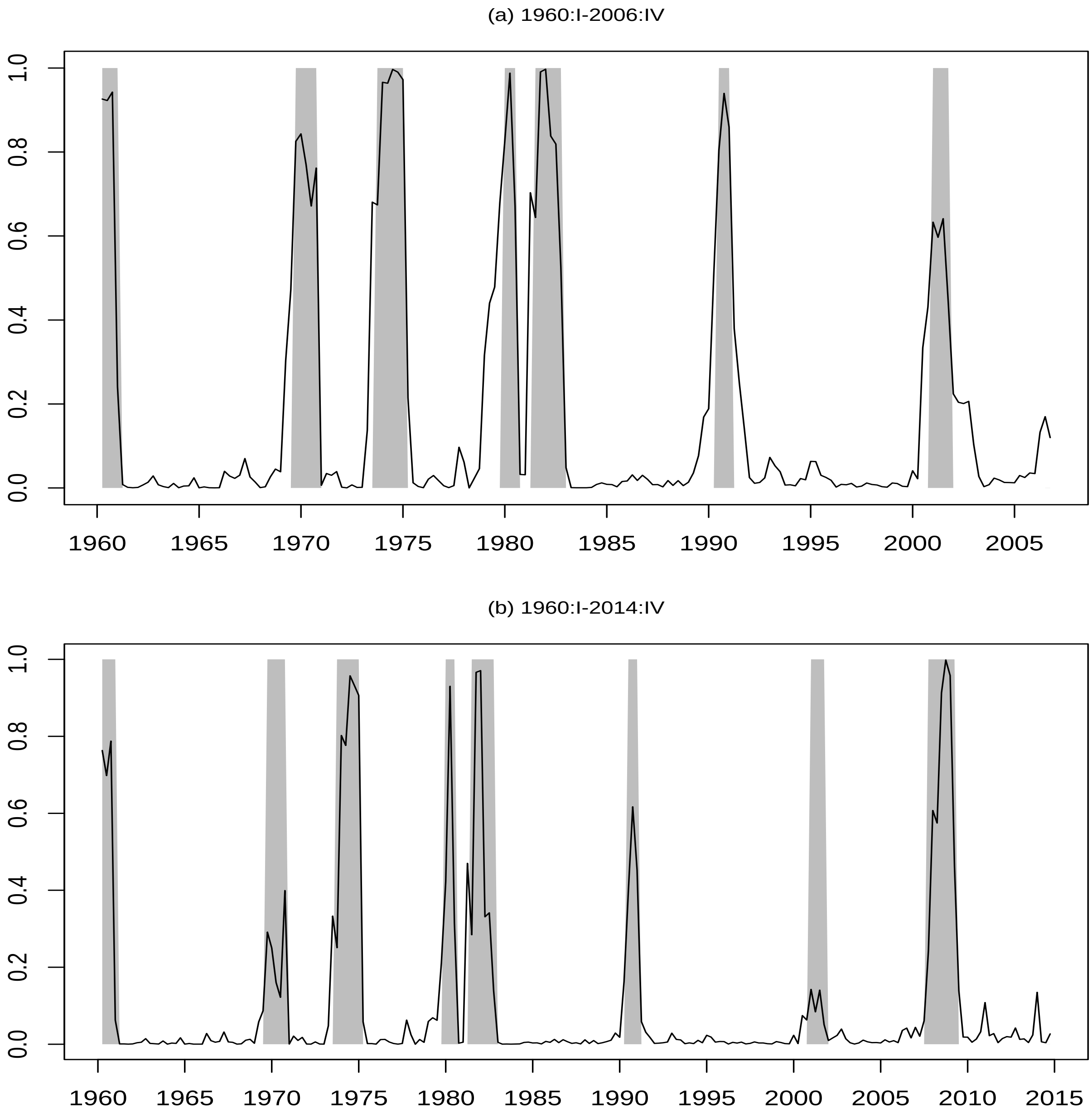
Figure 2. Test values over subsamples





Note. SupLR: the proposed test. QLR: the test of Cho and White (2007). supTS: the test of Carrasco, Hu and Ploberger (2014). The solid lines represent the test statistics and the dotted lines represent the 5% critical values.

Figure 3. Smoothed recession probabilities



Note. The solid lines represent the recession probabilities implied by the regime switching model. The shaded areas correspond to the NBER recessions.

Online Appendix: Proofs and Additional Results

The appendix is structured as follows. Section A.1 provides the proofs for the results in Section 4 of the paper. Section A.2 includes the proofs for Section 5, as well as some illustrations related to Assumptions 6 and 7, while Sections A.3 and A.4 present proofs for Subsections 6.1 and 6.2, respectively. Some figures appear at the end of the appendix. Because this appendix is long and sometimes technical, we include frequent remarks to highlight how the lemmas are related and their roles in the overall proof.

Throughout this appendix, $\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2)$, $f_t(\beta,\delta_1)$ and $f_t(\beta,\delta_2)$ are abbreviated as $\xi_{t|t-1}$, f_{1t} and f_{2t}, respectively. As stated prior to Lemma 1, for an expression A (e.g, $\xi_{t|t-1}$), "$\bar{A}$" denotes that it is evaluated at $\delta_1=\delta_2=\delta$, where δ is a generic value in Δ.

A.1 Proofs for Section 4

Proof of Lemma 1. The equation (12) can be written as

$$\xi_{t+1|t}=p+\rho\frac{A_t}{B_t}, \tag{A.1}$$

where ρ is as defined in the lemma, $A_t=f_{2t}(\xi_{t|t-1}-1)$ and $B_t=(f_{1t}-f_{2t})\xi_{t|t-1}+f_{2t}$.

Consider Lemma 1.1. Apply $\bar{f}_{1t}=\bar{f}_{2t}=\bar{f}_t$:

$$\bar{B}_t=\bar{f}_t \text{ and } \bar{A}_t=\bar{f}_t(\bar{\xi}_{t|t-1}-1). \tag{A.2}$$

Plugging this into (A.1), we obtain $\bar{\xi}_{t+1|t}=p+\rho(\bar{\xi}_{t|t-1}-1)$. This implies $\bar{\xi}_{2|1}=p+\rho(\bar{\xi}_{1|0}-1)=p+\rho\,(\xi_*-1)=\xi_*$, where the last equality follows from the definition of ρ and ξ_*. This process can be iterated forward, leading to $\bar{\xi}_{t+1|t}=\xi_*$ for all $t\geq 1$.

Consider Lemma 1.2. Differentiate (A.1) with respect to θ_j ($j=1,...,n_\beta+2n_\delta$):

$$\nabla_{\theta_j}\xi_{t+1|t}=\rho\left\{\frac{\nabla_{\theta_j}A_t}{B_t}-\frac{A_t\nabla_{\theta_j}B_t}{B_t^2}\right\}, \tag{A.3}$$

where

$$\begin{aligned}\nabla_{\theta_j}A_t &= \nabla_{\theta_j}f_{2t}(\xi_{t|t-1}-1)+f_{2t}\nabla_{\theta_j}\xi_{t|t-1},\\ \nabla_{\theta_j}B_t &= (\nabla_{\theta_j}f_{1t}-\nabla_{\theta_j}f_{2t})\xi_{t|t-1}+(f_{1t}-f_{2t})\nabla_{\theta_j}\xi_{t|t-1}+\nabla_{\theta_j}f_{2t}.\end{aligned}$$

Below, we evaluate the right hand side of (A.3) under three possible situations:

(1). If $j\in I_0$, then $\nabla_{\theta_j}\bar{f}_{1t}=\nabla_{\theta_j}\bar{f}_{2t}$ and $\bar{f}_{1t}=\bar{f}_{2t}=\bar{f}_t$, implying

$$\nabla_{\theta_j}\bar{A}_t=(\bar{\xi}_{t|t-1}-1)\nabla_{\theta_j}\bar{f}_{2t}+\bar{f}_t\nabla_{\theta_j}\bar{\xi}_{t|t-1} \text{ and } \nabla_{\theta_j}\bar{B}_t=\nabla_{\theta_j}\bar{f}_{2t} \tag{A.4}$$

Combining this with (A.2), we obtain $\nabla_{\theta_j}\bar{\xi}_{t+1|t}=\rho\nabla_{\theta_j}\bar{\xi}_{t|t-1}$. In particular, at $t=1$: $\nabla_{\theta_j}\bar{\xi}_{2|1}=\rho\nabla_{\theta_j}\bar{\xi}_{1|0}=\rho\nabla_{\theta_j}\xi_*=0$, where the last equality holds because ξ_* is independent of θ. This process can be iterated forward, leading to $\nabla_{\theta_j}\bar{\xi}_{t+1|t}=0$ for all $t\geq 1$.

(2). If $j\in I_1$, then $\nabla_{\theta_j}\bar{f}_{2t}=0$ and $\bar{f}_{1t}=\bar{f}_{2t}=\bar{f}_t$, implying

$$\nabla_{\theta_j}\bar{A}_t=\bar{f}_t\nabla_{\theta_j}\bar{\xi}_{t|t-1} \text{ and } \nabla_{\theta_j}\bar{B}_t=\bar{\xi}_{t|t-1}\nabla_{\theta_j}\bar{f}_{1t}. \tag{A.5}$$

Combining this with (A.2), we have $\nabla_{\theta_j}\bar{\xi}_{t+1|t}=\rho\nabla_{\theta_j}\bar{\xi}_{t|t-1}-\rho(\bar{\xi}_{t|t-1}-1)\bar{\xi}_{t|t-1}\nabla_{\theta_j}\log\bar{f}_{1t}=\rho\nabla_{\theta_j}\bar{\xi}_{t|t-1}-\rho\,(\xi_*-1)\,\xi_*\nabla_{\theta_j}\log\bar{f}_{1t}$. The result then follows because $r=\rho\,(1-\xi_*)\,\xi_*$. Note that $\nabla_{\theta_j}\bar{\xi}_{t+1|t}$ can also be written as

$$\nabla_{\theta_j}\bar{\xi}_{t+1|t}=r\sum_{s=0}^{t-1}\rho^s\nabla_{\theta_j}\log\bar{f}_{1(t-s)} \tag{A.6}$$

(3). If $j \in I_2$, then $\nabla_{\theta_j} \bar{f}_{1t} = 0$ and $\bar{f}_{1t} = \bar{f}_{2t} = \bar{f}_t$, implying

$$\nabla_{\theta_j} \bar{A}_t = \nabla_{\theta_j} \bar{f}_{2t}(\bar{\xi}_{t|t-1} - 1) + \bar{f}_t \nabla_{\theta_j} \bar{\xi}_{t|t-1} \quad \text{and} \quad \nabla_{\theta_j} \bar{B}_t = (1 - \bar{\xi}_{t|t-1}) \nabla_{\theta_j} \bar{f}_{2t}. \tag{A.7}$$

Therefore, $\nabla_{\theta_j} \bar{\xi}_{t+1|t} = \rho \nabla_{\theta_j} \bar{\xi}_{t|t-1} + \rho(\bar{\xi}_{t|t-1} - 1)\bar{\xi}_{t|t-1} \nabla_{\theta_j} \log \bar{f}_{2t} = -r \sum_{s=0}^{t-1} \rho^s \nabla_{\theta_j} \log \bar{f}_{2(t-s)}$. Because $\nabla_{\theta_j} \bar{f}_{2(t-s)} = \nabla_{\theta_{j-n_\delta}} \bar{f}_{1(t-s)}$ when $j \in I_2$, $\nabla_{\theta_j} \bar{\xi}_{t+1|t}$ equals the negative of (A.6).

Consider Lemma 1.3. Differentiating (A.3) with respect to θ_k:

$$\nabla_{\theta_j} \nabla_{\theta_k} \xi_{t+1|t} = \rho \left\{ \frac{\nabla_{\theta_j} \nabla_{\theta_k} A_t}{B_t} - \frac{\nabla_{\theta_j} A_t \nabla_{\theta_k} B_t}{B_t^2} - \frac{\nabla_{\theta_k} A_t \nabla_{\theta_j} B_t}{B_t^2} - \frac{A_t \nabla_{\theta_j} \nabla_{\theta_k} B_t}{B_t^2} + 2 \frac{A_t \nabla_{\theta_j} B_t \nabla_{\theta_k} B_t}{B_t^3} \right\}, \tag{A.8}$$

where

$$\begin{aligned}
\nabla_{\theta_j} \nabla_{\theta_k} A_t &= \nabla_{\theta_j} \nabla_{\theta_k} f_{2t}(\xi_{t|t-1} - 1) + \nabla_{\theta_j} f_{2t} \nabla_{\theta_k} \xi_{t|t-1} + \nabla_{\theta_k} f_{2t} \nabla_{\theta_j} \xi_{t|t-1} + f_{2t} \nabla_{\theta_j} \nabla_{\theta_k} \xi_{t|t-1}, \\
\nabla_{\theta_j} \nabla_{\theta_k} B_t &= (\nabla_{\theta_j} \nabla_{\theta_k} f_{1t} - \nabla_{\theta_j} \nabla_{\theta_k} f_{2t}) \xi_{t|t-1} + (\nabla_{\theta_j} f_{1t} - \nabla_{\theta_j} f_{2t}) \nabla_{\theta_k} \xi_{t|t-1} \\
&\quad + (\nabla_{\theta_k} f_{1t} - \nabla_{\theta_k} f_{2t}) \nabla_{\theta_j} \xi_{t|t-1} + (f_{1t} - f_{2t}) \nabla_{\theta_j} \nabla_{\theta_k} \xi_{t|t-1} + \nabla_{\theta_j} \nabla_{\theta_k} f_{2t}.
\end{aligned}$$

We now evaluate the right hand side of (A.8) at $\delta_1 = \delta_2 = \delta$ under six possible situations:

(**1**). If $j \in I_0$ and $k \in I_0$, then $\bar{f}_{1t} = \bar{f}_{2t} = \bar{f}_t$, $\nabla_{\theta_j} \bar{f}_{1t} = \nabla_{\theta_j} \bar{f}_{2t}, \nabla_{\theta_k} \bar{f}_{1t} = \nabla_{\theta_k} \bar{f}_{2t}, \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{1t} = \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t}$ and $\nabla_{\theta_j} \bar{\xi}_{t+1|t} = \nabla_{\theta_k} \bar{\xi}_{t+1|t} = 0$, implying $\nabla_{\theta_j} \nabla_{\theta_k} \bar{A}_t = \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t}(\bar{\xi}_{t|t-1} - 1) + \bar{f}_t \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1}$ and $\nabla_{\theta_j} \nabla_{\theta_k} \bar{B}_t = \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t}$. Combining them with (A.2) and (A.4), $\nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t+1|t}$ equals

$$\begin{aligned}
&\rho \left\{ \frac{\nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t}(\bar{\xi}_{t|t-1} - 1) + \bar{f}_t \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1}}{\bar{f}_t} - \frac{(\bar{\xi}_{t|t-1} - 1) \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t^2} \right. \\
&\left. - \frac{(\bar{\xi}_{t|t-1} - 1) \nabla_{\theta_k} \bar{f}_{2t} \nabla_{\theta_j} \bar{f}_{2t}}{\bar{f}_t^2} - \frac{(\bar{\xi}_{t|t-1} - 1) \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t} + 2 \frac{(\bar{\xi}_{t|t-1} - 1) \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t^2} \right\} = \rho \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1}.
\end{aligned}$$

Starting at $t = 1$ and iterating forward, we have $\nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t+1|t} = 0$ for all $t \geq 1$.

The proof for the remaining five cases uses similar arguments; we only outline the main steps.

(2). If $j \in I_0$ and $k \in I_1$, then $\nabla_{\theta_j} \bar{f}_{1t} = \nabla_{\theta_j} \bar{f}_{2t}, \nabla_{\theta_k} \bar{f}_{2t} = \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t} = \nabla_{\theta_j} \bar{\xi}_{t+1|t} = 0$, implying $\nabla_{\theta_j} \nabla_{\theta_k} \bar{A}_t = \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{\xi}_{t|t-1} + \bar{f}_t \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1}$ and $\nabla_{\theta_j} \nabla_{\theta_k} \bar{B}_t = \bar{\xi}_{t|t-1} \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{1t}$. Combining these two equations with (A.2), (A.4) and (A.5), $\nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t+1|t}$ equals

$$\begin{aligned}
&\rho \left\{ \frac{\nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{\xi}_{t|t-1} + \bar{f}_t \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1}}{\bar{f}_t} - \frac{\nabla_{\theta_j} \bar{f}_{2t} (\bar{\xi}_{t|t-1} - 1) \bar{\xi}_{t|t-1} \nabla_{\theta_k} \bar{f}_{1t}}{\bar{f}_t^2} - \frac{\nabla_{\theta_k} \bar{\xi}_{t|t-1} \nabla_{\theta_j} \bar{f}_{2t}}{\bar{f}_t} \right. \\
&\left. - \frac{(\bar{\xi}_{t|t-1} - 1) \bar{\xi}_{t|t-1} \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{1t}}{\bar{f}_t} + \frac{2(\bar{\xi}_{t|t-1} - 1) \bar{\xi}_{t|t-1} \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{f}_{1t}}{\bar{f}_t^2} \right\}
\end{aligned}$$

The result follows from rearranging the terms.

(3). If $j \in I_0$ and $k \in I_2$, then $\nabla_{\theta_j} \bar{f}_{1t} = \nabla_{\theta_j} \bar{f}_{2t}$ and $\nabla_{\theta_k} \bar{f}_{1t} = \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{1t} = \nabla_{\theta_j} \bar{\xi}_{t+1|t} = 0$, implying $\nabla_{\theta_j} \nabla_{\theta_k} \bar{A}_t = \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t}(\bar{\xi}_{t|t-1} - 1) + \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{\xi}_{t|t-1} + \bar{f}_t \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1}$ and $\nabla_{\theta_j} \nabla_{\theta_k} \bar{B}_t = (1 - \bar{\xi}_{t|t-1}) \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t}$. Combining these results with (A.2), (A.4) and (A.7), $\nabla_{\theta_j} \nabla_{\theta_k} \xi_{t+1|t}$ equals

$$\begin{aligned}
&\left\{ \frac{\nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t}(\bar{\xi}_{t|t-1} - 1) + \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{\xi}_{t|t-1} + \bar{f}_t \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1}}{\bar{f}_t} + \frac{(\bar{\xi}_{t|t-1} - 1)^2 \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t^2} \right. \\
&\left. - \frac{[\nabla_{\theta_k} \bar{f}_{2t}(\bar{\xi}_{t|t-1} - 1) + \bar{f}_t \nabla_{\theta_k} \bar{\xi}_{t|t-1}] \nabla_{\theta_j} \bar{f}_{2t}}{\bar{f}_t^2} + \frac{(\bar{\xi}_{t|t-1} - 1)^2 \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t} - \frac{2(\bar{\xi}_{t|t-1} - 1)^2 \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t^2} \right\}.
\end{aligned}$$

The result follows from rearranging the terms.

(4). If $j \in I_1$ and $k \in I_1$, then $\nabla_{\theta_j} \bar{f}_{2t} = \nabla_{\theta_k} \bar{f}_{2t} = \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t} = 0$, implying $\nabla_{\theta_j} \nabla_{\theta_k} \bar{A}_t = \bar{f}_t \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1}$ and $\nabla_{\theta_j} \nabla_{\theta_k} \bar{B}_t = \bar{\xi}_{t|t-1} \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{1t} + \nabla_{\theta_j} \bar{f}_{1t} \nabla_{\theta_k} \bar{\xi}_{t|t-1} + \nabla_{\theta_k} \bar{f}_{1t} \nabla_{\theta_j} \bar{\xi}_{t|t-1}$. Combining them with (A.2), (A.5), $\nabla_{\theta_j} \nabla_{\theta_k} \xi_{t+1|t}$ equals

$$\rho \Bigg\{ \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1} - \frac{\bar{\xi}_{t|t-1} \nabla_{\theta_j} \bar{\xi}_{t|t-1} \nabla_{\theta_k} \bar{f}_{1t}}{\bar{f}_t} - \frac{\bar{\xi}_{t|t-1} \nabla_{\theta_k} \bar{\xi}_{t|t-1} \nabla_{\theta_j} \bar{f}_{1t}}{\bar{f}_t}$$
$$- \frac{(\bar{\xi}_{t|t-1} - 1)[\bar{\xi}_{t|t-1} \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{1t} + \nabla_{\theta_j} \bar{f}_{1t} \nabla_{\theta_k} \bar{\xi}_{t|t-1} + \nabla_{\theta_j} \bar{\xi}_{t|t-1} \nabla_{\theta_k} \bar{f}_{1t}]}{\bar{f}_t} + \frac{2(\bar{\xi}_{t|t-1} - 1) \bar{\xi}_{t|t-1}^2 \nabla_{\theta_j} \bar{f}_{1t} \nabla_{\theta_k} \bar{f}_{1t}}{\bar{f}_t^2} \Bigg\}.$$

The result follows from rearranging the right hand side terms.

(5). If $j \in I_1$ and $k \in I_2$, then $\nabla_{\theta_j} \bar{f}_{2t} = \nabla_{\theta_k} \bar{f}_{1t} = \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{1t} = \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t} = 0$, implying $\nabla_{\theta_j} \nabla_{\theta_k} \bar{A}_t = \nabla_{\theta_k} \bar{f}_{2t} \nabla_{\theta_j} \bar{\xi}_{t|t-1} + \bar{f}_t \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1}$ and $\nabla_{\theta_j} \nabla_{\theta_k} \bar{B}_t = \nabla_{\theta_j} f_{1t} \nabla_{\theta_k} \bar{\xi}_{t|t-1} - \nabla_{\theta_k} \bar{f}_{2t} \nabla_{\theta_j} \bar{\xi}_{t|t-1}$. Combining them with (A.2), (A.5) and (A.7), $\nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t+1|t}$ equals

$$\rho \Bigg\{ \frac{\nabla_{\theta_k} \bar{f}_{2t} \nabla_{\theta_j} \bar{\xi}_{t|t-1} + \bar{f}_t \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1}}{\bar{f}_t} - \frac{(1 - \bar{\xi}_{t|t-1}) \nabla_{\theta_j} \bar{\xi}_{t|t-1} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t} - \frac{\left[\nabla_{\theta_k} \bar{f}_{2t} (\bar{\xi}_{t|t-1} - 1) + \bar{f}_t \nabla_{\theta_k} \bar{\xi}_{t|t-1}\right] \bar{\xi}_{t|t-1} \nabla_{\theta_j} \bar{f}_{1t}}{\bar{f}_t^2}$$
$$- \frac{(\bar{\xi}_{t|t-1} - 1) \left[\nabla_{\theta_j} f_{1t} \nabla_{\theta_k} \bar{\xi}_{t|t-1} - \nabla_{\theta_k} \bar{f}_{2t} \nabla_{\theta_j} \bar{\xi}_{t|t-1}\right]}{\bar{f}_t} - \frac{2(\bar{\xi}_{t|t-1} - 1)^2 \bar{\xi}_{t|t-1} \nabla_{\theta_j} \bar{f}_{1t} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t^2} \Bigg\}.$$

The result follows from rearranging the terms.

(**6**). If $j \in I_2$ and $k \in I_2$, then $\nabla_{\theta_j} \bar{f}_{1t} = \nabla_{\theta_k} \bar{f}_{1t} = \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{1t} = 0$, implying $\nabla_{\theta_j} \nabla_{\theta_k} \bar{A}_t = \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t} (\bar{\xi}_{t|t-1} - 1) + \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{\xi}_{t|t-1} + \nabla_{\theta_k} \bar{f}_{2t} \nabla_{\theta_j} \bar{\xi}_{t|t-1} + \bar{f}_t \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1}$ and $\nabla_{\theta_j} \nabla_{\theta_k} \bar{B}_t = (1 - \bar{\xi}_{t|t-1}) \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t} - \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{\xi}_{t|t-1} - \nabla_{\theta_k} \bar{f}_{2t} \nabla_{\theta_j} \bar{\xi}_{t|t-1}$. Combining them with (A.2) and (A.7), $\nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t+1|t}$ equals

$$\rho \Bigg\{ \frac{\nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t} (\bar{\xi}_{t|t-1} - 1) + \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{\xi}_{t|t-1} + \nabla_{\theta_k} \bar{f}_{2t} \nabla_{\theta_j} \bar{\xi}_{t|t-1} + \bar{f}_t \nabla_{\theta_j} \nabla_{\theta_k} \bar{\xi}_{t|t-1}}{\bar{f}_t}$$
$$- \frac{\left[\nabla_{\theta_k} \bar{f}_{2t} (\bar{\xi}_{t|t-1} - 1) + \bar{f}_t \nabla_{\theta_k} \bar{\xi}_{t|t-1}\right] (1 - \bar{\xi}_{t|t-1}) \nabla_{\theta_j} \bar{f}_{2t}}{\bar{f}_t^2} - \frac{\left[\nabla_{\theta_j} \bar{f}_{2t} (\bar{\xi}_{t|t-1} - 1) + \bar{f}_t \nabla_{\theta_j} \bar{\xi}_{t|t-1}\right] (1 - \bar{\xi}_{t|t-1}) \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t^2}$$
$$- \frac{(\bar{\xi}_{t|t-1} - 1)[(1 - \bar{\xi}_{t|t-1}) \nabla_{\theta_j} \nabla_{\theta_k} \bar{f}_{2t} - \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{\xi}_{t|t-1} - \nabla_{\theta_k} \bar{f}_{2t} \nabla_{\theta_j} \bar{\xi}_{t|t-1}]}{\bar{f}_t} + 2 \frac{(\bar{\xi}_{t|t-1} - 1)^3 \nabla_{\theta_j} \bar{f}_{2t} \nabla_{\theta_k} \bar{f}_{2t}}{\bar{f}_t^2} \Bigg\}.$$

The result follows from rearranging the terms.

Consider Lemma 1.4. Differentiating (A.8) with respect to θ_l:

$$\begin{aligned}
&\nabla_{\theta_j} \nabla_{\theta_k} \nabla_{\theta_l} \xi_{t+1|t} \\
&= \rho \Bigg\{ \frac{\nabla_{\theta_j} \nabla_{\theta_k} \nabla_{\theta_l} A_t}{B_t} - \frac{\nabla_{\theta_j} \nabla_{\theta_k} A_t \nabla_{\theta_l} B_t}{B_t^2} - \frac{\nabla_{\theta_j} \nabla_{\theta_l} A_t \nabla_{\theta_k} B_t}{B_t^2} - \frac{\nabla_{\theta_j} A_t \nabla_{\theta_k} \nabla_{\theta_l} B_t}{B_t^2} + \frac{2 \nabla_{\theta_j} A_t \nabla_{\theta_k} B_t \nabla_{\theta_l} B_t}{B_t^3} \\
&- \frac{\nabla_{\theta_k} \nabla_{\theta_l} A_t \nabla_{\theta_j} B_t}{B_t^2} - \frac{\nabla_{\theta_k} A_t \nabla_{\theta_j} \nabla_{\theta_l} B_t}{B_t^2} + \frac{2 \nabla_{\theta_k} A_t \nabla_{\theta_j} B_t \nabla_{\theta_l} B_t}{B_t^3} \\
&- \frac{\nabla_{\theta_l} A_t \nabla_{\theta_j} \nabla_{\theta_k} B_t}{B_t^2} - \frac{A_t \nabla_{\theta_j} \nabla_{\theta_k} \nabla_{\theta_l} B_t}{B_t^2} + \frac{2 A_t \nabla_{\theta_j} \nabla_{\theta_k} B_t \nabla_{\theta_l} B_t}{B_t^3} \\
&+ \frac{2 \nabla_{\theta_l} A_t \nabla_{\theta_j} B_t \nabla_{\theta_k} B_t}{B_t^3} + \frac{2 A_t \nabla_{\theta_j} \nabla_{\theta_l} B_t \nabla_{\theta_k} B_t}{B_t^3} + \frac{2 A_t \nabla_{\theta_j} B_t \nabla_{\theta_k} \nabla_{\theta_l} B_t}{B_t^3} - \frac{6 A_t \nabla_{\theta_j} B_t \nabla_{\theta_k} B_t \nabla_{\theta_l} B_t}{B_t^4} \Bigg\},
\end{aligned} \tag{A.9}$$

where

$$\begin{aligned}
\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}A_t &= \nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}f_{2t}(\xi_{t|t-1}-1)+\nabla_{\theta_j}\nabla_{\theta_l}f_{2t}\nabla_{\theta_k}\xi_{t|t-1}+\nabla_{\theta_k}\nabla_{\theta_l}f_{2t}\nabla_{\theta_j}\xi_{t|t-1} \\
&\quad +\nabla_{\theta_l}f_{2t}\nabla_{\theta_j}\nabla_{\theta_k}\xi_{t|t-1}+\nabla_{\theta_j}\nabla_{\theta_k}f_{2t}\nabla_{\theta_l}\xi_{t|t-1}+\nabla_{\theta_j}f_{2t}\nabla_{\theta_k}\nabla_{\theta_l}\xi_{t|t-1} \\
&\quad +\nabla_{\theta_k}f_{2t}\nabla_{\theta_j}\nabla_{\theta_l}\xi_{t|t-1}+f_{2t}\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\xi_{t|t-1}, \\
\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}B_t &= (\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}f_{1t}-\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}f_{2t})\xi_{t|t-1}+(\nabla_{\theta_j}\nabla_{\theta_l}f_{1t}-\nabla_{\theta_j}\nabla_{\theta_l}f_{2t})\nabla_{\theta_k}\xi_{t|t-1} \\
&\quad +(\nabla_{\theta_k}\nabla_{\theta_l}f_{1t}-\nabla_{\theta_k}\nabla_{\theta_l}f_{2t})\nabla_{\theta_j}\xi_{t|t-1}+(\nabla_{\theta_l}f_{1t}-\nabla_{\theta_l}f_{2t})\nabla_{\theta_j}\nabla_{\theta_k}\xi_{t|t-1} \\
&\quad +\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}f_{2t}+(\nabla_{\theta_j}\nabla_{\theta_k}f_{1t}-\nabla_{\theta_j}\nabla_{\theta_k}f_{2t})\nabla_{\theta_l}\xi_{t|t-1} \\
&\quad +(\nabla_{\theta_j}f_{1t}-\nabla_{\theta_j}f_{2t})\nabla_{\theta_k}\nabla_{\theta_l}\xi_{t|t-1}+(\nabla_{\theta_k}f_{1t}-\nabla_{\theta_k}f_{2t})\nabla_{\theta_j}\nabla_{\theta_l}\xi_{t|t-1} \\
&\quad +(f_{1t}-f_{2t})\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\xi_{t|t-1}
\end{aligned}$$

We now evaluate the above terms at $\delta_1=\delta_2=\delta$ for 10 possible cases. We only report the values of $\bar{\mathcal{E}}_{jkl,t}$ but omit the derivation details.

(**1**). If $j\in I_0, k\in I_0$ and $l\in I_0$, then $\bar{\mathcal{E}}_{jkl,t}=0$.

(**2**). If $j\in I_0, k\in I_0$ and $l\in I_1$, then $\bar{\mathcal{E}}_{jkl,t}$ equals

$$r\left\{-\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^2}-\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^2}-\frac{\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^2}+\frac{\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t}+2\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^3}\right\}.$$

(**3**). If $j\in I_0, k\in I_0$ and $l\in I_2$, then $\bar{\mathcal{E}}_{jkl,t}$ equals

$$r\left\{\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^2}+\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^2}+\frac{\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^2}-\frac{\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t}-2\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^3}\right\}.$$

(**4**). If $j\in I_0, k\in I_1$ and $l\in I_1$, then $\bar{\mathcal{E}}_{jkl,t}$ equals

$$\begin{aligned}
&\rho(1-2\xi_*)\left[\frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_j}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}\right. \\
&\left.-\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}-\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}\right]+r\left[\frac{\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t}-\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^2}\right] \\
&-2r\xi_*\left[\frac{\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t^2}+\frac{\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^2}-2\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^3}\right]
\end{aligned}$$

(**5**). If $j\in I_0, k\in I_1$ and $l\in I_2$, then $\bar{\mathcal{E}}_{jkl,t}$ equals

$$\begin{aligned}
&\rho(1-2\xi_*)\left[\frac{\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_j}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}-\frac{\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}\right. \\
&\left.-\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}+\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\xi_{t|t-1}}{\bar{f}_t}-\frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}\right] \\
&-r(1-2\xi_*)\left[\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^2}+\frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t^2}-2\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^3}\right].
\end{aligned}$$

(**6**). If $j\in I_0, k\in I_2$ and $l\in I_2$, then $\bar{\mathcal{E}}_{jkl,t}$ equals

$$\begin{aligned}
&-\rho(1-2\xi_*)\left[\frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_j}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}\right. \\
&\left.-\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}-\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}\right]-r\left[\frac{\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t}-\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t}\right]. \\
&+2r(1-\xi_*)\left[\frac{\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t^2}+\frac{\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^2}-2\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^3}\right]
\end{aligned}$$

(**7**). If $j \in I_1, k \in I_1$ and $l \in I_1$, then $\bar{\mathcal{E}}_{jkl,t}$ equals

$$\rho(1-2\xi_*)\left[\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}}{\bar{f}_t}\right.$$
$$\left.+\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_j}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}\right]$$
$$-2\rho\left[\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}\right]$$
$$+\rho(6\xi_*^2-4\xi_*)\left[\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}+\frac{\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}+\frac{\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}\right]$$
$$-2r\xi_*\left[\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^2}+\frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t^2}+\frac{\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t}\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t^2}\right]+r\frac{\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t}+6r\xi_*^2\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{1t}}{\bar{f}_t^3}$$

(**8**). If $j \in I_1, k \in I_1$ and $l \in I_2$, then $\bar{\mathcal{E}}_{jkl,t}$ equals

$$-\rho(1-2\xi_*)\left[\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t}-\frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{\xi}_{t|t-1}\nabla_{\theta_k}\bar{f}_{1t}}{\bar{f}_t}-\frac{\nabla_{\theta_k}\nabla_{\theta_l}\bar{\xi}_{t|t-1}\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t}-\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}\right]$$
$$-\rho(6\xi_*^2-6\xi_*+1)\left[\frac{\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}+\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}\right]$$
$$+2\rho\left[\frac{\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}-\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}-\frac{\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}\right]$$
$$-r(1-2\xi_*)\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^2}+(6\xi_*^2-4\xi_*)\left[\rho\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}-r\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{1t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^3}\right].$$

(**9**). If $j \in I_1, k \in I_2$ and $l \in I_2$, then $\bar{\mathcal{E}}_{jkl,t}$ equals

$$-\rho(1-2\xi_*)\left[\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t}+\frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{\xi}_{t|t-1}\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t}-\frac{\nabla_{\theta_k}\nabla_{\theta_l}\bar{\xi}_{t|t-1}\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t}+\frac{\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}}{\bar{f}_t}\right]$$
$$-\rho(6\xi_*^2-6\xi_*+1)\left[\frac{\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}+\frac{\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}\right]-r(1-2\xi_*)\frac{\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\bar{f}_{1t}}{\bar{f}_t^2}$$
$$+2\rho\left[\frac{\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}-\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}\right]$$
$$+[6(1-\xi_*)^2-4(1-\xi_*)]\left[\rho\frac{\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}+r\frac{\nabla_{\theta_j}\bar{f}_{1t}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^3}\right].$$

(10). If $j \in I_2, k \in I_2$ and $l \in I_2$, then $\bar{\mathcal{E}}_{jkl,t}$ equals

$$-\rho(1-2\xi_*)\left[\frac{\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}}{\bar{f}_t}\right.$$
$$\left.+\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_j}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}+\frac{\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t}\right]$$
$$+2r(1-\xi_*)\left[\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^2}+\frac{\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_j}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^2}+\frac{\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\nabla_{\theta_k}\bar{f}_{2t}}{\bar{f}_t^2}\right]$$
$$+\rho[6(1-\xi_*)^2-4(1-\xi_*)]\left[\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}+\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}+\frac{\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}\right]$$
$$+2\rho\left[\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}+\frac{\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}+\frac{\nabla_{\theta_l}\bar{f}_{2t}\nabla_{\theta_j}\bar{\xi}_{t|t-1}\nabla_{\theta_k}\bar{\xi}_{t|t-1}}{\bar{f}_t^2}\right]$$
$$-r\frac{\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t}-6r(1-\xi_*)^2\frac{\nabla_{\theta_j}\bar{f}_{2t}\nabla_{\theta_k}\bar{f}_{2t}\nabla_{\theta_l}\bar{f}_{2t}}{\bar{f}_t^3}.\ \blacksquare$$

The next lemma provides stochastic bounds for $\bar{\xi}_{t+1|t}$ and its derivatives.

Lemma A.1 *Suppose Assumption 4 hold. Then, there exists an open neighborhood of* (β_*, δ_*)*, denoted by* $B(\beta_*, \delta_*)$*, and a sequence of strictly stationary and ergodic random variables* $\{\lambda_t\}$ *satisfying* $E\lambda_t^{1+c} < M < \infty$ *for some* $c, M > 0$*, such that:*

$$\sup_{(\beta,\delta_1)\in B(\beta_*,\delta_*)} \left|\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{\xi}_{t+1|t}\right|^{\frac{\alpha(k)}{k}} < \lambda_t \quad (t = 1, ..., T)$$

for all $i_1, ..., i_k \in \{1, .., 2n_\delta + n_\beta\}$ *and* $k = 1, 2, 3$ *and* 4*, where* $\alpha(k) = 6$ *if* $k = 1, 2, 3$ *and* $\alpha(k) = 5$ *if* $k = 4$*. The above inequalities hold uniformly over* $\epsilon \leq p, q \leq 1 - \epsilon$ *with* ϵ *being an arbitrary number satisfying* $0 < \epsilon < 1/2$*.*

Proof of Lemma A.1. We use the difference equations in Lemma 1 to relate $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{\xi}_{t+1|t}$ to $\bar{f}_{1t}$ and $\bar{f}_{2t}$ and their derivatives. Because the higher order derivatives depend successively on the lower orders, we start with $k = 1$. Without loss of generality, suppose $j \in I_1$. Apply (A.6):

$$\left|\nabla_{\theta_j}\bar{\xi}_{t+1|t}\right|^6 \leq \left(\sum_{s=0}^{t-1}\left|r\rho^s\frac{\nabla_{\theta_j}\bar{f}_{1(t-s)}}{\bar{f}_{t-s}}\right|\right)^6 \leq \left(\sum_{s=0}^{\infty}|r\rho^s|\,\upsilon_{t-s}^{1/6}\right)^6 \leq \left(\sum_{s=0}^{\infty}(1-\epsilon)^s\upsilon_{t-s}^{1/6}\right)^6, \qquad \text{(A.10)}$$

where the second inequality follows from Assumption 4 and the last inequality uses $\rho = p + q - 1$. Because $\{\upsilon_t\}$ is stationary and ergodic, the right hand side is also stationary and ergodic (White, 2001, Theorem 3.35). Denote it by λ_t and apply Minkowski's inequality for an infinite sum:

$$\begin{aligned} E\lambda_t^{1+c} &= E\left[\sum_{s=0}^{\infty}(1-\epsilon)^s\upsilon_{t-s}^{1/6}\right]^{6(1+c)} \leq \left\{\sum_{s=0}^{\infty}\left[E((1-\epsilon)^s\upsilon_{t-s}^{1/6})^{6(1+c)}\right]^{\frac{1}{6(1+c)}}\right\}^{6(1+c)} \qquad \text{(A.11)} \\ &= \left\{\sum_{s=0}^{\infty}(1-\epsilon)^s\left[E\upsilon_{t-s}^{1+c}\right]^{\frac{1}{6(1+c)}}\right\}^{6(1+c)} \leq L\left\{\sum_{s=0}^{\infty}(1-\epsilon)^s\right\}^{6(1+c)}, \end{aligned}$$

where the last inequality holds because $E\upsilon_{t-s}^{1+c}$ is finite by Assumption 4. Because $\sum_{s=0}^{\infty}(1-\epsilon)^s = 1/\epsilon < \infty$, we have $E\lambda_t^{1+c} \leq L/\epsilon^{6(1+c)} < \infty$. This establishes the result for $k = 1$. Let $M = L/\epsilon^{6(1+c)}$.

The proof for $k > 1$ is similar. For $k = 2$, we have $|\nabla_{\theta_j}\nabla_{\theta_i}\bar{\xi}_{t+1|t}|^3 \leq (\sum_{s=0}^{\infty}|\rho^s\bar{\mathcal{E}}_{ji,t-s}|)^3$. We provide upper bounds for $|\bar{\mathcal{E}}_{ji,t}|$ for five possible cases. Specifically, if $j \in I_0$ and $i \in I_1$, then

$$\left|\bar{\mathcal{E}}_{ji,t}\right| = \left|-\frac{r\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\frac{\nabla_{\theta_i}\bar{f}_{1t}}{\bar{f}_t} + \frac{r\nabla_{\theta_j}\nabla_{\theta_i}\bar{f}_{1t}}{\bar{f}_t}\right| \leq \left|\frac{r\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\frac{\nabla_{\theta_i}\bar{f}_{1t}}{\bar{f}_t}\right| + \left|\frac{r\nabla_{\theta_j}\nabla_{\theta_i}\bar{f}_{1t}}{\bar{f}_t}\right| \leq 2\,|r|\,\upsilon_t^{1/3}.$$

The same bound holds if $j \in I_0$ and $i \in I_2$. If $j \in I_1$ and $i \in I_1$, then $|\bar{\mathcal{E}}_{ji,t}| \leq 2\,|\rho\,(1-2\xi_*)|\,\lambda_{t-1}^{1/6}\upsilon_t^{1/6} + 3\,|r|\,\upsilon_t^{1/3}$. If $j \in I_1$ and $i \in I_2$, then $|\bar{\mathcal{E}}_{ji,t}| \leq 2\,|\rho\,(1-2\xi_*)|\,\lambda_{t-1}^{1/6}\upsilon_t^{1/6} + |r(2\xi_*-1)|\,\upsilon_t^{1/3}$. If $j \in I_2$ and $k \in I_2$, then $|\bar{\mathcal{E}}_{ji,t}| \leq 2\,|\rho\,(1-2\xi_*)|\,\lambda_{t-1}^{1/6}\upsilon_t^{1/6} + (|r| + |2r(\xi_*-1)|)\,\upsilon_t^{1/3}$. Consequently, there exists a finite constant C_1, such that for all the five cases we have $|\bar{\mathcal{E}}_{ji,t}| \leq C_1(\lambda_{t-1}^{1/6}\upsilon_t^{1/6} + \upsilon_t^{1/3})$. This implies $\left|\nabla_{\theta_j}\nabla_{\theta_i}\bar{\xi}_{t+1|t}\right|^3 \leq \left(\sum_{s=0}^{\infty}C_1(1-\epsilon)^s(\lambda_{t-s-1}^{1/6}\upsilon_t^{1/6} + \upsilon_{t-s}^{1/3})\right)^3$. The right side is stationary and ergodic; we continue to denote it by λ_t. By Minkowski's inequality:

$$E\lambda_t^{1+c} \leq \left\{\sum_{s=0}^{\infty}\left[E\left(C_1(1-\epsilon)^s(\lambda_{t-s-1}^{1/6}\upsilon_t^{1/6} + \upsilon_{t-s}^{1/3})\right)^{3(1+c)}\right]^{\frac{1}{3(1+c)}}\right\}^{3(1+c)}. \qquad \text{(A.12)}$$

Apply Minkowski's inequality followed by the Cauchy–Schwarz inequality to the summands:

$$
\begin{aligned}
&E\left(C_1(1-\epsilon)^s(\lambda_{t-1}^{1/6}\upsilon_t^{1/6}+\upsilon_t^{1/3})\right)^{3(1+c)} \\
&\le (C_1(1-\epsilon)^s)^{3(1+c)}\left[\left(E\lambda_{t-1}^{(1+c)}E\upsilon_t^{(1+c)}\right)^{\frac{1}{6(1+c)}}+\left(E\upsilon_t^{(1+c)}\right)^{\frac{1}{3(1+c)}}\right]^{3(1+c)}.
\end{aligned}
$$

Because $E\lambda_{t-1}^{(1+c)}<M$ and $E\upsilon_t^{(1+c)}<L$, the last term in the preceding display is no greater than

$$
(1-\epsilon)^{3(1+c)s}C_1^{3(1+c)}\left[(ML)^{\frac{1}{6(1+c)}}+L^{\frac{1}{3(1+c)}}\right]^{3(1+c)}\le C_2(1-\epsilon)^{3(1+c)s}, \tag{A.13}
$$

where C_2 is a finite constant independent of p and q. Plug this into (A.12), we have $E\lambda_t^{1+c}\le C_2(\sum_{s=0}^{\infty}(1-\epsilon)^s)^{3(1+c)}=C_2/\epsilon^{3(1+c)}<\infty$. This proves the result for $k=2$.

Now, consider $k=3$. Inspecting the expressions of $\bar{\mathcal{E}}_{jil,t}$ reported in the proof of Lemma 1 shows that they comprise the following terms $(a,b,c=1,2)$:

$$
\begin{aligned}
&\frac{\nabla_{\theta_j}\nabla_{\theta_i}\bar{f}_{at}\nabla_{\theta_l}\bar{f}_{bt}}{\bar{f}_t^2},\ \frac{\nabla_{\theta_j}\nabla_{\theta_i}\nabla_{\theta_l}\bar{f}_{at}}{\bar{f}_t},\ \frac{\nabla_{\theta_j}\bar{f}_{at}\nabla_{\theta_i}\bar{f}_{bt}\nabla_{\theta_l}\bar{f}_{ct}}{\bar{f}_t^3},\ \frac{\nabla_{\theta_l}\bar{f}_{at}\nabla_{\theta_j}\nabla_{\theta_i}\bar{\xi}_{t|t-1}}{\bar{f}_t},\\
&\frac{\nabla_{\theta_j}\bar{f}_{at}\nabla_{\theta_l}\bar{f}_{bt}\nabla_{\theta_i}\bar{\xi}_{t|t-1}}{\bar{f}_t^2},\ \frac{\nabla_{\theta_j}\nabla_{\theta_i}\bar{f}_{at}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t},\ \frac{\nabla_{\theta_j}\bar{f}_{at}\nabla_{\theta_i}\bar{\xi}_{t|t-1}\nabla_{\theta_l}\bar{\xi}_{t|t-1}}{\bar{f}_t}.
\end{aligned} \tag{A.14}
$$

By Assumption 4 and the above results for $k=1$ and 2, the quantities in (A.14) are bounded, respectively, by $\upsilon_t^{1/2},\upsilon_t^{1/2},\upsilon_t^{1/2},\upsilon_t^{1/6}\lambda_{t-1}^{1/3},\upsilon_t^{1/3}\lambda_{t-1}^{1/6},\upsilon_t^{1/3}\lambda_{t-1}^{1/6}$ and $\upsilon_t^{1/6}\lambda_{t-1}^{1/3}$. Therefore, the ten cases specified in Lemma 1 all satisfy $\left|\bar{\mathcal{E}}_{jil,t}\right|\le C_3(\upsilon_t^{1/2}+\upsilon_t^{1/6}\lambda_{t-1}^{1/3}+\upsilon_t^{1/3}\lambda_{t-1}^{1/6})$, where C_3 is finite and independent of (p,q). This implies $\left|\nabla_{\theta_j}\nabla_{\theta_i}\nabla_{\theta_l}\bar{\xi}_{t+1|t}\right|^2\le\left|\sum_{s=0}^{\infty}(1-\epsilon)^sC_3(\upsilon_{t-s}^{1/2}+\upsilon_{t-s}^{1/6}\lambda_{t-s-1}^{1/3}+\upsilon_{t-s}^{1/3}\lambda_{t-s-1}^{1/6})\right|^2$. Denote the right hand side by λ_t and proceed along the same lines as between (A.12) and (A.13). It then follows that $E\lambda_t^{1+c}<\infty$. For $k=4$, the expressions of $\bar{\mathcal{E}}_{jilm,t}$, although omitted here, include terms as in (A.14) but with the orders of derivatives sum to 4 instead of 3. Using the same arguments as between (A.12) and (A.13), it can be shown that $E\lambda_t^{1+c}<\infty$ holds. ■

The next lemma establishes stochastic orders of some quantities related to $\xi_{t|t-1}$, f_{1t} and f_{2t}. The quantities are all evaluated at $(\widetilde{\beta},\widetilde{\delta},\widetilde{\delta})$.

Lemma A.2 *Let i_s,j_s,l_s,m_s,n_s be arbitrary integers satisfying $1\le i_s,j_s,l_s,m_s,n_s\le 2n_\delta+n_\beta$ for $s\in\{1,2,3,4\}$. The following results hold uniformly over $\epsilon\le p,q\le 1-\epsilon$ with ϵ being an arbitrary number satisfying $0<\epsilon<1/2$:*

1. *For any $a\in\{1,2\},u\in\{1,2,3,4\}$ and $v\in\{0,1,2,3\}$ satisfying $u+v\le 4$, we have (interpret $\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\widetilde{\xi}_{t|t-1}$ as 1 when $v=0$)*

$$
\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\widetilde{f}_{at}}{\widetilde{f}_t}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\widetilde{\xi}_{t|t-1}=o_p(1), \tag{A.15}
$$

   Further, if $u+v\le 3$, then the result holds with $o_p(1)$ replaced by $O_p(T^{-1/2})$.

2. *For any $(a,b,c)\in\{1,2\},(u,w)\in\{1,2,3\}$ and $v\in\{0,1,2\}$ satisfying $u+v+w\le 4$:*

$$
\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\widetilde{f}_{at}}{\widetilde{f}_t}\frac{\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\widetilde{f}_{bt}}{\widetilde{f}_t}\frac{\nabla_{\theta_{l_1}}...\nabla_{\theta_{l_w}}\widetilde{f}_{ct}}{\widetilde{f}_t}=O_p(1).
$$

3. *For any* $(a,b,c) \in \{1,2\}, (u,w) \in \{1,2,3\}$ *and* $(v,z) \in \{0,1\}$ *satisfying* $u+v+w+z \leq 3:$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\widetilde{f}_{at}}{\widetilde{f}_t}\frac{\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_u}}\widetilde{f}_{bt}}{\widetilde{f}_t}\frac{\nabla_{\theta_{l_1}}...\nabla_{\theta_{l_v}}\widetilde{f}_{ct}}{\widetilde{f}_t}\nabla_{\theta_{m_1}}...\nabla_{\theta_{m_w}}\widetilde{\xi}_{t|t-1}\nabla_{\theta_{n_1}}...\nabla_{\theta_{n_z}}\widetilde{\xi}_{t|t-1} = O_p(1).$$

Proof of Lemma A.2. By the mean value theorem, the left hand side of (A.15) equals

$$T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}f_{at}^*}{f_t^*}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\xi_{t|t-1}^* + \left\{T^{-3/2}\sum_{t=1}^{T}\nabla_{\theta'}\left(\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\bar{f}_{at}}{\bar{f}_{at}}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\bar{\xi}_{t|t-1}\right)\right\}T^{1/2}\left(\widetilde{\theta}-\theta_*\right), \tag{A.16}$$

where "$*$" and "$-$" denote that the relevant quantities are evaluated at the true values $\theta'_* = (\beta'_*, \delta'_*, \delta'_*)$ and $\bar{\theta}' = \left(\bar{\beta}', \bar{\delta}', \bar{\delta}'\right)$, where $\bar{\theta} = \theta_* + c(\widetilde{\theta} - \theta_*)$ for some $c \in (0,1)$. The first summation is over terms that are stationary and ergodic, which are bounded by $\lambda_t^{v/\alpha(k)}\upsilon_t^{u/\alpha(k)}$ by Assumption 4 and Lemma A.1. Apply Hölder's inequality:

$$\begin{aligned} E(\lambda_t^{v/\alpha(k)}\upsilon_t^{u/\alpha(k)})^{1+c} &\leq \left(E\left(\lambda_t^{v(1+c)/\alpha(k)}\right)^{\alpha(k)/v}\right)^{\frac{v}{\alpha(k)}}\left(E\left(\upsilon_t^{u(1+c)/\alpha(k)}\right)^{\alpha(k)/(\alpha(k)-v)}\right)^{\frac{\alpha(k)-v}{\alpha(k)}} \\ &\leq \left(E\lambda_t^{1+c}\right)^{\frac{v}{\alpha(k)}}\left(E\upsilon_t^{1+c}\right)^{\frac{\alpha(k)-v}{\alpha(k)}}, \end{aligned}$$

where the last inequality follows because $u+v < \alpha(k)$. Both terms on the right hand side are finite by Assumption 4 and Lemma A.1. Therefore, the first term in the display (A.16) is $o_p(1)$ by Theorem 3.34 in White (2001). Now turn to the second term in the display (A.16). We have, for any $k \in \{1,...,2n_\delta + n_\beta\}:$

$$\begin{aligned} &\left|T^{-3/2}\sum_{t=1}^{T}\nabla_{\theta_k}\left(\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\bar{f}_{at}}{\bar{f}_{at}}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\bar{\xi}_{t|t-1}\right)\right| \\ \leq &\left|T^{-3/2}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\nabla_{\theta_k}\bar{f}_{at}}{\bar{f}_{at}}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\bar{\xi}_{t|t-1}\right| + \left|T^{-3/2}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\bar{f}_{at}}{\bar{f}_{at}}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\nabla_{\theta_k}\bar{\xi}_{t|t-1}\right| \\ &+\left|T^{-3/2}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\bar{f}_{at}}{\bar{f}_{at}}\frac{\nabla_{\theta_k}\bar{f}_{at}}{\bar{f}_{at}}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\bar{\xi}_{t|t-1}\right| \\ \leq &\; T^{-3/2}\sum_{t=1}^{T}\left\{2\upsilon_t^{(u+1)/\alpha(k)}\lambda_t^{v/\alpha(k)} + \upsilon_t^{u/\alpha(k)}\lambda_t^{(v+1)/\alpha(k)}\right\} = O_p\left(T^{-1/2}\right), \end{aligned}$$

where the equality follows from Assumption 4, Lemma A.1 and $u+v+1 \leq 5$. In addition, $T^{1/2}(\widetilde{\theta}-\theta_*) = O_p(1)$, which follows from Assumptions 1, 2(i), 3(i), and a first order Taylor expansion of the score function around the true value. Therefore, the display (A.16) is $o_p(1)$.

Now we consider the cases with $u+v \leq 3$. If $u+v < 3$, then the terms inside the first summation of (A.16) are bounded by $\lambda_t^{v/6}\upsilon_t^{u/6}$. We have

$$E(\lambda_t^{v/6}\upsilon_t^{u/6})^{2(1+c)} \leq \left(E(\lambda_t^{v(1+c)/3})^{\frac{3}{v}}\right)^{v/3}\left(E(\upsilon_t^{u(1+c)/3})^{\frac{3}{(3-v)}}\right)^{(3-v)/3} \leq \left(E\lambda_t^{1+c}\right)^{v/3}\left(E\upsilon_t^{1+c}\right)^{(3-v)/3}.$$

The right hand side is finite. If $u+v=3$, i.e., $u=3$ and $v=0$, then $E(\lambda_t^{v/6}\upsilon_t^{u/6})^{2(1+c)} = E\upsilon_t^{(1+c)} < \infty$. Apply the central limit theorem. It follows that the left hand side of (A.15) is $O_p\left(T^{-1/2}\right)$.

Lemma A.2.2 and A.2.3 can be proved using the same arguments, i.e., first applying the mean value theorem and then obtaining bounds for the two resulting terms separately. It follows that the left hand side quantity in Lemma A.2.2 is bounded by $T^{-1}\sum_{t=1}^{T}\upsilon_t^{(u+v+w)/\alpha(k)} + O_p(T^{-1/2})$, while that in Lemma A.2.3 is bounded by $T^{-1}\sum_{t=1}^{T}\upsilon_t^{(1+u+v)/\alpha(k)}\lambda_t^{(w+z)/\alpha(k)} + O_p(T^{-1/2})$. The two leading terms both satisfy the law of large numbers, therefore are $O_p(1)$. ■

We state some notations to be used in subsequent proofs. Define

$$\hat{\theta}(\delta_2) = (\hat{\beta}(\delta_2)', \hat{\delta}_1(\delta_2)', \delta_2')',$$

where $\hat{\beta}(\delta_2)$ and $\hat{\delta}_1(\delta_2)$ are defined in (14). Let $\hat{\xi}_{t+1|t}$, $\hat{f}_{1t}$ and $\hat{f}_{2t}$ denote $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$, $f_t(\beta,\delta_1)$ and $f_t(\beta,\delta_2)$ evaluated at $(\beta,\delta_1,\delta_2) = (\hat{\beta}(\delta_2),\hat{\delta}_1(\delta_2),\delta_2)$. Also, let $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\hat{\xi}_{t+1|t}$, $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\hat{f}_{1t}$ and $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\hat{f}_{2t}$ denote the k-th order derivatives of $\xi_{t+1|t}$, f_{1t} and f_{2t} with respect to the i_1-th,...i_k-th elements of θ evaluated at $(\hat{\beta}(\delta_2),\hat{\delta}_1(\delta_2),\delta_2)$.

Remark 3 *Lemma A.3 generalizes Lemma B2(a)-(d) in Cho and White (2007) to Markov switching models. The results quantify how δ_1 and β need to change in order to maximize the likelihood when δ_2 is moved away from $\widetilde{\delta}$. They provide the necessary inputs for the chain rule when computing the derivatives $\mathcal{L}^{(k)}_{i_1,...,i_k}(p,q,\delta)$ $(k=1,2,3,4)$ in (16).*

Lemma A.3 *Let the null hypothesis and Assumptions 1-4 hold. For any $k,l,m\in\{1,...,n_\delta\}$:*

1. *Let e_k be an n_δ-dimensional unit vector whose k-th element equals 1, then*

$$\begin{bmatrix} \nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta}) \\ \xi_*\nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} = (\xi_*-1)\begin{bmatrix} 0 \\ e_k \end{bmatrix} + O_p(T^{-1/2}).$$

2. *The second order derivatives satisfy*

$$\begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) \\ \xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} = -\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\widetilde{D}_{kl,t} + O_p(T^{-1/2}).$$

3. *The third order derivatives satisfy*

$$\begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta}) \\ \xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} = O_p(1).$$

Proof of Lemma A.3. As the proof is long, we organize it into three parts, corresponding to Lemma A.3.1, A.3.2 and A.3.3 respectively.

Proof of the first result in Lemma A.3. By construction, $\hat{\theta}(\delta_2)$ satisfies

$$\mathcal{M}_j^{(1)}(p,q,\delta_2) = T^{-1}\sum_{t=1}^{T}\frac{\hat{M}_{jt}}{\hat{B}_t} = 0 \quad (j=1,...,n_\beta+n_\delta), \tag{A.17}$$

where

$$\hat{B}_t = (\hat{f}_{1t} - \hat{f}_{2t})\hat{\xi}_{t|t-1} + \hat{f}_{2t}, \tag{A.18}$$
$$\hat{M}_{jt} = (\nabla_{\theta_j}\hat{f}_{1t} - \nabla_{\theta_j}\hat{f}_{2t})\hat{\xi}_{t|t-1} + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1} + \nabla_{\theta_j}\hat{f}_{2t}.$$

Because (A.17) holds for all $\delta_2 \in \Delta$, its derivatives with respect to δ_2 must equal zero. The proof makes use of this property. It proceeds in three steps. For an arbitrary $k \in \{1, ..., n_\delta\}$, the first step differentiates the $n_\beta + n_\delta$ equations in (A.17) with respect to δ_{2k} to obtain a system of $n_\beta + n_\delta$ linear equations, with $\nabla_{\delta_{2k}}\hat{\beta}(\delta_2)$ and $\nabla_{\delta_{2k}}\hat{\delta}_1(\delta_2)$ being the unknowns. The second step evaluates these equations at $\delta_2 = \widetilde{\delta}$ and provides approximations to them. The third step solves these approximating equations to obtain explicit expressions for $\nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta})$ and $\nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta})$. These three steps are then repeated for all $k \in \{1, ..., n_\delta\}$ to prove Lemma A.3.1. The idea of differentiating the first order conditions is inspired by Cho and White (2007). At the same time the proof here is more complex due to the presence of $\xi_{t+1|t}$ and the allowance for multiple parameters to be affected by regime switching.

Step 1 for proving Lemma A.3.1. Consider an arbitrary $k \in \{1, ..., n_\delta\}$ and an arbitrary $j \in \{1, ..., n_\beta + n_\delta\}$. Taking the first order derivative of the j-th equation (A.17) with respect to the δ_{2k} (Here, view $\hat{B}_t$ and $\hat{M}_{jt}$ as functions of p, q and δ_2; note that β and δ_1 are now functions of these three elements.):

$$\mathcal{M}^{(2)}_{jk}(p, q, \delta_2) = \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{M}_{jt}}{\hat{B}_t} - \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t^2}\hat{M}_{jt} = 0, \tag{A.19}$$

where

$$\begin{aligned}\nabla_{\delta_{2k}}\hat{M}_{jt} = {} & \Big\{\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{1t} + (1 - \hat{\xi}_{t|t-1})\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{2t} + (\nabla_{\theta_j}\hat{f}_{1t} - \nabla_{\theta_j}\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1} \\ & +(\nabla_{\theta_j}\xi_{t|t-1})(\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t}) + (\hat{f}_{1t} - \hat{f}_{2t})(\nabla_{\theta_j}\nabla_{\theta'}\hat{\xi}_{t|t-1})\Big\}\nabla_{\delta_{2k}}\hat{\theta}(\delta_2),\end{aligned} \tag{A.20}$$

and

$$\nabla_{\delta_{2k}}\hat{B}_t = \Big\{\hat{\xi}_{t|t-1}\nabla_{\theta'}\hat{f}_{1t} + (1 - \hat{\xi}_{t|t-1})\nabla_{\theta'}\hat{f}_{2t} + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\theta}(\delta_2) \tag{A.21}$$

with

$$\nabla_{\delta_{2k}}\hat{\theta}(\delta_2) = \begin{bmatrix}\nabla_{\delta_{2k}}\hat{\beta}(\delta_2) \\ \nabla_{\delta_{2k}}\hat{\delta}_1(\delta_2) \\ e_k\end{bmatrix}, \tag{A.22}$$

where e_k is an n_δ-dimensional vector whose k-th element equals 1 and otherwise zero. We view (A.19) as a linear equation with the first $(n_\beta + n_\delta)$ elements of $\nabla_{\delta_{2k}}\hat{\theta}(\delta_2)$ being the unknowns. The above differentiation can be carried for all $j = 1, ..., n_\beta + n_\delta$, while keeping k fixed at the same value. This delivers $n_\beta + n_\delta$ equations with the same number of unknowns specified in (A.22).

Step 2 for proving Lemma A.3.1. We first evaluate $T^{-1}\sum_{t=1}^{T}(\nabla_{\delta_{2k}}\hat{B}_t/\hat{B}_t^2)\hat{M}_{jt}$ in (A.19) at $\delta_2 = \widetilde{\delta}$ for an arbitrary $j \in \{1, ..., n_\beta + n_\delta\}$. It equals (using $\widetilde{f}_{1t} = \widetilde{f}_{2t} = \widetilde{f}_t$ and $\widetilde{\xi}_{t|t-1} = \xi_*$)

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\widetilde{f}_{1t} + (1 - \xi_*)\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t^2}[\xi_*\nabla_{\theta'}\widetilde{f}_{1t} + (1 - \xi_*)\nabla_{\theta'}\widetilde{f}_{2t}]\nabla_{\delta_{2k}}\hat{\theta}(\widetilde{\delta}).$$

Using (A.22), this can be rewritten as

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t^2}\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}\left[\begin{array}{c}\nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta})\\ \xi_*\nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]$$
$$+\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t^2}(1-\xi_*)\nabla_{\delta_{2k}}\widetilde{f}_{2t},$$

where $\nabla_{\beta'}\widetilde{f}_{1t}$ denotes the derivative of $f_t(\beta,\delta_1)$ with respect to β evaluated at $\hat{\beta}(\widetilde{\delta})$ and $\hat{\delta}_1(\widetilde{\delta})$; $\nabla_{\delta_1'}\widetilde{f}_{1t}$ and $\nabla_{\delta_{2k}}\widetilde{f}_{2t}$ are defined analogously. Further, if $j\in\{1,...,n_\beta\}$, the preceding display equals

$$\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{cc}\frac{\nabla_{\theta_j}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & \frac{\nabla_{\theta_j}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{array}\right]\left[\begin{array}{c}\nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta})\\ \xi_*\nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]+\frac{1}{T}(1-\xi_*)\sum_{t=1}^{T}\frac{\nabla_{\theta_j}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{2k}}\widetilde{f}_{2t}}{\widetilde{f}_t}. \quad \text{(A.23)}$$

Meanwhile, if $j\in\{n_\beta+1,...,n_\beta+n_\delta\}$, then the same display equals (using $\nabla_{\theta_j}\widetilde{f}_{2t}=0$) ξ_* times (A.23). Let D be a diagonal matrix whose first n_β diagonal elements equal 1 and the rest ξ_*. Then the above two cases for j can be combined, leading to

$$D\widetilde{I}\left[\begin{array}{c}\nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta})\\ \xi_*\nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]+D\left[\begin{array}{c}(1-\xi_*)\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\beta}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{2k}}\widetilde{f}_{2t}}{\widetilde{f}_t}\\ (1-\xi_*)\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_1}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{2k}}\widetilde{f}_{2t}}{\widetilde{f}_t}\end{array}\right], \quad \text{(A.24)}$$

where $\widetilde{I}$ is defined in (17).

Now consider the first term in (A.19). It equals (using $\widetilde{f}_{1t}=\widetilde{f}_{2t}=\widetilde{f}_t$ and $\widetilde{\xi}_{t|t-1}=\xi_*$)

$$\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\xi_*\nabla_{\theta_j}\nabla_{\theta'}\widetilde{f}_{1t}}{\widetilde{f}_t}+\frac{(1-\xi_*)\nabla_{\theta_j}\nabla_{\theta'}\widetilde{f}_{2t}}{\widetilde{f}_t}+\frac{\nabla_{\theta_j}\widetilde{f}_{1t}-\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t}\nabla_{\theta'}\widetilde{\xi}_{t|t-1}+\nabla_{\theta_j}\widetilde{\xi}_{t|t-1}\frac{\nabla_{\theta'}\widetilde{f}_{1t}-\nabla_{\theta'}\widetilde{f}_{2t}}{\widetilde{f}_t}\right]\right\}\nabla_{\delta_{2k}}\hat{\theta}(\widetilde{\delta}). \quad \text{(A.25)}$$

All the terms inside the curly brackets are $O_p(T^{-1/2})$ by Lemma A.2.1. Their effects are dominated by $\widetilde{I}$, which is positive definite in large samples. Combining this with (A.24) and (A.19), we have:

$$\widetilde{I}\left[\begin{array}{c}\nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta})\\ \xi_*\nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]=-\left[\begin{array}{c}(1-\xi_*)\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\beta}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{2k}}\widetilde{f}_{2t}}{\widetilde{f}_t}\\ (1-\xi_*)\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_1}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{2k}}\widetilde{f}_{2t}}{\widetilde{f}_t}\end{array}\right]+O_p(T^{-1/2}). \quad \text{(A.26)}$$

The preceding display provides $(n_\beta+n_\delta)$ linear equations with the same number of unknowns.

Step 3 for proving Lemma A.3.1. We show how to solve (A.26) for $k=n_\delta$. Consider the following partition of the system (A.26) with $\widetilde{I}_{22},\widetilde{\phi}_2$ and $\widetilde{B}_2$ being scalars:

$$\left[\begin{array}{cc}\widetilde{I}_{11} & \widetilde{I}_{12}\\ \widetilde{I}_{21} & \widetilde{I}_{22}\end{array}\right]\left[\begin{array}{c}\widetilde{\phi}_1\\ \widetilde{\phi}_2\end{array}\right]=\left[\begin{array}{c}\widetilde{B}_1\\ \widetilde{B}_2\end{array}\right]+O_p(T^{-1/2}).$$

This implies

$$\left[\begin{array}{cc}\widetilde{I}_{11} & \widetilde{I}_{12}\\ 0 & \widetilde{I}_{22}-\widetilde{I}_{21}\widetilde{I}_{11}^{-1}\widetilde{I}_{12}\end{array}\right]\left[\begin{array}{c}\widetilde{\phi}_1\\ \widetilde{\phi}_2\end{array}\right]=\left[\begin{array}{c}\widetilde{B}_1\\ \widetilde{B}_2-\widetilde{I}_{21}\widetilde{I}_{11}^{-1}\widetilde{B}_1\end{array}\right]+O_p(T^{-1/2}), \quad \text{(A.27)}$$

which further implies $\widetilde{\phi}_2 = [\widetilde{B}_2 - \widetilde{I}_{21}\widetilde{I}_{11}^{-1}\widetilde{B}_1]/[\widetilde{I}_{22} - \widetilde{I}_{21}\widetilde{I}_{11}^{-1}\widetilde{I}_{12}] + O_p(T^{-1/2})$. Because $\widetilde{B}_1 = (\xi_* - 1)\,\widetilde{I}_{12}$ and $\widetilde{B}_2 = (\xi_* - 1)\,\widetilde{I}_{22}$, after cancellation we have $\widetilde{\phi}_2 = \xi_* - 1 + O_p(T^{-1/2})$. Plugging this result into the first set of equations in (A.27), we obtain $\widetilde{\phi}_1 = \widetilde{I}_{11}^{-1}\widetilde{B}_1 - \widetilde{I}_{11}^{-1}\widetilde{I}_{12}\widetilde{\phi}_2 + O_p(T^{-1/2}) = (\xi_* - 1)\,\widetilde{I}_{11}^{-1}\widetilde{I}_{12} - \widetilde{I}_{11}^{-1}\widetilde{I}_{12}\left[(\xi_* - 1) + O_p(T^{-1/2})\right] + O_p(T^{-1/2}) = O_p(T^{-1/2})$. This completes the proof for the case $k = n_\delta$. For other values of k, the same argument can be used after exchanging the k- and n_δ-th columns of $\widetilde{I}$ and the k- and n_δ-th elements of $\widetilde{\phi}$ and $\widetilde{B}$. ■

Proof of the second result in Lemma A.3. View the quantities in (A.19) as functions of δ_2, p and q and differentiate them with respect to the l-th element of δ_2 $(l = 1, ..., n_\delta)$:

$$\begin{aligned}\mathcal{M}_{jkl}^{(3)}(p,q,\delta_2) \;=\; \frac{1}{T}\sum_{t=1}^{T}\Bigg\{ & \frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{M}_{jt}}{\hat{B}_t} - \frac{\nabla_{\delta_{2k}}\hat{M}_{jt}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2} - \frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2}\hat{M}_{jt} \\ & -\frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t^2}\nabla_{\delta_{2l}}\hat{M}_{jt} + 2\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^3}\hat{M}_{jt}\Bigg\} = 0,\end{aligned} \tag{A.28}$$

where

$$\begin{aligned}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{M}_{jt} \;=\; & \sum_{s=1}^{n_\beta+2n_\delta}\Bigg\{\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{1t} + \hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t} \\ & -\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{2t} + (1-\hat{\xi}_{t|t-1})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t} \\ & +(\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1} + (\nabla_{\theta_j}\hat{f}_{1t} - \nabla_{\theta_j}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\ & +(\nabla_{\theta_s}\hat{f}_{1t} - \nabla_{\theta_s}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta'}\hat{\xi}_{t|t-1} + (\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1} \\ & +(\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\hat{\xi}_{t|t-1} + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Bigg\}\times \\ & \nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}(\delta_2) \\ & +\Bigg\{\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{1t} + (1-\hat{\xi}_{t|t-1})\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{2t} + (\nabla_{\theta_j}\hat{f}_{1t} - \nabla_{\theta_j}\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1} \\ & +(\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1} + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Bigg\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\delta_2),\end{aligned} \tag{A.29}$$

and

$$\begin{aligned}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t \;=\; & \sum_{s=1}^{n_\beta+2n_\delta}\Bigg\{\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{1t} + \hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{2t} \\ & +(1-\hat{\xi}_{t|t-1})\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t} + (\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1} + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Bigg\} \\ & \times\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}(\delta_2) \\ & +\left\{\hat{\xi}_{t|t-1}\nabla_{\theta'}\hat{f}_{1t} + (1-\hat{\xi}_{t|t-1})\nabla_{\theta'}\hat{f}_{2t} + (\hat{f}_{1t} - \hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1}\right\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\delta_2).\end{aligned} \tag{A.30}$$

We now apply (A.20), (A.21), (A.29) and (A.30) to analyze the five terms in (A.28). Start with the third term $T^{-1}\sum_{t=1}^{T}[\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t/\hat{B}_t^2]\hat{M}_{jt}$. At $\delta_2=\widetilde{\theta}$, it equals

$$\sum_{u=1}^{n_\beta+2n_\delta}\sum_{s=1}^{n_\beta+2n_\delta}\left\{\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t^2}[\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}\nabla_{\theta_s}\widetilde{f}_{1t}+\xi_*\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}\nabla_{\theta_s}\widetilde{f}_{2t}\right.$$
$$\left.+(1-\xi_*)\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t}+(\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\theta_s}\widetilde{\xi}_{t|t-1}]\right\}\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta}) \quad \text{(A.31)}$$
$$+\sum_{s=1}^{n_\beta+n_\delta}\left\{\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t^2}[\xi_*\nabla_{\theta_s}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\widetilde{f}_{2t}]\right\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta}).$$

Because $\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})$ and $\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})$ are $O_p\left(T^{-1/2}\right)$ except when $s\in\{n_\beta+k,n_\beta+n_\delta+k\}$ and $u\in\{n_\beta+l,n_\beta+n_\delta+l\}$, the preceding display equals

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t^2}\{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}+\xi_*\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}-\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{2t}$$
$$+(1-\xi_*)\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{2t}+(\nabla_{\delta_{1l}}\widetilde{f}_{1t}-\nabla_{\delta_{1l}}\widetilde{f}_{2t})\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}\}\nabla_{\delta_{2l}}\hat{\delta}_{1l}(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\widetilde{\delta})$$
$$+\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t^2}\{\nabla_{\delta_{2l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}+\xi_*\nabla_{\delta_{1k}}\nabla_{\delta_{2l}}\widetilde{f}_{1t}-\nabla_{\delta_{2l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{2t}$$
$$+(1-\xi_*)\nabla_{\delta_{1k}}\nabla_{\delta_{2l}}\widetilde{f}_{2t}+(\nabla_{\delta_{2l}}\widetilde{f}_{1t}-\nabla_{\delta_{2l}}\widetilde{f}_{2t})\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}\}\nabla_{\delta_{2l}}\hat{\delta}_{2l}(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\widetilde{\delta})$$
$$+\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t^2}\{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{2k}}\widetilde{f}_{1t}+\xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}-\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{2k}}\widetilde{f}_{2t}$$
$$+(1-\xi_*)\nabla_{\delta_{2k}}\nabla_{\delta_{1l}}\widetilde{f}_{2t}+(\nabla_{\delta_{1l}}\widetilde{f}_{1t}-\nabla_{\delta_{1l}}\widetilde{f}_{2t})\nabla_{\delta_{2k}}\widetilde{\xi}_{t|t-1}\}\nabla_{\delta_{2l}}\hat{\delta}_{1l}(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\delta}_{2k}(\widetilde{\delta})$$
$$+\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t^2}\{\nabla_{\delta_{2l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{2k}}\widetilde{f}_{1t}+\xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\widetilde{f}_{1t}-\nabla_{\delta_{2l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{2k}}\widetilde{f}_{2t}$$
$$+(1-\xi_*)\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\widetilde{f}_{2t}+(\nabla_{\delta_{2l}}\widetilde{f}_{1t}-\nabla_{\delta_{2l}}\widetilde{f}_{2t})\nabla_{\delta_{2k}}\widetilde{\xi}_{t|t-1}\}\nabla_{\delta_{2l}}\hat{\delta}_{2l}(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\delta}_{2k}(\widetilde{\delta})$$
$$+\sum_{s=1}^{n_\beta+n_\delta}\left\{\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t^2}[\xi_*\nabla_{\theta_s}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\widetilde{f}_{2t}]\right\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})+O_p(T^{-1/2}).$$

Apply $\nabla_{\delta_{2l}}\hat{\delta}_{1l}(\widetilde{\delta})=(\xi_*-1)/\xi_*+O_p\left(T^{-1/2}\right)$, $\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\widetilde{\delta})=(\xi_*-1)/\xi_*+O_p\left(T^{-1/2}\right)$ and $\nabla_{\delta_{2l}}\hat{\delta}_{2l}(\widetilde{\delta})=\nabla_{\delta_{2l}}\hat{\delta}_{2l}(\widetilde{\delta})=1$ and rearrange the terms, the preceding display reduces to

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t^2}\left\{\left(\frac{1-\xi_*}{\xi_*}\right)\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}\right. \quad \text{(A.32)}$$
$$\left.+\frac{1}{\xi_*^2}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}+\frac{1}{\xi_*^2}\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}\right\}$$
$$+\sum_{s=1}^{n_\beta+n_\delta}\left\{\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_*\nabla_{\theta_j}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_j}\widetilde{f}_{2t}}{\widetilde{f}_t^2}[\xi_*\nabla_{\theta_s}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\widetilde{f}_{2t}]\right\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})+O_p(T^{-1/2}).$$

As in the proof of Lemma A.3.1, the above display leads to $(n_\beta + n_\delta)$ equations with j taking values between 1 and $(n_\beta + n_\delta)$. These equations can be written collectively as

$$D\widetilde{I}\left[\begin{array}{c}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})\\ \xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]+D\left[\begin{array}{c}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_\beta\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{U}_{kl,t}\\ \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_1}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{U}_{kl,t}\end{array}\right]+O_p(T^{-1/2}).$$

This completes the analysis for the third term in (A.28). Below we show the other terms in (A.28) are all asymptotically negligible.

Consider the first term in (A.28). Applying the expression (A.29) to (A.28) leads to quantities of the following form: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\widetilde{\xi}_{t|t-1})$, where $a\in\{1,2\}, u\in\{1,2,3\}$ and $v\in\{0,1,2\}$ with $1\le u+v\le 3$. They are all $O_p(T^{-1/2})$ because of Lemma A.2.1. Therefore, this term is negligible. Consider the second term in (A.28). At $\delta_2=\widetilde{\delta}$, $\nabla_{\delta_{2k}}\hat{B}_t$ can be rewritten as

$$\begin{aligned}&\sum_{s=1}^{n_\beta}\nabla_{\theta_s}\widetilde{f}_{1t}\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})+\sum_{s=n_\beta+1}^{n_\beta+n_\delta}\xi_*\nabla_{\theta_s}\widetilde{f}_{1t}\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta}) \qquad (A.33)\\ &+\sum_{s=n_\beta+n_\delta+1}^{n_\beta+2n_\delta}(1-\xi_*)\nabla_{\theta_s}\widetilde{f}_{2t}\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\\ =\;&\sum_{s=1}^{n_\beta}\nabla_{\beta_s}\widetilde{f}_{1t}\nabla_{\delta_{2k}}\hat{\beta}_s(\widetilde{\delta})+\sum_{s=1,s\neq k}^{n_\delta}\xi_*\nabla_{\delta_{1s}}\widetilde{f}_{1t}\nabla_{\delta_{2k}}\hat{\delta}_{1s}(\widetilde{\delta})+\nabla_{\delta_{1k}}\widetilde{f}_{1t}\left(\xi_*\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\widetilde{\delta})+(1-\xi_*)\right).\end{aligned}$$

The preceding display is $O_p(T^{-1/2})$ because $\nabla_{\delta_{2k}}\hat{\beta}_s(\widetilde{\delta})=O_p(T^{-1/2})$ and $\nabla_{\delta_{2k}}\hat{\delta}_{1s}(\widetilde{\delta})=O_p(T^{-1/2})$ for $s\neq k$, and $\xi_*\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\widetilde{\delta})+(1-\xi_*)=O_p(T^{-1/2})$. Therefore, the second term in (A.28) is also negligible. The fourth and fifth terms are also $O_p(T^{-1/2})$ after applying (A.33) to $\nabla_{\delta_{2k}}\hat{B}_t$.

Combining the above results for the five terms, we can rewrite (A.28) as

$$\widetilde{I}\left[\begin{array}{c}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})\\ \xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]=-\left[\begin{array}{c}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_\beta\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{U}_{kl,t}\\ \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_1}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{U}_{kl,t}\end{array}\right]+O_p\left(T^{-1/2}\right).$$

Because of Assumption 2(i) (the uniqueness of the maximizer), Assumption 3(i) (the uniform convergence of the criterion function), and the compactness of the parameter space, the restricted MLE $\widetilde{\theta}$ is a consistent estimator of θ_*. Thus, for any $\varepsilon>0$, $P(||\widetilde{\theta}-\theta_*||>\varepsilon)\to 0$ as $T\to\infty$. Given this result and Assumption 3(i) (i.e., $T^{-1}\sum_{t=1}^{T}\nabla_{(\beta',\delta')'}\log f_t(\beta,\delta)\nabla_{(\beta',\delta')}\log f_t(\beta,\delta)$ is positive definite in an open neighborhood of θ_* for large T), we have $\widetilde{I}^{-1}=O_p(1)$. Multiplying both sides of the preceding displayed equation by $\widetilde{I}^{-1}$, we obtain the desired result. ■

Proof of the third result in Lemma A.3. View the quantities in (A.28) as functions of δ_2, p

and q and differentiate them with respect to the h-th element of δ_2 ($h = 1, ..., n_\delta$):

$$
\begin{aligned}
\mathcal{M}^{(4)}_{jklh}(p,q,\delta_2) \quad = \quad & \frac{1}{T}\sum_{t=1}^{T}\left\{ \frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{M}_{jt}}{\hat{B}_t} - \frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{M}_{jt}\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^2} \right. \qquad \text{(A.34)} \\
& -\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2h}}\hat{M}_{jt}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2} - \frac{\nabla_{\delta_{2k}}\hat{M}_{jt}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^2} + 2\frac{\nabla_{\delta_{2k}}\hat{M}_{jt}\nabla_{\delta_{2l}}\hat{B}_t\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^3} \\
& -\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^2}\hat{M}_{jt} - \frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2}\nabla_{\delta_{2h}}\hat{M}_{jt} + 2\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^3}\hat{M}_{jt} \\
& -\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^2}\nabla_{\delta_{2l}}\hat{M}_{jt} - \frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t^2}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{M}_{jt} + 2\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^3}\nabla_{\delta_{2l}}\hat{M}_{jt} \\
& +2\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2h}}\hat{B}_t\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^3}\hat{M}_{jt} + 2\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^3}\hat{M}_{jt} \\
& \left. +2\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^3}\nabla_{\delta_{2h}}\hat{M}_{jt} - 6\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2l}}\hat{B}_t\nabla_{\delta_{2h}}\hat{B}_t}{\hat{B}_t^4}\hat{M}_{jt}. \right\} = 0.
\end{aligned}
$$

Among the fifteen terms, only the 1st and the 6th terms involve third order derivatives. They will be analyzed later. Among the remaining terms, we have the following five cases: (1) The 4th, 7th and 9th terms involve second order derivatives of $\hat{B}_t$ and first order derivatives of $\hat{M}_{jt}$, which lead to: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\widetilde{f}_{bt}/\widetilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{i_2}}\widetilde{f}_{bt}/\widetilde{f}_t)\nabla_{\theta_{j_1}}\widetilde{\xi}_{t|t-1}\nabla_{\theta_{j_2}}\widetilde{\xi}_{t|t-1}$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\widetilde{f}_{bt}/\widetilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\widetilde{f}_{bt}/\widetilde{f}_t)\nabla_{\theta_{j_2}}\widetilde{\xi}_{t|t-1}$ and $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\widetilde{f}_{at}/\widetilde{f}_t)\nabla_{\theta_{j_1}}\widetilde{\xi}_{t|t-1}$, where $1 \le i_1, i_2, j_1, j_2 \le n_\beta + 2n_\delta$, $a = 1, 2$ and $b = 1, 2$. They are all $O_p(1)$ by Lemma A.2. (2) The 2rd, 3rd and 10th terms consist of first order derivatives of $\hat{B}_t$ and second order derivatives of $\hat{M}_{jt}$. They lead to: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\nabla_{\theta_{j_3}}\widetilde{f}_{bt}/\widetilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\widetilde{f}_{bt}/\widetilde{f}_t)\nabla_{\theta_{j_2}}\widetilde{\xi}_{t|t-1}$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{i_2}}\widetilde{f}_{bt}/\widetilde{f}_t)\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\widetilde{\xi}_{t|t-1}$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\widetilde{f}_{bt}/\widetilde{f}_t)$ and $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\widetilde{f}_{bt}/\widetilde{f}_t)\nabla_{\theta_{j_2}}\widetilde{\xi}_{t|t-1}$, which are all $O_p(1)$. These three terms are thus $O_p(T^{-1/2})$ after applying (A.33) to the first order derivatives of $\hat{B}_t$. (3) The 5th, 11th and 14th terms consist of: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{i_2}}\widetilde{f}_{bt}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\widetilde{f}_{ct}/\widetilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{i_2}}\widetilde{f}_{bt}/\widetilde{f}_t)(\nabla_{\theta_{i_3}}\widetilde{f}_{ct}/\widetilde{f}_t)\nabla_{\theta_{j_1}}\widetilde{\xi}_{t|t-1}$, which are all $O_p(1)$. Consequently, these three terms are $O_p(T^{-1/2})$ after applying (A.33). (4) The 8th, 12th and 13th terms lead to: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\widetilde{f}_{bt}/\widetilde{f}_t)(\nabla_{\theta_{j_2}}\widetilde{f}_{ct}/\widetilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{i_2}}\widetilde{f}_{bt}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\widetilde{f}_{ct}/\widetilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{i_2}}\widetilde{f}_{bt}/\widetilde{f}_t)(\nabla_{\theta_{i_3}}\widetilde{f}_{ct}/\widetilde{f}_t)\nabla_{\theta_{j_1}}\widetilde{\xi}_{t|t-1}$, which are all $O_p(1)$. (5) The 15th term consists of $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{i_2}}\widetilde{f}_{bt}/\widetilde{f}_t)(\nabla_{\theta_{i_3}}\widetilde{f}_{ct}/\widetilde{f}_t)(\nabla_{\theta_{i_4}}\widetilde{f}_{ct}/\widetilde{f}_{ct})$. This term is $O_p\left(T^{-1/2}\right)$ after applying (A.33).

To analyze the remaining two terms in (A.34), we need third order derivatives of $\hat{M}_{jt}$ and $\hat{B}_t$:

$$
\begin{aligned}
&\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{M}_{jt} \qquad (A.35)\\
= &\sum_{u=1}^{n_\beta+2n_\delta}\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}+\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{1t}+\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}\\
&+\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t}-\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{2t}\\
&+(1-\hat{\xi}_{t|t-1})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t}-\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t}\\
&+(\nabla_{\theta_j}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_j}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\nabla_{\theta_j}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_j}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}+(\nabla_{\theta_j}\hat{f}_{1t}-\nabla_{\theta_j}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_u}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}\\
&+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)\nabla_{\delta_{2h}}\hat{\theta}(\delta_2)\\
&+\sum_{u=1}^{n_\beta+2n_\delta}\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{1t}+\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{2t}\\
&+(1-\hat{\xi}_{t|t-1})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t}+(\nabla_{\theta_j}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_j}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta_j}\hat{f}_{1t}-\nabla_{\theta_j}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_u}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1}+(\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\hat{\xi}_{t|t-1}\\
&+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}\Big\}[\nabla_{\delta_{2k}}\nabla_{\delta_{2h}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)+\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\theta}_u(\delta_2)]\\
&+\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}+\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{1t}+(1-\hat{\xi}_{t|t-1})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t}\\
&-\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\hat{f}_{2t}+(\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\nabla_{\theta_j}\hat{f}_{1t}-\nabla_{\theta_j}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2h}}\hat{\theta}(\delta_2)\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\delta_2),\\
&+\Big\{\hat{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{1t}+(1-\xi_{t|t-1})\nabla_{\theta_j}\nabla_{\theta'}\hat{f}_{2t}+(\nabla_{\theta_j}\hat{f}_{1t}-\nabla_{\theta_j}\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_j}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_j}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\theta}(\delta_2),
\end{aligned}
$$

and

$$
\begin{aligned}
& \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{B}_t \\
= & \sum_{s=1}^{n_\beta+2n_\delta}\sum_{u=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{1t}+\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}+\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t} \\
& +\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{2t}-\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t}+(1-\hat{\xi}_{t|t-1})\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t} \\
& -\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t}+(\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
& +(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2h}}\hat{\theta}(\delta_2) \\
& +\sum_{s=1}^{n_\beta+2n_\delta}\sum_{u=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{1t}+\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{2t}+(1-\hat{\xi}_{t|t-1})\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t} \\
& +(\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}\Big\}\times \\
& \Big[\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)+\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)\nabla_{\delta_{2h}}\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\Big] \\
& +\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{1t}+\hat{\xi}_{t|t-1}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}+(1-\hat{\xi}_{t|t-1})\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t}-\nabla_{\theta'}\hat{\xi}_{t|t-1}\nabla_{\theta_s}\hat{f}_{2t} \\
& +(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2h}}\hat{\theta}(\delta_2)\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\delta_2) \\
& +\Big[\hat{\xi}_{t|t-1}\nabla_{\theta'}\hat{f}_{1t}+(1-\hat{\xi}_{t|t-1})\nabla_{\theta'}\hat{f}_{2t}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big]\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\theta}(\delta_2).
\end{aligned}
$$

Consider the 1st term in (A.34). In the expression of $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{M}_{jt}$, only the last two lines involve third order derivatives of $\hat{\theta}(\delta_2)$. These derivatives are multiplied by (after division by $\widetilde{f}_t$): $T^{-1}\sum_{t=1}^{T}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\widetilde{f}_{at}/\widetilde{f}_t$ and $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)\nabla_{\theta_{j_1}}\widetilde{\xi}_{t|t-1}$, where $a=1,2$. They are $O_p(T^{-1/2})$ by Lemma A.2. The remaining components of $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{M}_{jt}$ lead to: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\widetilde{f}_{at}/\widetilde{f}_t)$ for $a=1,2$ and $k\leq 4$ and $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\widetilde{\xi}_{t|t-1})$ for $a=1,2$ and $k+m\leq 4$. They are all $o_p(1)$ by Lemma A.2. Therefore the contribution of $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{M}_{jt}$ to (A.34) is $o_p(1)$. Finally, we turn to the 6th term in (A.34). In the expression for $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{B}_t$, only the final line involves third order derivatives of $\hat{\theta}(\delta_2)$. It can be analyzed in the same way as the second term in (A.19); see Step 2 of the proof there. The remaining components, multiplied by $\hat{M}_{jt}$, lead to: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{i_2}}\widetilde{f}_{bt}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\widetilde{\xi}_{t|t-1})$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\widetilde{f}_{bt}/\widetilde{f}_t)$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\widetilde{f}_{bt}/\widetilde{f}_t)(\nabla_{\theta_{j_2}}\widetilde{\xi}_{t|t-1})$, $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\widetilde{f}_{bt}/\widetilde{f}_t)$ and $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{i_2}}\widetilde{f}_{bt}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}\widetilde{\xi}_{t|t-1})$ for $a=1,2$ and $b=1,2$. They are all $O_p(1)$ by Lemma A.2. This implies the desired result. ■

Proof of Lemma 2. The first order derivative with respect to the j-th element of δ_2 satisfies

$$\begin{aligned}
\mathcal{L}_j^{(1)}(p,q,\delta_2) &= \sum_{t=1}^{T} \frac{1}{\hat{B}_t}\left(\nabla_{\theta'}\hat{f}_{1t}\hat{\xi}_{t|t-1} + \nabla_{\theta'}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1}) + (\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta'}\hat{\xi}_{t|t-1}\right)\nabla_{\delta_{2j}}\hat{\theta}(\delta_2) \\
&= \sum_{s=1}^{n_\beta+n_\delta}\left\{\sum_{t=1}^{T}\frac{1}{\hat{B}_t}\left(\nabla_{\theta_s}\hat{f}_{1t}\hat{\xi}_{t|t-1} + \nabla_{\theta_s}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1}) + (\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\theta_s}\hat{\xi}_{t|t-1}\right)\right\}\nabla_{\delta_{2j}}\hat{\theta}_s(\delta_2) \\
&\quad + \sum_{t=1}^{T}\frac{1}{\hat{B}_t}\left(\nabla_{\delta_{2j}}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1}) + (\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}\right),
\end{aligned}$$

where the second equality follows from the definition of $\nabla_{\delta_{2j}}\hat{\theta}(\delta_2)$; see (A.22). The term inside the curly brackets equals zero because of the first order conditions determining $\hat{\beta}(\delta_2)$ and $\hat{\delta}_1(\delta_2)$. Therefore, we can write

$$\mathcal{L}_j^{(1)}(p,q,\delta_2) = \sum_{t=1}^{T}\frac{\hat{L}_{jt}}{\hat{B}_t},$$

where $\hat{B}_t$ is defined in (A.18) and $\hat{L}_{jt} = \nabla_{\delta_{2j}}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1}) + (\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}$. The following results hold at $\delta_2 = \widetilde{\delta}$: $\widetilde{\xi}_{t|t-1} = \xi_*, \hat{\delta}_1(\widetilde{\delta}) = \widetilde{\delta}$ and $\hat{\beta}(\widetilde{\delta}) = \widetilde{\beta}$. Consequently, $\mathcal{L}_j^{(1)}(p,q,\widetilde{\delta}) = (1-\xi_*)\sum_{t=1}^{T}(\nabla_{\delta_{2j}}\widetilde{f}_{2t}/\widetilde{f}_t) = 0$, where the last equality follows because $\widetilde{\delta}$ is the MLE of the null likelihood. This proves the first result in the lemma.

Now consider the second result. Because $\widetilde{\xi}_{t|t-1} = \xi_*$, the following identity holds at $\delta_2 = \widetilde{\delta}$:

$$\hat{L}_{jt} = [(1-\xi_*)/\xi_*]\hat{M}_{(n_\beta+j)t}. \tag{A.36}$$

This relationship generalizes an analogous result in Cho and White (2007, p. 1683-1684, c.f. the relationship between $h_t(\theta_2)$ and $k_t(\theta_2)$) to Markov switching models. It allows us to relate $\mathcal{L}_{jk}^{(2)}(p,q,\delta_2)$ to $\mathcal{M}_{(n_\beta+j)k}^{(2)}(p,q,\delta_2)$ when analyzing the former's properties. Specifically, we differentiate $\mathcal{L}_j^{(1)}(p,q,\delta_2)$ with respect to the k-th element of δ_2:

$$\mathcal{L}_{jk}^{(2)}(p,q,\delta_2) = \sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t} - \sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t^2}\hat{L}_{jt},$$

where

$$\begin{aligned}
\nabla_{\delta_{2k}}\hat{L}_{jt} &= \{\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1}) - \nabla_{\delta_{2j}}\hat{f}_{2t}\nabla_{\theta'}\hat{\xi}_{t|t-1} \\
&\quad + (\nabla_{\theta'}\hat{f}_{1t} - \nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1} + (\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{\xi}_{t|t-1}\}\nabla_{\delta_{2k}}\hat{\theta}(\delta_2).
\end{aligned} \tag{A.37}$$

Because $\mathcal{M}_{(n_\beta+j)k}^{(2)}(p,q,\delta_2) = 0$, we have

$$\begin{aligned}
T^{-1/2}\mathcal{L}_{jk}^{(2)}(p,q,\delta_2) &= T^{-1/2}\mathcal{L}_{jk}^{(2)}(p,q,\delta_2) - T^{1/2}\left(\frac{1-\xi_*}{\xi_*}\right)\mathcal{M}_{(n_\beta+j)k}^{(2)}(p,q,\delta_2) \\
&= T^{-1/2}\sum_{t=1}^{T}\left\{\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t} - \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\} \\
&\quad - T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t^2}\left\{\hat{L}_{jt} - \left(\frac{1-\xi_*}{\xi_*}\right)\hat{M}_{(n_\beta+j)t}\right\},
\end{aligned} \tag{A.38}$$

where the second summation on the right hand side equals 0 at $\delta_2 = \widetilde{\delta}$ because of (A.36). Now consider the two terms in the first summation separately. At $\delta_2 = \widetilde{\delta}$, $T^{-1/2}\sum_{t=1}^{T}\nabla_{\delta_{2k}}\hat{L}_{jt}/\hat{B}_t$ equals

$$T^{-1/2}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\left\{(1-\xi_*)\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t}+\frac{1}{\xi_*}\nabla_{\delta_{1j}}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}+\frac{1}{\xi_*}\nabla_{\delta_{1k}}\widetilde{f}_{1t}\nabla_{\delta_{2j}}\widetilde{\xi}_{t|t-1}\right\}+O_p(T^{-1/2}).$$

Meanwhile, at $\delta_2 = \widetilde{\delta}$, $T^{-1/2}\sum_{t=1}^{T}\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}/\hat{B}_t$ equals

$$-T^{-1/2}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\left\{(1-\xi_*)\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t}+\frac{1}{\xi_*}\nabla_{\delta_{1j}}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}+\frac{1}{\xi_*}\nabla_{\delta_{1k}}\widetilde{f}_{1t}\nabla_{\delta_{2j}}\widetilde{\xi}_{t|t-1}\right\}+O_p(T^{-1/2}).$$

The result follows by combining the above two displays.

Consider the third order derivatives. Using (A.38), we have

$$\begin{aligned}
&T^{-3/4}\mathcal{L}^{(3)}_{jkl}(p,q,\delta_2)-T^{1/4}\left(\frac{1-\xi_*}{\xi_*}\right)\mathcal{M}^{(3)}_{(n_\beta+j)kl}(p,q,\delta_2) \qquad\qquad \text{(A.39)}\\
=\;&T^{-3/4}\sum_{t=1}^{T}\left\{\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
&-T^{-3/4}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
&-T^{-3/4}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2l}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2l}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
&-T^{-3/4}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2}\left\{\hat{L}_{jt}-\left(\frac{1-\xi_*}{\xi_*}\right)\hat{M}_{(n_\beta+j)t}\right\}\\
&+2T^{-3/4}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^3}\left\{\hat{L}_{jt}-\left(\frac{1-\xi_*}{\xi_*}\right)\hat{M}_{(n_\beta+j)t}\right\},
\end{aligned}$$

where the last two summations equal 0 because of (A.36) and

$$\begin{aligned}
&\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{L}_{jt} \qquad\qquad \text{(A.40)}\\
=\;&\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1})-\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{2t}\nabla_{\theta'}\hat{\xi}_{t|t-1}-\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{2t}\nabla_{\theta_s}\hat{\xi}_{t|t-1}\\
&-\nabla_{\delta_{2j}}\hat{f}_{2t}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}(\delta_2)\\
&+\Big\{\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1})-\nabla_{\delta_{2j}}\hat{f}_{2t}\nabla_{\theta'}\hat{\xi}_{t|t-1}+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}\\
&+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\delta_2).
\end{aligned}$$

The first summation in (A.39) consists of the following: $T^{-3/4}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_u}}\widetilde{f}_{at}/\widetilde{f}_t)\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_v}}\widetilde{\xi}_{t|t-1}$ with $u+v\leq 3$. They are $O_p(T^{-1/4})$ by the first result in Lemma A.2. Combining this result with Lemma A.3, it follows that this summation is $O_p(T^{-1/4})$. The remaining two summations in (A.39) have the same structure. They are both $O_p(T^{-1/4})$ after applying (A.33).

Consider the fourth order derivatives. Applying (A.39) and omitting terms that are zero implied by (A.36), we have

$$\begin{aligned}
& T^{-1}\mathcal{L}^{(4)}_{jklm}(p,q,\delta_2)-\left(\frac{1-\xi_*}{\xi_*}\right)\mathcal{M}^{(4)}_{(n_\beta+j)klm}(p,q,\delta_2) \qquad \text{(A.41)}\\
= \; & T^{-1}\sum_{t=1}^{T}\left\{\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2m}}\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& +2T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2m}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2m}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2l}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2l}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& -T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\}\\
& +2T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2l}}\hat{L}_{jt}\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t^2}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2m}}\hat{B}_t\nabla_{\delta_{2l}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t^2}\right\}\\
& +2T^{-3/4}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\hat{B}_t\nabla_{\delta_{2l}}\hat{B}_t}{\hat{B}_t^2}\left\{\frac{\nabla_{\delta_{2m}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2m}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\},
\end{aligned}$$

where

$$
\begin{aligned}
&\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{L}_{jt} \qquad \text{(A.42)}\\
= &\sum_{u=1}^{n_\beta+2n_\delta}\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1})+\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&-\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t}\nabla_{\theta_u}\hat{\xi}_{t|t-1}-\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{2t}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&-\nabla_{\delta_{2j}}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t}\nabla_{\theta_s}\hat{\xi}_{t|t-1}-\nabla_{\delta_{2j}}\nabla_{\theta_u}\hat{f}_{2t}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&-\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{2t}\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}-\nabla_{\delta_{2j}}\hat{f}_{2t}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_u}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_u}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\times\\
&\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)\nabla_{\delta_{2m}}\hat{\theta}(\delta_2)\\
&+\sum_{u=1}^{n_\beta+2n_\delta}\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1})-\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{2t}\nabla_{\theta_u}\hat{\xi}_{t|t-1}-\nabla_{\delta_{2j}}\nabla_{\theta_u}\hat{f}_{2t}\nabla_{\theta_s}\hat{\xi}_{t|t-1}\\
&-\nabla_{\delta_{2j}}\hat{f}_{2t}\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_u}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta_u}\hat{f}_{1t}-\nabla_{\theta_u}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\hat{\xi}_{t|t-1}\Big\}\times\\
&\Big[\nabla_{\delta_{2m}}\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\hat{\theta}_u(\delta_2)+\nabla_{\delta_{2k}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_u(\delta_2)\Big]\\
&+\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1})-\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{2t}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&-\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{f}_{2t}\nabla_{\theta_s}\hat{\xi}_{t|t-1}-\nabla_{\delta_{2j}}\hat{f}_{2t}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta'}\hat{\xi}_{t|t-1}\\
&+(\nabla_{\theta'}\hat{f}_{1t}-\nabla_{\theta'}\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta'}\hat{\xi}_{t|t-1}\Big\}\times\\
&\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2m}}\hat{\theta}(\delta_2)\\
&+\sum_{s=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{f}_{2t}(1-\hat{\xi}_{t|t-1})-\nabla_{\delta_{2j}}\hat{f}_{2t}\nabla_{\theta_s}\hat{\xi}_{t|t-1}+(\nabla_{\theta_s}\hat{f}_{1t}-\nabla_{\theta_s}\hat{f}_{2t})\nabla_{\delta_{2j}}\hat{\xi}_{t|t-1}\\
&+(\hat{f}_{1t}-\hat{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\hat{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\delta_2).
\end{aligned}
$$

We consider the terms in (A.41) separately. The first summation involves the following quantities: $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\widetilde{f}_{at}/\widetilde{f}_t)$ for $k=2,3,4$ and $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\widetilde{f}_{at}/\widetilde{f}_t)(\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\widetilde{\xi}_{t|t-1})$ for $2\leq k+m\leq 4$. They are all $o_p(1)$. Consequently the first summation is also $o_p(1)$. The 2nd, 4th, 5th, 8th, 9th and 10th terms involve first order derivatives of $\hat{B}_t$, and are $o_p(1)$ because of the relationship (A.33). The remaining three terms have the same structure. It suffices to analyze the first of them:

$$-T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t}\left\{\frac{\nabla_{\delta_{2k}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right\} \tag{A.43}$$

Further, for $\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{B}_t$, it suffices to consider $((1-\xi_*)/\xi_*)\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}+(1/\xi_*^2)\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1m}}\widetilde{\xi}_{t|t-1}+(1/\xi_*^2)\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1m}}\widetilde{f}_{1t}+\sum_{s=1}^{n_\beta+n_\delta}(\xi_*\nabla_{\theta_s}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\widetilde{f}_{2t})\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})$. For $\nabla_{\delta_{2k}}\hat{L}_{jt}$, it suffices to consider $(1-\xi_*)\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t}+(1/\xi_*)\nabla_{\delta_{1j}}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}+(1/\xi_*)\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}$. For $\nabla_{\delta_{2k}}\hat{M}_{(n_\beta+j)t}$, it suffices to consider $-(1-\xi_*)\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t}-(1/\xi_*)\nabla_{\delta_{1j}}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}-(1/\xi_*)\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}$. Combining the above three formulas, we have, at $\widetilde{\delta}$, (A.43) equals

$$\begin{aligned}
& -T^{-1}\sum_{t=1}^{T}\left[\widetilde{U}_{lm,t}+\sum_{s=1}^{n_\beta+n_\delta}\frac{\xi_*\nabla_{\theta_s}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\widetilde{f}_{2t}}{\widetilde{f}_t}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})\right]\widetilde{U}_{jk,t}+o_p(1)\\
= & -T^{-1}\sum_{t=1}^{T}\widetilde{U}_{lm,t}\widetilde{U}_{jk,t}-T^{-1}\sum_{t=1}^{T}\left\{\widetilde{U}_{jk,t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\right\}\widetilde{I}^{-1}\widetilde{D}_{lm}+o_p(1)\\
= & -\left\{\widetilde{V}_{jklm}-\widetilde{D}_{jk}'\widetilde{I}^{-1}\widetilde{D}_{lm}\right\}+o_p(1),
\end{aligned}$$

where the first equality uses Lemma A.3.2 and the second applies (17). Consequently,

$$\begin{aligned}
& T^{-1}\mathcal{L}_{jklm}^{(4)}(p,q,\widetilde{\delta})-T^{-1}\left(\frac{1-\xi_*}{\xi_*}\right)\mathcal{M}_{(n_\beta+j)klm}^{(4)}(p,q,\widetilde{\delta})\\
= & -\left\{\widetilde{V}_{jklm}-\widetilde{D}_{jk}'\widetilde{I}^{-1}\widetilde{D}_{lm}+\widetilde{V}_{jmkl}-\widetilde{D}_{jm}'\widetilde{I}^{-1}\widetilde{D}_{kl}+\widetilde{V}_{kmjl}-\widetilde{D}_{km}'\widetilde{I}^{-1}\widetilde{D}_{jl}\right\}+o_p(1).
\end{aligned}$$

This proves the final result of the lemma. ■

A.2 Proofs and illustrations for Section 5

The next lemma is used in the proof of Lemma A.5 to establish the stochastic equicontinuity. We use "*" to signify the true parameter value.

Lemma A.4 *Let Assumptions 1-5 and the null hypothesis hold. Let* $z_t(\rho)=T^{-1/2}\sum_{s=1}^{t-1}\rho^{t-s}\varepsilon_{js}\varepsilon_{it}$, *where* $\varepsilon_{it}=\nabla_{\delta_{1i}}f_{1t}^*/f_t^*$ *and* $\varepsilon_{js}=\nabla_{\delta_{1j}}f_{1s}^*/f_s^*$. *Then, for any* ρ, ρ_1 *and* ρ_2 *satisfying* $\epsilon-1\le\rho_1\le\rho\le\rho_2\le 1-\epsilon$, *we have*

$$E\left(\left|\sum_{t=1}^{T}[z_t(\rho)-z_t(\rho_1)]\right|^2\left|\sum_{t=1}^{T}[z_t(\rho_2)-z_t(\rho)]\right|^2\right)\le C(\rho_2-\rho_1)^2, \tag{A.44}$$

where C *is a finite constant that depends only on* $0<\epsilon<1/2$ *and the moments of* ε_{it} *and* ε_{js} *up to the fourth order.*

Proof. Let $c_{t-s}(\rho) = T^{-1/2}\rho^{t-s}$, $c_{t-s}(\rho_1, \rho) = c_{t-s}(\rho) - c_{t-s}(\rho_1)$ and $z_t(s, r) = z_t(r) - z_t(s)$. We first show that the left hand side of (A.44) is bounded from above by

$$C\left(\sum_{t=1}^{T}\sum_{s=1}^{t-1} c_{t-s}(\rho_1,\rho)^2\right)\left(\sum_{t=1}^{T}\sum_{h=1}^{t-1} c_{t-h}(\rho,\rho_2)^2\right). \tag{A.45}$$

Because ε_{js} and ε_{it} are martingale differences, the left hand side equals

$$\begin{aligned}
& E\sum_{t=1}^{T} z_t(\rho_1,\rho)^2 z_t(\rho,\rho_2)^2 + E\sum_{t=1}^{T}\sum_{k=1,k\neq t}^{T} z_t(\rho_1,\rho)^2 z_k(\rho,\rho_2)^2 \\
& +2E\sum_{t=1}^{T}\sum_{l=1,l\neq t}^{T} z_t(\rho_1,\rho) z_l(\rho_1,\rho) z_t(\rho,\rho_2) z_l(\rho,\rho_2) \\
= \;& \text{(T.1)+(T.2)+(T.3)}.
\end{aligned}$$

We analyze the three terms separately:

$$\begin{aligned}
\text{(T.2)} \;=\;& E\sum_{t=1}^{T}\varepsilon_{it}^2\left(\sum_{s=1}^{t-1} c_{t-s}(\rho_1,\rho)\varepsilon_{js}\right)^2 \sum_{k=1}^{t-1}\varepsilon_{ik}^2\left(\sum_{h=1}^{k-1} c_{k-h}(\rho,\rho_2)\varepsilon_{jh}\right)^2 \\
& +E\sum_{k=1}^{T}\varepsilon_{ik}^2\left(\sum_{h=1}^{k-1} c_{k-h}(\rho,\rho_2)\varepsilon_{jh}\right)^2 \sum_{t=1}^{k-1}\varepsilon_{it}^2\left(\sum_{s=1}^{t-1} c_{t-s}(\rho_1,\rho)\varepsilon_{js}\right)^2.
\end{aligned}$$

Due to symmetry, it suffices to consider the first term on the right hand side, which equals

$$\begin{aligned}
& C_1\sum_{t=1}^{T}\sum_{k=1}^{t-1} E\left(\sum_{s=1}^{t-1} c_{t-s}(\rho_1,\rho)^2\varepsilon_{js}^2\right)\varepsilon_{ik}^2\left(\sum_{h=1}^{k-1} c_{k-h}(\rho,\rho_2)^2\varepsilon_{jh}^2\right) && \text{(I)} \\
& +C_1\sum_{t=1}^{T}\sum_{k=1}^{t-1}\sum_{s=1}^{k-1}\sum_{h=1,h\neq s}^{k-1} c_{t-s}(\rho_1,\rho)c_{t-h}(\rho_1,\rho)c_{k-s}(\rho,\rho_2)c_{k-h}(\rho,\rho_2) && \text{(II)}
\end{aligned}$$

for some $0 < C_1 < \infty$, where C_1 depends only on $E\varepsilon_{it}^2$ and $E\varepsilon_{jt}^2$. (Below, the finite constants C_s $(s = 2, 3, 4, 5)$ also depend only on the moments of ε_{jt} and ε_{it}, up to the fourth order). Term (I) is further bounded by

$$C_2\sum_{t=1}^{T}\sum_{k<t}(\sum_{s=1}^{t-1} c_{t-s}(\rho_1,\rho)^2)(\sum_{h=1}^{k-1} c_{k-h}(\rho,\rho_2)^2) \leq C_2\sum_{t=1}^{T}(\sum_{s=1}^{t-1} c_{t-s}(\rho_1,\rho)^2)\sum_{t=1}^{T}(\sum_{h=1}^{t-1} c_{t-h}(\rho,\rho_2)^2). \tag{A.46}$$

Applying the Cauchy-Schwarz inequality to the elements of (II), we have $\sum_{s=1}^{k-1}|c_{t-s}(\rho_1,\rho)c_{k-s}(\rho,\rho_2)| \leq (\sum_{s=1}^{k-1} c_{t-s}(\rho_1,\rho)^2)^{1/2}(\sum_{s=1}^{k-1} c_{k-s}(\rho,\rho_2)^2)^{1/2}$ and $\sum_{h=1}^{k-1}|c_{t-h}(\rho_1,\rho)c_{k-h}(\rho,\rho_2)| \leq (\sum_{h=1}^{k-1} c_{t-h}(\rho_1,\rho)^2)^{1/2}$ $(\sum_{h=1}^{k-1} c_{k-h}(\rho,\rho_2)^2)^{1/2}$. Combining these two inequalities:

$$\begin{aligned}
|\text{(II)}| \;\leq\;& C_1\sum_{t=1}^{T}\sum_{k<t}\left(\sum_{s=1}^{k-1} c_{t-s}(\rho_1,\rho)^2\right)\left(\sum_{h=1}^{k-1} c_{k-h}(\rho,\rho_2)^2\right) \qquad \text{(A.47)} \\
\leq\;& C_1\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1} c_{t-s}(\rho_1,\rho)^2\right)\sum_{t=1}^{T}\left(\sum_{h=1}^{t-1} c_{t-h}(\rho,\rho_2)^2\right),
\end{aligned}$$

which is proportional to (A.46). Hence,

$$|(\mathrm{I}){+}(\mathrm{II})| \leq C_3 \sum_{t=1}^{T} \left(\sum_{s=1}^{t-1} c_{t-s}(\rho_1, \rho)^2 \right) \sum_{t=1}^{T} \left(\sum_{h=1}^{t-1} c_{t-h}(\rho, \rho_2)^2 \right) . \tag{A.48}$$

Apply the Cauchy-Schwarz inequality to (T.3):

$$\begin{aligned} |(\mathrm{T.3})| &\leq E \left(\sum_{t=1}^{T} z_t(\rho_1, \rho)^2 \right) \left(\sum_{t=1}^{T} z_t(\rho, \rho_2)^2 \right) \\ &= E \left(\sum_{t=1}^{T} z_t(\rho_1, \rho)^2 \sum_{k=1, k \neq t}^{T} z_k(\rho, \rho_2)^2 \right) + E \sum_{t=1}^{T} z_t(\rho_1, \rho)^2 z_t(\rho, \rho_2)^2, \end{aligned} \tag{A.49}$$

where the first term is the same as (T.2) and the second term equals (T.1). Consequently, a separate analysis of (T.3) is not needed.

Finally, we turn to (T.1). It equals

$$\begin{aligned} & E\varepsilon_{it}^4 \sum_{t=1}^{T} \left(E \sum_{s=1}^{t-1} \sum_{k=1}^{t-1} \sum_{h=1}^{t-1} \sum_{l=1}^{t-1} c_{t-s}(\rho_1, \rho) c_{t-k}(\rho_1, \rho) c_{t-h}(\rho, \rho_2) c_{t-l}(\rho, \rho_2) \varepsilon_{js} \varepsilon_{jk} \varepsilon_{jh} \varepsilon_{jl} \right) \\ = \; & E\varepsilon_{it}^4 E\varepsilon_{jt}^4 \sum_{t=1}^{T} \left(\sum_{s=1}^{t-1} c_{t-s}(\rho_1, \rho)^2 c_{t-s}(\rho, \rho_2)^2 \right) + E\varepsilon_{it}^4 \left(E\varepsilon_{jt}^2 \right)^2 \sum_{t=1}^{T} \left(\sum_{s=1}^{t-1} \sum_{h=1, h \neq s}^{t-1} c_{t-s}(\rho_1, \rho)^2 c_{t-h}(r, r_2)^2 \right) \\ & + 2E\varepsilon_{it}^4 \left(E\varepsilon_{jt}^2 \right)^2 \sum_{t=1}^{T} \left(\sum_{s=1}^{t-1} c_{t-s}(\rho_1, \rho) c_{t-s}(\rho, \rho_2) \sum_{k=1, k \neq s}^{t-1} c_{t-k}(\rho_1, \rho) c_{t-k}(\rho, \rho_2) \right) \\ \leq \; & C_4 \sum_{t=1}^{T} \left(\sum_{s=1}^{t-1} \sum_{h=1}^{t-1} c_{t-s}(\rho_1, \rho)^2 c_{t-h}(\rho, \rho_2)^2 \right) \qquad \text{(III)} \\ & + C_4 \sum_{t=1}^{T} \left(\sum_{s=1}^{t-1} c_{t-s}(\rho_1, \rho) c_{t-s}(\rho, \rho_2) \sum_{k=1, k \neq s}^{t-1} c_{t-k}(\rho_1, \rho) c_{t-k}(\rho, \rho_2) \right) . \quad \text{(IV)} \end{aligned}$$

As in (A.47), we have $|(\mathrm{IV})| \leq C_4 \sum_{t=1}^{T} \sum_{s=1}^{t-1} c_{t-s}(\rho_1, \rho)^2 \sum_{k=1}^{t-1} c_{t-k}(\rho, \rho_2)^2$. Hence,

$$|(\mathrm{III}){+}(\mathrm{IV})| \leq C_5 \left(\sum_{t=1}^{T} \sum_{s=1}^{t-1} c_{t-s}(\rho_1, \rho)^2 \right) \left(\sum_{t=1}^{T} \sum_{h=1}^{t-1} c_{t-h}(\rho, \rho_2)^2 \right) . \tag{A.50}$$

Combining (A.48), (A.49), and (A.50) leads to (A.45).

By the mean value theorem: $c_{t-s}(\rho_1, \rho) = T^{-1/2}(\rho^{t-s} - \rho_1^{t-s}) \leq T^{-1/2}(t-s)(1-2\epsilon)^{t-s-1}(\rho - \rho_1)$. The right hand side of (A.45) is therefore bounded by $C\{T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{t-1} (t-s)^2 (1-2\epsilon)^{2(t-s-1)}\}^2 (\rho_2 - \rho_1)^2$. The term in the curly brackets is finite; the result follows after redefining the constant C. ■

Lemma A.5 *Let the null hypothesis and Assumptions 1-5 hold. Then, over* $\epsilon \leq p, q \leq 1 - \epsilon$ *:*

$$T^{-1/2} \mathcal{L}^{(2)}(p, q, \widetilde{\delta}) \Rightarrow G(p, q),$$

where the elements of $G(p, q)$ *are mean zero continuous Gaussian processes satisfying* $\mathrm{Cov}[G_{jk}(p_r, q_r), G_{lm}(p_s, q_s)] = \omega_{jklm}(p_r, q_r; p_s, q_s)$ *for* $j, k, l, m \in \{1, 2, \ldots, n_\delta\}$*, where* $\omega_{jklm}(p_r, q_r; p_s, q_s)$ *is given by (20).*

Proof of Lemma A.5. We prove the result by first verifying the finite-dimensional convergence and then the stochastic equicontinuity. Apply the mean value theorem:

$$T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_{jk,t} = T^{-1/2}\sum_{t=1}^{T}U_{jk,t} + \left\{T^{-1}\sum_{t=1}^{T}\nabla_{\theta'}\bar{U}_{jk,t}\right\}T^{1/2}(\widetilde{\theta}-\theta_*), \tag{A.51}$$

where $U_{jk,t}$ and $\bar{U}_{jk,t}$ have the same definition as $\widetilde{U}_{jk,t}$ but evaluated at the true value θ_* and some value $\bar{\theta} = \theta_* + c(\widetilde{\theta}-\theta_*)$ for some $c \in (0,1)$, respectively.

We establish the weak convergence of the first term of (A.51) in two steps. First, for any $\epsilon \le p$, $q \le 1-\epsilon$, $T^{-1/2}\sum_{t=1}^{T}U_{jk,t}$ satisfies the central limit theorem. Second, to verify its stochastic equicontinuity, it suffices to consider the second component in its definition (17). This term equals

$$T^{-1/2}\frac{1}{\xi_*^2}\sum_{t=1}^{T}\nabla_{\delta_{1j}}\xi_{t|t-1}\frac{\nabla_{\delta_{1k}}f_{1t}}{f_t} = \left(\frac{1-p}{1-q}\right)\left\{T^{-1/2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\rho^s\frac{\nabla_{\delta_{1j}}f_{1(t-s)}}{f_{t-s}}\right)\frac{\nabla_{\delta_{1k}}f_{1t}}{f_t}\right\}.$$

where the quantities are all evaluated at the true value θ_*, and the equality follows from (A.6) and (2). Denote the quantity inside the curly brackets as $W(\rho)$. Note that we have $|\rho| \le 1-2\epsilon$. Then, Lemma A.4 implies, for any $\rho_1 \le \rho \le \rho_2$, we have $E[|W(\rho_1)-W(\rho)|^2\,|W(\rho)-W(\rho_2)|^2] \le C\,(\rho_1-\rho_2)^2$, where C is a finite constant. This fulfills the condition required in Theorem 13.5 in Billingsley (1999; c.f. the Display (13.14) in p. 143). This shows that $W(\rho)$ is stochastic equicontinuous.

The second term in (A.51) equals, by the mean value theorem,

$$\begin{aligned} & -\left\{T^{-1}\sum_{t=1}^{T}\left(\frac{\nabla_{(\beta',\delta')}f_{1t}}{f_t}\right)U_{jk,t}\right\}I^{-1}\left\{T^{-1/2}\sum_{t=1}^{T}\left(\frac{\nabla_{(\beta',\delta')'}f_{1t}}{f_t}\right)\right\} + o_p(1) \\ = \quad & -D_{jk}I^{-1}\left\{T^{-1/2}\sum_{t=1}^{T}\left(\frac{\nabla_{(\beta',\delta')'}f_{1t}}{f_t}\right)\right\} + o_p(1), \end{aligned}$$

where the quantities are all evaluated at the true value θ_*, and the second equality holds because of the uniform law of large numbers. The term inside the last curly brackets is independent of p and q. It satisfies the central limit theorem. From these results, it follows that $T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_{jk,t}$ converges weakly over $\epsilon \le p$, $q \le 1-\epsilon$. It is simple to verify that its covariance function satisfies the stated formula; we omit the details.■

An analytical illustration of $\Omega(p_r, q_r; p_s, q_s)$ and $\Omega(p,q)$. We consider the following model for which $\Omega(p_r, q_r; p_s, q_s)$ can be computed analytically:

$$y_t = w_t\gamma_1\mathbf{1}_{\{s_t=1\}} + w_t\gamma_2\mathbf{1}_{\{s_t=2\}} + u_t,$$

where $u_t \sim i.i.d.N(0,\sigma_*^2)$ and w_t is a scalar variable that is strictly exogenous or equal to y_{t-1}. We use an asterisk to denote the true parameter value. Below, we derive the expressions of $\Omega(p_r, q_r; p_s, q_s)$ for five cases. Recall $\Omega(p,q) = \Omega(p,q;p,q)$. Let $\rho_r = p_r + q_r - 1$ and $\rho_s = p_s + q_s - 1$.

First, we consider the situation where only γ is allowed to switch and σ_*^2 is unknown. When w_t is strictly exogenous, the covariance function (20) equals

$$\frac{2(1-p_r)(1-p_s)}{(1-q_r)(1-q_s)}\frac{Var(w_t^2) + 2\sum_{k=1}^{\infty}(\rho_r\rho_s)^k E(w_t^2 w_{t-k}^2)}{\sigma_*^4}. \tag{A.52}$$

When $w_t = y_{t-1}$, it equals

$$\frac{(1-p_r)(1-p_s)}{(1-q_r)(1-q_s)}\left(\frac{1}{1-\gamma_*^2}\right)\left\{\frac{4}{1-\gamma_*^2}+\frac{4\rho_r\rho_s}{1-\gamma_*^2}\left(\frac{2}{1-\rho_r\rho_s\gamma_*^2}+\frac{1}{1-\rho_r\rho_s}\right)\right. \tag{A.53}$$
$$\left.+\frac{16\rho_r^2\rho_s\gamma_*^2}{(1-\rho_r\gamma_*^2)\,(1-\rho_r\rho_s\gamma_*^2)}+\frac{16\rho_r\rho_s^2\gamma_*^2}{(1-\rho_s\gamma_*^2)\,(1-\rho_r\rho_s\gamma_*^2)}-\frac{16\rho_r\rho_s\gamma_*^2}{(1-\rho_r\gamma_*^2)(1-\rho_s\gamma_*^2)}\right\}.$$

These two functions are different even when $w_t \sim i.i.d.N(0,1)$ and $\gamma_* = 0$. This is because $\nabla_{\gamma_1}\xi_{t|t-1}$ is independent of $\nabla_{\gamma_1} f_{1t}$ when w_t is strictly exogenous. This difference shows that the covariance function is not invariant to the dynamic properties of the model.

Now, we consider the same situations but with σ_*^2 being known. Then, when w_t is strictly exogenous, the covariance function equals

$$\frac{2(1-p_r)(1-p_s)}{(1-q_r)(1-q_s)}\frac{E(w_t^4)+2\sum_{k=1}^{\infty}(\rho_r\rho_s)^k E(w_t^2 w_{t-k}^2)}{\sigma_*^4}. \tag{A.54}$$

When $w_t = y_{t-1}$, it equals

$$\frac{(1-p_r)(1-p_s)}{(1-q_r)(1-q_s)}\left(\frac{1}{1-\gamma_*^2}\right)\left\{\frac{6}{1-\gamma_*^2}+\frac{4\rho_r\rho_s}{1-\gamma_*^2}\left(\frac{2}{1-\rho_r\rho_s\gamma_*^2}+\frac{1}{1-\rho_r\rho_s}\right)\right. \tag{A.55}$$
$$\left.+\frac{16\rho_r^2\rho_s\gamma_*^2}{(1-\rho_r\gamma_*^2)\,(1-\rho_r\rho_s\gamma_*^2)}+\frac{16\rho_r\rho_s^2\gamma_*^2}{(1-\rho_s\gamma_*^2)\,(1-\rho_r\rho_s\gamma_*^2)}-\frac{16\rho_r\rho_s\gamma_*^2}{(1-\rho_r\gamma_*^2)(1-\rho_s\gamma_*^2)}\right\}.$$

These two functions are different from (A.52) and (A.53). This difference shows that the presence of nuisance parameters can affect the covariance function.

Next, we consider the situation where σ_*^2 is allowed to switch with γ_* being unknown. Then, irrespective of whether the regressor is strictly or weakly exogenous, we always have

$$\text{Cov}\left(G\left(p_r,q_r\right),G\left(p_s,q_s\right)\right)=\frac{(1-p_r)(1-p_s)}{(1-q_r)(1-q_s)}\frac{1}{\sigma_*^8}\left\{\frac{3}{2}+\left(\frac{\rho_r\rho_s}{1-\rho_r\rho_s}\right)\right\}. \tag{A.56}$$

This function is different from (A.52) and (A.53). Therefore, the covariance function is not invariant to the null hypothesis, i.e., to which parameter is allowed to switch.

We report some simulation results to complement the above analysis. Under strict exogeneity, w_t is generated as $w_t = 0.5w_{t-1}+\varepsilon_t$ with $\varepsilon_t \sim i.i.d.N(0,1)$, where ε_t is independent of u_s at all leads and lags. We let $\gamma_* = 0.5$ and $\sigma_*^2 = 1$ and consider $(p_r,q_r) = (0.6,0.9)$ and $(p_s,q_s) = (0.6,x)$ with x varying between 0.1 and 0.9. Figure S1 displays the five correlations functions given by (A.52)-(A.56). The solid lines starting from the top correspond to (A.56), (A.54), (A.52), (A.55), and (A.53), respectively. These functions show clearly the dependence on the three factors highlighted above. The figure also displays the empirical correlations computed using simulations (the dashed lines). They are generated by simulating samples of 250 observations using the same parameter value as above, computing $T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_{jk,t}$ using each series, and repeating 10,000 times to obtain the empirical correlations. The values are close to their asymptotic approximations in all five cases.

We now show that $\Omega\left(p,q\right)$ is strictly positive in the five cases. The cases (A.52), (A.54), and (A.56) are clear because $|\rho| < 1$. For (A.53), the expression inside the braces is minimized at $\rho = 0$. The minimized value equals $4/(1-\gamma_*^2)$, which is strictly positive because of $|\gamma_*| < 1$ due to stationarity. The case (A.55) can be studied in the same way as (A.53). ■

An illustration of $\Omega(p,q)$ in the context of the model (8). We first obtain an alternative representation for $(x^{\otimes 2})'\Omega(p,q)(x^{\otimes 2})$. Then, using this representation, we explain why Assumption 6 is not restrictive. Without loss of generality, we assume that only the regression coefficients are allowed to switch between the regimes.

Recall that the $(j+(k-1)n_\delta, l+(m-1)n_\delta)$-th element of $\Omega(p,q)$ is given by

$$\omega_{jklm}(p_r,q_r;p_s,q_s) = E\left[U_{jk,t}(p_r,q_r)\,U_{lm,t}(p_s,q_s)\right] - E[U_{jk,t}(p_r,q_r)\,S_t']\left(ES_tS_t'\right)^{-1}E[S_tU_{lm,t}(p_s,q_s)],$$

where

$$\begin{aligned} S_t &= \left[\begin{array}{ccc} \frac{z_t'u_t}{\sigma_*^2} & \frac{1}{2\sigma_*^2}\left(\frac{u_t^2}{\sigma_*^2}-1\right) & \frac{w_t'u_t}{\sigma_*^2} \end{array}\right]', \\ U_{jk,t} &= \left(\frac{1-\xi_*}{\xi_*}\right)\left\{\frac{w_{jt}w_{kt}}{\sigma_*^2}\left(\frac{u_t^2}{\sigma_*^2}-1\right) + \sum_{s=1}^{t-1}\rho^s\left(\frac{w_{j(t-s)}u_{t-s}}{\sigma_*^2}\right)\left(\frac{w_{kt}u_t}{\sigma_*^2}\right) + \sum_{s=1}^{t-1}\rho^s\left(\frac{w_{k(t-s)}u_{t-s}}{\sigma_*^2}\right)\left(\frac{w_{jt}u_t}{\sigma_*^2}\right)\right\}. \end{aligned}$$

Let $R_{jk,t}(p_r,q_r)$ be the residual from the population regression of $U_{jk,t}(p_r,q_r)$ on S_t: $R_{jk,t}(p_r,q_r) = U_{jk,t}(p_r,q_r) - E[U_{jk,t}(p_r,q_r)\,S_t']\,(ES_tS_t')^{-1}S_t$. Then, $\omega_{jklm}(p_r,q_r;p_s,q_s)$ can be rewritten as

$$\omega_{jklm}(p_r,q_r;p_s,q_s) = ER_{jk,t}(p_r,q_r)\,R_{lm,t}(p_s,q_s).$$

Some rows of $\Omega(p,q)$ are identical because $R_{jk,t}(p,q) = R_{kj,t}(p,q)$ for all $j,k\in\{1,...,n_\delta\}$. This feature makes the positiveness of $(x^{\otimes 2})'\Omega(p,q)(x^{\otimes 2})$ unclear. To address this issue, let $M_t(p,q)$ be a n_δ-dimensional square matrix whose (j,k)-th element is $R_{jk,t}(p,q)$. Let $V_t(p,q)$ be an $n_\delta(n_\delta+1)/2$ dimensional vector which includes the upper triangular elements of $M_t(p,q)$ in the lexicographical order. For example, when $n_\delta = 3$,

$$V_t(p,q) = \left[\begin{array}{cccccc} R_{11,t}(p,q) & R_{12,t}(p,q) & R_{13,t}(p,q) & R_{22,t}(p,q) & R_{23,t}(p,q) & R_{33,t}(p,q) \end{array}\right]'.$$

Define

$$\Sigma(p,q) = EV_t(p,q)\,V_t(p,q)'.$$

Similarly, for any $x = (x_1,...,x_{n_\delta})' \in R^{n_\delta}$, define $W(x) = xx'$, and let $\eta(x)$ be an $n_\delta(n_\delta+1)/2$ dimensional vector that includes the upper triangular elements of $W(x)$ in the lexicographical order with the off-diagonal elements multiplied by 2. For example, when $n_\delta = 3$, $\eta(x) = \left[\begin{array}{cccccc} x_1^2 & 2x_1x_2 & 2x_1x_3 & x_2^2 & 2x_2x_3 & x_3^2 \end{array}\right]'$. Using $\Sigma(p,q)$ and $\eta(x)$, we have

$$\left(x^{\otimes 2}\right)'\Omega(p,q)\left(x^{\otimes 2}\right) = \eta(x)'\Sigma(p,q)\,\eta(x).$$

For any nonzero $x\in R^{n_\delta}$,

$$\frac{\left(x^{\otimes 2}\right)'\Omega(p,q)\left(x^{\otimes 2}\right)}{\|x^{\otimes 2}\|^2} = \left(\frac{\|\eta(x)\|^2}{\|x^{\otimes 2}\|^2}\right)\frac{\eta(x)'\Sigma(p,q)\,\eta(x)}{\|\eta(x)\|^2} \geq \frac{\eta(x)'\Sigma(p,q)\,\eta(x)}{\|\eta(x)\|^2},$$

where the last inequality holds because $\|\eta(x)\|^2 = \sum_{i=1}^{n_\delta}x_i^4 + 4\sum_{i=1}^{n_\delta}\sum_{j=i+1}^{n_\delta}x_i^2x_j^2$ and $\left\|x^{\otimes 2}\right\|^2 = \sum_{i=1}^{n_\delta}x_i^4 + 2\sum_{i=1}^{n_\delta}\sum_{j=i+1}^{n_\delta}x_i^2x_j^2$. Further,

$$\min_{x\in R^{n_\delta},\|x^{\otimes 2}\|=1}\frac{\eta(x)'\Sigma(p,q)\,\eta(x)}{\|\eta(x)\|^2} \geq \min_{x\in R^{n_\delta},x\neq 0}\frac{\eta(x)'\Sigma(p,q)\,\eta(x)}{\|\eta(x)\|^2} \geq \min_{\|z\|=1}z'\Sigma(p,q)\,z.$$

The right hand side equals the smallest eigenvalue of $\Sigma(p,q)$. Thus, the expression in Assumption 6 is bounded from below by the smallest eigenvalue of $\Sigma(p,q)$.

Note that, by construction, $\Sigma(p,q)$ is the covariance matrix of the residuals from the population regressions of $U_{jk,t}(p,q)$ (those with $j \leq k$) on S_t. Thus, its smallest eigenvalue can not be negative. This eigenvalue equals zero only when: (a) some residuals have zero variance, i.e., some regressions have a perfect fit, or (2) some residuals are perfectly correlated with each other. Given the expressions of S_t and $U_{jk,t}(p,q)$, neither (a) nor (b) is expected to happen as long as the transition probability is bounded away from $p+q=1$. This implies that Assumption 6 is not restrictive for the model (8).

For other models, although the expressions of S_t and $U_{jk,t}(p,q)$ are different, the regression interpretation does not change. As a result, $\min_{x\in R^{n_\delta},\|x^{\otimes 2}\|=1}\left(x^{\otimes 2}\right)'\Omega(p,q)\left(x^{\otimes 2}\right)$ is still bounded from below by the smallest eigenvalue of $\Sigma(p,q)$. This suggests that Assumption (6) is not restrictive in general situations. ■

Lemma A.6 *Let $\hat{\delta}_2(p,q) = \arg\max_{\delta_2\in\Delta}\mathcal{L}(p,q,\delta_2)$. Under the null hypothesis and Assumptions 1-6, $T^{1/4}(\hat{\delta}_2(p,q)-\widetilde{\delta}) = O_p(1)$ uniformly over Λ_ϵ.*

Proof of Lemma A.6. Because of Assumptions 2(i), 3(i), and the compactness of the parameter space, the restricted MLE $\widetilde{\delta}$ is a consistent estimator of δ_*. Similarly, because of Assumptions 2(ii), 3(ii), and the compactness, $\hat{\delta}_2(p,q)$ converges in probability to δ_* uniformly over Λ_ϵ. Thus, $\hat{\delta}_2(p,q)-\widetilde{\delta} = (\hat{\delta}_2(p,q)-\delta_*) + (\delta_* - \widetilde{\delta}) = o_p(1)$ uniformly over Λ_ϵ. This implies, for any $\upsilon > 0$,

$$P\left(\inf_{(p,q)\in\Lambda_\epsilon}||\hat{\delta}_2(p,q)-\widetilde{\delta}|| > \upsilon\right) \to 0 \text{ as } T\to\infty. \tag{A.57}$$

Below, we choose υ sufficiently small so that Assumption 5 can be applied.

Let δ_2 denote a generic value in the parameter space satisfying $||\delta_2-\widetilde{\delta}|| \leq \upsilon$. Let $\eta = T^{1/4}(\delta_2-\widetilde{\delta})$ and η_j be the j-th element of η. We consider the following fifth order Taylor expansion of $\mathcal{L}(p,q,\delta_2)$ around $\widetilde{\delta}$ over the set $\{\delta_2\in\Delta : ||\delta_2-\widetilde{\delta}|| \leq \upsilon\}$:

$$\begin{aligned} & \mathcal{L}(p,q,\delta_2) - \mathcal{L}(p,q,\widetilde{\delta}) \qquad\qquad (A.58)\\ = & \frac{1}{2!}T^{-1/2}\sum_{j,k=1}^{n_\delta}\mathcal{L}^{(2)}_{jk}(p,q,\widetilde{\delta})\eta_j\eta_k + \frac{1}{3!}T^{-3/4}\sum_{j,k,l=1}^{n_\delta}\mathcal{L}^{(3)}_{jkl}(p,q,\widetilde{\delta})\eta_j\eta_k\eta_l \\ & +\frac{1}{4!}T^{-1}\sum_{j,k,l,m=1}^{n_\delta}\mathcal{L}^{(4)}_{jklm}(p,q,\widetilde{\delta})\eta_j\eta_k\eta_l\eta_m + \frac{1}{5!}T^{-5/4}\sum_{j,k,l,m,n=1}^{n_\delta}\mathcal{L}^{(5)}_{jklmn}(p,q,\bar{\delta})\eta_j\eta_k\eta_l\eta_m\eta_n \\ \equiv & (L2)+(L3)+(L4)+(L5), \end{aligned}$$

where $\bar{\delta} = \widetilde{\delta} + c(\delta_2-\widetilde{\delta})$ for some $c\in(0,1)$. The following inequality holds

$$\mathcal{L}(p,q,\delta_2) - \mathcal{L}(p,q,\widetilde{\delta}) \leq |(L2)| + |(L3)| + |(L5)| + (L4).$$

Below, we derive upper bounds in terms of $||\eta||$ for $|(L2)|, |(L3)|, |(L5)|$, and $(L4)$. These bounds hold uniformly over Λ_ϵ. Then, using these bounds, we show that for any $\varepsilon > 0$, there exists $M > 0$ and $\xi > 0$, such that

$$P\left(\sup_{(p,q)\in\Lambda_\epsilon}\sup_{T^{-1/4}M\leq||\delta_2-\widetilde{\delta}||\leq\upsilon}\mathcal{L}(p,q,\delta_2) - \mathcal{L}(p,q,\widetilde{\delta}) \leq -\xi\right) \geq 1-\varepsilon \tag{A.59}$$

for any sufficiently large T. Because, by construction, the maximized log likelihood ratio can not be negative, this proves $T^{1/4}(\hat{\delta}_2(p,q)-\widetilde{\delta}) = O_p(1)$ uniformly over Λ_ϵ in light of (A.57). Note that by the definition of the Euclidean norm, we have $|\eta_j| \le ||\eta||$ for all $j \in \{1, ..., n_\delta\}$ and $||\eta^{\otimes 2}|| \ge ||\eta||^2/n_\delta$.

By the triangle inequality,

$$|(L2)| \le \frac{1}{2!}\sum_{j,k=1}^{n_\delta} \left| T^{-1/2}\mathcal{L}_{jk}^{(2)}(p,q,\widetilde{\delta})\right| |\eta_j||\eta_k| \le ||\eta||^2 \frac{1}{2!}\sum_{j,k=1}^{n_\delta}\left| T^{-1/2}\mathcal{L}_{jk}^{(2)}(p,q,\widetilde{\delta})\right|.$$

On the right hand side, $T^{-1/2}\mathcal{L}_{jk}^{(2)}(p,q,\widetilde{\delta})$ is $O_p(1)$ uniformly over Λ_ϵ by Lemmas 2.2 and A.5. Thus, for any $\varepsilon > 0$, there exists $C_2 > 0$ independent of υ, such that

$$P\left(|(L2)| \le C_2||\eta||^2\right) \ge 1 - \frac{\varepsilon}{4}$$

for large T uniformly over Λ_ϵ.

$$|(L3)| \le \frac{1}{3!}T^{-1/4}\sum_{j,k,l=1}^{n_\delta}\left| T^{-1/2}\mathcal{L}_{jkl}^{(3)}(p,q,\widetilde{\delta})\right| |\eta_j\eta_k\eta_l| \le ||\eta||^2||T^{-1/4}\eta||\frac{1}{3!}\sum_{j,k,l=1}^{n_\delta}\left| T^{-1/2}\mathcal{L}_{jkl}^{(3)}(p,q,\widetilde{\delta})\right|.$$

By Lemmas 2.3, $T^{-1/2}\mathcal{L}_{jkl}^{(3)}(p,q,\widetilde{\delta}) = O_p(1)$ uniformly over Λ_ϵ. In addition, $||T^{-1/4}\eta|| = ||\delta_2 - \widetilde{\delta}|| \le \upsilon$. Thus, for any $\varepsilon > 0$, there exists $C_3 > 0$ independent of υ, such that

$$P\left(|(L3)| \le \upsilon C_3||\eta||^2\right) \ge 1 - \frac{\varepsilon}{4}$$

for large T uniformly over Λ_ϵ. Similarly to $(L3)$,

$$|(L5)| \le \frac{1}{5!}||T^{-1/4}\eta||||\eta||^4\sum_{j,k,l,m,n=1}^{n_\delta}\left\| T^{-1}\mathcal{L}_{jklmn}^{(5)}(p,q,\bar{\delta})\right\| \le \upsilon||\eta||^4\frac{1}{5!}\sum_{j,k,l,m,n=1}^{n_\delta}\left\| T^{-1}\mathcal{L}_{jklmn}^{(5)}(p,q,\bar{\delta})\right\|.$$

By Assumption 5, $T^{-1}\mathcal{L}_{jklmn}^{(5)}(p,q,\bar{\delta}) = O_p(1)$ uniformly over Λ_ϵ. Thus, for any $\varepsilon > 0$, there exists $C_5 > 0$ that is non-decreasing in υ, such that

$$P\left(|(L5)| \le \upsilon C_5||\eta||^4\right) \ge 1 - \frac{\varepsilon}{4}$$

for large T uniformly over Λ_ϵ.

The analysis of $(L4)$ is slightly more complex. We have, by Lemma 2.4,

$$(L4) = -\frac{1}{8}\left(\eta^{\otimes 2}\right)'\widetilde{\Omega}(p,q)\left(\eta^{\otimes 2}\right) + \frac{1}{4!}T^{-1}\sum_{j,k,l,m=1}^{n_\delta} R_{jklm}(p,q)\eta_j\eta_k\eta_l\eta_m,$$

where $\widetilde{\Omega}(p,q)$ is an n_δ^2-dimensional square matrix whose $(j+(k-1)n_\delta, l+(m-1)n_\delta)$-th element equals $\widetilde{V}_{jklm} - \widetilde{D}'_{jk}\widetilde{I}^{-1}\widetilde{D}_{lm}$, and $R_{jklm}(p,q)$ denotes the $o_p(1)$ remainder term in Lemma 2.4. Because $R_{jklm}(p,q)$ is $o_p(1)$ and $\widetilde{\Omega}(p,q)$ converges in probability to $\Omega(p,q)$ uniformly over Λ_ϵ,

$$(L4) = -\frac{1}{8}\left(\eta^{\otimes 2}\right)'\Omega(p,q)\left(\eta^{\otimes 2}\right) + ||\eta||^4 * o_p(1).$$

Further, by Assumption 6,

$$\left(\eta^{\otimes 2}\right)' \Omega(p,q)\left(\eta^{\otimes 2}\right) \geq L||\eta^{\otimes 2}||^2 \geq \frac{L}{n_\delta^2}||\eta||^4.$$

Therefore, for any $\varepsilon > 0$,

$$P\left((L4) \leq -\frac{L}{16n_\delta^2}||\eta||^4\right) \geq 1 - \frac{\varepsilon}{4}$$

for large T uniformly over Λ_ϵ.

Let $C = \max(C_2, C_3, C_5)$. The above bounds for $|(L2)|, |(L3)|, |(L5)|$, and $(L4)$ imply

$$P\left(\sup_{(p,q)\in\Lambda_\epsilon}\sup_{||\delta_2-\widetilde{\delta}||\leq\upsilon}\mathcal{L}(p,q,\delta_2)-\mathcal{L}(p,q,\widetilde{\delta}) \leq \sup_{||\eta||\leq T^{1/4}\upsilon} C||\eta||^2 + \upsilon C||\eta||^2 + \upsilon C||\eta||^4 - \frac{L}{16n_\delta^2}||\eta||^4\right)$$
$$\geq 1 - \varepsilon$$

for sufficiently large T. By (A.57), we can choose a small υ so that $\upsilon \leq \min(1, \frac{L}{32n_\delta^2 C})$. Then,

$$\sup_{||\eta||\leq T^{1/4}\upsilon} C||\eta||^2 + \upsilon C||\eta||^2 + \upsilon C||\eta||^4 - \frac{L}{16n_\delta^2}||\eta||^4 \leq \sup_{||\eta||\leq T^{1/4}\upsilon}\left(2C - \frac{L}{32n_\delta^2}||\eta||^2\right)||\eta||^2.$$

The right hand side is negative and strictly decreasing in $||\eta||^2$ when $||\eta||^2 > 64Cn_\delta^2/L$. Let $M = \max(\sqrt{128Cn_\delta^2/L}, \sqrt{C})$, then

$$\sup_{M\leq||\eta||\leq T^{1/4}\upsilon}\left(2C - \frac{L}{32n_\delta^2}||\eta||^2\right)||\eta||^2 \leq -2C^2 < 0 \text{ when } T \text{ is sufficiently large.}$$

Therefore, (A.59) holds with $\xi = -2C^2$. This shows that when $||\eta|| \geq M$ and M is sufficiently large, the fourth order term asymptotically dominates the remaining terms in the fifth order Taylor expansion and, as a result, the log likelihood ratio is strictly negative with probability close to one in large samples. This proves $T^{1/4}(\hat{\delta}_2(p,q) - \widetilde{\delta}) = O_p(1)$ uniformly over Λ_ϵ. ■

Proof of Proposition 1. The proof takes two steps. In the first step, we study the likelihood ratio under an additional restriction: $||\delta_2(p,q) - \widetilde{\delta}|| \leq T^{-1/4}M$ for some $M < \infty$. In the second step, we study the effect of this restriction on the approximation, and establish the weak convergence.

Step 1. Let $\hat{\delta}_2(p,q) = \arg\max_{\delta_2\in\Delta}\mathcal{L}(p,q,\delta_2)$. Lemma A.6 implies that for any $\varepsilon > 0$, there exists an $M < \infty$, such that

$$P\left(\sup_{(p,q)\in\Lambda_\epsilon}||\hat{\delta}_2(p,q) - \widetilde{\delta}|| \leq T^{-1/4}M\right) \geq 1 - \varepsilon \text{ for sufficiently large } T. \tag{A.60}$$

Let δ_2 denote a generic value in the parameter space satisfying $||\delta_2 - \widetilde{\delta}|| \leq T^{-1/4}M$ and $\eta = T^{1/4}(\delta_2 - \widetilde{\delta})$. Recall that in Lemma A.6, we study the Taylor expansion (A.58) over $\{\delta_2 \in \Delta : ||\delta_2 - \widetilde{\delta}|| \leq \upsilon\}$ with υ being a small constant. Below, we further study this expansion over the set $\{\delta_2 \in \Delta : ||\delta_2 - \widetilde{\delta}|| \leq T^{-1/4}M\}$. We consider the terms $(L2), (L3), (L4)$, and $(L5)$ separately.

$$|(L3)| \leq T^{-1/4}||\eta||^3\frac{1}{3!}\sum_{j,k,l=1}^{n_\delta}\left|T^{-1/2}\mathcal{L}_{jkl}^{(3)}(p,q,\widetilde{\delta})\right| = O_p(T^{-1/4}),$$

where the equality holds because $||\eta|| \leq M$ and $T^{-1/2}\mathcal{L}^{(3)}_{jkl}(p,q,\widetilde{\delta}) = O_p(1)$ by Lemma 2.3. Applying Lemma 2.4 and $||\eta|| \leq M$, we have $(L4) = -(1/8)\left(\eta^{\otimes 2}\right)'\widetilde{\Omega}(p,q)\left(\eta^{\otimes 2}\right)+o_p(1)$, where $\widetilde{\Omega}(p,q)$ is an n_δ^2-dimensional square matrix whose $(j+(k-1)n_\delta, l+(m-1)n_\delta)$-th element equals $\widetilde{V}_{jklm}-\widetilde{D}'_{jk}\widetilde{I}^{-1}\widetilde{D}_{lm}$. Further, because $\widetilde{\Omega}(p,q)$ converges in probability to $\Omega(p,q)$ uniformly over Λ_ϵ,

$$(L4) = -\frac{1}{8}\left(\eta^{\otimes 2}\right)'\Omega(p,q)\left(\eta^{\otimes 2}\right) + o_p(1).$$

The analysis of $(L5)$ is similar to that of $(L3)$:

$$|(L5)| \leq T^{-1/4}||\eta||^5\frac{1}{5!}\sum_{j,k,l,m,n=1}^{n_\delta}\left\|T^{-1}\mathcal{L}^{(5)}_{jklmn}(p,q,\bar{\delta})\right\| = O_p(T^{-1/4}),$$

which holds because $||\eta|| \leq M$ and $T^{-1}\mathcal{L}^{(5)}_{jklmn}(p,q,\bar{\delta}) = O_p(1)$ by Assumption 5.

In all three cases, the orders are uniform over Λ_ϵ. Thus,

$$\begin{aligned}
& \sup_{(p,q)\in\Lambda_\epsilon}\ \sup_{||\delta_2-\widetilde{\delta}||\leq T^{-1/4}M} 2[\mathcal{L}(p,q,\delta_2)-\mathcal{L}(p,q,\widetilde{\delta})] \\
= & \sup_{(p,q)\in\Lambda_\epsilon}\ \sup_{||\eta||\leq|M} \left\{\left(\eta^{\otimes 2}\right)'\left[T^{-1/2}\operatorname{vec}\mathcal{L}^{(2)}(p,q,\widetilde{\delta})\right] - \frac{1}{4}\left(\eta^{\otimes 2}\right)'\Omega(p,q)\left(\eta^{\otimes 2}\right)\right\} + o_p(1) \\
\Longrightarrow & \sup_{(p,q)\in\Lambda_\epsilon}\ \sup_{\|\eta\|\leq M} \left\{\left(\eta^{\otimes 2}\right)'\operatorname{vec}G(p,q) - \frac{1}{4}\left(\eta^{\otimes 2}\right)'\Omega(p,q)\left(\eta^{\otimes 2}\right)\right\} \\
= & \sup_{(p,q)\in\Lambda_\epsilon}\ \sup_{\|\eta\|\leq M} \mathcal{W}^{(2)}(p,q,\eta).
\end{aligned}$$

where the weak convergence holds because of Lemma A.5 and the fact that the supremum operator is continuous when taken over a compact set.

Step 2. We now study the effect of the restriction $\|\eta\| \leq M$ on the approximation. Define $\hat{\eta}(p,q) = \operatorname{argmax}_{\eta\in R^{n_\delta}} \mathcal{W}^{(2)}(p,q,\eta)$. We have $\left(\eta^{\otimes 2}\right)'\operatorname{vec}G(p,q) = ||\eta||^2 * O_p(1)$ and $-\left(\eta^{\otimes 2}\right)'\Omega(p,q)\left(\eta^{\otimes 2}\right) \leq -(L/n_\delta^2)||\eta||^4$ uniformly over Λ_ϵ by Assumption 6. Therefore, to ensure $\sup_{\eta\in R^{n_\delta}}\mathcal{W}^{(2)}(p,q,\eta)$ is nonnegative, for any $\varepsilon > 0$, there must exist a $K < \infty$ such that

$$P\left(\|\hat{\eta}(p,q)\| \leq K\right) \geq 1-\varepsilon \tag{A.61}$$

for any sufficiently large T uniformly over Λ_ϵ. Without loss of generality, we assume $K \leq M$. Otherwise, M can be increased until this is satisfied. Let $F_j(\cdot)$ $(j = 1,2,3,4)$ denote the CDFs of the following variables: $\text{SupLR}(\Lambda_\epsilon)$, $\sup_{(p,q)\in\Lambda_\epsilon}\sup_{||\delta_2-\widetilde{\delta}||\leq T^{-1/4}M} 2[\mathcal{L}(p,q,\delta_2)-\mathcal{L}(p,q,\widetilde{\delta})]$, $\sup_{(p,q)\in\Lambda_\epsilon}\sup_{\|\eta\|\leq M}\mathcal{W}^{(2)}(p,q,\eta)$, and $\sup_{(p,q)\in\Lambda_\epsilon}\sup_{\eta\in R^{n_\delta}}\mathcal{W}^{(2)}(p,q,\eta)$. Then,

$$|F_1(x)-F_4(x)| \leq |F_1(x)-F_2(x)| + |F_2(x)-F_3(x)| + |F_3(x)-F_4(x)|.$$

Among the terms on the right hand side,

$$|F_1(x)-F_2(x)| \leq P\left(\sup_{(p,q)\in\Lambda_\epsilon}||\hat{\delta}_2(p,q)-\widetilde{\delta}|| > T^{-1/4}M\right),$$

and

$$|F_3(x) - F_4(x)| \leq P\left(\|\hat{\eta}(p,q)\| > M\right)$$

uniformly over $x \in R$, both of which are bounded by ε for large T by (A.60) and (A.61). Thus,

$$|F_1(x) - F_4(x)| \leq \varepsilon + |F_2(x) - F_3(x)| + \varepsilon$$

for large T over $x \in R$. The convergence in Step 1 implies $F_2(x) - F_3(x) \to 0$ for any $M < \infty$ over $x \in R$. Because ε can be made arbitrarily small, the result of Proposition 1 holds. ■

Lemma A.7 *Suppose that the null hypothesis and Assumptions 1-8 hold with (24) satisfied for all* $i_1, i_2 \in \{1, ..., n_\delta\}$*. The following results hold uniformly over* $\{(p,q) : \epsilon \leq p, q \leq 1-\epsilon, p+q=1\}$ *for any* $k, l \in \{1, ..., n_\delta\}$ *:*

1. *Let* e_k *be an* n_δ *dimensional unit vector whose k-th element equals 1, then*

$$\begin{bmatrix} \nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta}) \\ \xi_* \nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} = (\xi_* - 1)\begin{bmatrix} 0 \\ e_k \end{bmatrix}$$

2. *The second order derivatives satisfy*

$$\begin{aligned} \begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) \\ \xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} &= -\left(\frac{1-\xi_*}{\xi_*}\right)\left\{\alpha_{kl} - \widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\begin{bmatrix} \nabla_\beta\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t} \\ \left(\frac{1-\xi_*}{\xi_*}\right)\nabla_{\delta_1}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t} \end{bmatrix}\right. \\ &\left. +\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\begin{bmatrix} \nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}\alpha_{kl}^{(1)} + \nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}\alpha_{kl}^{(2)} \\ \nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}\alpha_{kl}^{(1)} \end{bmatrix}\right\} + o_p(T^{-1/2}), \end{aligned}$$

where $\widetilde{I}$ *is defined in (17) and* $\alpha_{kl}' = (\alpha_{kl}^{(1)\prime}, \alpha_{kl}^{(2)\prime})$.

Proof of Lemma A.7. When $p+q=1$, the derivatives of $\xi_{t|t-1}$ with respect to θ are all equal to zero when evaluated at $\delta_1 = \delta_2 = \delta$. The first result of the lemma follows from this feature and the argument in the proof of Lemma A.3. We thus omit the details. The proof for second result is more complex; we present it below.

Consider (A.28). There, only the summations over the first and the third terms are nonzero by the relationship (A.33) and the first result of this lemma. Evaluating these two terms at the restricted MLE for $j \in \{1, ..., n_\beta + n_\delta\}$, we obtain,

$$D\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_\beta\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t} \\ \left(\frac{1-\xi_*}{\xi_*}\right)^2\frac{\nabla_{\delta_1}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t} \end{bmatrix} + D\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) + \frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) + \frac{\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix}$$

and

$$D\widetilde{I}\begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) \\ \xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} + D\left(\frac{1-\xi_*}{\xi_*}\right)\begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_\beta\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t} \\ \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_1}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t} \end{bmatrix},$$

where D has the same definition as in (A.24). Combining the preceding two displays, we obtain

$$\begin{aligned}
&\begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) \\ \xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} \\
= &-\left(\frac{1-\xi_*}{\xi_*}\right)\widetilde{I}^{-1}\begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_\beta \widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t} \\ \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_1}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{bmatrix} + \widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \left(\frac{1-\xi_*}{\xi_*}\right)\frac{1}{\widetilde{f}_t}\nabla_{\theta_\beta}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t} \\ \left(\frac{1-\xi_*}{\xi_*}\right)^2\frac{1}{\widetilde{f}_t}\nabla_{\theta_{\delta_1}}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}\end{bmatrix} \\
&+\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) + \frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) + \frac{\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}.
\end{aligned}$$

Because $\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t} = \alpha_{jk}'\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}$ and $\sum_{t=1}^{T}(\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}/\widetilde{f}_t) = 0$, the preceding display equals

$$\begin{aligned}
&\begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) \\ \xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} \qquad \text{(A.62)}\\
= &-\left(\frac{1-\xi_*}{\xi_*}\right)\alpha_{kl} + \widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \left(\frac{1-\xi_*}{\xi_*}\right)\frac{1}{\widetilde{f}_t}\nabla_{\theta_\beta}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t} \\ \left(\frac{1-\xi_*}{\xi_*}\right)^2\frac{1}{\widetilde{f}_t}\nabla_{\theta_{\delta_1}}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}\end{bmatrix} \\
&+\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) + \frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})\end{bmatrix}.
\end{aligned}$$

The variables $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})$ and $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta})$ appear on both sides of the display. We solve for them in two steps. First, because the last two terms on the right hand side are $O_p(T^{-1/2})$, we have

$$\begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) \\ \xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} = -\left(\frac{1-\xi_*}{\xi_*}\right)\begin{bmatrix}\alpha_{kl}^{(1)} \\ \alpha_{kl}^{(2)}\end{bmatrix} + O_p(T^{-1/2}).$$

Next, we apply this result to the third term on the right hand side of (A.62) to rewrite it as

$$-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{1}{T}\widetilde{I}^{-1}\sum_{t=1}^{T}\begin{bmatrix} \frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{kl}^{(1)} + \frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{kl}^{(2)} \\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{kl}^{(1)}\end{bmatrix} + o_p\left(T^{-1/2}\right).$$

The result follows after applying this expression to (A.62). ■

Proof of Lemma 3. When $p+q=1$, the likelihood corresponds to that of a mixture model. The proof below relies heavily on the arguments in Lemmas C2, 3 and 4 in Cho and White (2007).

Consider the first result. Among the summations on the right hand side of (A.39), only the first one is nonzero. When evaluated at the restricted MLE, $T^{-1/2}\sum_{t=1}^{T}(\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{L}_{jt}/\hat{B}_t)$ and $((1-\xi_*)/\xi_*)T^{-1/2}\sum_{t=1}^{T}(\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{M}_{(n_\beta+j)t}/\hat{B}_t)$ equal

$$(1-\xi_*)T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t} + (1-\xi_*)T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_{1j}}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}),$$

$$\frac{(1-\xi_*)^3}{\xi_*^2}T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t} + (1-\xi_*)T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_{1j}}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}).$$

Taking their difference leads to

$$T^{-1/2}\mathcal{L}^{(3)}_{jkl}(p,q,\widetilde{\delta}) = -\frac{(1-\xi_*)(1-2\xi_*)}{\xi_*^2}T^{-1/2}\sum_{T=1}^{T}\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t} \Rightarrow G^{(3)}_{jkl}.$$

Next, we consider $T^{-1/2}\mathcal{L}^{(4)}_{jklm}(p,q,\widetilde{\delta})$. In (A.41), only the 1st, 3rd, 6th and 7th summations on the right hand side are nonzero. For the 1st one, $T^{-1/2}\sum_{t=1}^{T}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{L}_{jt}/\hat{B}_t$ at $\widetilde{\delta}$ equals

$$\begin{aligned}&(1-\xi_*)T^{-1/2}\textstyle\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\Big\{\nabla_{\delta_{2j}}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\widetilde{f}_{2t}+\nabla_{\delta_{2j}}\nabla_{\delta_{2l}}\nabla_{\beta'}\widetilde{f}_{2t}\nabla_{\delta_{2m}}\nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta})\\&+\nabla_{\delta_{2j}}\nabla_{\delta_{2k}}\nabla_{\beta'}\widetilde{f}_{2t}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})+\nabla_{\delta_{2j}}\nabla_{\delta_{2m}}\nabla_{\beta'}\widetilde{f}_{2t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})+\nabla_{\delta_{2j}}\nabla_{\beta'}\widetilde{f}_{2t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})\Big\}.\end{aligned}$$

Meanwhile, $((1-\xi_*)/\xi_*)T^{-1/2}\sum_{t=1}^{T}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{M}_{(n_\beta+j)t}/\hat{B}_t$ evaluated at $\widetilde{\delta}$ equals

$$\begin{aligned}&((1-\xi_*)/\xi_*)T^{-1/2}\textstyle\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\Big\{\frac{(\xi_*-1)^3}{\xi_*^2}\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}\\&+(\xi_*-1)\nabla_{\delta_{1j}}\nabla_{\delta_{1l}}\nabla_{\beta'}\widetilde{f}_{1t}\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})+(\xi_*-1)\nabla_{\delta_{1j}}\nabla_{\delta_{1l}}\nabla_{\delta_1'}\widetilde{f}_{1t}\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\\&+(\xi_*-1)\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\beta'}\widetilde{f}_{1t}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})+(\xi_*-1)\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\delta_1'}\widetilde{f}_{1t}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\\&+(\xi_*-1)\nabla_{\delta_{1j}}\nabla_{\delta_{1m}}\nabla_{\beta'}\widetilde{f}_{1t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})+(\xi_*-1)\nabla_{\delta_{1j}}\nabla_{\delta_{1m}}\nabla_{\delta_1'}\widetilde{f}_{1t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta})\\&+\xi_*\nabla_{\delta_{1j}}\nabla_{\beta'}\widetilde{f}_{1t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\beta}(\delta_2)\Big\}.\end{aligned}$$

Their difference equals

$$\begin{aligned}&(1-\xi_*)\left(1+\left(\tfrac{1-\xi_*}{\xi_*}\right)^3\right)T^{-1/2}\textstyle\sum_{t=1}^{T}\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\\&-\left(\tfrac{1-\xi_*}{\xi_*}\right)^2T^{-1/2}\textstyle\sum_{t=1}^{T}\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1l}}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(1)}_{km}-\left(\tfrac{1-\xi_*}{\xi_*}\right)^3T^{-1/2}\textstyle\sum_{t=1}^{T}\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1l}}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(2)}_{km}\\&-\left(\tfrac{1-\xi_*}{\xi_*}\right)^2T^{-1/2}\textstyle\sum_{t=1}^{T}\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(1)}_{lm}-\left(\tfrac{1-\xi_*}{\xi_*}\right)^3T^{-1/2}\textstyle\sum_{t=1}^{T}\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(2)}_{lm}\\&-\left(\tfrac{1-\xi_*}{\xi_*}\right)^2T^{-1/2}\textstyle\sum_{t=1}^{T}\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1m}}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(1)}_{kl}-\left(\tfrac{1-\xi_*}{\xi_*}\right)^3T^{-1/2}\textstyle\sum_{t=1}^{T}\frac{\nabla_{\delta_{1j}}\nabla_{\delta_{1m}}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(2)}_{kl}+o_p(1).\end{aligned}$$

The preceding display is $O_p(1)$ by Lemma A.3 and Assumption 4. The 3rd, 6th and 7th summation in (A.41) have the same structure. Applying Lemma A.7.2, the 3rd term equals

$$\begin{aligned}&\left(\tfrac{1-\xi_*}{\xi_*}\right)^2T^{-1/2}\textstyle\sum_{t=1}^{T}\frac{\nabla_{\beta'}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(1)}_{jk}+\left(\tfrac{1-\xi_*}{\xi_*}\right)^3T^{-1/2}\textstyle\sum_{t=1}^{T}\frac{\nabla_{\delta_1'}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(2)}_{jk}\\&-\left(\tfrac{1-\xi_*}{\xi_*}\right)^2T^{-1/2}\textstyle\sum_{t=1}^{T}(\alpha^{(1)}_{jk})'\frac{\nabla_{\beta}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(2)}_{lm}-\left(\tfrac{1-\xi_*}{\xi_*}\right)^2T^{-1/2}\textstyle\sum_{t=1}^{T}(\alpha^{(1)}_{jk})'\frac{\nabla_{\beta}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(1)}_{lm}\\&-\left(\tfrac{1-\xi_*}{\xi_*}\right)^2T^{-1/2}\textstyle\sum_{t=1}^{T}(\alpha^{(2)}_{jk})'\frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(1)}_{lm}=O_p(1).\end{aligned}$$

We now consider the fifth order derivative. The components of

$$T^{-1/2}\mathcal{L}^{(5)}_{jklmn}(p,q,\delta_2)-T^{-1/2}\left(\frac{1-\xi_*}{\xi_*}\right)\mathcal{M}^{(5)}_{(n_\beta+j)klmn}(p,q,\delta_2) \tag{A.63}$$

can be grouped into three subsets according to whether they depend on the first, second or third order derivatives of $\hat{B}_t$, c.f. (A.39). First, those depending on the first order derivatives are identically zero using the relationship (A.33). Second, we apply the first result of Lemma A.7 to (A.30). When evaluated at $\widetilde{\delta}$, $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{B}_t/\widetilde{f}_t$ equals

$$\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{kl}+\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left[\begin{array}{c}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})\\ \xi_*\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]. \tag{A.64}$$

Applying the second result in Lemma A.7, the term involving $[(1-\xi_*)/\xi_*]\alpha_{kl}$ gets canceled and the remainder term is of lower order. Consequently, in (A.63), the terms depending on the second order derivatives of $\hat{B}_t$ are all $O_p(1)$. Third, the terms depending on the third order derivatives of $\hat{B}_t$ are of the following form:

$$T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t}\left(\frac{\nabla_{\delta_{2n}}\hat{L}_{jt}}{\hat{B}_t}-\left(\frac{1-\xi_*}{\xi_*}\right)\frac{\nabla_{\delta_{2n}}\hat{M}_{(n_\beta+j)t}}{\hat{B}_t}\right). \tag{A.65}$$

When evaluated at $\widetilde{\delta}$, $\nabla_{\delta_{2n}}\hat{L}_{jt}/\hat{B}_t$ and $(\nabla_{\delta_{2n}}\hat{M}_{(n_\beta+j)t}/\hat{B}_t)$ are representable as linear functions of $(\hat{M}_{it}/\hat{B}_t)$ $(i=1,...,n_\beta+n_\delta)$ because $\nabla_{\delta_{1j}}\nabla_{\delta_{1n}}\widetilde{f}_{1t}=\alpha'_{jn}\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}$, c.f. (A.37), (A.20) and (A.18). This insightful observation is due to Cho and White (2007). This implies that, at $\widetilde{\delta}$, the stochastic order of (A.65) is the same as that of

$$T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{B}_t}{\hat{B}_t}\frac{\hat{M}_{it}}{\hat{B}_t},\ i=1,...,n_\beta+n_\delta. \tag{A.66}$$

The order of (A.66) can be found by analyzing (A.34). There, the terms depending on the 0th, 1st and 2rd order derivatives of $\hat{B}_t$ are all of order $O_p(T^{-1/2})$ after applying (A.33) and (A.64). The only term that remains is (A.66). Therefore, for (A.34) to equal zero, (A.66) must be of order $O_p(1)$ when evaluated at $\widetilde{\delta}$. This implies (A.65) is $O_p\left(1\right)$.

For the sixth order derivatives, we need to obtain explicit expressions for $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\beta}(\widetilde{\delta})$ and $\xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\delta}_1(\widetilde{\delta})$ by analyzing (A.34). The effects of the terms other than (A.66) are negligible. Writing out the expression for (A.66) explicitly, we obtain

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\Bigg\{\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t^2}\left[\frac{(\xi_*-1)^2}{\xi_*^2}(\xi_*-1)+(1-\xi_*)\right]\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1h}}\widetilde{f}_{1t}\\
&+\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t^2}\sum_{u=1}^{n_\beta+2n_\delta}[(\xi_*-1)\nabla_{\delta_{1k}}\nabla_{\theta_u}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\delta_{2k}}\nabla_{\theta_u}\widetilde{f}_{2t}]\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\\
&+\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t^2}\sum_{s=1}^{n_\beta+2n_\delta}[(\xi_*-1)\nabla_{\theta_s}\nabla_{\delta_{1l}}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\nabla_{\delta_{2l}}\widetilde{f}_{2t}]\nabla_{\delta_{2h}}\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\\
&+\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t^2}\sum_{s=1}^{n_\beta+2n_\delta}[(\xi_*-1)\nabla_{\theta_s}\nabla_{\delta_{1h}}\widetilde{f}_{1t}+(1-\xi_*)\nabla_{\theta_s}\nabla_{\delta_{2h}}\widetilde{f}_{2t}]\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})\\
&+\widetilde{I}\left[\begin{array}{c}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\beta}(\widetilde{\delta})\\ \xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]\Bigg\}=o_p\left(1\right).
\end{aligned}$$

Equivalently,

$$\begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\beta}(\widetilde{\delta}) \\ \xi_*\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} \tag{A.67}$$

$$\begin{aligned} = & -\widetilde{I}^{-1}T^{-1}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t^2}\left[\frac{(\xi_*-1)^2}{\xi_*^2}(\xi_*-1)+(1-\xi_*)\right]\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1h}}\widetilde{f}_{1t} \\ & -(\xi_*-1)\sum_{u=1}^{n_\delta}[\alpha_{ku}\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\delta}_{1u}(\widetilde{\delta})+\alpha_{lu}\nabla_{\delta_{2h}}\nabla_{\delta_{2k}}\hat{\delta}_{1u}(\widetilde{\delta})+\alpha_{hu}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_{1u}(\widetilde{\delta})]+o_p(1)\,. \end{aligned}$$

Applying the above expression, $T^{-1}\mathcal{L}^{(6)}_{jklmnr}(p,q,\widetilde{\delta})$ equals, by the same argument as in Cho and White (2007, l.13-24 in p. 1713),

$$T^{-1}\sum_{(i_1,i_2,\ldots,i_6)\in IND_1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\nabla_{\delta_{2i_6}}\hat{B}_t}{\hat{B}_t^2}\left(\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\hat{L}_{i_1t}-\frac{1-\xi_*}{\xi_*}\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\hat{M}_{(n_\beta+i_1)t}\right)+o_p(1)\,,$$

where all the quantities are evaluated at $\delta_2=\widetilde{\delta}$. Further, at $\delta_2=\widetilde{\delta}$, $\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\nabla_{\delta_{2i_6}}\hat{B}_t$ equals

$$\begin{aligned} & \left[\frac{(\xi_*-1)^2}{\xi_*^2}(\xi_*-1)+(1-\xi_*)\right]\nabla_{\delta_{1i_4}}\nabla_{\delta_{1i_5}}\nabla_{\delta_{1i_6}}\widetilde{f}_{1t} \\ & +(\xi_*-1)\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}\sum_{u=1}^{n_\delta}[\alpha_{i_4u}\nabla_{\delta_{2i_5}}\nabla_{\delta_{2i_6}}\hat{\delta}_{1u}(\widetilde{\delta})+\alpha_{i_5u}\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_6}}\hat{\delta}_{1u}(\widetilde{\delta})+\alpha_{i_6u}\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\hat{\delta}_{1u}(\widetilde{\delta})] \\ & +\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}\begin{bmatrix} \nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\nabla_{\delta_{2i_6}}\hat{\beta}(\widetilde{\delta}) \\ \xi_*\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\nabla_{\delta_{2i_6}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix}. \end{aligned}$$

Because of (A.67), the above display equals

$$\frac{(\xi_*-1)(1-2\xi_*)}{\xi_*^2}\left\{\nabla_{\delta_{1i_4}}\nabla_{\delta_{1i_5}}\nabla_{\delta_{1i_6}}\widetilde{f}_{1t}-\left(\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}\right)\widetilde{I}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t^2}\nabla_{\delta_{1i_4}}\nabla_{\delta_{1i_5}}\nabla_{\delta_{1i_6}}\widetilde{f}_{1t}\right]\right\}$$

The result follows because, when evaluated at $\widetilde{\delta}$, $\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\hat{L}_{i_1t}-[(1-\xi_*)/\xi_*]\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\hat{M}_{(n_\beta+i_1)t}$ equals $[(\xi_*-1)(1-2\xi_*)/\xi_*^2]\nabla_{\delta_{1i_1}}\nabla_{\delta_{1i_2}}\nabla_{\delta_{1i_3}}\widetilde{f}_{1t}$.

When $p=q=1/2$, the results for the 3rd to the 6th order derivatives follow from the above proofs. The proof for $T^{-1/2}\mathcal{L}^{(7)}_{i_1,\ldots,i_7}(1/2,1/2,\widetilde{\delta})$ is similar to that of $T^{-1/2}\mathcal{L}^{(5)}_{i_1,\ldots,i_5}(p,1-q,\widetilde{\delta})$. The proof for $T^{-1/2}\mathcal{L}^{(8)}_{i_1,\ldots,i_8}(1/2,1/2,\widetilde{\delta})$ is similar to that of $T^{-1/2}\mathcal{L}^{(6)}_{i_1,\ldots,i_6}(p,q,\widetilde{\delta})$. We omit the details.■

A.2.1 Discussions related to Assumption 7

We use the three models in Subsection 5.2 to explain how to check Assumption 7 in practice.

Linear regression with normal errors. The model under the null hypothesis is

$$y_t=\gamma+z_t'\alpha+u_t \text{ with } u_t\sim i.i.d.N(0,\sigma^2). \tag{A.68}$$

The conditional density of y_t, evaluated at $(\gamma, \alpha', \sigma^2)$, is given by

$$f_t(\gamma, \alpha, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_t - \gamma - z_t'\alpha)^2}{2\sigma^2}\right).$$

We check Assumption 7 for three separate cases, according to which parameter is allowed to switch. The steps for each case are the same.

Case 1: only γ is allowed to switch. To check Assumption 7, we need to compute the first order derivatives of $f_t(\gamma, \alpha, \sigma^2)$ with respect to γ, α, and σ^2, and its second order derivative with respect to γ. They are all straightforward to obtain:

$$\begin{aligned}
\nabla_\alpha f_t(\gamma, \alpha, \sigma^2) &= f_t(\gamma, \alpha, \sigma^2)\left(\frac{z_t(y_t - \gamma - z_t'\alpha)}{\sigma^2}\right), \qquad \text{(A.69)}\\
\nabla_{\sigma^2} f_t(\gamma, \alpha, \sigma^2) &= f_t(\gamma, \alpha, \sigma^2)\left(\frac{(y_t - \gamma - z_t'\alpha)^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right),\\
\nabla_\gamma f_t(\gamma, \alpha, \sigma^2) &= f_t(\gamma, \alpha, \sigma^2)\left(\frac{y_t - \gamma - z_t'\alpha}{\sigma^2}\right),\\
\nabla_\gamma \nabla_\gamma f_t(\gamma, \alpha, \sigma^2) &= f_t(\gamma, \alpha, \sigma^2)\left(\frac{(y_t - \gamma - z_t'\alpha)^2}{\sigma^4} - \frac{1}{\sigma^2}\right).
\end{aligned}$$

Now, we evaluate the four equations at the MLE, $(\widetilde{\gamma}, \widetilde{\alpha}, \widetilde{\sigma}^2)$, and apply the notation in the paper (i.e., $\widetilde{f}_{1t} = f_t(\widetilde{\gamma}, \widetilde{\alpha}, \widetilde{\sigma}^2)$ and $\nabla_\gamma \widetilde{f}_{1t} = \nabla_\gamma f_t(\widetilde{\gamma}, \widetilde{\alpha}, \widetilde{\sigma}^2)$):

$$\begin{aligned}
\nabla_\alpha \widetilde{f}_{1t} &= \widetilde{f}_{1t}\frac{z_t\widetilde{u}_t}{\widetilde{\sigma}^2}, \qquad \text{(A.70)}\\
\nabla_{\sigma^2} \widetilde{f}_{1t} &= \widetilde{f}_{1t}\frac{\widetilde{u}_t^2 - \widetilde{\sigma}^2}{2\widetilde{\sigma}^4},\\
\nabla_\gamma \widetilde{f}_{1t} &= \widetilde{f}_{1t}\frac{\widetilde{u}_t}{\widetilde{\sigma}^2},\\
\nabla_\gamma \nabla_\gamma \widetilde{f}_{1t} &= \widetilde{f}_{1t}\frac{(\widetilde{u}_t^2 - \widetilde{\sigma}^2)}{\widetilde{\sigma}^4}.
\end{aligned}$$

From the second and fourth equations, we obtain $\nabla_\gamma \nabla_\gamma \widetilde{f}_{1t} = 2\nabla_{\sigma^2} \widetilde{f}_{1t}$ for any t. Therefore, Assumption 7 holds, and a refinement is needed. The equality $\nabla_\gamma \nabla_\gamma \widetilde{f}_{1t} = 2\nabla_{\sigma^2} \widetilde{f}_{1t}$ does not depend on whether the null hypothesis holds. It leads to a singularity because

$$\sum_{t=1}^T \frac{\nabla_\gamma \nabla_\gamma \widetilde{f}_{1t}}{\widetilde{f}_{1t}} = 2\sum_{t=1}^T \frac{\nabla_{\sigma^2} \widetilde{f}_{1t}}{\widetilde{f}_{1t}} = \frac{1}{\widetilde{\sigma}^4}\sum_{t=1}^T \left(\widetilde{u}_t^2 - \widetilde{\sigma}^2\right) = 0,$$

where the last equality follows because $\widetilde{\sigma}^2 = T^{-1}\sum_{t=1}^T \widetilde{u}_t^2$.

Case 2: α or σ^2 or both are allowed to switch. Without loss of generality, we assume that only α is allowed to switch. We need the first order derivatives of $f_t(\gamma, \alpha, \sigma^2)$ with respect to γ, α, and σ^2, which are already computed in (A.69), and the second order derivative with respect to α:

$$\nabla_\alpha \nabla_{\alpha'} \widetilde{f}_{1t} = \widetilde{f}_{1t}\frac{z_t z_t'(\widetilde{u}_t^2 - \widetilde{\sigma}^2)}{\widetilde{\sigma}^4}.$$

Because of the presence of $z_t z_t'$, the elements of $\nabla_\alpha \nabla_{\alpha'} \widetilde{f}_{1t}$ can no longer be expressed as a linear function of the first order derivatives with respect to γ, α, and σ^2 for all t; c.f., the first three equations in (A.70). Therefore, in this case, Assumption 7 is violated and a refinement is not needed.

Case 3: γ and some other parameters are allowed to switch. In this case, because $\nabla_\gamma \nabla_\gamma \widetilde{f}_{1t} = 2\nabla_{\sigma^2} \widetilde{f}_{1t}$, Assumption 7 is satisfied, and a refinement is needed.

Generalized linear model. The model is

$$y_t = g(\gamma + z_t'\alpha + u_t), \tag{A.71}$$

where $g(\cdot)$ is a smooth invertible function and $u_t \sim i.i.d.N(0, \sigma^2)$. Without loss of generality, we assume that $g(\cdot)$ is strictly increasing. An example of (A.71) in macroeconomics is a backward looking monetary policy rule: $R_t = R^*(\pi_{t-1}/\pi^*)^\alpha \exp(u_t)$, where R_t is the gross nominal interest rate, π_{t-1} is the lagged inflation (i.e, $\pi_t = P_t/P_{t-1}$ with P_t the price level at time t), and R^* and π^* are the steady-state levels of interest rate and inflation, respectively. This policy rule can be rewritten as (A.71), with $g(\cdot) = \exp(\cdot)$, $\gamma = \log R^* - \alpha \log \pi^*$, and $z_t = \log \pi_{t-1}$. The conditional density of y_t in (A.71), evaluated at $(\gamma, \alpha', \sigma^2)$, is

$$f_t(\gamma, \alpha, \sigma^2) = \frac{dg^{-1}(y_t)}{dy_t} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(g^{-1}(y_t) - \gamma - z_t'\alpha)^2}{2\sigma^2}\right).$$

In the policy rule example, $f_t(\gamma, \alpha, \sigma^2) = 1/(\sqrt{2\pi}\sigma R_t) \exp\left(-(\log R_t - \gamma - \alpha \log \pi_{t-1})^2/(2\sigma^2)\right)$. As in (A.68), we consider three cases, according to which parameter is allowed to switch.

Case 1: only γ is allowed to switch. We need to compute the first order derivatives of $f_t(\gamma, \alpha, \sigma^2)$ with respect to γ, α, and σ^2, and its second order derivative with respect to γ. Because $dg^{-1}(y_t)/dy_t$ does not depend on any parameter, the expressions in (A.69) continue to hold after replacing y_t by $g^{-1}(y_t)$. That is, we have

$$\begin{aligned}
\nabla_\alpha f_t(\gamma, \alpha, \sigma^2) &= f_t(\gamma, \alpha, \sigma^2)\left(\frac{z_t(g^{-1}(y_t) - \gamma - z_t'\alpha)}{\sigma^2}\right), \\
\nabla_{\sigma^2} f_t(\gamma, \alpha, \sigma^2) &= f_t(\gamma, \alpha, \sigma^2)\left(\frac{(g^{-1}(y_t) - \gamma - z_t'\alpha)^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right), \\
\nabla_\gamma f_t(\gamma, \alpha, \sigma^2) &= f_t(\gamma, \alpha, \sigma^2)\left(\frac{g^{-1}(y_t) - \gamma - z_t'\alpha}{\sigma^2}\right), \\
\nabla_\gamma \nabla_\gamma f_t(\gamma, \alpha, \sigma^2) &= f_t(\gamma, \alpha, \sigma^2)\left(\frac{(g^{-1}(y_t) - \gamma - z_t'\alpha)^2}{\sigma^4} - \frac{1}{\sigma^2}\right).
\end{aligned}$$

We obtain the same equations as in (A.70), which imply $\nabla_\gamma \nabla_\gamma \widetilde{f}_{1t} = 2\nabla_{\sigma^2} \widetilde{f}_{1t}$ for all t. Therefore, Assumption 7 holds, and a refinement is needed.

For Case 2 (α or σ^2 or both are allowed to switch) and Case 3 (γ and some other parameters are allowed to switch), the computation is similar to Cases 2 and 3 of (A.68), respectively, and the conclusions are the same. The details are thus omitted.

Stationary Gaussian vector autoregression. The model is

$$Y_t = \gamma + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + ... + \Phi_p Y_{t-p} + u_t,$$

where Y_t is k-dimensional random vector, γ is a k-dimensional parameter vector, $\Phi_1, ..., \Phi_p$ are k-by-k coefficient matrices, and $u_t \sim i.i.d.N(0, \Sigma)$ with Σ nonsingular.

Without loss of generality, we assume $p = 1$. The conditional density of Y_t is

$$f_t(\gamma, \Phi_1, \Sigma) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}(Y_t - \gamma - \Phi_1 Y_{t-1})'\Sigma^{-1}(Y_t - \gamma - \Phi_1 Y_{t-1})\right).$$

We consider three cases according to which parameter is allowed to switch.

Case 1: only γ is allowed to switch. We need to compute the first order derivatives of $f_t(\gamma, \Phi_1, \Sigma)$ with respect to γ, Φ_1, and Σ, and its second order derivative with respect to γ. They are given by

$$\begin{aligned}
\nabla_{\Phi_1} f_t(\gamma, \Phi_1, \Sigma) &= f_t(\gamma, \Phi_1, \Sigma)\Sigma^{-1}u_t Y_{t-1}', \\
\nabla_{\Sigma} f_t(\gamma, \Phi_1, \Sigma) &= \frac{1}{2} f_t(\gamma, \Phi_1, \Sigma)\{\Sigma^{-1}(u_t u_t' - \Sigma)\Sigma^{-1}\}, \\
\nabla_{\gamma} f_t(\gamma, \Phi_1, \Sigma) &= f_t(\gamma, \Phi_1, \Sigma)\Sigma^{-1}u_t, \\
\nabla_{\gamma}\nabla_{\gamma'} f_t(\gamma, \Phi_1, \Sigma) &= f_t(\gamma, \Phi_1, \Sigma)\{\Sigma^{-1}(u_t u_t' - \Sigma)\Sigma^{-1}\},
\end{aligned}$$

where $u_t = Y_t - \gamma - \Phi_1 Y_{t-1}$. Evaluating the equations at the MLE, we obtain

$$\nabla_{\gamma}\nabla_{\gamma'} f_t(\widetilde{\gamma}, \widetilde{\Phi}_1, \widetilde{\Sigma}) = 2\nabla_{\Sigma} f_t(\widetilde{\gamma}, \widetilde{\Phi}_1, \widetilde{\Sigma}) \text{ for any } t.$$

Therefore, a refinement is needed.

For Case 2 (Φ_1 or Σ or both are allowed to switch) and Case 3 (γ and some other parameters are allowed to switch), the computation is similar to Cases 2 and 3 of (A.68), respectively, and the conclusions are the same. The details are omitted.

A.3 Proofs for Subsection 6.1

In this section, we present three sets of results and then apply them to prove Lemma A.17, Proposition 2, and Corollary 2. The first set of results (Lemmas A.8-A.11) pertain to the first to the eighth order derivatives of $\xi_{t|t-1}(p_T, q_T, \beta, \delta_1, \delta_2)$ with respect to $(\beta', \delta_1', \delta_2')$. They are designed to be more informative than Lemmas 1, A.1, and A.2 when $\rho_T \to 0$. The second set of results (Lemmas A.12-A.14) provide approximations to the first to the third order derivatives of $\hat{\beta}(\delta_2)$ and $\hat{\delta}_1(\delta_2)$ with respect to δ_2. Their remainder terms are sharper than those in Lemma A.3 when $\rho_T \to 0$. The third set of results (Lemmas A.15-A.16) provide approximations to the fourth to the seventh order derivatives of $\hat{\beta}(\delta_2)$ and $\hat{\delta}_1(\delta_2)$ with respect to δ_2. They are needed to study the fifth to the eighth order terms in the Taylor expansion when $\rho_T \to 0$.

Define

$$\eta_{t|t-1}(p, q, \beta, \delta_1, \delta_2) = \frac{\xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2)}{\rho}. \tag{A.72}$$

For any $i_1, ..., i_k \in \{1, ..., n_\beta + 2n_\delta\}$, we write $\eta_{t|t-1}(p, q, \beta, \delta_1, \delta_2)$ and $\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_k}} \eta_{t|t-1}(p, q, \beta, \delta_1, \delta_2)$ as $\eta_{t|t-1}$ and $\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_k}} \eta_{t|t-1}$ to shorten the expressions. We continue to use "$\bar{A}$" to mean that the expression A is evaluated at $(\beta', \delta', \delta')$, where (β, δ) is a generic value in $\Theta\times\Delta$. Similarly, we use "$\widetilde{A}$" to mean that A is evaluated at the restricted MLE $\widetilde{\theta} = (\widetilde{\beta}', \widetilde{\delta}', \widetilde{\delta}')'$. For example, $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{\eta}_{t|t-1}$ and $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\widetilde{\eta}_{t|t-1}$ represent $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\eta_{t|t-1}$ evaluated at $(\beta', \delta', \delta')$ and $(\widetilde{\beta}', \widetilde{\delta}', \widetilde{\delta}')$, respectively.

Remark 4 *Lemmas A.8-A.11, presented below, hold uniformly over $\epsilon \le p, q \le 1-\epsilon$. In particular, they hold for any (p_T, q_T) satisfying $\epsilon \le p_T, q_T \le 1-\epsilon$ for all T and $p_T + q_T - 1 \to 0$ as $T \to \infty$.*

The next lemma can be compared with Lemma 1.

Lemma A.8 *For any $0 \le p, q \le 1$, $(\beta, \delta) \in \Theta \times \Delta$, $j, k, l \in \{1, ..., n_\beta + 2n_\delta\}$, and $t = 1, ..., T-1$:*

$$\begin{aligned}
&\nabla_{\theta_j}\bar{\eta}_{t+1,t} = \rho\nabla_{\theta_j}\bar{\eta}_{t|t-1} + \bar{\mathcal{V}}_{j,t},\\
&\nabla_{\theta_j}\nabla_{\theta_k}\bar{\eta}_{t+1|t} = \rho\nabla_{\theta_j}\nabla_{\theta_k}\bar{\eta}_{t|t-1} + \bar{\mathcal{V}}_{jk,t},\\
&\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\bar{\eta}_{t+1|t} = \rho\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\bar{\eta}_{t|t-1} + \bar{\mathcal{V}}_{jkl,t},
\end{aligned}$$

where $\bar{\mathcal{V}}_{j,t}, \bar{\mathcal{V}}_{jk,t}, \bar{\mathcal{V}}_{jkl,t}$ equal $\bar{\mathcal{E}}_{j,t}, \bar{\mathcal{E}}_{jk,t}, \bar{\mathcal{E}}_{jkl,t}$ in Lemma 1 after ρ and r in their expressions are replaced by 1 and $\xi_*(1-\xi_*)$, respectively.

Proof of Lemma A.8. The three results follow from Lemma 1 and the definition (A.72). For the first result, we only need to consider $j \in I_1$. (The sets I_0, I_1,and I_2 are defined following (13).) By Lemma 1.1, we have $\nabla_{\theta_j}\bar{\xi}_{t+1|t} = \rho\nabla_{\theta_j}\bar{\xi}_{t|t-1} + \rho\xi_*(1-\xi_*)\nabla_{\theta_j}\log\bar{f}_{1t}$. By (A.72), this equation is equivalent to $\rho\nabla_{\theta_j}\bar{\eta}_{t+1|t} = \rho^2\nabla_{\theta_j}\bar{\eta}_{t|t-1} + \rho\xi_*(1-\xi_*)\nabla_{\theta_j}\log\bar{f}_{1t}$. Dividing both sides by ρ, we obtain

$$\nabla_{\theta_j}\bar{\eta}_{t+1|t} = \rho\nabla_{\theta_j}\bar{\eta}_{t|t-1} + \xi_*(1-\xi_*)\nabla_{\theta_j}\log\bar{f}_{1t} = \rho\nabla_{\theta_j}\bar{\eta}_{t|t-1} + \bar{\mathcal{V}}_{j,t}. \tag{A.73}$$

This proves the first result of the Lemma. The remaining two results can be proved in the same way, i.e., starting with the equations for $\nabla_{\theta_j}\nabla_{\theta_k}\bar{\xi}_{t+1|t}$ and $\nabla_{\theta_j}\nabla_{\theta_k}\nabla_{\theta_l}\bar{\xi}_{t+1|t}$ in the proof of Lemma 1, applying (A.72), and then dividing both sides by ρ. The details are omitted. ■

We now study $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{\eta}_{t+1|t}$ for $k = 4, ..., 7$. Because this involves a large number of cases depending on $(i_1, ..., i_k)$, some consolidation is desirable. To achieve this, for any given $i_1, ..., i_k \in \{1, ..., n_\beta + 2n_\delta\}$, we define $\mathbb{A}_{i_1\cdots i_k}$ to be a set that includes $(\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_s}}\bar{f}_{1t})/\bar{f}_t$, $(\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_s}}\bar{f}_{2t})/\bar{f}_t$, and $\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_{s-1}}}\bar{\eta}_{t|t-1}$ for all $j_1, ..., j_s \in \{i_1, ..., i_k\}$ and $s \in \{1, ..., k\}$. We define another set $\mathbb{B}_{i_1\cdots i_k}$, which includes the products of the elements in $\mathbb{A}_{i_1\cdots i_k}$ with the requirement that the orders of their derivatives must sum up to k. To illustrate the structure of $\mathbb{B}_{i_1\cdots i_k}$, we suppose $k = 4$. Then, the following elements belong to $\mathbb{B}_{i_1\cdots i_k}$: $(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}\nabla_{\theta_{i_4}}\bar{f}_{1t})/\bar{f}_t$, $(\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}\bar{f}_{1t}\nabla_{\theta_{i_4}}\bar{f}_{2t})/\bar{f}_t^2$, $(\nabla_{\theta_{i_1}}\bar{f}_{1t}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}\bar{f}_{2t}\nabla_{\theta_{i_4}}\bar{\eta}_{t|t-1})/\bar{f}_t^2$, and $(\nabla_{\theta_{i_1}}\bar{f}_{1t}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}\bar{\eta}_{t|t-1}\nabla_{\theta_{i_4}}\bar{\eta}_{t|t-1})/\bar{f}_t^2$. The number of elements in $\mathbb{B}_{i_1\cdots i_k}$ is clearly finite for any finite k. We denote this number by M_k.

Lemma A.9 *For any $0 \le p, q \le 1, (\beta, \delta) \in \Theta \times \Delta, i_1, ..., i_k \in \{1, ..., n_\beta + 2n_\delta\}$, and $k \in \{4, 5, 6, 7\}$, we have*

$$\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{\eta}_{t+1|t} = \rho\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{\eta}_{t|t-1} + \bar{\mathcal{V}}_{i_1...i_k,t}$$

with

$$\bar{\mathcal{V}}_{i_1...i_k,t} = \sum_{j=1}^{M_k} a_j b_{i_1\cdots i_k,j,t},$$

where $b_{i_1\cdots i_k,j,t}$ denotes the j-th element of $\mathbb{B}_{i_1\cdots i_k}$ at time t, and a_i is finite for any $i = 1, ..., M_k$.

Proof of Lemma A.9. We start with the $k=4$ case. By (A.9) and (A.72),

$$
\begin{aligned}
&\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}\eta_{t+1|t} \qquad\qquad (A.74)\\
=\;& \frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}A_t}{B_t} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}A_t\nabla_{\theta_{i_4}}B_t}{B_t^2} \\
&+\nabla_{\theta_{i_4}}\Bigg\{-\frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}A_t\nabla_{\theta_{i_3}}B_t}{B_t^2} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_3}}A_t\nabla_{\theta_{i_2}}B_t}{B_t^2} - \frac{\nabla_{\theta_{i_1}}A_t\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}B_t}{B_t^2} \\
&+\frac{2\nabla_{\theta_{i_1}}A_t\nabla_{\theta_{i_2}}B_t\nabla_{\theta_{i_3}}B_t}{B_t^3} - \frac{\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}A_t\nabla_{\theta_{i_1}}B_t}{B_t^2} - \frac{\nabla_{\theta_{i_2}}A_t\nabla_{\theta_{i_1}}\nabla_{\theta_{i_3}}B_t}{B_t^2} + \frac{2\nabla_{\theta_{i_2}}A_t\nabla_{\theta_{i_1}}B_t\nabla_{\theta_{i_3}}B_t}{B_t^3} \\
&-\frac{\nabla_{\theta_{i_3}}A_t\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}B_t}{B_t^2} - \frac{A_t\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}B_t}{B_t^2} + \frac{2A_t\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}B_t\nabla_{\theta_{i_3}}B_t}{B_t^3} + \frac{2\nabla_{\theta_{i_3}}A_t\nabla_{\theta_{i_1}}B_t\nabla_{\theta_{i_2}}B_t}{B_t^3} \\
&+\frac{2A_t\nabla_{\theta_{i_1}}\nabla_{\theta_{i_3}}B_t\nabla_{\theta_{i_2}}B_t}{B_t^3} + \frac{2A_t\nabla_{\theta_{i_1}}B_t\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}B_t}{B_t^3} - \frac{6A_t\nabla_{\theta_{i_1}}B_t\nabla_{\theta_{i_2}}B_t\nabla_{\theta_{i_3}}B_t}{B_t^4}\Bigg\},
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}A_t}{B_t} =\;& \frac{1}{B_t}\Big\{(\xi_{t|t-1}-1)\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}\nabla_{\theta_{i_4}}f_{2t} + \nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{2t}\nabla_{\theta_{i_4}}\xi_{t|t-1} \qquad (A.75)\\
&+\nabla_{\theta_{i_1}}\nabla_{\theta_{i_3}}\nabla_{\theta_{i_4}}f_{2t}\nabla_{\theta_{i_2}}\xi_{t|t-1} + \nabla_{\theta_{i_1}}\nabla_{\theta_{i_3}}f_{2t}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_4}}\xi_{t|t-1} \\
&+\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}\nabla_{\theta_{i_4}}f_{2t}\nabla_{\theta_{i_1}}\xi_{t|t-1} + \nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{2t}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_4}}\xi_{t|t-1} \\
&+\nabla_{\theta_{i_3}}\nabla_{\theta_{i_4}}f_{2t}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\xi_{t|t-1} + \nabla_{\theta_{i_3}}f_{2t}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_4}}\xi_{t|t-1} \\
&+\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_4}}f_{2t}\nabla_{\theta_{i_3}}\xi_{t|t-1} + \nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}f_{2t}\nabla_{\theta_{i_3}}\nabla_{\theta_{i_4}}\xi_{t|t-1} \\
&+\nabla_{\theta_{i_1}}\nabla_{\theta_{i_4}}f_{2t}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}\xi_{t|t-1} + \nabla_{\theta_{i_1}}f_{2t}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}\nabla_{\theta_{i_4}}\xi_{t|t-1} \\
&+\nabla_{\theta_{i_2}}\nabla_{\theta_{i_4}}f_{2t}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_3}}\xi_{t|t-1} + \nabla_{\theta_{i_2}}f_{2t}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_3}}\nabla_{\theta_{i_4}}\xi_{t|t-1} \\
&+\nabla_{\theta_{i_4}}f_{2t}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}\xi_{t|t-1} + f_{2t}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}\xi_{t|t-1}\Big\}.
\end{aligned}
$$

On the right hand side of (A.74), only the first term involves the fourth order derivative of $\xi_{t|t-1}$. By (A.75), it includes 16 elements. The first 15 elements belong to $\mathbb{B}_{i_1\cdots i_4}$ (up to multiplication by constants) when evaluated at $\delta_1=\delta_2=\delta$. The 16th element, when evaluated at $\delta_1=\delta_2=\delta$, equals $\bar{f}_{2t}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}\bar{\xi}_{t|t-1}/\bar{B}_t = \nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}\bar{\xi}_{t|t-1} = \rho\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}\bar{\eta}_{t|t-1}$, which holds because $\bar{B}_t = \bar{f}_{2t} = \bar{f}_{1t} = \bar{f}_t$. The remaining components of (A.74) all belong to the set $\mathbb{B}_{i_1\cdots i_4}$ (up to multiplication by constants) when evaluated at $\delta_1=\delta_2=\delta$. This proves the Lemma for the $k=4$ case.

For the $k=5$ case, we compute the derivative of (A.74) with respect to θ_{i_5}. Then, we observe that only $(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_5}}A_t)/B_t$ involves the fifth derivative of $\eta_{t|t-1}$, in the form of $\rho\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_5}}\bar{\eta}_{t|t-1}$, and that the remaining terms of $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_5}}\bar{\eta}_{t+1|t}$ all belong to $\mathbb{B}_{i_1\cdots i_5}$ (up to multiplication by constants). This process can be continued for $k=6,7$, and 8. We omit the details. ■

The next lemma can be compared with Lemma A.1.

Lemma A.10 *Suppose Assumptions 8 and 9(i) hold. Then, there exists an open neighborhood of (β_*,δ_*), denoted by $B(\beta_*,\delta_*)$, and a sequence of strictly stationary and ergodic random variables $\{\lambda_t\}$ satisfying $E\lambda_t^{1+c}<M<\infty$ for some $c,M>0$, such that*

$$\sup_{(\beta,\delta_1)\in B(\beta_*,\delta_*)}\left|\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{\eta}_{t+1|t}\right|^{\frac{\alpha(k)}{k}} < \lambda_t$$

for any $t \in \{1,...,T\}$, $i_1,...,i_k \in \{1,..,2n_\delta + n_\beta\}$, *and* $k \in \{1,...,8\}$, *where* $\alpha(k) = 12$ *for* $k \in \{1,...,7\}$, *and* $\alpha(8) = 8$. *This inequality holds uniformly over* $\epsilon \leq p, q \leq 1-\epsilon$ *for any* $0 < \epsilon < 1/2$.

Proof of Lemma A.10. The proof is similar to that of Lemma A.1. We first consider $k = 1$ and then apply the results cumulatively to study $k = 2, 3, ..., 8$.

When $k = 1$, without loss of generality, suppose $j \in I_1$. By (A.73),

$$\left|\nabla_{\theta_j}\bar{\eta}_{t+1|t}\right|^{12} \leq \left(\sum_{s=0}^{t-1}\left|\xi_*(1-\xi_*)\rho^s\frac{\nabla_{\theta_j}\bar{f}_{1(t-s)}}{\bar{f}_{t-s}}\right|\right)^{12} \leq \left(\sum_{s=0}^{\infty}|\rho^s|\,\upsilon_{t-s}^{1/12}\right)^{12} \leq \left(\sum_{s=0}^{\infty}(1-\epsilon)^s\upsilon_{t-s}^{1/12}\right)^{12}.$$

The upper bound is the same as that in (A.10), expect that 6 is replaced by 12. Denote this upper bound by λ_t. Then, the inequality (A.11) holds after 6 is replaced by 12, which implies $E\lambda_t^{1+c} < \infty$. This shows that the result of the Lemma holds when $k = 1$.

When $k = 2$, by Lemma A.8, $|\nabla_{\theta_j}\nabla_{\theta_i}\bar{\eta}_{t+1|t}|^6 \leq (\sum_{s=0}^{\infty}|\rho^s\bar{\mathcal{V}}_{ji,t-s}|)^6$. The expression of $\bar{\mathcal{V}}_{ji,t-s}$ depends on $(i,\ j)$. If $j \in I_0$ and $i \in I_1$,

$$|\bar{\mathcal{V}}_{ji,t}| \leq \left|-\frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\frac{\nabla_{\theta_i}\bar{f}_{1t}}{\bar{f}_t} + \frac{\nabla_{\theta_j}\nabla_{\theta_i}\bar{f}_{1t}}{\bar{f}_t}\right| \leq \left|\frac{\nabla_{\theta_j}\bar{f}_{2t}}{\bar{f}_t}\frac{\nabla_{\theta_i}\bar{f}_{1t}}{\bar{f}_t}\right| + \left|\frac{\nabla_{\theta_j}\nabla_{\theta_i}\bar{f}_{1t}}{\bar{f}_t}\right| \leq 2\upsilon_t^{1/6}.$$

The other cases of $(i,\ j)$ can be studied in the same way. As a result, $|\bar{\mathcal{V}}_{ji,t}| \leq C(\lambda_{t-1}^{1/12}\upsilon_t^{1/12} + \upsilon_t^{1/6})$ for any $i, j \in \{1,..,2n_\delta + n_\beta\}$ with C being a finite constant. This implies

$$\left|\nabla_{\theta_j}\nabla_{\theta_i}\bar{\eta}_{t+1|t}\right|^6 \leq \left(\sum_{s=0}^{\infty}C(1-\epsilon)^s(\lambda_{t-1}^{1/12}\upsilon_t^{1/12} + \upsilon_t^{1/6})\right)^6.$$

Denote the right hand side by λ_t. The inequalities (A.12)-(A.13) hold after 3 and 6 in their expressions are replaced by 6 and 12, respectively. Thus, $E\lambda_t^{1+c} < \infty$, which implies that the result of the Lemma holds when $k = 2$.

When $k \in \{3, ..., 7\}$, Lemmas A.8 or A.9 imply $|\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{\eta}_{t+1|t}|^{12/k} \leq (\sum_{s=0}^{\infty}|\rho^s\bar{\mathcal{V}}_{i_1...i_k,t-s}|)^{12/k}$. Using the definition of $\bar{\mathcal{V}}_{i_1...i_k,t-s}$ and the results for lower values of k, we have

$$|\bar{\mathcal{V}}_{i_1...i_k,t}| \leq C\sum_{i=1}^{k}\upsilon_t^{i/12}\lambda_{t-1}^{(k-i)/12}.$$

Therefore,

$$|\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\bar{\eta}_{t+1|t}|^{12/k} \leq \left(\sum_{s=0}^{\infty}C(1-\epsilon)^s\sum_{i=1}^{k}\upsilon_{t-s}^{i/12}\lambda_{t-s-1}^{(k-i)/12}\right)^{12/k}.$$

Denote the right hand side by λ_t. Then, the inequalities (A.12)-(A.13) hold, after 3 and 6 are replaced by $12/k$ and 12, respectively. This implies $E\lambda_t^{1+c} < \infty$.

Finally, for $k = 8$, we have

$$\begin{aligned}|\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_8}}\bar{\eta}_{t+1|t}| &\leq \left(\sum_{s=0}^{\infty}|\rho^s\bar{\mathcal{V}}_{i_1...i_8,t-s}|\right)\\ &\leq C\sum_{s=0}^{\infty}(1-\epsilon)^s\sum_{i=1}^{7}(\upsilon_{t-s}^{i/12}\lambda_{t-s-1}^{(8-i)/12}) + C\sum_{s=0}^{\infty}(1-\epsilon)^s\upsilon_{t-s}.\end{aligned}$$

Denote the right hand side by λ_t. Its first term can be studied as in (A.12)-(A.13). Its second term satisfies $E(\sum_{s=0}^{\infty}(1-\epsilon)^s \upsilon_{t-s})^{1+c} < C$ by Assumption 9(i) and the Minkowski's inequality. This implies $E\lambda_t^{1+c} < \infty$ for $k = 8$. ■

Lemma A.11 *Under Assumptions 8 and 9(i), for any* $i_1, ..., i_k \in \{1, .., 2n_\delta + n_\beta\}$ *and* $k \in \{1, ..., 8\}$, *we have* $T^{-1}\sum_{t=1}^{T} \widetilde{b}_{i_1 \cdots i_k, j, t} = O_p(1)$, *where* $\widetilde{b}_{i_1 \cdots i_k, j, t}$ *is the* j*-th element of* $\mathbb{B}_{i_1 \cdots i_k}$ *at time* t *evaluated at* $\widetilde{\theta}$. *Further, if* $\widetilde{b}_{i_1 \cdots i_k, j, t} = (\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_s}} \widetilde{f}_{it}/\widetilde{f}_{it})(\nabla_{\theta_{i_{s+1}}} ... \nabla_{\theta_{i_k}} \widetilde{\eta}_{t|t-1})$ *with* $i \in \{1, 2\}$,

$$T^{-1}\sum_{t=1}^{T} \widetilde{b}_{i_1 \cdots i_k, j, t} = \begin{cases} O_p\left(T^{-1/2}\right) & \text{if } \ k \in \{1, ..., 6\} \\ o_p(1) & \text{if } \ k = 7 \\ O_p(1) & \text{if } \ k = 8 \end{cases},$$

These results hold uniformly over $\epsilon \leq p, q \leq 1 - \epsilon$ *for any* $0 < \epsilon < 1/2$.

Proof of Lemma A.11. The proof is similar to that of Lemma A.2. We focus on the following element of $\mathbb{B}_{i_1 \cdots i_k}$ evaluated at $\widetilde{\theta}$: $\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_s}} \widetilde{f}_{1t} \nabla_{\theta_{i_{s+1}}} ... \nabla_{\theta_{i_k}} \widetilde{\eta}_{t|t-1}/\widetilde{f}_{1t}$. By the Taylor expansion,

$$\begin{aligned} & T^{-1}\sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_s}} f_{1t}^*}{f_t^*} \nabla_{\theta_{i_{s+1}}} ... \nabla_{\theta_{i_k}} \eta_{t|t-1}^* \\ & + T^{-1/2}\left\{T^{-1}\sum_{t=1}^{T} \nabla_{\theta'}\left(\frac{\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_s}} \bar{f}_{1t}}{\bar{f}_{1t}} \nabla_{\theta_{i_{s+1}}} ... \nabla_{\theta_{i_k}} \bar{\eta}_{t|t-1}\right)\right\} T^{1/2}\left(\widetilde{\theta} - \theta_*\right), \end{aligned} \tag{A.76}$$

where "$*$" and "$-$" mean that the expressions are evaluated at $\theta_*' = (\beta_*', \delta_*', \delta_*')$ and $\bar{\theta}' = (\bar{\beta}', \bar{\delta}', \bar{\delta}')$, respectively, where $\bar{\theta} = \widetilde{\theta} + c(\theta_* - \widetilde{\theta})$ for some $c \in (0, 1)$.

Between the two terms of (A.76), the first has the following properties: When $k \in \{1, ..., 6\}$, the $(2 + c)$th moment of its summand is finite by Lemma A.10. The CLT holds, implying that this term is $O_p\left(T^{-1/2}\right)$. When $k = 7$, the $(1 + c)$th moment of its summands is finite. The LLN holds, implying that this term is $o_p(1)$. At the same time, the expression inside the braces is $O_p(1)$ when $k \in \{1, ..., 7\}$ by Lemma A.10, and $T^{1/2}(\widetilde{\theta} - \theta_*) = O_p(1)$ because $\widetilde{\theta}$ is the restrictive MLE. These results imply that (A.76) is $O_p\left(T^{-1/2}\right)$ when $k \in \{1, ..., 6\}$ and $o_p(1)$ when $k = 7$. The uniformity with respect to (p, q) follows from the compactness of $[\epsilon, \leq 1 - \epsilon]$, the continuity of $\widetilde{b}_{i_1 \cdots i_k, j, t}$ with respect to ρ, and the proof of Lemma A.4.

When $k = 8$, Lemma A.10 can be directly applied to $T^{-1}\sum_{t=1}^{T} \nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_s}} \widetilde{f}_{1t} \nabla_{\theta_{i_{s+1}}} ... \nabla_{\theta_{i_8}} \widetilde{\eta}_{t|t-1}/\widetilde{f}_{1t}$, which implies that it is $O_p(1)$. ■

Remark 5 *To illustrate Lemma A.11, we consider* $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_s}} \widetilde{f}_{1t}/\widetilde{f}_t)(\nabla_{\theta_{i_{s+1}}} ... \nabla_{\theta_{i_k}} \widetilde{\xi}_{t|t-1})$. *By (A.72) and Lemma A.11, it equals*

$$\rho_T \frac{1}{T}\sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} ... \nabla_{\theta_{i_s}} \widetilde{f}_{1t} \nabla_{\theta_{i_{s+1}}} ... \nabla_{\theta_{i_k}} \widetilde{\eta}_{t|t-1}}{\widetilde{f}_t} = \begin{cases} O_p\left(|\rho_T| T^{-1/2}\right) & \text{if } \ k \in \{1, ..., 6\} \\ o_p(|\rho_T|) & \text{if } \ k = 7 \\ O_p(|\rho_T|) & \text{if } \ k = 8 \end{cases}.$$

The stochastic orders depend explicitly on $|\rho_T|$. *This feature enables us to identify the non-negligible terms in the Taylor expansion of the likelihood ratio under various drifting sequences of* ρ_T.

The next three lemmas pertain to the derivatives of $\hat{\beta}(\delta_2)$ and $\hat{\delta}_1(\delta_2)$ with respect to δ_2 evaluated at $\delta_2 = \widetilde{\delta}$. The dependence of $\hat{\beta}(\widetilde{\delta})$ and $\hat{\delta}_1(\widetilde{\delta})$ on (p_T, q_T) is suppressed to shorten the expressions. There are two key difference between them and Lemma A.3. First, they require Assumption 7. Second, their remainder terms are expressed in terms of both ρ_T and T, not just T. The second difference is important for them to be informative when $\rho_T \to 0$.

Lemma A.12 *Suppose that the null hypothesis and Assumptions 1-4 and 7 hold with (24) satisfied for all $i_1, i_2 \in \{1, ..., n_\delta\}$ and that $\epsilon \leq p_T, q_T \leq 1-\epsilon$ is satisfied for all T. Then, for any $k \in \{1, ..., n_\delta\}$,*

$$\begin{bmatrix} \nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta}) \\ \xi_T \nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} = (\xi_T - 1)\begin{bmatrix} 0 \\ e_k \end{bmatrix} + O_p(T^{-1/2}|\rho_T|),$$

where e_k is an n_δ-dimensional unit vector whose k-th element equals 1.

Proof of Lemma A.12. In the proof of Lemma A.3, the expressions between Display (A.17) and Display (A.24) do not involve approximations. We start with (A.25).

If $j \in \{1, ..., n_\beta\}$, (A.25) can be rewritten as

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\theta_j}\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\begin{bmatrix} \nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta}) \\ \xi_T \nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} + \frac{1}{T}\sum_{t=1}^{T}\frac{(1-\xi_T)\nabla_{\theta_j}\nabla_{\delta_{2k}'}\widetilde{f}_{2t}}{\widetilde{f}_t}.$$

If $j \in \{n_\beta + 1, ..., n_\beta + n_\delta\}$, it can be rewritten as

$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\xi_T\nabla_{\theta_j}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} \quad \frac{\nabla_{\theta_j}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t} + \frac{\nabla_{\theta_j}\widetilde{f}_{1t}\nabla_{\delta_1'}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T} + \frac{\nabla_{\theta_j}\widetilde{\xi}_{t|t-1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T}\right]\begin{bmatrix} \nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta}) \\ \xi_T \nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix}$$
$$-\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\nabla_{\theta_j}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t} + \frac{\nabla_{\theta_j}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t}\right).$$

The above $n_\beta + n_\delta$ cases can be expressed jointly as a vector:

$$\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & \frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t} \\ \frac{\xi_T\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & \frac{\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t} + \frac{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_1'}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T} + \frac{\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T} \end{bmatrix}\begin{bmatrix} \nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta}) \\ \xi_T \nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} \quad \text{(A.77)}$$
$$+\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \frac{(1-\xi_T)\nabla_\beta\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{f_t} \\ -\frac{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t} - \frac{\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t} \end{bmatrix}.$$

Combining this with (A.24) and applying Assumption 7, we obtain

$$\left\{\widetilde{I} - \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{f} & \frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t} \\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & \frac{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_1'}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2} + \frac{\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2} \end{bmatrix}\right\}\begin{bmatrix} \nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta}) \\ \xi_T \nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix}$$
$$= \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} -\frac{(1-\xi_T)\nabla_\beta\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t} + \frac{(1-\xi_T)\nabla_\beta\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{f_t} \\ -\frac{(1-\xi_T)\nabla_{\delta_1}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t} - \frac{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T} - \frac{\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T} \end{bmatrix},$$

which leads to, after rearranging the terms,

$$\left\{\widetilde{I}-\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{cc}\frac{\nabla_{\beta}\nabla_{\beta'}\widetilde{f}_{1t}}{f} & \frac{\nabla_{\beta}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & 0\end{array}\right]\right\}\left[\begin{array}{c}\nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right] \tag{A.78}$$

$$= \frac{1-\xi_T}{T}\sum_{t=1}^{T}\left[\begin{array}{c}-\frac{\nabla_{\beta}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t}+\frac{\nabla_{\beta}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{f_t}\\ -\frac{\nabla_{\delta_1}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{array}\right]$$

$$+\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{c}0\\ \left(\frac{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_1'}\widetilde{\xi}_{t|t-1}}{\xi_T\widetilde{f}_t}+\frac{\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\xi_T\widetilde{f}_t}\right)\nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta})-\left(\frac{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}}{\xi_T\widetilde{f}_t}+\frac{\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\xi_T\widetilde{f}_t}\right)\end{array}\right].$$

By Lemmas A.11 and A.3, the second term on the right hand side is $O_p\left(T^{-1/2}\left|\rho_T\right|\right)$. Thus,

$$\left\{\widetilde{I}-\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{cc}\frac{\nabla_{\beta}\nabla_{\beta'}\widetilde{f}_{1t}}{f} & \frac{\nabla_{\beta}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & 0\end{array}\right]\right\}\left[\begin{array}{c}\nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right] \tag{A.79}$$

$$= \frac{1-\xi_T}{T}\sum_{t=1}^{T}\left[\begin{array}{c}-\frac{\nabla_{\beta}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t}+\frac{\nabla_{\beta}\nabla_{\delta_{1k}'}f}{f}\\ -\frac{\nabla_{\delta_1}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{array}\right]+\left[\begin{array}{c}0\\ O_p\left(T^{-1/2}\left|\rho_T\right|\right)\end{array}\right].$$

This represents a system of $n_\beta+n_\delta$ equations with $\nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta})$ and $\xi_T\nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta})$ being unknowns. Its structure is the same as (A.26), except that the remainder term is $O_p(T^{-1/2}\left|\rho_T\right|)$ instead of $O_p(T^{-1/2})$. Therefore, it can be solved in the same way as in Step 3 of Lemma 2.1; c.f. (A.27). This produces the same leading term as that of Lemma 2.1 with a remainder term of $O_p\left(T^{-1/2}\left|\rho_T\right|\right)$.■

Lemma A.13 *Suppose that the null hypothesis and Assumptions 1-4 and 7 hold with (24) satisfied for all $i_1, i_2 \in \{1,...,n_\delta\}$, and that $\epsilon \le p_T, q_T \le 1-\epsilon$ is satisfied for all T. Then, for any $k, l \in \{1,...,n_\delta\}$,*

$$\left[\begin{array}{c}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]$$

$$= -\left(\frac{1-\xi_T}{\xi_T}\right)\left[\begin{array}{c}\alpha_{kl}^{(1)}\\ \alpha_{kl}^{(2)}\end{array}\right]-\widetilde{I}^{-1}\frac{1}{T\xi_T^2}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}+\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t}\right)$$

$$+\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{c}\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\beta}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}\\ \left(\frac{1-\xi_T}{\xi_T}\right)^2\frac{\nabla_{\delta_1}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{array}\right]-\left(\frac{1-\xi_T}{\xi_T}\right)\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{c}\frac{\nabla_{\beta}\nabla_{\beta'}\widetilde{f}_{1t}\alpha_{kl}^{(1)}}{\widetilde{f}_t}+\frac{\nabla_{\beta}\nabla_{\delta_1'}\widetilde{f}_{1t}\alpha_{kl}^{(2)}}{\widetilde{f}_t}\\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}\alpha_{kl}^{(1)}}{\widetilde{f}_t}\end{array}\right]$$

$$+O_p\left(T^{-1/2}\left|\rho_T\right|+T^{-1}\right)$$

Remark 6 *Among the first four terms on the right hand side, the second is $O_p\left(\left|\rho_T\right|\right)$, and the third and fourth are $O_p\left(T^{-1/2}\right)$. In the proof of Lemma A.17, their effects on the likelihood function* *depend on the relationship between ρ_T and $T^{-1/4}$. In particular, the effect of the second term on the fourth order derivative of the log likelihood dominates (is dominated by) that of the other two terms when ρ_T converges to 0 slower (faster) than $O(T^{-1/4})$. The three terms all have a first order effect on the fourth order derivative of the likelihood when the converging rate of ρ_T is exactly $T^{-1/4}$.*

Proof of Lemma A.13. We study the five terms in (A.28) separately. The expressions between (A.29) and (A.31) do not involve approximations. They hold after ξ_* is replaced by ξ_T.

We consider the first term in (A.28). Because of (A.29),

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\widetilde{M}_{jt}}{\widetilde{f}_t} \\
= \; & \frac{1}{T}\sum_{s=1}^{n_\beta+2n_\delta}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\Big\{\nabla_{\theta'}\widetilde{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\widetilde{f}_{1t}+\xi_T\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\widetilde{f}_{1t}-\nabla_{\theta'}\widetilde{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\widetilde{f}_{2t} \\
& +(1-\xi_T)\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\widetilde{f}_{2t}+(\nabla_{\theta_j}\nabla_{\theta'}\widetilde{f}_{1t}-\nabla_{\theta_j}\nabla_{\theta'}\widetilde{f}_{2t})\nabla_{\theta_s}\widetilde{\xi}_{t|t-1}+(\nabla_{\theta_j}\widetilde{f}_{1t}-\nabla_{\theta_j}\widetilde{f}_{2t})\nabla_{\theta_s}\nabla_{\theta'}\widetilde{\xi}_{t|t-1} \\
& +(\nabla_{\theta_s}\widetilde{f}_{1t}-\nabla_{\theta_s}\widetilde{f}_{2t})\nabla_{\theta_j}\nabla_{\theta'}\widetilde{\xi}_{t|t-1}+(\nabla_{\theta_s}\nabla_{\theta'}\widetilde{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta'}\widetilde{f}_{2t})\nabla_{\theta_j}\widetilde{\xi}_{t|t-1} \\
& +(\nabla_{\theta'}\widetilde{f}_{1t}-\nabla_{\theta'}\widetilde{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\widetilde{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2l}}\hat{\theta}(\widetilde{\delta}) \\
& +\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\Big\{\xi_T\nabla_{\theta_j}\nabla_{\theta'}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_j}\nabla_{\theta'}\widetilde{f}_{2t}+(\nabla_{\theta_j}\widetilde{f}_{1t}-\nabla_{\theta_j}\widetilde{f}_{2t})\nabla_{\theta'}\widetilde{\xi}_{t|t-1} \\
& +(\nabla_{\theta'}\widetilde{f}_{1t}-\nabla_{\theta'}\widetilde{f}_{2t})\nabla_{\theta_j}\widetilde{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\widetilde{\delta}).
\end{aligned}$$

Using Lemma A.11, we obtain the following approximation to it

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\widetilde{M}_{jt}}{\widetilde{f}_t} \\
= \; & \frac{1}{T}\sum_{s=1}^{n_\beta+2n_\delta}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\Big\{\xi_T\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta'}\widetilde{f}_{2t}\Big\}\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2l}}\hat{\theta}(\widetilde{\delta}) \\
& +\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\Big\{\xi_T\nabla_{\theta_j}\nabla_{\theta'}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_j}\nabla_{\theta'}\widetilde{f}_{2t}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\widetilde{\delta})+O_p\Big(T^{-1/2}\left|\rho_T\right|\Big) \\
= \; & \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\left\{\frac{(\xi_T-1)^2}{\xi_T}\nabla_{\theta_j}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_j}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\widetilde{f}_{2t}\right\} \\
& +\frac{1}{T}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\Big\{\xi_T\nabla_{\theta_j}\nabla_{\theta'}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_j}\nabla_{\theta'}\widetilde{f}_{2t}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\widetilde{\delta})+O_p\Big(T^{-1/2}\left|\rho_T\right|\Big),
\end{aligned}$$

where the last equality is due to Lemma A.12. Because this holds for any $j \in \{1,..,n_\beta+n_\delta\}$, the resulting approximations can be represented jointly as

$$D\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_\beta\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}\\ \left(\frac{1-\xi_T}{\xi_T}\right)^2\frac{\nabla_{\delta_1}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{bmatrix} \tag{A.80}$$

$$+D\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}\frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})+\frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta})\\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})+\frac{\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}+O_p\left(T^{-1/2}\left|\rho_T\right|\right),$$

where D is a diagonal matrix whose first n_β diagonal elements equal 1 and the rest ξ_T.

Next, consider the fourth term in (A.28). Applying (A.33) to $\nabla_{\delta_{2k}}\widetilde{B}_t$, we obtain

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\widetilde{B}_t\nabla_{\delta_{2l}}\widetilde{M}_{jt}}{\widetilde{B}_t^2} \\ =\;&\sum_{s=1}^{n_\beta}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\beta_s}\widetilde{f}_{1t}\nabla_{\delta_{2l}}\widetilde{M}_{jt}}{\widetilde{f}_t^2}\right)\nabla_{\delta_{2k}}\hat{\beta}_s(\widetilde{\delta})+\sum_{s=1,s\neq k}^{n_\delta}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\xi_T\nabla_{\delta_{1s}}\widetilde{f}_{1t}\nabla_{\delta_{2l}}\widetilde{M}_{jt}}{\widetilde{f}_t^2}\right)\nabla_{\delta_{2k}}\hat{\delta}_{1s}(\widetilde{\delta}) \\ &+\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}\nabla_{\delta_{2l}}\widetilde{M}_{jt}}{\widetilde{f}_t^2}\right)\left(\xi_T\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\widetilde{\delta})+(1-\xi_T)\right)=O_p(T^{-1/2}\left|\rho_T\right|),\end{aligned} \tag{A.81}$$

where the second equality follows from Lemma A.12. Because the second and fifth terms in (A.28) depend on the first order derivative of B_t, they are also $O_p(T^{-1/2}\left|\rho_T\right|)$.

It remains to study the third term in (A.28). Applying (A.31), we have

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t^2}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\widetilde{B}_t \\ =\;&\sum_{s,u=1}^{n_\beta+2n_\delta}\Bigg\{\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t^2}\Big[\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}\nabla_{\theta_s}\widetilde{f}_{1t}+\xi_T\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}\nabla_{\theta_s}\widetilde{f}_{2t} \\ &+(1-\xi_T)\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t}+(\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\theta_s}\widetilde{\xi}_{t|t-1}\Big]\Bigg\}\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta}) \\ &+\sum_{s=1}^{n_\beta+n_\delta}\left\{\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t^2}\left[\xi_T\nabla_{\theta_s}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_s}\widetilde{f}_{2t}\right]\right\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta}).\end{aligned}$$

By Lemma A.12, $\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})$ and $\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})$ are $O_p\left(T^{-1/2}\left|\rho_T\right|\right)$ except when $s\in\{n_\beta+k, n_\beta+n_\delta+k\}$

and $u \in \{n_\beta + l, n_\beta + n_\delta + l\}$. Thus,

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T} \frac{\widetilde{M}_{jt}}{\widetilde{B}_t^2} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \widetilde{B}_t \\
= \; & \frac{1}{T}\sum_{t=1}^{T} \frac{\xi_T \nabla_{\theta_j} \widetilde{f}_{1t} + (1-\xi_T)\nabla_{\theta_j} \widetilde{f}_{2t}}{\widetilde{f}_t^2} \left\{ \left(\frac{1-\xi_T}{\xi_T}\right) \nabla_{\delta_{1k}} \nabla_{\delta_{1l}} \widetilde{f}_{1t} \right. \\
& \left. + \frac{1}{\xi_T^2} \nabla_{\delta_{1l}} \widetilde{\xi}_{t|t-1} \nabla_{\delta_{1k}} \widetilde{f}_{1t} + \frac{1}{\xi_T^2} \nabla_{\delta_{1l}} \widetilde{f}_{1t} \nabla_{\delta_{1k}} \widetilde{\xi}_{t|t-1} \right\} \\
& + \sum_{s=1}^{n_\beta + n_\delta} \left\{ \frac{1}{T}\sum_{t=1}^{T} \frac{\xi_T \nabla_{\theta_j} \widetilde{f}_{1t} + (1-\xi_T)\nabla_{\theta_j} \widetilde{f}_{2t}}{\widetilde{f}_t^2} [\xi_T \nabla_{\theta_s} \widetilde{f}_{1t} + (1-\xi_T)\nabla_{\theta_s} \widetilde{f}_{2t}] \right\} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\theta}_s(\widetilde{\delta}) \\
& + O_p(T^{-1/2} |\rho_T|).
\end{aligned}
$$

The right hand side is the same as (A.32), except that the remainder term is $O_p(T^{-1/2} |\rho_T|)$ instead of $O_p(T^{-1/2})$. These $n_\beta + n_\delta$ equations for $j = 1, ..., n_\beta + n_\delta$ can be represented jointly as

$$
D\widetilde{I} \begin{bmatrix} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\beta}(\widetilde{\delta}) \\ \xi_T \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} + D\frac{1}{T}\sum_{t=1}^{T} \frac{\nabla_{(\beta', \delta_1')'} \widetilde{f}_{1t}}{\widetilde{f}_t} \widetilde{U}_{kl,t} + O_p(T^{-1/2} |\rho_T|),
$$

where $\widetilde{U}_{kl,t}$ is defined in (17) with ξ_* replaced by ξ_T.

Combining the above results, we can write (A.28) as

$$
\begin{aligned}
\widetilde{I} \begin{bmatrix} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\beta}(\widetilde{\delta}) \\ \xi_T \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} = \; & -\frac{1}{T}\sum_{t=1}^{T} \frac{\nabla_{(\beta', \delta_1')'} \widetilde{f}_{1t}}{\widetilde{f}_t} \widetilde{U}_{kl,t} + \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \left(\frac{1-\xi_T}{\xi_T}\right) \frac{\nabla_\beta \nabla_{\delta_{1k}} \nabla_{\delta_{1l}} \widetilde{f}_{1t}}{\widetilde{f}_t} \\ \left(\frac{1-\xi_T}{\xi_T}\right)^2 \frac{\nabla_{\delta_1} \nabla_{\delta_{1k}} \nabla_{\delta_{1l}} \widetilde{f}_{1t}}{\widetilde{f}_t} \end{bmatrix} \\
& + \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \frac{\nabla_\beta \nabla_{\beta'} \widetilde{f}_{1t}}{\widetilde{f}_t} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\beta}(\widetilde{\delta}) + \frac{\nabla_\beta \nabla_{\delta_1'} \widetilde{f}_{1t}}{\widetilde{f}_t} \xi_T \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\delta}_1(\widetilde{\delta}) \\ \frac{\nabla_{\delta_1} \nabla_{\beta'} \widetilde{f}_{1t}}{\widetilde{f}_t} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\beta}(\widetilde{\delta}) + \frac{\nabla_{\delta_1} \nabla_{\delta_1'} \widetilde{f}_{1t}}{\widetilde{f}_t} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} \\
& + O_p\left(T^{-1/2} |\rho_T|\right).
\end{aligned}
$$

Dividing both sides by $\widetilde{I}$:

$$
\begin{aligned}
& \begin{bmatrix} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\beta}(\widetilde{\delta}) \\ \xi_T \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} \qquad \text{(A.82)} \\
= \; & -\widetilde{I}^{-1} \frac{1}{T}\sum_{t=1}^{T} \frac{\nabla_{(\beta', \delta_1')'} \widetilde{f}_{1t}}{\widetilde{f}_t} \widetilde{U}_{kl,t} + \widetilde{I}^{-1} \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \left(\frac{1-\xi_T}{\xi_T}\right) \frac{\nabla_\beta \nabla_{\delta_{1k}} \nabla_{\delta_{1l}} \widetilde{f}_{1t}}{\widetilde{f}_t} \\ \left(\frac{1-\xi_T}{\xi_T}\right)^2 \frac{\nabla_{\delta_1} \nabla_{\delta_{1k}} \nabla_{\delta_{1l}} \widetilde{f}_{1t}}{\widetilde{f}_t} \end{bmatrix} \\
& + \widetilde{I}^{-1} \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \frac{\nabla_\beta \nabla_{\beta'} \widetilde{f}_{1t}}{\widetilde{f}_t} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\beta}(\widetilde{\delta}) + \frac{\nabla_\beta \nabla_{\delta_1'} \widetilde{f}_{1t}}{\widetilde{f}_t} \xi_T \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\delta}_1(\widetilde{\delta}) \\ \frac{\nabla_{\delta_1} \nabla_{\beta'} \widetilde{f}_{1t}}{\widetilde{f}_t} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\beta}(\widetilde{\delta}) \end{bmatrix} + O_p\left(T^{-1/2} |\rho_T|\right).
\end{aligned}
$$

By Assumption 7 and the definition of $\widetilde{U}_{kl,t}$, we have

$$
\begin{aligned}
& \widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{U}_{kl,t} \\
= & \left(\frac{1-\xi_T}{\xi_T}\right)\alpha_{kl}+\widetilde{I}^{-1}\frac{1}{T\xi_T^2}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}+\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t}\right).
\end{aligned}
$$

Apply this representation to the first right hand side term of (A.82), we have

$$
\begin{aligned}
& \begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) \\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix}+\left(\frac{1-\xi_T}{\xi_T}\right)\alpha_{kl} \qquad \text{(A.83)} \\
& +\widetilde{I}^{-1}\frac{1}{T\xi_T^2}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}+\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t}\right) \\
= & \widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_\beta\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t} \\ \left(\frac{1-\xi_T}{\xi_T}\right)^2\frac{\nabla_{\delta_1}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t} \end{bmatrix} \\
& +\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})+\frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) \end{bmatrix}+O_p\left(T^{-1/2}\left|\rho_T\right|\right).
\end{aligned}
$$

This system of equations can be solved in two steps. First, because the two terms on the right hand side are $O_p\left(T^{-1/2}\right)$, we have

$$
\begin{aligned}
& \begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) \\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} \\
= & \left(\frac{\xi_T-1}{\xi_T}\right)\alpha_{kl}-\widetilde{I}^{-1}\frac{1}{T\xi_T^2}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}+\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t}\right)+O_p\left(T^{-1/2}\right).
\end{aligned}
$$

Second, applying this solution to the second term on the right hand side of (A.83), we have

$$
\begin{aligned}
& \begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) \\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} \\
= & \left(\frac{\xi_T-1}{\xi_T}\right)\alpha_{kl}-\widetilde{I}^{-1}\frac{1}{T\xi_T^2}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}+\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t}\right) \\
& +\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_\beta\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t} \\ \left(\frac{1-\xi_T}{\xi_T}\right)^2\frac{\nabla_{\delta_1}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t} \end{bmatrix}-\left(\frac{1-\xi_T}{\xi_T}\right)\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{kl}^{(1)}+\frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{kl}^{(2)} \\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{kl}^{(1)} \end{bmatrix} \\
& +O_p\left(T^{-1/2}\left|\rho_T\right|+T^{-1}\right). \qquad \blacksquare
\end{aligned}
$$

Lemma A.14 *Suppose that the null and Assumptions 1-4 and 7 hold with (24) satisfied for all $i_1, i_2 \in \{1, ..., n_\delta\}$, and that $\epsilon \leq p_T, q_T \leq 1-\epsilon$ is satisfied for all T. Then, for any $k, l, h \in \{1, ..., n_\delta\}$,*

$$\begin{aligned}
&\begin{bmatrix} \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\beta}(\widetilde{\delta}) \\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix} \\
= \;& \frac{(1-2\xi_T)(1-\xi_T)}{\xi_T^2}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t} \\
&+(1-\xi_T)\sum_{u=1}^{n_\delta}\left(\alpha_{ku}\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\delta}_{1u}(\widetilde{\delta})+\alpha_{lu}\nabla_{\delta_{2h}}\nabla_{\delta_{2k}}\hat{\delta}_{1u}(\widetilde{\delta})+\alpha_{hu}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_{1u}(\widetilde{\delta})\right)+O_p(|\rho_T|+T^{-1/2}).
\end{aligned}$$

Remark 7 *This expression is needed to study the sixth order derivative of the log likelihood.*

Proof of Lemma A.14. We divide the fifteen components in the summation of (A.34) into 4 subsets, such that those in the i-th subset depend on the $(i-1)$-th order derivative of B_t, but not on its lower order derivatives. We study these 4 subsets separately. The components in the first two subsets are $O_p\left(T^{-1/2}\right)$ and $O_p(T^{-1/2}|\rho_T|)$ respectively, by the CLT and Lemma A.12. Let R_t denote a remainder term that can differ between cases.

We show that the elements of the third subset are $O_p\left(T^{-1/2}+\rho_T^2\right)$. Consider the following representative element

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\widetilde{M}_{jt}}{\widetilde{B}_t}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\widetilde{B}_t}{\widetilde{B}_t} \\
= \;& \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\widetilde{M}_{jt}}{\widetilde{f}_t}\left(\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{1h}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}+\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right) \\
&+\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\widetilde{M}_{jt}}{\widetilde{f}_t}\sum_{s=1}^{n_\beta+n_\delta}\left(\frac{\xi_T\nabla_{\theta_s}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_s}\widetilde{f}_{2t}}{\widetilde{f}_t}\right)\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})+O_p\left(T^{-1/2}|\rho_T|\right),
\end{aligned}$$

where $O_p\left(T^{-1/2}|\rho_T|\right)$ follows because, by Lemma A.12, $\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})$ and $\nabla_{\delta_{2h}}\hat{\theta}_s(\widetilde{\delta})$ are $O_p\left(T^{-1/2}|\rho_T|\right)$ except when $s\in\{n_\beta+h, n_\beta+n_\delta+h\}$ and $u\in\{n_\beta+l, n_\beta+n_\delta+l\}$. Thus,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\widetilde{M}_{jt}}{\widetilde{B}_t}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\widetilde{B}_t}{\widetilde{B}_t} \\
= \;& \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\widetilde{M}_{jt}}{\widetilde{f}_t}\left\{\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{1h}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}+\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right. \\
&\left.+\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\begin{bmatrix} \nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta}) \\ \xi_T\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix}\right\}+O_p\left(T^{-1/2}|\rho_T|\right).
\end{aligned} \tag{A.84}$$

Applying Lemma A.13 to the second term on its right hand side, we have

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\widetilde{M}_{jt}}{\widetilde{B}_t}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\widetilde{B}_t}{\widetilde{B}_t} \qquad \text{(A.85)}\\
= \;& \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\widetilde{M}_{jt}}{\widetilde{B}_t}\left\{\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right.\\
&-\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1h}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right)\\
&+\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{c}\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\beta}\nabla_{\delta_{1h}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}\\ \left(\frac{1-\xi_T}{\xi_T}\right)^2\frac{\nabla_{\delta_1}\nabla_{\delta_{1h}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{array}\right]\\
&\left.-\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{1-\xi_T}{\xi_T}\right)\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{c}\frac{\nabla_{\beta}\nabla_{\beta'}\widetilde{f}_{1t}\alpha_{lh}^{(1)}}{\widetilde{f}_t}+\frac{\nabla_{\beta}\nabla_{\delta_1'}\widetilde{f}_{1t}\alpha_{lh}^{(2)}}{\widetilde{f}_t}\\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}\alpha_{lh}^{(1)}}{\widetilde{f}_t}\end{array}\right]\right\}+O_p(T^{-1/2}\left|\rho_T\right|+T^{-1}).
\end{aligned}
$$

The last two summations satisfy a CLT. Therefore, the right hand side equals

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\widetilde{M}_{jt}}{\widetilde{B}_t}\left\{\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right. \qquad \text{(A.86)}\\
&\left.-\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1h}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right)\right\}+O_p(T^{-1/2}).
\end{aligned}
$$

We now study the structure of $\nabla_{\delta_{2k}}\widetilde{M}_{jt}/\widetilde{B}_t$ to obtain further approximations. By (A.77),

$$\begin{bmatrix} \frac{\nabla_{\delta_{2k}}\widetilde{M}_{1t}}{\widetilde{B}_t} \\ \cdots \\ \frac{\nabla_{\delta_{2k}}\widetilde{M}_{(n_\beta+n_\delta)t}}{\widetilde{B}_t} \end{bmatrix} \tag{A.87}$$

$$= \begin{bmatrix} \frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & \frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t} \\ \frac{\xi_T\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & \frac{\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t} + \frac{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_1'}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T} + \frac{\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T} \end{bmatrix} \begin{bmatrix} \nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta}) \\ \xi_T\nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta}) \end{bmatrix}$$

$$+ \begin{bmatrix} \frac{(1-\xi_T)\nabla_\beta\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t} \\ -\frac{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t} - \frac{\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t} \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ \frac{\nabla_{\delta_1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t}\xi_T\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\widetilde{\delta}) + \left(\frac{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t} + \frac{\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t}\right)(\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\widetilde{\delta}) - 1) \end{bmatrix}$$

$$+ \begin{bmatrix} \frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta}) + \sum_{s=1,s\neq k}^{n_\delta}\frac{\nabla_\beta\nabla_{\delta_{1s}}\widetilde{f}_{1t}}{\widetilde{f}_t}\xi_T\nabla_{\delta_{2k}}\hat{\delta}_{1s}(\widetilde{\delta}) \\ \frac{\xi_T\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\hat{\beta}(\widetilde{\delta}) + \sum_{s=1,s\neq k}^{n_\delta}\left(\frac{\xi_T\nabla_{\delta_1}\nabla_{\delta_{1s}}\widetilde{f}_{1t}}{\widetilde{f}_t} + \frac{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_{1s}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t} + \frac{\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1s}}\widetilde{f}_{1t}}{\widetilde{f}_t}\right)\nabla_{\delta_{2k}}\hat{\delta}_{1s}(\widetilde{\delta}) \end{bmatrix}$$

$$+ \begin{bmatrix} \frac{\nabla_\beta\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t}((1-\xi_T) + \xi_T\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\widetilde{\delta})) \\ 0 \end{bmatrix}.$$

If the last two terms on the right hand side are omitted from the expression, it will produce an approximation error that is of order $O_p\left(T^{-1/2}|\rho_T|\right)$. Thus,

$$\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \frac{\nabla_{\delta_{2k}}\widetilde{M}_{1t}}{\widetilde{B}_t} \\ \cdots \\ \frac{\nabla_{\delta_{2k}}\widetilde{M}_{(n_\beta+n_\delta)t}}{\widetilde{B}_t} \end{bmatrix}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\widetilde{B}_t}{\widetilde{B}_t} \tag{A.88}$$

$$= \frac{1}{T}\sum_{t=1}^{T}\left\{\begin{bmatrix} 0 \\ \frac{\nabla_{\delta_1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t}\xi_T\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\widetilde{\delta}) \end{bmatrix} + \begin{bmatrix} 0 \\ \left(\frac{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t} + \frac{\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t}\right)(\nabla_{\delta_{2k}}\hat{\delta}_{1k}(\widetilde{\delta}) - 1) \end{bmatrix}\right\}$$

$$\times\left(\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2} + \frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right.$$

$$\left.\left. -\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2} + \frac{\nabla_{\delta_{1h}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right)\right)\right) + O_p\left(T^{-1/2}\right).$$

Because of Assumption 7, the term in the first brackets can be omitted because its leads to a zero. The products of the remaining terms are $O_p\left(\rho_T^2\right)$ because of Lemma A.10. Therefore, the preceding display is $O_p\left(T^{-1/2} + \rho_T^2\right)$. This shows that elements in the third subset are $O_p\left(T^{-1/2} + \rho_T^2\right)$.

The fourth subset has only one component. Because the elements in the first three subsets are $O_p(T^{-1/2} + \rho_T^2)$, for the expression (A.34) to equal zero, this component must satisfy

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\widetilde{B}_t}{\widetilde{f}_t}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t} = O_p\left(T^{-1/2} + \rho_T^2\right). \tag{A.89}$$

We study $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\widetilde{B}_t$ to determine the leading terms of (A.89).

$$\begin{aligned}
&\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\widetilde{B}_t}{\widetilde{f}_t} \\
= &\sum_{s,u,v=n_\beta+1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\Big\{\xi_T\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta_v}\widetilde{f}_{1t} + (1-\xi_T)\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta_v}\widetilde{f}_{2t} + (\nabla_{\theta_s}\nabla_{\theta_v}\widetilde{f}_{1t} - \nabla_{\theta_s}\nabla_{\theta_v}\widetilde{f}_{2t})\nabla_{\theta_u}\widetilde{\xi}_{t|t-1} \\
&+(\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t} - \nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\theta_v}\widetilde{\xi}_{t|t-1} + (\nabla_{\theta_u}\nabla_{\theta_v}\widetilde{f}_{1t} - \nabla_{\theta_u}\nabla_{\theta_v}\widetilde{f}_{2t})\nabla_{\theta_s}\widetilde{\xi}_{t|t-1} \\
&+(\nabla_{\theta_u}\widetilde{f}_{1t} - \nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_v}\widetilde{\xi}_{t|t-1} + (\nabla_{\theta_v}\widetilde{f}_{1t} - \nabla_{\theta_v}\widetilde{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1} \\
&+(\nabla_{\theta_s}\widetilde{f}_{1t} - \nabla_{\theta_s}\widetilde{f}_{2t})\nabla_{\theta_u}\nabla_{\theta_v}\widetilde{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2h}}\hat{\theta}_v(\widetilde{\delta}) \\
&+\sum_{s,u=1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\Big\{\xi_T\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t} + (1-\xi_T)\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t} \\
&+(\nabla_{\theta_s}\widetilde{f}_{1t} - \nabla_{\theta_s}\widetilde{f}_{2t})\nabla_{\theta_u}\widetilde{\xi}_{t|t-1} + (\nabla_{\theta_u}\widetilde{f}_{1t} - \nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\theta_s}\widetilde{\xi}_{t|t-1}\Big\} \\
&\times\Big\{\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta}) + \nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2h}}\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta}) + \nabla_{\delta_{2h}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})\Big\} \\
&+\sum_{s=1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\{\xi_T\nabla_{\theta_s}\widetilde{f}_{1t} + (1-\xi_T)\nabla_{\theta_s}\widetilde{f}_{2t}\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\theta}_s(\widetilde{\delta}) + R_t \\
= &\; Q_{1t} + Q_{2t} + Q_{3t} + R_t,
\end{aligned}$$

where the three summations are denoted by Q_{1t}, Q_{2t}, and Q_{3t}, respectively. The remainder term R_t arises because the first summation starts at $s,u,v = n_\beta + 1$ instead of $s,u,v = 1$. By Lemma A.12, the effect of R_t on (A.89) is $O_p(T^{-1/2}|\rho_T|)$.

We study Q_{1t}, Q_{2t}, and Q_{3t} sequentially. First, $Q_{1t} = D_{1,klh,t} + R_t$, where $D_{1,klh,t} = \{(\xi_T - 1)^3/\xi_T^2 + (1-\xi_T)\}\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1h}}\widetilde{f}_{1t}/\widetilde{f}_t$ and the effect of R_t on (A.89) is $O_p(|\rho_T(p_T - q_T)|)$ by the expressions in Lemma 1. Next,

$$\begin{aligned}
Q_{2t} \;=\; &\sum_{s,u=1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\{\xi_T\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t} + (1-\xi_T)\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t}\} \\
&\times\Big(\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta}) + \nabla_{\delta_{2h}}\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta}) + \nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\Big) + R_t,
\end{aligned} \tag{A.90}$$

where R_t arises because the components depending on the first order derivative of the density are omitted. The effect of R_t on (A.89) is $O_p(|\rho_T|)$. Further,

$$\begin{aligned}
&\sum_{s,u=1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\{\xi_T\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t}\}\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\\
=\;& -(1-\xi_T)\sum_{u=1}^{n_\beta+2n_\delta}\frac{\nabla_{\delta_{1k}}\nabla_{\theta_u}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})+(1-\xi_T)\sum_{u=1}^{n_\beta+2n_\delta}\frac{\nabla_{\delta_{2k}}\nabla_{\theta_u}\widetilde{f}_{2t}}{\widetilde{f}_t}\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})+R_t\\
=\;& (\xi_T-1)\sum_{u=1}^{n_\delta}\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1u}}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\delta}_{1u}(\widetilde{\delta})+R_t=(\xi_T-1)\sum_{u=1}^{n_\delta}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{ku}\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\delta}_{1u}(\widetilde{\delta})+R_t,
\end{aligned}$$

where the three equalities are due to Lemma A.12, cancellation, and Assumption 7, respectively. The effect of R_t on (A.89) is $O_p(T^{-1/2}|\rho_T|)$ by Lemma A.12. Thus, $Q_{2t}=D_{2,klh,t}+R_t$, where

$$D_{2,klh,t}=(\xi_T-1)\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\sum_{u=1}^{n_\delta}[\alpha_{ku}\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\delta}_{1u}(\widetilde{\delta})+\alpha_{lu}\nabla_{\delta_{2h}}\nabla_{\delta_{2k}}\hat{\delta}_{1u}(\widetilde{\delta})+\alpha_{hu}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_{1u}(\widetilde{\delta})],$$

and the effect of R_t on (A.89) is $O_p\left(T^{-1/2}|\rho_T|\right)$. Finally, Q_{3t} can be rewritten as

$$Q_{3t}=\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\begin{bmatrix}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}.$$

Combining these results for Q_{1t},Q_{2t}, and Q_{3t}, we obtain

$$\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\widetilde{B}_t}{\widetilde{f}_t}=D_{1,klh,t}+D_{2,klh,t}+\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\begin{bmatrix}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}+R_t, \tag{A.91}$$

which implies

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\widetilde{B}_t}{\widetilde{f}_t} \qquad\qquad (A.92)\\
=\;&\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\{D_{1,klh,t}+D_{2,klh,t}\}+\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\begin{bmatrix}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}\\
&+O_p\left(|\rho_T|+|\rho_T(p_T-q_T)|\right).
\end{aligned}$$

Because this holds for any $j\in\{1,...,n_\beta+n_\delta\}$, we can express the results jointly as a vector:

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}\widetilde{M}_{1t}\\ \cdots\\ \widetilde{M}_{(n_\beta+n_\delta)t}\end{bmatrix}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\widetilde{B}_t}{\widetilde{f}_t^2}\\
=\;&\frac{(1-2\xi_T)(\xi_T-1)}{\xi_T^2}D\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t}\\
&+(\xi_T-1)D\widetilde{I}\sum_{u=1}^{n_\delta}[\alpha_{ku}\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\delta}_{1u}(\widetilde{\delta})+\alpha_{lu}\nabla_{\delta_{2h}}\nabla_{\delta_{2k}}\hat{\delta}_{1u}(\widetilde{\delta})+\alpha_{hu}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_{1u}(\widetilde{\delta})]\\
&+D\widetilde{I}\begin{bmatrix}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}+O_p\left(|\rho_T|+|\rho_T(p_T-q_T)|\right).
\end{aligned}$$

Applying (A.89) and then dividing both sides by $\widetilde{I}$, we obtain the expression in the Lemma. ■

Remark 8 *In the above proof, the components of $\mathcal{M}^{(4)}_{jklh}(p_T, q_T, \widetilde{\delta})$ are divided into four subsets. This shortens the proof by limiting the number of cases to consider. The same strategy is applied below to study $\mathcal{M}^{(k)}_{i_1,...,i_k}(p_T, q_T, \widetilde{\delta})$ and $T^{-1}\mathcal{L}^{(k)}_{i_1,...,i_k}(p_T, q_T, \widetilde{\delta}) - ((1-\xi_T)/\xi_T)\,\mathcal{M}^{(k)}_{(n_\beta+i_1)i_2,...,i_k}(p_T, q_T, \widetilde{\delta})$ for $k > 4$. There, we divide their components into k subsets, such that those in the i-th ($i = 1, ..., k$) subset depend on the $(i-1)$-th order derivative of B_t, but not on its lower order derivatives. We then study these k subsets separately.*

Lemma A.15 *Suppose that the null and Assumptions 1-9 hold with (24) satisfied for all $i_1, i_2 \in \{1, ..., n_\delta\}$, and that $\epsilon \le p_T, q_T \le 1-\epsilon$ is satisfied for all T. Then, for any $k, l, h, m \in \{1, ..., n_\delta\}$,*

$$\begin{aligned}\left[\begin{array}{c}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta}) \\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right] &= -\frac{1}{T}\widetilde{I}^{-1}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{k}_{klhm,t}(p_T, q_T) - (C_3 + C_4) \\ &\quad + O_p\left(T^{-1/2} + |\rho_T|\right),\end{aligned}$$

where $C_3 = \frac{\xi_T}{2}\sum_{u,s=1}^{n_\delta}\alpha_{uv}\sum_{(i_1,i_2,i_3,i_4)\in S}\nabla_{\delta_{2i_3}}\nabla_{\delta_{2i_4}}\hat{\delta}_{1u}(\delta_2)\nabla_{\delta_{2i_1}}\nabla_{\delta_{2i_2}}\hat{\delta}_{1v}(\delta_2)$ *and* $C_4 = (\xi_T - 1)\sum_{u=1}^{n_\delta}$ $\{\alpha_{hu}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\delta}_{1u}(\widetilde{\delta}) + \alpha_{ku}\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\delta}_{1u}(\widetilde{\delta}) + \alpha_{lu}\nabla_{\delta_{2k}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_{1u}(\widetilde{\delta}) + \alpha_{mu}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\delta}_{1u}(\widetilde{\delta})\}$.

Proof of Lemma A.15. We compute the derivative of (A.34) with respect to the m-th element of δ_2. The components in the summation can be grouped into 5 subsets as explained in Remark 8.

The first subset comprises just one component: $T^{-1}\sum_{t=1}^{T}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\widetilde{M}_{jt}/\widetilde{B}_t$, which is $O_p(T^{-1/2})$ by Lemma A.11. The second subset involves the first order derivative of B_t. They are $O_p\left(T^{-1/2}|\rho_T|\right)$ by the same argument as in (A.81).

A representative element in the third subset is

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\widetilde{M}_{jt}}{\widetilde{B}_t}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\widetilde{B}_t}{\widetilde{B}_t} \qquad\qquad (A.93)\\ =\;& \frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\widetilde{M}_{jt}}{\widetilde{B}_t}\left\{\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2} + \frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right.\\ &\left. - \frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2} + \frac{\nabla_{\delta_{1h}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right)\right\} + O_p(T^{-1/2}),\end{aligned}$$

where the equality follows from (A.86). If $j \in \{1, ..., n_\beta\}$, then by (A.29) this is equal to

$$\begin{aligned}&-\frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\nabla_{\beta_j}\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2} + \frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}\nabla_{\delta_{1m}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right)\right\}\\ &\times\left\{\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2} + \frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right.\\ &\left. - \frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2} + \frac{\nabla_{\delta_{1h}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right)\right\} + O_p(T^{-1/2} + \rho_T^2)\\ =\;& O_p(T^{-1/2} + \rho_T^2),\end{aligned}$$

where the equality follows from Lemma A.13 and Assumption 7. If $j \in \{n_\beta + 1, ..., n_\beta + n_\delta\}$, then this is equal to

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left\{\frac{(1-\xi_T)^2}{\xi_T}\frac{\nabla_{\theta_j}\nabla_{\delta_{1k}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t} + \xi_T\frac{\nabla_{\theta_j}\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\begin{bmatrix}\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}\right.\\
&\left. + \left(\xi_T - \xi_T^2\right)\frac{\nabla_{\theta_j}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\right\}\left\{\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2} + \frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right.\\
&\left. -\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2} + \frac{\nabla_{\delta_{1h}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right)\right\} + O_p(T^{-1/2}+\rho_T^2)\\
= \; & O_p(T^{-1/2}+\rho_T^2),
\end{aligned}$$

where the equality follows from Lemma A.13 and Assumptions 7 and 9. Therefore, the elements in the third subset are $O_p(T^{-1/2}+\rho_T^2)$.

A representative element in the fourth subset is

$$T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\widetilde{M}_{jt}}{\widetilde{f}_t}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\widetilde{B}_t}{\widetilde{f}_t}. \tag{A.94}$$

If $j \in \{1, ..., n_\beta\}$, then by the expression of $\nabla_{\delta_{2k}}\widetilde{M}_{jt}$ in (A.87) this is $O_p\left(T^{-1/2}|\rho_T|\right)$. If $j \in \{n_\beta+1, ..., n_\beta+n_\delta\}$, $\nabla_{\delta_{2k}}\widetilde{M}_{jt}$ can be expressed as a linear function of $(\widetilde{M}_{it}/\widetilde{f}_t)$ $(i = 1, ..., n_\beta + n_\delta)$ and $(\nabla_{\theta_s}\widetilde{f}_{1t}/\widetilde{f}_t)\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}$ $(s, u = n_\beta + 1, ..., n_\beta + 2n_\delta)$. Thus, because of (A.89), this component is $O_p\left(T^{-1/2}+\rho_T^2\right)$ if

$$T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\widetilde{B}_t}{\widetilde{f}_t}\frac{\nabla_{\theta_s}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1} = O_p\left(T^{-1/2}+\rho_T^2\right). \tag{A.95}$$

for $s, u = n_\beta + 1, ..., n_\beta + 2n_\delta$. To verify (A.95), note that by (A.91) and Lemma A.14, we have

$$\begin{aligned}
\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\widetilde{B}_t}{\widetilde{f}_t} = \; & \frac{(1-2\xi_T)(\xi_T-1)}{\xi_T^2}\left\{\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\right.\\
& \left. -\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\right\} + R_t,
\end{aligned}$$

where the effect of R_t on (A.95) is $O_p\left(T^{-1/2}+\rho_T^2\right)$. From this expression and Assumption 9, (A.95) follows. Therefore, the components in the fourth subsets are $O_p\left(\rho_T^2+T^{-1/2}\right)$.

The fifth subset comprises just one component. Note that because the components in the first four subsets are all $O_p\left(\rho_T^2+T^{-1/2}\right)$, this component must satisfy

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\widetilde{B}_t}{\widetilde{f}_t}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t} = O_p\left(\rho_T^2+T^{-1/2}\right). \tag{A.96}$$

We now examine $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\widetilde{B}_t$ to further study the left hand side of (A.96). We have

$$
\begin{aligned}
&\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\widetilde{B}_t}{\widetilde{f}_t} \qquad\qquad \text{(A.97)}\\
=\;& \sum_{s,u,v,w=1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\{\xi_T\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta_v}\nabla_{\theta_w}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta_v}\nabla_{\theta_w}\widetilde{f}_{2t}\}\\
&\quad\times\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2h}}\hat{\theta}_v(\widetilde{\delta})\nabla_{\delta_{2m}}\hat{\theta}_w(\widetilde{\delta})\\
&+\sum_{s,u,v=1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\{\xi_T\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta_v}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_s}\nabla_{\theta_u}\nabla_{\theta_v}\widetilde{f}_{2t}\}\\
&\times\Big\{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2h}}\hat{\theta}_v(\widetilde{\delta})\nabla_{\delta_{2m}}\hat{\theta}_u(\widetilde{\delta})+\nabla_{\delta_{2k}}\nabla_{\delta_{2h}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2m}}\hat{\theta}_v(\widetilde{\delta})\\
&+\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2h}}\hat{\theta}_v(\widetilde{\delta})+\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2m}}\hat{\theta}_v(\widetilde{\delta})\\
&+\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2h}}\hat{\theta}_v(\widetilde{\delta})+\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\theta}_v(\widetilde{\delta})\Big\}\\
&+\sum_{s,v=1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\{\xi_T\nabla_{\theta_s}\nabla_{\theta_v}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_s}\nabla_{\theta_v}\widetilde{f}_{2t}\}\\
&\times\Big\{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\theta}_v(\widetilde{\delta})+\nabla_{\delta_{2k}}\nabla_{\delta_{2h}}\hat{\theta}_s(\delta_2)\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_v(\widetilde{\delta})\\
&\quad+\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\theta}_v(\widetilde{\delta})\Big\}\\
&+\sum_{s,u=1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\{\xi_T\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t}\}\\
&\times\Big\{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2h}}\hat{\theta}_u(\widetilde{\delta})+\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2m}}\hat{\theta}_u(\widetilde{\delta})\\
&\quad+\nabla_{\delta_{2h}}\nabla_{\delta_{2k}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})+\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2k}}\hat{\theta}_u(\widetilde{\delta})\Big\}\\
&+\sum_{v=1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\{\xi_T\nabla_{\theta_v}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_v}\widetilde{f}_{2t}\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\theta}_v(\widetilde{\delta})\;+R_t\\
=\;& S_{1t}+S_{2t}+S_{3t}+S_{4t}+S_{5t}+R_t,
\end{aligned}
$$

where S_{1t},...,S_{5t} denote the five summations on the right hand side, and R_t represents the omitted terms, which depend on the derivatives of $\xi_{t|t-1}$ whose effect on (A.96) is $O_p\left(|\rho_T|\right)$.

We study S_{1t},...,S_{5t} sequentially. For S_{1t},

$$
\begin{aligned}
&\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}S_{1t}\\
&=\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\left\{\frac{(1-\xi_T)^4\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1h}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^3}+\frac{(1-\xi_T)\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\widetilde{f}_{2t}}{\widetilde{f}_t}\right\}+O_p\left(T^{-1/2}\left|\rho_T\right|\right)\\
&\equiv\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}E_{1t}+O_p(T^{-1/2}\left|\rho_T\right|),
\end{aligned}
$$

where $E_{1t}=(1-\xi_T)\left((1-\xi_T)^3/\xi_T^3+1\right)\nabla_{\delta_{1k}}\nabla_{\delta_{1l}}\nabla_{\delta_{1h}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}/\widetilde{f}_t$ and the $O_p(T^{-1/2}\left|\rho_T\right|)$ is due to Lemma A.12. Among the six components of S_{2t}, it suffices to study the first. By Lemma A.12, it leads to

$$\begin{aligned}
&\xi_T\left(\frac{\xi_T-1}{\xi_T}\right)^2\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\sum_{s=1}^{n_\beta+n_\delta}\frac{\nabla_{\theta_s}\nabla_{\delta_{1h}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})\\
&+(1-\xi_T)\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\sum_{s=1}^{n_\beta}\frac{\nabla_{\theta_s}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\widetilde{f}_{2t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})+O_p\left(T^{-1/2}\left|\rho_T\right|\right)\\
=\;&-\left(\frac{1-\xi_T}{\xi_T}\right)^2\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\frac{\nabla_{\beta'}\nabla_{\delta_{1h}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{kl}^{(1)}-\left(\frac{1-\xi_T}{\xi_T}\right)^3\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\frac{\nabla_{\delta_1'}\nabla_{\delta_{1h}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{kl}^{(2)}\\
&+O_p\left(T^{-1/2}+|\rho_T|\right),
\end{aligned}$$

where the equality is due to Lemma A.13. Therefore,

$$\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}S_{2t}=\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}E_{2t}+O_p\left(T^{-1/2}+|\rho_T|\right),$$

where $E_{2t}=-((1-\xi_T)^2/\xi_T^2)\sum_{(i_1,i_2,i_3,i_4)\in S}\left(\nabla_{\beta'}\nabla_{\delta_{1i_1}}\nabla_{\delta_{1i_2}}\widetilde{f}_{1t}\alpha_{i_3i_4}^{(1)}+\left(\frac{1-\xi_T}{\xi_T}\right)\nabla_{\delta_1'}\nabla_{\delta_{1i_1}}\nabla_{\delta_{1i_2}}\widetilde{f}_{1t}\alpha_{i_3i_4}^{(2)}\right)/\widetilde{f}_t$.

Among the three components of S_{3t}, it suffices to study the first:

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\sum_{s=1}^{n_\beta+n_\delta}\left(\xi_T\frac{\nabla_{\theta_s}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}+(1-\xi_T)\frac{\nabla_{\theta_s}\nabla_{\beta'}\widetilde{f}_{2t}}{\widetilde{f}_t}\right)\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})\\
&+\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\sum_{s=1}^{n_\beta+n_\delta}\xi_T\nabla_{\theta_s}\nabla_{\delta_1'}\widetilde{f}_{1t}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})\\
=\;&\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\delta_2)'\left(\xi_T\frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}+(1-\xi_T)\frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{2t}}{\widetilde{f}_t}\right)\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})\\
&+\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\delta_2)'\frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})\\
&+\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\delta_2)'\frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\xi_T\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\\
&+\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\delta_2)'\frac{\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\xi_T\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\\
=\;&\left(\frac{\xi_T-1}{\xi_T}\right)^2\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\left(\alpha_{kl}^{(1)\prime}\frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{hm}^{(1)}+\alpha_{kl}^{(2)\prime}\frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{hm}^{(1)}+\alpha_{kl}^{(1)\prime}\frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{hm}^{(2)}\right)\\
&+\frac{1}{T}\sum_{t=1}^T\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\delta_2)'\frac{\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\xi_T\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_1(\delta_2)+O_p\left(T^{-1/2}+|\rho_T|\right),
\end{aligned}$$

where the first equality is simply an equivalent representation in vector notation and the second is because of Lemma A.13. The last two lines of the preceding display can be written as

$$\left(\frac{\xi_T-1}{\xi_T}\right)^2\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\alpha_{kl}'\frac{\nabla_{(\beta',\delta_1')'}\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{hm}-\left(\frac{\xi_T-1}{\xi_T}\right)^2\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\alpha_{kl}^{(2)\prime}\frac{\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{hm}^{(2)}$$
$$+\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\delta_2)'\frac{\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\xi_T\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_1(\delta_2)+O_p\left(T^{-1/2}+|\rho_T|\right). \tag{A.98}$$

Therefore,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}S_{3t}\\
=&\left(\frac{\xi_T-1}{\xi_T}\right)^2\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\sum_{(i_1,i_2,i_3,i_4)\in S}\left(\frac{1}{2}\alpha_{i_1i_2}^{(1)\prime}\frac{\nabla_{\beta}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{i_3i_4}^{(1)}+\alpha_{i_1i_2}^{(2)\prime}\frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{i_3i_4}^{(1)}\right)\\
&+\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\sum_{u,s=1}^{n_\delta}\frac{\xi_T\nabla_{\delta_{1u}}\nabla_{\delta_{1v}}\widetilde{f}_{1t}}{2\widetilde{f}_t}\sum_{(i_1,i_2,i_3,i_4)\in S}\nabla_{\delta_{2i_3}}\nabla_{\delta_{2i_4}}\hat{\delta}_{1u}(\widetilde{\delta})\nabla_{\delta_{2i_1}}\nabla_{\delta_{2i_2}}\hat{\delta}_{1v}(\widetilde{\delta})+O_p\left(T^{-1/2}+|\rho_T|\right)\\
=&\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}E_{3t}+\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}C_3+O_p\left(T^{-1/2}+|\rho_T|\right),
\end{aligned}$$

where

$$\begin{aligned}
E_{3t}&=\left(\tfrac{\xi_T-1}{\xi_T}\right)^2\textstyle\sum_{(i_1,i_2,i_3,i_4)\in S}\left(\tfrac{1}{2}\alpha_{i_1i_2}^{(1)\prime}\tfrac{\nabla_{\beta}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{i_3i_4}^{(1)}+\alpha_{i_1i_2}^{(2)\prime}\tfrac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{i_3i_4}^{(1)}\right),\\
C_3&=\tfrac{\xi_T}{2}\left(\textstyle\sum_{u,s=1}^{n_\delta}\alpha_{uv}\sum_{(i_1,i_2,i_3,i_4)\in S}\nabla_{\delta_{2i_3}}\nabla_{\delta_{2i_4}}\hat{\delta}_{1u}(\delta_2)\nabla_{\delta_{2i_1}}\nabla_{\delta_{2i_2}}\hat{\delta}_{1v}(\delta_2)\right).
\end{aligned}$$

The four components of S_{4t} have the same structure, it suffices to study the first. This leads to

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\sum_{s,v=1}^{n_\beta+2n_\delta}\xi_T\frac{\nabla_{\theta_s}\nabla_{\theta_v}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2h}}\hat{\theta}_v(\widetilde{\delta})\\
&+\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\sum_{s,v=1}^{n_\beta+2n_\delta}(1-\xi_T)\frac{\nabla_{\theta_s}\nabla_{\theta_v}\widetilde{f}_{2t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2h}}\hat{\theta}_v(\widetilde{\delta})\\
=&(\xi_T-1)\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\sum_{s=1}^{n_\beta+2n_\delta}\frac{\nabla_{\theta_s}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})\\
&+(1-\xi_T)\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\sum_{s=1}^{n_\beta+2n_\delta}\frac{\nabla_{\theta_s}\nabla_{\delta_{2h}}\widetilde{f}_{2t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})+O_p\left(T^{-1/2}|\rho_T|\right)\\
=&(\xi_T-1)\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\sum_{s=1+n_\beta}^{n_\beta+n_\delta}\frac{\nabla_{\theta_s}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\theta}_s(\widetilde{\delta})+O_p\left(T^{-1/2}|\rho_T|\right)\\
=&\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\sum_{u=1}^{n_\delta}(\xi_T-1)\,\alpha_{hu}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\delta}_{1u}(\widetilde{\delta})\right)+O_p\left(T^{-1/2}|\rho_T|\right),
\end{aligned}$$

where the first equality is because of Lemma A.13, the second is due to cancellation, and the third is because of Assumption 7. Therefore,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M_{jt}}}{\widetilde{f}_t}S_{4t}=\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M_{jt}}}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}C_4+O_p\left(T^{-1/2}\left|\rho_T\right|\right),$$

where

$$\begin{aligned}C_4 &= (\xi_T-1)\sum_{u=1}^{n_\delta}\Big\{\alpha_{hu}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\delta}_{1u}(\widetilde{\delta})+\alpha_{ku}\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\delta}_{1u}(\widetilde{\delta}) \\ &\quad +\alpha_{lu}\nabla_{\delta_{2k}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_{1u}(\widetilde{\delta})+\alpha_{mu}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\hat{\delta}_{1u}(\widetilde{\delta})\Big\}.\end{aligned}$$

Finally, S_{5t} can be rewritten as

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M_{jt}}}{\widetilde{f}_t}S_{5t}=\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M_{jt}}}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left[\begin{array}{c}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right].$$

The above results for S_{1t} to S_{5t} imply

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M_{jt}}}{\widetilde{f}_t}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\widetilde{B}_t}{\widetilde{f}_t} \qquad \text{(A.99)}\\ =& \frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M_{jt}}}{\widetilde{f}_t}\left\{E_{1t}+E_{2t}+E_{3t}\right\}+\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M_{jt}}}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(C_3+C_4\right)\\ &+\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M_{jt}}}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left[\begin{array}{c}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]+O_p\left(T^{-1/2}+\left|\rho_T\right|\right).\end{aligned}$$

Combining this expression with (A.96), we have, for any $j\in\{1,...,n_\beta+n_\delta\}$,

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M_{jt}}}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left[\begin{array}{c}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]\\ =& -\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M_{jt}}}{\widetilde{f}_t}\left(E_{1t}+E_{2t}+E_{3t}\right)-\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M_{jt}}}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(C_3+C_4\right)+O_p\left(T^{-1/2}+\left|\rho_T\right|\right),\end{aligned}$$

The resulting $(n_\beta+n_\delta)$ equations can be expressed jointly as

$$\begin{aligned}&\widetilde{I}\left[\begin{array}{c}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]\\ =& -\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(E_{1t}+E_{2t}+E_{3t}\right)-\widetilde{I}\left(C_3+C_4\right)+O_p\left(T^{-1/2}+\left|\rho_T\right|\right),\end{aligned}$$

which implies

$$\begin{aligned}&\left[\begin{array}{c}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2h}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]\\ =& -\frac{1}{T}\widetilde{I}^{-1}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(E_{1t}+E_{2t}+E_{3t}\right)-\left(C_3+C_4\right)+O_p\left(T^{-1/2}+\left|\rho_T\right|\right).\end{aligned}$$

Because $\widetilde{k}_{klhm,t}\left(p_T,q_T\right)=E_{1t}+E_{2t}+E_{3t}$, the Lemma follows. ■

Lemma A.16 *Suppose that the null hypothesis and Assumptions 1-9 hold with (24) satisfied for all* $i_1,i_2\in\{1,...,n_\delta\}$*, and that* $\epsilon\leq p_T,q_T\leq 1-\epsilon$ *is satisfied for all* T*. Then, for any* $i_1,i_2,...,i_k\in\{1,...,n_\delta\}$ *and* $k\in\{5,6,7\}$*, we have* $\nabla_{\delta_{2i_1}}\cdots\nabla_{\delta_{2i_k}}\hat{\beta}(\widetilde{\delta})=O_p\left(1\right)$ *and* $\nabla_{\delta_{2i_1}}\cdots\nabla_{\delta_{2i_k}}\hat{\delta}_1(\widetilde{\delta})=O_p\left(1\right)$.

Proof of Lemma A.16. The proof is simpler than for lower order derivatives because the expressions of the leading terms are not needed.

For $k=5$, we compute second order derivative of (A.34) with respect to δ_2, and evaluate the resulting expression at $\delta_2=\widetilde{\delta}$. Because of Lemma A.11, the summands are all $O_p\left(1\right)$. Thus, $T^{-1}\sum_{t=1}^{T}\widetilde{M}_{jt}\nabla_{\delta_{2i_1}}\nabla_{\delta_{2i_2}}...\nabla_{\delta_{2i_5}}\widetilde{B}_t/\widetilde{f}_t^2=O_p\left(1\right)$ for all $j\in\{1,...,n_\beta+n_\delta\}$, which implies

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\widetilde{M}_{jt}}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\begin{bmatrix}\nabla_{\delta_{2i_1}}\cdots\nabla_{\delta_{2i_5}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2i_1}}\cdots\nabla_{\delta_{2i_5}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}=O_p\left(1\right)$$

for all $j\in\{1,...,n_\beta+n_\delta\}$. These $n_\beta+n_\delta$ cases can be expressed jointly as

$$\widetilde{I}\begin{bmatrix}\nabla_{\delta_{2i_1}}\cdots\nabla_{\delta_{2i_5}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2i_1}}\cdots\nabla_{\delta_{2i_5}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}=O_p\left(1\right).$$

Dividing both sides by $\widetilde{I}$, we obtain $\nabla_{\delta_{2i_1}}\cdots\nabla_{\delta_{2i_5}}\hat{\beta}(\widetilde{\delta})=O_p\left(1\right)$ and $\nabla_{\delta_{2i_1}}\cdots\nabla_{\delta_{2i_5}}\hat{\delta}_1(\widetilde{\delta})=O_p\left(1\right)$.

The results for $k=6,7$ can be proved in the same way, i.e., by computing the third and fourth order derivatives of (A.34), and applying the same argument as above. We omit the details. ■

Let $\widetilde{U}_{jk,t}(p_T,q_T)$ equal (17), with ρ and ξ_* replaced with ρ_T and ξ_T, respectively. The next lemma presents asymptotic approximations to $\mathcal{L}^{(k)}_{i_1,...,i_k}(p_T,q_T,\widetilde{\delta})$ for $k=2,...,8$. Note that their remainder terms are formulated in terms of ρ_T, p_T-q_T, and T. This important feature enables the results to be used as a basis for studying $LR(p_T,q_T)$ under various drifting sequences of (p_T,q_T).

Lemma A.17 *Suppose that the null hypothesis and Assumptions 1-9 hold, with (24) satisfied for all* i_1*,*$i_2\in\{1,...,n_\delta\}$*, and* $\epsilon\leq p_T,q_T\leq 1-\epsilon$ *for any* T*. Then, for any* $j,k,l,m,n,r,s,u\in\{1,...,n_\delta\}$*,*

$$T^{-1/2}\mathcal{L}^{(2)}_{jk}(p_T,q_T,\widetilde{\delta})=T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_{jk,t}(p_T,q_T)+O_p(T^{-1/2}\rho_T^2),$$

$$T^{-1/2}\mathcal{L}^{(3)}_{jkl}(p_T,q_T,\widetilde{\delta})=T^{-1/2}\sum_{t=1}^{T}\widetilde{s}_{jkl,t}(p_T,q_T)+O_p(|\rho_T|),$$

$$T^{-1/2}\mathcal{L}^{(4)}_{jklm}(p_T,q_T,\widetilde{\delta})=-T^{1/2}\widetilde{\omega}^{(2)}_{jklm}(p_T,q_T)+T^{-1/2}\sum_{t=1}^{T}\widetilde{k}_{jklm,t}(p_T,q_T)+O_p(T^{-1/2}+|\rho_T|),$$

$$T^{-1}\mathcal{L}^{(5)}_{jklmn}(p_T,q_T,\widetilde{\delta})=O_p(T^{-1/2}+\rho_T^2),$$

$$T^{-1}\mathcal{L}^{(6)}_{jklmnr}(p_T,q_T,\widetilde{\delta})=-\sum_{(i_1,...,i_6)\in IND_1}\omega^{(3)}_{i_1i_2...i_6}(p_T,q_T)+O_p(T^{-1/2}+\rho_T^2+(p_T-q_T)^2|\rho_T|),$$

$$T^{-1}\mathcal{L}^{(7)}_{jklmnrs}(p_T,q_T,\widetilde{\delta})=O_p(T^{-1/2}+|\rho_T|+|p_T-q_T|),$$

$$T^{-1}\mathcal{L}^{(8)}_{jklmnrsu}(p_T,q_T,\widetilde{\delta})=-\sum_{(i_1,...,i_8)\in IND_2}\omega^{(4)}_{i_1i_2...i_8}\left(p_T,q_T\right)+O_p(|\rho_T|+|p_T-q_T|)+o_p\left(1\right),$$

where IND_1 and IND_2 are defined in Lemma 3. In addition, $T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_{jk,t}(p_T,q_T)=O_p(|\rho_T|)$, $T^{-1/2}\sum_{t=1}^{T}\widetilde{s}_{jkl,t}(p_T,q_T)=O_p\left(|p_T-q_T|\right)$, $\widetilde{\omega}_{jklm}^{(2)}(p_T,q_T)=\widetilde{V}_{jklm}-\widetilde{D}_{jk}'\widetilde{I}^{-1}\widetilde{D}_{lm}+\widetilde{V}_{jmkl}-\widetilde{D}_{jm}'\widetilde{I}^{-1}\widetilde{D}_{kl}+\widetilde{V}_{jlkm}-\widetilde{D}_{jl}'\widetilde{I}^{-1}\widetilde{D}_{km}=O_p\left(\rho_T^2\right)$, *and* $T^{-1/2}\sum_{t=1}^{T}\widetilde{k}_{jklm,t}(p_T,q_T)=O_p\left(1\right)$.

Lemma A.17 acts as a bridge between Lemmas 2 and 3, where the transition probabilities belong to two disjoint subsets. Because the second order derivatives $T^{-1/2}\mathcal{L}_{jk}^{(2)}(p_T,q_T,\widetilde{\delta})$ are $O_p(|\rho_T|)$, they stop being the the leading terms of (35) when $|\rho_T|$ converges sufficiently fast to zero. Because the third order derivatives $T^{-1/2}\mathcal{L}_{jkl}^{(3)}(p_T,q_T,\widetilde{\delta})$ are $O_p\left(|\rho_T|+|p_T-q_T|\right)$, whether they can serve as the leading terms of (35) depends on $|p_T-q_T|$, i.e., the closeness of (p_T,q_T) to $(0.5,0.5)$. Finally, when $|\rho_T|$ and $|p_T-q_T|$ both converge to zero at fast rates, $T^{-1/2}\sum_{t=1}^{T}\widetilde{k}_{jklm,t}(p_T,q_T)$ emerge as the leading terms of the expansion.

Proof of Lemma A.17. We prove the seven displayed results in the lemma separately.

Part 1: Proof of the first result of Lemma A.17. This result follows from Lemmas A.11 and A.12 and the proof of Lemma 2.1. By (A.20) and (A.37),

$$\begin{aligned}&\nabla_{\delta_{2k}}\widetilde{L}_{jt}-\left(\frac{1-\xi_T}{\xi_T}\right)\nabla_{\delta_{2k}}\widetilde{M}_{(n_\beta+j)t}\\ =\;&(1-\xi_T)\left(\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t}-\nabla_{\delta_{1j}}\nabla_{\delta_1'}\widetilde{f}_{1t}\nabla_{\delta_{2k}}\hat{\delta}_1(\widetilde{\delta})\right)\\ &-\frac{1}{\xi_T}\left\{(\nabla_{\theta'}\widetilde{f}_{1t}-\nabla_{\theta'}\widetilde{f}_{2t})\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}+\nabla_{\delta_{2j}}\widetilde{f}_{2t}\nabla_{\theta'}\widetilde{\xi}_{t|t-1}\right\}\nabla_{\delta_{2k}}\hat{\theta}(\widetilde{\delta}).\end{aligned}\tag{A.100}$$

Together with (A.38) and Assumption 7, this implies

$$\begin{aligned}&T^{-1/2}\mathcal{L}_{jk}^{(2)}(p_T,q_T,\widetilde{\delta})\\ =\;&-\frac{1}{\xi_T}\left\{T^{-1/2}\sum_{t=1}^{T}\frac{(\nabla_{\theta'}\widetilde{f}_{1t}-\nabla_{\theta'}\widetilde{f}_{2t})\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}+\nabla_{\delta_{2j}}\widetilde{f}_{2t}\nabla_{\theta'}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t}\right\}\nabla_{\delta_{2k}}\hat{\theta}(\widetilde{\delta}).\end{aligned}\tag{A.101}$$

The expression inside the braces is $O_p(|\rho_T|)$ by Lemma A.11. In addition, $\nabla_{\delta_{2k}}\hat{\theta}_j(\widetilde{\delta})=O_p(T^{-1/2}|\rho_T|)$ except when $j\in\{n_\beta+k,n_\beta+n_\delta+k\}$ by Lemma A.12. Therefore,

$$\begin{aligned}T^{-1/2}\mathcal{L}_{jk}^{(2)}(p_T,q_T,\widetilde{\delta})\;=\;&\left(\frac{1-\xi_T}{\xi_T^2}\right)T^{-1/2}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\left\{\nabla_{\delta_{1k}}\widetilde{f}_{1t}\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}+\nabla_{\delta_{2j}}\widetilde{f}_{2t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}\right\}\\ &+\frac{1}{\xi_T}T^{-1/2}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\left\{\nabla_{\delta_{2k}}\widetilde{f}_{2t}\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}-\nabla_{\delta_{2j}}\widetilde{f}_{2t}\nabla_{\delta_{2k}}\widetilde{\xi}_{t|t-1}\right\}+O_p(T^{-1/2}\rho_T^2).\end{aligned}$$

The result follows because $\nabla_{\delta_{1k}}\widetilde{f}_{1t}=\nabla_{\delta_{2k}}\widetilde{f}_{2t}$ and $\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}=-\nabla_{\delta_{2k}}\widetilde{\xi}_{t|t-1}$.

Part 2: Proof of the second result of Lemma A.17. We consider Display (A.39) with the scaling factor $T^{-3/4}$ replaced by $T^{-1/2}$. The last two terms on the right hand side equal zero when $\delta_2=\widetilde{\delta}$. The second and third terms are are $O_p(|\rho_T|)$ by (A.81). The first term has a more complex structure. We now study it in detail.

By (A.29) and (A.40), $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\widetilde{M}_{jt}$ and $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\widetilde{L}_{jt}$ can be expressed as

$$\begin{aligned}
&\sum_{s,u=1}^{n_\beta+2n_\delta}\Big\{\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\widetilde{f}_{1t}+\xi_T\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}\nabla_{\theta_j}\nabla_{\theta_s}\widetilde{f}_{2t} \\
&+(1-\xi_T)\nabla_{\theta_j}\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t}+(\nabla_{\theta_j}\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_j}\nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\theta_s}\widetilde{\xi}_{t|t-1} \\
&+(\nabla_{\theta_j}\widetilde{f}_{1t}-\nabla_{\theta_j}\widetilde{f}_{2t})\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}+(\nabla_{\theta_s}\widetilde{f}_{1t}-\nabla_{\theta_s}\widetilde{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1} \\
&+(\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\theta_j}\widetilde{\xi}_{t|t-1}+(\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\theta_j}\nabla_{\theta_s}\widetilde{\xi}_{t|t-1}\Big\} \\
&\times\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta}) \\
&+\Big\{\xi_T\nabla_{\theta_j}\nabla_{\theta'}\widetilde{f}_{1t}+(1-\xi_T)\nabla_{\theta_j}\nabla_{\theta'}\widetilde{f}_{2t}+(\nabla_{\theta_j}\widetilde{f}_{1t}-\nabla_{\theta_j}\widetilde{f}_{2t})\nabla_{\theta'}\widetilde{\xi}_{t|t-1} \\
&+(\nabla_{\theta'}\widetilde{f}_{1t}-\nabla_{\theta'}\widetilde{f}_{2t})\nabla_{\theta_j}\widetilde{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\widetilde{\delta})
\end{aligned}\tag{A.102}$$

and

$$\begin{aligned}
&\sum_{s,u=1}^{n_\beta+2n_\delta}\Big\{(1-\xi_T)\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t}-\nabla_{\delta_{2j}}\nabla_{\theta_s}\widetilde{f}_{2t}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}-\nabla_{\delta_{2j}}\nabla_{\theta_u}\widetilde{f}_{2t}\nabla_{\theta_s}\widetilde{\xi}_{t|t-1} \\
&-\nabla_{\delta_{2j}}\widetilde{f}_{2t}\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}+(\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\delta_{2j}}\widetilde{\xi}_{t|t-1} \\
&+(\nabla_{\theta_s}\widetilde{f}_{1t}-\nabla_{\theta_s}\widetilde{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}+(\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\widetilde{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta}) \\
&+\Big\{(1-\xi_T)\nabla_{\delta_{2j}}\nabla_{\theta'}\widetilde{f}_{2t}-\nabla_{\delta_{2j}}\widetilde{f}_{2t}\nabla_{\theta'}\widetilde{\xi}_{t|t-1}+(\nabla_{\theta'}\widetilde{f}_{1t}-\nabla_{\theta'}\widetilde{f}_{2t})\nabla_{\delta_{2j}}\widetilde{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\widetilde{\delta}).
\end{aligned}\tag{A.103}$$

Combining these two expressions, we obtain

$$T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\widetilde{L}_{jt}}{\widetilde{B}_t}-\left(\frac{1-\xi_T}{\xi_T}\right)T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\widetilde{M}_{(n_\beta+j)t}}{\widetilde{B}_t} \tag{A.104}$$

$$\begin{aligned}
= \; & T^{-1/2}\sum_{s,u=n_\beta+1}^{n_\beta+2n_\delta}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\Big\{\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t}(1-\xi_T)-(1-\xi_T)\nabla_{\delta_{1j}}\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t} && \text{(M0)}\\
& -\nabla_{\delta_{2j}}\nabla_{\theta_s}\widetilde{f}_{2t}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}-\left(\frac{1-\xi_T}{\xi_T}\right)\nabla_{\delta_{1j}}\nabla_{\theta_s}\widetilde{f}_{1t}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1} && \text{(M1)}\\
& -\nabla_{\delta_{2j}}\nabla_{\theta_u}\widetilde{f}_{2t}\nabla_{\theta_s}\widetilde{\xi}_{t|t-1}-\left(\frac{1-\xi_T}{\xi_T}\right)\nabla_{\delta_{1j}}\nabla_{\theta_u}\widetilde{f}_{1t}\nabla_{\theta_s}\widetilde{\xi}_{t|t-1} && \text{(M2)}\\
& -\nabla_{\delta_{2j}}\widetilde{f}_{2t}\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}-\left(\frac{1-\xi_T}{\xi_T}\right)\nabla_{\delta_{1j}}\widetilde{f}_{1t}\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1} && \text{(M3)}\\
& +(\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\delta_{2j}}\widetilde{\xi}_{t|t-1}-\left(\frac{1-\xi_T}{\xi_T}\right)(\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1} && \text{(M4)}\\
& +(\nabla_{\theta_s}\widetilde{f}_{1t}-\nabla_{\theta_s}\widetilde{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}-\left(\frac{1-\xi_T}{\xi_T}\right)(\nabla_{\theta_s}\widetilde{f}_{1t}-\nabla_{\theta_s}\widetilde{f}_{2t})\nabla_{\delta_{1j}}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1} && \text{(M5)}\\
& +(\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\delta_{2j}}\nabla_{\theta_s}\widetilde{\xi}_{t|t-1}-\left(\frac{1-\xi_T}{\xi_T}\right)(\nabla_{\theta_u}\widetilde{f}_{1t}-\nabla_{\theta_u}\widetilde{f}_{2t})\nabla_{\delta_{1j}}\nabla_{\theta_s}\widetilde{\xi}_{t|t-1}\Big\} && \text{(M6)}\\
& \times\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2l}}\hat{\theta}_u(\widetilde{\delta})\\
& +(1-\xi_T)T^{-1/2}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\Big\{\nabla_{\delta_{2j}}\nabla_{\theta'}\widetilde{f}_{2t}-\nabla_{\delta_{1j}}\nabla_{\theta'}\widetilde{f}_{1t}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\widetilde{\delta})\\
& -\frac{1}{\xi_T}T^{-1/2}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\Big\{\nabla_{\delta_{2j}}\widetilde{f}_{2t}\nabla_{\theta'}\widetilde{\xi}_{t|t-1}-(\nabla_{\theta'}\widetilde{f}_{1t}-\nabla_{\theta'}\widetilde{f}_{2t})\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}\Big\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}(\widetilde{\delta})+O_p\left(T^{-1/2}\left|\rho_T\right|\right),
\end{aligned}$$

where the $O_p\left(T^{-1/2}\left|\rho_T\right|\right)$ term arises because the first summation begins at $n_\beta+1$ instead of 1.

We now study the effect of (M1)-(M6) on the first summation, followed by that of (M0). The expressions in Lemma 1.3 play an important role for this analysis. First,

$$\begin{aligned}
\text{(M1)} &= \left((1-2\xi_T)/\xi_T^2\right)T^{-1/2}\sum_{t=1}^{T}\sum_{u=n_\beta+1}^{n_\beta+2n_\delta}\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}/\widetilde{f}_t+O_p\left(T^{-1/2}\rho_T^2\right)\\
&= O_p\left(\left|\rho_T\left(p_T-q_T\right)\right|+T^{-1/2}\rho_T^2\right),
\end{aligned}$$

where the first equality follows from Lemma A.12, and the second from Lemmas A.11. Similarly,

$$\text{(M2)}=O_p\left(T^{-1/2}\rho_T^2+\left|\rho_T\left(p_T-q_T\right)\right|\right).$$

Next,

$$
\begin{aligned}
\text{(M3)} \quad = \quad & -\frac{1}{\xi_T} T^{-1/2} \sum_{t=1}^{T} \sum_{s,u=n_\beta+1}^{n_\beta+n_\delta} \frac{\nabla_{\delta_{1j}} \widetilde{f}_{1t}}{\widetilde{f}_t} \nabla_{\theta_s} \nabla_{\theta_u} \widetilde{\xi}_{t|t-1} \nabla_{\delta_{2k}} \hat{\theta}_s(\widetilde{\delta}) \nabla_{\delta_{2l}} \hat{\theta}_u(\widetilde{\delta}) \\
& -\frac{1}{\xi_T} T^{-1/2} \sum_{t=1}^{T} \sum_{s,u=n_\beta+n_\delta+1}^{n_\beta+2n_\delta} \frac{\nabla_{\delta_{1j}} \widetilde{f}_{1t}}{\widetilde{f}_t} \nabla_{\theta_s} \nabla_{\theta_u} \widetilde{\xi}_{t|t-1} \nabla_{\delta_{2k}} \hat{\theta}_s(\widetilde{\delta}) \nabla_{\delta_{2l}} \hat{\theta}_u(\widetilde{\delta}) \\
& -\frac{2}{\xi_T} T^{-1/2} \sum_{t=1}^{T} \sum_{s=n_\beta+1}^{n_\beta+n_\delta} \sum_{u=n_\beta+n_\delta+1}^{n_\beta+2n_\delta} \frac{\nabla_{\delta_{1j}} \widetilde{f}_{1t}}{\widetilde{f}_t} \nabla_{\theta_s} \nabla_{\theta_u} \widetilde{\xi}_{t|t-1} \nabla_{\delta_{2k}} \hat{\theta}_s(\widetilde{\delta}) \nabla_{\delta_{2l}} \hat{\theta}_u(\widetilde{\delta}).
\end{aligned}
$$

Among the three terms on the right hand side, the third term is $O_p(T^{-1/2}\rho_T^2 + |\rho_T(p_T - q_T)|)$ by the expression for the (I_1, I_2) case in Lemma 1.3. The first two terms lead to

$$
\begin{aligned}
& -T^{-1/2} \sum_{t=1}^{T} \frac{\nabla_{\delta_{1j}} \widetilde{f}_{1t}}{\widetilde{f}_t \xi_T} \left\{ \left(\frac{1-\xi_T}{\xi_T} \right)^2 \nabla_{\delta_{1k}} \nabla_{\delta_{1l}} \widetilde{\xi}_{t|t-1} + \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \widetilde{\xi}_{t|t-1} \right\} + O_p(T^{-1/2}\rho_T^2) \quad \text{(A.105)} \\
= \quad & O_p(T^{-1/2}\rho_T^2 + |\rho_T(p_T - q_T)|),
\end{aligned}
$$

where the equality follows from the expressions for (I_1, I_1) and (I_2, I_2) in Lemma 1.3. Thus,

$$
\text{(M3)} = O_p(T^{-1/2}\rho_T^2 + |\rho_T(p_T - q_T)|).
$$

The structure of (M4) is similar to that of (M1):

$$
\begin{aligned}
\text{(M4)} \quad = \quad & \frac{1}{\xi_T} T^{-1/2} \sum_{t=1}^{T} \sum_{s,u=n_\beta+1}^{n_\beta+2n_\delta} \frac{\nabla_{\theta_s} \nabla_{\theta_u} \widetilde{f}_{1t} - \nabla_{\theta_s} \nabla_{\theta_u} \widetilde{f}_{2t}}{\widetilde{f}_t} \nabla_{\delta_{2j}} \widetilde{\xi}_{t|t-1} \nabla_{\delta_{2k}} \hat{\theta}_s(\widetilde{\delta}) \nabla_{\delta_{2l}} \hat{\theta}_u(\widetilde{\delta}) \\
= \quad & \frac{1}{\xi_T} T^{-1/2} \sum_{t=1}^{T} \left(\left(\frac{1-\xi_T}{\xi_T} \right)^2 - 1 \right) \frac{\nabla_{\delta_{1k}} \nabla_{\delta_{1l}} \widetilde{f}_{1t}}{\widetilde{f}_t} \nabla_{\delta_{2j}} \widetilde{\xi}_{t|t-1} + O_p(T^{-1/2}\rho_T^2) \\
= \quad & O_p(T^{-1/2}\rho_T^2 + |\rho_T(p_T - q_T)|).
\end{aligned}
$$

For (M5), we apply the expression for the (I_1, I_2) case in Lemma 1.3:

$$
\begin{aligned}
\text{(M5)} \quad = \quad & T^{-1/2} \sum_{t=1}^{T} \sum_{s=n_\beta+1}^{n_\beta+2n_\delta} \frac{\nabla_{\theta_s} \widetilde{f}_{1t} - \nabla_{\theta_s} \widetilde{f}_{2t}}{\widetilde{f}_t} \sum_{u=n_\beta+n_\delta+1}^{n_\beta+2n_\delta} \nabla_{\delta_{2j}} \nabla_{\theta_u} \widetilde{\xi}_{t|t-1} \nabla_{\delta_{2k}} \hat{\theta}_s(\widetilde{\delta}) \nabla_{\delta_{2l}} \hat{\theta}_u(\widetilde{\delta}) \\
& - \left(\frac{1-\xi_T}{\xi_T} \right) T^{-1/2} \sum_{t=1}^{T} \sum_{s=n_\beta+1}^{n_\beta+2n_\delta} \frac{\nabla_{\theta_s} \widetilde{f}_{1t} - \nabla_{\theta_s} \widetilde{f}_{2t}}{\widetilde{f}_t} \sum_{u=n_\beta+1}^{n_\beta+n_\delta} \nabla_{\delta_{1j}} \nabla_{\theta_u} \widetilde{\xi}_{t|t-1} \nabla_{\delta_{2k}} \hat{\theta}_s(\widetilde{\delta}) \nabla_{\delta_{2l}} \hat{\theta}_u(\widetilde{\delta}) \\
& + O_p(T^{-1/2}\rho_T^2 + |\rho_T(p_T - q_T)|).
\end{aligned}
$$

The two leading terms on the right hand side lead to

$$
T^{-1/2} \sum_{t=1}^{T} \sum_{s=n_\beta+1}^{n_\beta+2n_\delta} \frac{\nabla_{\theta_s} \widetilde{f}_{1t} - \nabla_{\theta_s} \widetilde{f}_{2t}}{\widetilde{f}_t} \left\{ \left(\frac{\xi_T - 1}{\xi_T} \right)^2 \nabla_{\delta_{1j}} \nabla_{\delta_{1l}} \widetilde{\xi}_{t|t-1} + \nabla_{\delta_{2j}} \nabla_{\delta_{2l}} \widetilde{\xi}_{t|t-1} \right\} \nabla_{\delta_{2k}} \hat{\theta}_s(\widetilde{\delta}) + O_p(T^{-1/2}\rho_T^2),
$$

which reduces to (A.105). Therefore, (M5) $= O_p(T^{-1/2}\rho_T^2 + |\rho_T (p_T - q_T)|)$. Similarly, (M6) $= O_p(T^{-1/2}\rho_T^2 + |\rho_T (p_T - q_T)|)$. Therefore, (M1)-(M6) are all $O_p(T^{-1/2}\rho_T^2 + |\rho_T (p_T - q_T)|)$. Finally,

$$(\text{M0}) = -\frac{(1-\xi_T)(1-2\xi_T)}{\xi_T^2} T^{-1/2} \sum_{t=1}^{T} \frac{\nabla_{\delta_{1j}} \nabla_{\delta_{1k}} \nabla_{\delta_{1l}} \widetilde{f}_{1t}}{\widetilde{f}_t} + O_p(T^{-1/2}|\rho_T|).$$

The last two terms in (A.104) depend on $\nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\theta}(\widetilde{\delta})$. They can be rewritten as

$$\begin{aligned}
& T^{-1/2} \sum_{t=1}^{T} \sum_{s=n_\beta+1}^{n_\beta+2n_\delta} \frac{1}{\widetilde{f}_t} \left\{ (1-\xi_T) \nabla_{\delta_{2j}} \nabla_{\theta_s} \widetilde{f}_{2t} - (1-\xi_T) \nabla_{\delta_{1j}} \nabla_{\theta_s} \widetilde{f}_{1t} \right\} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\theta}_s(\widetilde{\delta}) \\
& -\frac{1}{\xi_T} T^{-1/2} \sum_{t=1}^{T} \sum_{s=1}^{n_\beta+2n_\delta} \frac{1}{\widetilde{f}_t} \left\{ \nabla_{\delta_{1j}} \widetilde{f}_{1t} \nabla_{\theta_s} \widetilde{\xi}_{t|t-1} + (\nabla_{\theta_s} \widetilde{f}_{1t} - \nabla_{\theta_s} \widetilde{f}_{2t}) \nabla_{\delta_{1j}} \widetilde{\xi}_{t|t-1} \right\} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\theta}_s(\widetilde{\delta}).
\end{aligned}$$

The first double summation equals 0 by Assumption 7. The second double summation equals

$$-\frac{1}{\xi_T} T^{-1/2} \sum_{t=1}^{T} \sum_{s=1}^{n_\delta} \frac{1}{\widetilde{f}_t} \left\{ \nabla_{\delta_{1j}} \widetilde{f}_{1t} \nabla_{\delta_{1s}} \widetilde{\xi}_{t|t-1} + \nabla_{\delta_{1s}} \widetilde{f}_{1t} \nabla_{\delta_{1j}} \widetilde{\xi}_{t|t-1} \right\} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \hat{\delta}_{1s}(\widetilde{\delta}) = O_p(|\rho_T|).$$

The Lemma follows after combining this result with those for (M0)-(M6).

Part 3: Proof of the third result of Lemma A.17. We divide the components of (A.41) into three subsets as explained in Remark 8. The components in the second subset depend on the first order derivative of B_t. They are $O_p(|\rho_T|)$ by Lemmas A.12 and (A.81).

The first subset has only one component:

$$\begin{aligned}
& T^{-1/2} \sum_{t=1}^{T} \frac{\nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \nabla_{\delta_{2m}} \widetilde{L}_{jt}}{\widetilde{B}_t} - \left(\frac{1-\xi_T}{\xi_T}\right) T^{-1/2} \sum_{t=1}^{T} \frac{\nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \nabla_{\delta_{2m}} \widetilde{M}_{(n_\beta+j)t}}{\widetilde{B}_t} && \text{(A.106)} \\
= \; & T^{-1/2} \sum_{t=1}^{T} \widetilde{k}_{1,jklm,t} + O_p(|\rho_T|),
\end{aligned}$$

where $O_p(|\rho_T|)$ follows from Lemmas A.12 and A.13 and

$$\begin{aligned}
\widetilde{k}_{1,jklm,t} = \; & \frac{1}{\widetilde{f}_t} \left\{ (1-\xi_T) \left(1 + \left(\frac{1-\xi_T}{\xi_T}\right)^3 \right) \nabla_{\delta_{1j}} \nabla_{\delta_{1k}} \nabla_{\delta_{1l}} \nabla_{\delta_{1m}} \widetilde{f}_{1t} \right. && \text{(A.107)} \\
& - \left(\frac{1-\xi_T}{\xi_T}\right)^2 \nabla_{\delta_{1j}} \nabla_{\delta_{1l}} \nabla_{\beta'} \widetilde{f}_{1t} \alpha_{km}^{(1)} - \left(\frac{1-\xi_T}{\xi_T}\right)^3 \nabla_{\delta_{1j}} \nabla_{\delta_{1l}} \nabla_{\delta_1'} \widetilde{f}_{1t} \alpha_{km}^{(2)} \\
& - \left(\frac{1-\xi_T}{\xi_T}\right)^2 \nabla_{\delta_{1j}} \nabla_{\delta_{1k}} \nabla_{\beta'} \widetilde{f}_{1t} \alpha_{lm}^{(1)} - \left(\frac{1-\xi_T}{\xi_T}\right)^3 \nabla_{\delta_{1j}} \nabla_{\delta_{1k}} \nabla_{\delta_1'} \widetilde{f}_{1t} \alpha_{lm}^{(2)} \\
& \left. - \left(\frac{1-\xi_T}{\xi_T}\right)^2 \nabla_{\delta_{1j}} \nabla_{\delta_{1m}} \nabla_{\beta'} \widetilde{f}_{1t} \alpha_{kl}^{(1)} - \left(\frac{1-\xi_T}{\xi_T}\right)^3 \nabla_{\delta_{1j}} \nabla_{\delta_{1m}} \nabla_{\delta_1'} \widetilde{f}_{1t} \alpha_{kl}^{(2)} \right\}.
\end{aligned}$$

The third subset has three components. Because they have the same structure, it suffices to consider the first:

$$T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\widetilde{B}_t}{\widetilde{B}_t}\left\{\frac{\nabla_{\delta_{2k}}\widetilde{L}_{jt}}{\widetilde{B}_t}-\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{2k}}\widetilde{M}_{(n_\beta+j)t}}{\widetilde{B}_t}\right\}$$

$$= T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\widetilde{B}_t}{\widetilde{f}_t}\left\{\frac{(1-\xi_T)\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T}+\frac{\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1j}}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right\}+O_p\left(|\rho_T|\right),$$

where the equality follows because of (A.100) and Lemma A.12. By applying (A.84) to $\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\widetilde{B}_t/\widetilde{B}_t$, this expression can be written as

$$T^{-1/2}\sum_{t=1}^{T}\left(\frac{(1-\xi_T)\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T}+\frac{\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1j}}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right)\times$$
$$\left\{\frac{(1-\xi_T)\nabla_{\delta_{1h}}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T}+\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1h}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1h}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right.$$
$$\left.+\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\begin{bmatrix}\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}\right\}+O_p\left(|\rho_T|\right).$$

Further, because $\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})$ and $\xi_T\nabla_{\delta_{2h}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta})$ satisfy (A.83), the above expression equals

$$T^{-1/2}\sum_{t=1}^{T}\left(\frac{(1-\xi_T)\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T}+\frac{\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1j}}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right)$$
$$\times\left\{\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1m}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right. \tag{N1}$$
$$-\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left[\frac{\nabla_{\delta_{1m}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1m}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right] \tag{N2}$$
$$+\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_\beta\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\\ \left(\frac{1-\xi_T}{\xi_T}\right)^2\frac{\nabla_{\delta_1}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{bmatrix} \tag{N3}$$
$$\left.-\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{1-\xi_T}{\xi_T}\right)\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}\frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & \frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & \frac{\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{bmatrix}\begin{bmatrix}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}\right\} \tag{N4}$$
$$+O_p\left(|\rho_T|\right),$$

where (N1)-(N4) denote the four components in the braces. Below, we study them separately.

The product of the three terms in the parentheses with (N1) leads to

$$
\begin{aligned}
&T^{-1/2}\sum_{t=1}^{T}\left\{\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}\frac{\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1j}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}\right\} \qquad \text{(A.108)}\\
&\times\left\{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\frac{\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}\nabla_{\delta_{1m}}\widetilde{\xi}_{t|t-1}\right\}\\
&+\left(\frac{1-\xi_T}{\xi_T}\right)T^{-1/2}\sum_{t=1}^{T}\alpha_{jk}'\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left\{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\frac{\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}\nabla_{\delta_{1m}}\widetilde{\xi}_{t|t-1}\right\}.
\end{aligned}
$$

Its product with (N2) leads to

$$
\begin{aligned}
&-T^{-1/2}\sum_{t=1}^{T}\left\{\frac{\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1k}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1j}}\widetilde{f}_{1t}\nabla_{\delta_{1k}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right\} \qquad \text{(A.109)}\\
&\times\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left\{\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1m}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right\}\\
&-\left(\frac{1-\xi_T}{\xi_T}\right)T^{-1/2}\sum_{t=1}^{T}\alpha_{jk}'\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left\{\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1m}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right\}.
\end{aligned}
$$

Note that second term in (A.109) and that in (A.108) are opposite to each other, therefore cancel out when taking the summation. The product with (N3) leads to

$$
\begin{aligned}
&\left(\frac{1-\xi_T}{\xi_T}\right)\alpha_{jk}'T^{-1/2}\sum_{t=1}^{T}\begin{bmatrix}\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_\beta\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\\ \left(\frac{1-\xi_T}{\xi_T}\right)^2\frac{\nabla_{\delta_1}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{bmatrix}+O_p\left(|\rho_T|\right) \qquad \text{(A.110)}\\
=\;&\left(\frac{1-\xi_T}{\xi_T}\right)^2T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\beta'}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{jk}^{(1)}+\left(\frac{1-\xi_T}{\xi_T}\right)^3T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_1'}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{jk}^{(2)}+O_p\left(|\rho_T|\right).
\end{aligned}
$$

Finally, the product with (N4) leads to

$$
\begin{aligned}
&-\left(\frac{1-\xi_T}{\xi_T}\right)^2T^{-1/2}\sum_{t=1}^{T}\alpha_{jk}'\begin{bmatrix}\frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & \frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & \frac{\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{bmatrix}\begin{bmatrix}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix} \qquad \text{(A.111)}\\
&+O_p\left(|\rho_T|\right)\\
=\;&-\left(\frac{1-\xi_T}{\xi_T}\right)^2T^{-1/2}\sum_{t=1}^{T}\left(\alpha_{jk}^{(1)}\right)'\frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{lm}^{(1)}-\left(\frac{1-\xi_T}{\xi_T}\right)^2T^{-1/2}\sum_{t=1}^{T}\left(\alpha_{jk}^{(1)}\right)'\frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{lm}^{(2)}\\
&-\left(\frac{1-\xi_T}{\xi_T}\right)^2T^{-1/2}\sum_{t=1}^{T2}\left(\alpha_{jk}^{(2)}\right)'\frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{lm}^{(1)}+O_p\left(|\rho_T|+T^{-1/2}\right),
\end{aligned}
$$

where the equality holds because of Lemma A.13.

The result follows from taking the summation of (A.106) and (A.108)-(A.111). In particular, the sum of (A.108) and (A.109) equals $T^{1/2}\widetilde{\omega}_{jklm}^{(2)}(p_T,q_T)$. The sum of (A.106), (A.110), and (A.111) equals $T^{-1/2}\sum_{t=1}^{T}\widetilde{k}_{jklm,t}(p_T,q_T)+O_p(|\rho_T|+T^{-1/2})$.

Part 4: Proof of the fourth result of Lemma A.17. As explained in Remark (8), we divide the components of $T^{-1}\mathcal{L}^{(5)}_{jklmn}(p_T, q_T, \widetilde{\delta}) - ((1-\xi_T)/\xi_T)\,\mathcal{M}^{(5)}_{(n_\beta+j)klmn}(p_T, q_T, \widetilde{\delta})$ into 4 subsets. Those in the first two subsets are $O_p(T^{-1/2})$ and $O_p(T^{-1/2}|\rho_T|)$ respectively, by the CLT, Lemma A.12, and (A.81). For the third subset, we consider the following representative element:

$$T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\widetilde{B}_t}{\widetilde{f}_t}\left\{\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2n}}\widetilde{L}_{jt}}{\widetilde{f}_t}-\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2n}}\widetilde{M}_{(n_\beta+j)t}}{\widetilde{f}_t}\right\}.$$

Applying (A.83) as in (A.112), we rewrite it as

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2n}}\widetilde{L}_{jt}}{\widetilde{f}_t}-\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2n}}\widetilde{M}_{(n_\beta+j)t}}{\widetilde{f}_t}\right) \\
&\times\left\{\frac{\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1l}}\widetilde{f}_{1t}\nabla_{\delta_{1m}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right. \\
&-\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{1m}}\widetilde{\xi}_{t|t-1}\nabla_{\delta_{1l}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T^2}+\frac{\nabla_{\delta_{1m}}\widetilde{f}_{1t}\nabla_{\delta_{1l}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}\right) \\
&+\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{c}\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_\beta\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\\ \left(\frac{1-\xi_T}{\xi_T}\right)^2\frac{\nabla_{\delta_1}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{array}\right] \\
&\left.-\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{1-\xi_T}{\xi_T}\right)\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{cc}\frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & \frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\\ \frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t} & \frac{\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\end{array}\right]\left[\begin{array}{c}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\hat{\delta}_1(\widetilde{\delta})\end{array}\right]\right\} \\
&+O_p\left(T^{-1/2}|\rho_T|\right).
\end{aligned}\tag{A.112}$$

Further, by (A.103) and (A.102), we have

$$\begin{aligned}
&\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2n}}\widetilde{L}_{jt}}{\widetilde{B}_t}-\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2n}}\widetilde{M}_{(n_\beta+j)t}}{\widetilde{B}_t} \\
=\;&(1-\xi_T)\sum_{s,u=1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\left\{\nabla_{\delta_{2j}}\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t}-\nabla_{\delta_{1j}}\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t}\right\}\nabla_{\delta_{2k}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2n}}\hat{\theta}_u(\widetilde{\delta}) \\
&-(1-\xi_T)\sum_{v=n_\beta+1}^{n_\beta+n_\delta}\frac{\nabla_{\delta_{1j}}\nabla_{\theta_v}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2k}}\nabla_{\delta_{2n}}\hat{\theta}_v(\widetilde{\delta})+R_t,
\end{aligned}\tag{A.113}$$

where R_t arises because of omitting the terms that depend on the derivatives of $\xi_{t|t-1}$. The effect of R_t on (A.112) is $O_p(\rho_T^2+T^{-1/2})$. From (A.113) and Assumption 9, it follows that (A.112) is $O_p(T^{-1/2}+\rho_T^2)$.

For the fourth subset, a representative component is

$$T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\nabla_{\delta_{2m}}\widetilde{B}_t}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{2n}}\widetilde{L}_{jt}}{\widetilde{f}_t}-\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{2n}}\widetilde{M}_{(n_\beta+j)t}}{\widetilde{f}_t}\right).\tag{A.114}$$

As we have noted when studying (A.65) and (A.94), $\nabla_{\delta_{2n}} \widetilde{L}_{jt}/\widetilde{f}_t$ and $\nabla_{\delta_{2n}} \widetilde{M}_{(n_\beta+j)t}/\widetilde{f}_t$ can be represented as a linear function of $\widetilde{M}_{it}/\widetilde{f}_t$ $(i = 1, ..., n_\beta + n_\delta)$ and $\nabla_{\theta_s} \widetilde{f}_{1t} \nabla_{\theta_u} \widetilde{\xi}_{t|t-1}/\widetilde{f}_t$ $(s, u = n_\beta + 1, ..., n_\beta + 2n_\delta)$. By (A.89), $T^{-1} \sum_{t=1}^T \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \nabla_{\delta_{2m}} \widetilde{B}_t \widetilde{M}_{it}/\widetilde{f}_t^2 = O_p(T^{-1/2} + \rho_T^2)$ for $i = 1, ..., n_\beta + n_\delta$. By (A.95), $T^{-1} \sum_{t=1}^T \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \nabla_{\delta_{2m}} \widetilde{B}_t \nabla_{\theta_s} \widetilde{f}_{1t} \nabla_{\theta_u} \widetilde{\xi}_{t|t-1}/\widetilde{f}_t^2 = O_p(T^{-1/2} + \rho_T^2)$ for any $s, u \in \{n_\beta + 1, ..., n_\beta + 2n_\delta\}$. Therefore, (A.114) is $O_p(T^{-1/2} + \rho_T^2)$.

Part 5: Proof of the fifth result of Lemma A.17. The proof is similar to Part 4. We divide the components of $T^{-1}\mathcal{L}^{(6)}_{jklmnr}(p, q, \widetilde{\delta}) - ((1 - \xi_T)/\xi_T)\mathcal{M}^{(6)}_{(n_\beta+j)klmnr}(p, q, \widetilde{\delta})$ into five subsets. The proof below shows that the components in the fourth subset lead to $\sum_{(i_1,i_2,...,i_6)\in IND_1} \omega^{(3)}_{i_1 i_2 \dots i_6}(p_T, q_T; p_T, q_T)$, and that the remaining ones are $O_p(T^{-1/2} + \rho_T^2 + |\rho_T| \, (p_T - q_T)^2)$.

The components in the first two subsets are $O_p(T^{-1/2})$ and $O_p(T^{-1/2}|\rho_T|)$ for the same reason as in Part 3. For the third subset, it suffices to consider

$$T^{-1} \sum_{t=1}^T \left(\frac{\nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \nabla_{\delta_{2m}} \widetilde{L}_{jt}}{\widetilde{B}_t} - \left(\frac{1 - \xi_T}{\xi_T} \right) \frac{\nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \nabla_{\delta_{2m}} \widetilde{M}_{(n_\beta+j)t}}{\widetilde{B}_t} \right) \frac{\nabla_{\delta_{2n}} \nabla_{\delta_{2r}} \widetilde{B}_t}{\widetilde{B}_t}. \tag{A.115}$$

By (A.83), this equals

$$\begin{aligned} &\frac{1}{T} \sum_{t=1}^T \left(\frac{\nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \nabla_{\delta_{2m}} \widetilde{L}_{jt}}{\widetilde{f}_t} - \left(\frac{1 - \xi_T}{\xi_T} \right) \frac{\nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \nabla_{\delta_{2m}} \widetilde{M}_{(n_\beta+j)t}}{\widetilde{f}_t} \right) \qquad \text{(A.116)} \\ &\times \left\{ \frac{\nabla_{\delta_{1n}} \widetilde{\xi}_{t|t-1} \nabla_{\delta_{1r}} \widetilde{f}_{1t}}{\widetilde{f}_t \xi_T^2} + \frac{\nabla_{\delta_{1r}} \widetilde{f}_{1t} \nabla_{\delta_{1n}} \widetilde{\xi}_{t|t-1}}{\widetilde{f}_t \xi_T^2} \right. \\ &\left. - \frac{\nabla_{(\beta',\delta_1')} \widetilde{f}_{1t}}{\widetilde{f}_t} \widetilde{I}^{-1} \frac{1}{T} \sum_{t=1}^T \frac{\nabla_{(\beta',\delta_1')'} \widetilde{f}_{1t}}{\widetilde{f}_t} \left(\frac{\nabla_{\delta_{1r}} \widetilde{\xi}_{t|t-1} \nabla_{\delta_{1n}} \widetilde{f}_{1t}}{\widetilde{f}_t \xi_T^2} + \frac{\nabla_{\delta_{1r}} \widetilde{f}_{1t} \nabla_{\delta_{1n}} \widetilde{\xi}_{t|t-1}}{\widetilde{f}_t \xi_T^2} \right) \right\} + O_p(T^{-1/2}). \end{aligned}$$

Further, because of (A.35) and (A.42) and Lemma A.13,

$$\begin{aligned} &\frac{\nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \nabla_{\delta_{2m}} \widetilde{L}_{jt}}{\widetilde{f}_t} - \left(\frac{1 - \xi_T}{\xi_T} \right) \frac{\nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \nabla_{\delta_{2m}} \widetilde{M}_{(n_\beta+j)t}}{\widetilde{f}_t} \qquad \text{(A.117)} \\ = \; &\widetilde{k}_{1,jklm,t} - (1 - \xi_T) \nabla_{\delta_{1j}} \nabla_{\delta_1'} \widetilde{f}_{1t} \nabla_{\delta_{2k}} \nabla_{\delta_{2l}} \nabla_{\delta_{2m}} \hat{\delta}_1(\widetilde{\delta}) + R_t, \end{aligned}$$

where $\widetilde{k}_{1,jklm,t}$ is defined in (A.107), and R_t arises because the terms involving the derivatives of $\xi_{t|t-1}$ are omitted, and because of the application of Lemmas A.12 and A.13. The effect of R_t on (A.115) is $O_p(\rho_T^2 + T^{-1/2})$. From this result and Assumption 9, it follows that (A.115) is $O_p(\rho_T^2 + T^{-1/2})$.

The fourth subset has 10 components. Their overall effect on $T^{-1}\mathcal{L}^{(6)}_{jklmnr}(p_T, q_T, \widetilde{\delta})$ is

$$-\frac{1}{T} \sum_{(i_1,...,i_6)\in IND_1} \sum_{t=1}^T \frac{\nabla_{\delta_{2i_4}} \nabla_{\delta_{2i_5}} \nabla_{\delta_{2i_6}} \widetilde{B}_t}{\widetilde{B}_t} \left\{ \frac{\nabla_{\delta_{2i_2}} \nabla_{\delta_{2i_3}} \hat{L}_{i_1 t}}{\widetilde{B}_t} - \left(\frac{1 - \xi_T}{\xi_T} \right) \frac{\nabla_{\delta_{2i_2}} \nabla_{\delta_{2i_3}} \widetilde{M}_{(n_\beta+i_1)t}}{\widetilde{B}_t} \right\}. \tag{A.118}$$

Using (A.103) and (A.102), the expression in the braces can be represented as

$$
\begin{aligned}
&(1-\xi_T)\sum_{s,u=n_\beta+1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\left\{\nabla_{\delta_{2i_1}}\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{2t}-\nabla_{\delta_{1i_1}}\nabla_{\theta_s}\nabla_{\theta_u}\widetilde{f}_{1t}\right\}\nabla_{\delta_{2i_2}}\hat{\theta}_s(\widetilde{\delta})\nabla_{\delta_{2i_3}}\hat{\theta}_u(\widetilde{\delta})\\
&-(1-\xi_T)\sum_{s=n_\beta+1}^{n_\beta+n_\delta}\frac{\nabla_{\delta_{1j}}\nabla_{\theta_s}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\hat{\theta}_s(\widetilde{\delta})\\
&-\sum_{s=n_\beta+1}^{n_\beta+n_\delta}\left\{\frac{\nabla_{\delta_{2j}}\widetilde{f}_{2t}\nabla_{\theta_s}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T}-\frac{(\nabla_{\theta_s}\widetilde{f}_{1t}-\nabla_{\theta_s}\widetilde{f}_{2t})\nabla_{\delta_{1j}}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T}\right\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})\\
&+T^{-1/2}R_{1t}+\left|\rho_T\left(p_T-q_T\right)\right|R_{2t},
\end{aligned}
$$

where $T^{-1/2}R_{1t}$ and $\left|\rho_T\left(p_T-q_T\right)\right|R_{2t}$ represent the following terms which are omitted from the first summation: (a) those with $s\in\{1,..,n_\beta\}$ or $u\in\{1,..,n_\beta\}$, and (b) those involving derivatives of $\xi_{t|t-1}$ for $s,u\in\{n_\beta+1,...,n_\beta+2n_\delta\}$. We consider the product of the three leading terms in the above displayed equation with $\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\nabla_{\delta_{2i_6}}\widetilde{B}_t$ in (A.118). The product of the first term with $\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\nabla_{\delta_{2i_6}}\widetilde{B}_t$ can be studied using Lemma A.12, the product of the second term with $\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\nabla_{\delta_{2i_6}}\widetilde{B}_t$ can be studied using (A.89) because of Assumption 7, and the product of the third term with $\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\nabla_{\delta_{2i_6}}\widetilde{B}_t$ can be studied using (A.114). As a result, (A.118) equals

$$
\begin{aligned}
&((1-\xi_T)(1-2\xi_T)/\xi_T^2)\sum_{(i_1,\ldots,i_6)\in IND_1}T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\nabla_{\delta_{2i_6}}\widetilde{B}_t\nabla_{\delta_{1i_1}}\nabla_{\delta_{1i_2}}\nabla_{\delta_{1i_3}}\widetilde{f}_{1t}}{\widetilde{f}_t^2}\\
&+O_p(T^{-1/2}+\rho_T^2+\left|\rho_T\right|(p_T-q_T)^2).
\end{aligned}
$$

Further, because of (A.91) and Lemma A.14, we have

$$
\begin{aligned}
&\frac{\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\nabla_{\delta_{2i_6}}\widetilde{B}_t}{\widetilde{B}_t} \qquad\qquad (A.119)\\
=\;&\frac{(\xi_T-1)(1-2\xi_T)}{\xi_T^2}\left\{\frac{\nabla_{\delta_{1i_4}}\nabla_{\delta_{1i_5}}\nabla_{\delta_{1i_6}}\widetilde{f}_{1t}}{\widetilde{f}_t}-\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{I}^{-1}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\frac{\nabla_{\delta_{1i_4}}\nabla_{\delta_{1i_5}}\nabla_{\delta_{1i_6}}\widetilde{f}_{1t}}{\widetilde{f}_t}\right\}\\
&+|\rho_T(p_T-q_T)|R_{3t}+T^{-1/2}R_{4t},
\end{aligned}
$$

where $|\rho_T(p_T-q_T)|R_{3t}$ arises because the terms involving the first and second order derivatives of the density (e.g., $\nabla_{\delta_{1i_4}}\widetilde{f}_{1t}\nabla_{\delta_{1i_5}}\widetilde{\xi}_{t|t-1}$ and $\nabla_{\delta_{1i_4}}\nabla_{\delta_{1i_5}}\widetilde{f}_{1t}\nabla_{\delta_{1i_6}}\widetilde{\xi}_{t|t-1})$) are omitted. From these two displayed equations, it follows that (A.118) equals $\sum_{(i_1,i_2,\ldots,i_6)\in IND_1}\omega^{(3)}_{i_1i_2\ldots i_6}(p_T,q_T;p_T,q_T)+O_p(T^{-1/2}+\left|\rho_T\right|(p_T-q_T)^2+\rho_T^2)$.

Finally, for the fifth subset, it is sufficient to consider

$$
T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{1l}}\nabla_{\delta_{2m}}\nabla_{\delta_{2n}}\nabla_{\delta_{2r}}\widetilde{B}_t}{\widetilde{B}_t}\left(\frac{\nabla_{\delta_{1k}}\widetilde{L}_{jt}}{\widetilde{B}_t}-\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{1k}}\widetilde{M}_{(n_\beta+j)t}}{\widetilde{B}_t}\right). \qquad (A.120)
$$

For this, Lemma A.15 and the analysis (A.97)-(A.99) imply

$$\frac{\nabla_{\delta_{1l}}\nabla_{\delta_{2m}}\nabla_{\delta_{2n}}\nabla_{\delta_{2r}}\widetilde{B}_t}{\widetilde{B}_t} \tag{A.121}$$
$$= \widetilde{k}_{lmnr,t} - \frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left\{\frac{1}{T}\widetilde{I}^{-1}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{k}_{lmnr,t}\right\} + \left(T^{-1/2} + |\rho_T|\right)R_t.$$

Applying this approximation, we can study (A.120) in the same way as (A.114). From this, it follows that (A.120) is $O_p\left(T^{-1/2} + \rho_T^2\right)$.

Part 6: Proof of the sixth result of Lemma A.17. We divide the components of $T^{-1}\mathcal{L}^{(7)}_{jklmnrs}(p_T, q_T, \widetilde{\delta}) - ((1-\xi_T)/\xi_T)\,\mathcal{M}^{(7)}_{(n_\beta+j)klmnrs}(p_T, q_T, \widetilde{\delta})$ into 6 subsets as explained in Remark 8. Those in the first two subsets are $O_p(T^{-1/2} + |\rho_T|)$ and $O_p(T^{-1/2}\,|\rho_T|)$, respectively, for the same reason as in Part 3. Those in the third subset are $O_p\left(|\rho_T| + T^{-1/2}\right)$ because of (A.85).

Let $i_1 = j$ and $i_2, ..., i_7 \in \{k, l, m, n, r, s\}$. For the fourth subset, a representative component is $T^{-1}\sum_{t=1}^{T}\nabla_{\delta_{2i_5}}\nabla_{\delta_{2i_6}}\nabla_{\delta_{2i_7}}\widetilde{B}_t\{\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\nabla_{\delta_{2i_4}}\widetilde{L}_{i_1t} - ((1-\xi_T)/\xi_T)\,\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\nabla_{\delta_{2i_4}}\widetilde{M}_{(n_\beta+i_1)t}\}/\widetilde{B}_t^2$, which is $O_p(|\rho_T| + |p_T - q_T| + T^{-1/2})$ because of (A.119) and (A.117). For the fifth subset, a representative component is

$$T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_4}}...\nabla_{\delta_{2i_7}}\widetilde{B}_t}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\widetilde{L}_{i_1t}}{\widetilde{f}_t} - \left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\widetilde{M}_{(n_\beta+i_1)t}}{\widetilde{f}_t}\right). \tag{A.122}$$

By (A.113), this can be written as

$$\frac{(\xi_T-1)(1-2\xi_T)}{\xi_T^2}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_4}}...\nabla_{\delta_{2i_7}}\widetilde{B}_t}{\widetilde{f}_t}\frac{\nabla_{\delta_{1i_1}}\nabla_{\delta_{1i_2}}\nabla_{\delta_{1i_3}}\widetilde{f}_{1t}}{\widetilde{f}_t} \tag{A.123}$$
$$-(1-\xi_T)\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{n_\delta}\frac{\nabla_{\delta_{2i_4}}...\nabla_{\delta_{2i_7}}\widetilde{B}_t}{\widetilde{f}_t}\frac{\nabla_{\delta_{1i_1}}\nabla_{\delta_{1s}}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\hat{\delta}_{1s}(\widetilde{\delta}) + O_p(|\rho_T|),$$

where first term is $O_p(|p_T - q_T|)$ and the second is $O_p(\rho_T^2 + T^{-1/2})$ by (A.96). Therefore, the components in this subset are $O_p(|\rho_T| + |p_T - q_T| + T^{-1/2})$. For the sixth subset, a representative component is $T^{-1}\sum_{t=1}^{T}\nabla_{\delta_{2i_3}}...\nabla_{\delta_{2i_7}}\widetilde{B}_t(\nabla_{\delta_{2i_2}}\widetilde{L}_{i_1t} - ((1-\xi_T)/\xi_T)\,\nabla_{\delta_{2i_2}}\widetilde{M}_{(n_\beta+i_1)t})/\widetilde{B}_t^2$. This is $O_p(|\rho_T| + (p_T - q_T)^2 + T^{-1/2})$ by Lemma A.18 proved below.

Part 7: Proof of the seventh result of Lemma A.17. We divide the components of

$$T^{-1}\mathcal{L}^{(8)}_{jklmnrsu}(p, q, \widetilde{\delta}) - ((1-\xi_T)/\xi_T)\,\mathcal{M}^{(8)}_{(n_\beta+j)klmnrsu}(p, q, \widetilde{\delta}) \tag{A.124}$$

into 7 subsets as explained in Remark 8. The proof below shows that those in the fifth and seventh subsets lead to $-\sum_{(i_1,i_2,...,i_8)\in IND_2}\omega^{(4)}_{i_1i_2...i_8}(p_T, q_T; p_T, q_T)$, and that the remaining ones are $O_p(|\rho_T| + |(p_T - q_T)|) + o_p(1)$.

The components in the first three subsets are $O_p(|\rho_T|) + o_p(1)$, $O_p(T^{-1/2}|\rho_T|)$, and $O_p(|\rho_T| + T^{-1/2})$, respectively, for the same reason as in Part 6. Those in the fourth subset are $O_p(|\rho_T| + |p_T - q_T| + T^{-1/2})$ because of (A.119). A representative component of the sixth subset is

$$T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_4}}...\nabla_{\delta_{2i_8}}\widetilde{B}_t}{\widetilde{B}_t^2}\{\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\widetilde{L}_{i_1t} - ((1-\xi_T)/\xi_T)\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\widetilde{M}_{(n_\beta+i_1)t}\},$$

where $i_1 = j$ and $i_2, ..., i_8 \in \{k, l, m, n, r, s, u\}$. This can be studied in the same way as (A.122), except that (A.96) is replaced by Lemma A.18 proved below. From this, it follows that the components of this subset are $O_p(|\rho_T| + |p_T - q_T| + T^{-1/2})$.

The fifth subset consists of 35 components. Their overall effect on (A.124) is

$$-\sum_{(i_1,...,i_8)\in IND_2}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_5}}...\nabla_{\delta_{2i_8}}\widetilde{B}_t}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\nabla_{\delta_{2i_4}}\widetilde{L}_{i_1t}}{\widetilde{f}_t} - \left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\nabla_{\delta_{2i_4}}\widetilde{M}_{(n_\beta+i_1)t}}{\widetilde{f}_t}\right).$$

By (A.97)-(A.99), Lemma A.15, and (A.117), this equals

$$\begin{aligned} -\sum_{(i_1,...,i_8)\in IND_2}\frac{1}{T}\sum_{t=1}^{T}&\left\{\widetilde{k}_{i_5...i_8,t} - \frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{1}{T}\widetilde{I}^{-1}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{k}_{i_5...i_8,t}\right)\right\} \\ &\times\widetilde{k}_{1,i_1...i_4,t} + O_p\left(T^{-1/2} + |\rho_T|\right), \end{aligned} \quad \text{(A.125)}$$

where $\widetilde{k}_{1,i_1...i_4,t}$ is defined in (A.107). The 7th subset has 7 components. Their effect on (A.124) is

$$-\sum_{(i_1,...,i_8)\in IND_3}\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_3}}...\nabla_{\delta_{2i_8}}\widetilde{B}_t}{\widetilde{f}_t}\left(\frac{\nabla_{\delta_{2i_2}}\widetilde{L}_{i_1t}}{\widetilde{f}_t} - \left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{2i_2}}\widetilde{M}_{(n_\beta+i_1)t}}{\widetilde{f}_t}\right),$$

where the elements of IND_3 are defined as follows: $i_1 = j, i_2 \in \{k, l, m, n, r, s, u\}$, and $i_3, ..., i_8$ are the remaining elements of $\{k, l, m, n, r, s, u\}$ once i_2 is selected, the ordering of which does not matter. Applying (A.100) and Assumption 7, this summation can be rewritten as

$$\begin{aligned} &-\left(\frac{1-\xi_T}{\xi_T}\right)\sum_{(i_1,...,i_8)\in IND_3}T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_3}}...\nabla_{\delta_{2i_8}}\widetilde{B}_t}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{i_1i_2} + O_p(|\rho_T|) \quad \text{(A.126)} \\ =\ &\left(\frac{1-\xi_T}{\xi_T}\right)^2\frac{1}{T}\sum_{(i_1,i_2)\in IND_3}\sum_{(j_3,...,j_8)\in IND_4}\sum_{t=1}^{T}\left\{\widetilde{k}_{j_5...j_8,t} - \frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{1}{T}\widetilde{I}^{-1}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{k}_{j_5...j_8,t}\right)\right\} \\ &\times\left\{\frac{\nabla_{\delta_{1j_3}}\nabla_{\delta_{1j_4}}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(1)}_{i_1i_2} + \left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{1j_3}}\nabla_{\delta_{1j_4}}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(2)}_{i_1i_2} - \alpha^{(1)\prime}_{i_1i_2}\frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(1)}_{j_3j_4}\right. \\ &\left.-\alpha^{(1)}_{i_1i_2}\frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(2)}_{j_3j_4} - \alpha^{(2)\prime}_{i_1i_2}\frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(1)}_{j_3j_4}\right\} + O_p\left(T^{-1/2} + |\rho_T| + |p_T - q_T|\right), \end{aligned}$$

where the equality follows from Lemma A.19 proved below. Computing the summations in (A.125) and (A.126) and re-arranging the terms, we obtain the expression in the Lemma. This completes the proof of Lemma A.17. ■

The next two auxiliary lemmas are used in Parts 6 and 7 of the preceding proof.

Lemma A.18 *Suppose that the null hypothesis and Assumptions 1-4 and 7 hold with (24) satisfied for all* $i_1, i_2 \in \{1, ..., n_\delta\}$*, and that* $\epsilon \le p_T, q_T \le 1-\epsilon$ *is satisfied for all* T*. Then,*

$$T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_1}}...\nabla_{\delta_{2i_5}}\widetilde{B}_t}{\widetilde{B}_t}\frac{\widetilde{M}_{jt}}{\widetilde{B}_t} = O_p(T^{-1/2} + |\rho_T| + (p_T - q_T)^2)$$

for any $j \in \{1, ..., n_\beta + n_\delta\}$ *and any* $i_1, ..., i_5 \in \{1, ..., n_\delta\}$*.*

Proof of Lemma A.18. The proof is similar to that of (A.96). We divide the components of $\mathcal{M}^{(6)}_{ji_1...i_5}(p, q, \widetilde{\delta})$ into 6 subsets as explained in Remark 8. The elements of the first three subsets are $O_p(T^{-1/2}), O_p(T^{-1/2}|\rho_T|)$, and $O_p(T^{-1/2} + |\rho_T|)$, respectively, as in Part 6 of the preceding proof.

A representative element of the fourth subset is $T^{-1}\sum_{t=1}^{T}(\nabla_{\delta_{2i_1}}\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\widetilde{B}_t\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\widetilde{M}_{jt})/\widetilde{B}_t^2$. If $j \in \{n_\beta + 1, ..., n_\beta + n_\delta\}$, it equals

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_1}}\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\widetilde{B}_t}{\widetilde{f}_t}\\ &\times\left\{\frac{(1-\xi_T)\nabla_{\beta_j}\nabla_{\delta_{1i_4}}\nabla_{\delta_{1i_5}}\widetilde{f}_{1t}}{\widetilde{f}_t\xi_T} + \frac{\nabla_{\beta_j}\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\begin{bmatrix}\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{2i_4}}\nabla_{\delta_{2i_5}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}\right\} + O_p(|\rho_T|),\end{aligned}$$

which is $O_p(|\rho_T| + T^{-1/2})$ by Lemma A.13. If $j \in \{n_\beta + 1, ..., n_\beta + n_\delta\}$, it equals

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_1}}\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\widetilde{B}_t}{\widetilde{f}_t}\\ &\times\left\{\frac{(1-\xi_T)^2\nabla_{\theta_j}\nabla_{\delta_{1i_4}}\nabla_{\delta_{1i_5}}\widetilde{f}_{1t}}{\xi_T\widetilde{f}_t} + \frac{\xi_T\nabla_{\theta_j}\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\begin{bmatrix}\nabla_{\delta_{1i_4}}\nabla_{\delta_{1i_5}}\hat{\beta}(\widetilde{\delta})\\ \xi_T\nabla_{\delta_{1i_4}}\nabla_{\delta_{1i_5}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}\right\}\\ &+(\xi_T - \xi_T^2)\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_1}}\nabla_{\delta_{2i_2}}\nabla_{\delta_{2i_3}}\widetilde{B}_t}{\widetilde{f}_t}\frac{\nabla_{\theta_j}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\nabla_{\delta_{1i_4}}\nabla_{\delta_{1i_5}}\hat{\delta}_1(\widetilde{\delta}) + O_p(|\rho_T|),\end{aligned}$$

where the first term is $O_p(T^{-1/2}+|\rho_T|+(p_T-q_T)^2)$ by Lemma A.13 and (A.119), and the second term is $O_p(T^{-1/2}+\rho_T^2)$ by (A.89). Thus, the elements of the fourth subset are $O_p(T^{-1/2}+|\rho_T|+(p_T-q_T)^2)$. The elements of the fifth subset are $O_p(T^{-1/2} + |\rho_T|)$ by (A.87) and (A.96).

The sixth subset comprises only one component, i.e., the left hand side of the equation in the Lemma. Because the elements in the first five subsets are $O_p(T^{-1/2}+|\rho_T|+(p_T - q_T)^2)$, this element must be of the same order, otherwise $\mathcal{M}^{(6)}_{ji_1...i_5}(p, q, \widetilde{\delta})$ will be nonzero with positive probability. ■

Lemma A.19 *Suppose that the null hypothesis and Assumptions 1-4 and 7 hold with (24) satisfied for all* $i_1, i_2 \in \{1, ..., n_\delta\}$*, and that* $\epsilon \le p_T, q_T \le 1-\epsilon$ *is satisfied for all* T*. Then, for any* $j \in$

$\{1,...,n_\beta+n_\delta\}$ *and any* $i_1,...,i_8\in\{1,...,n_\delta\}$,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_3}}...\nabla_{\delta_{2i_8}}\widetilde{B}_t}{\widetilde{f}_t}\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha_{i_1i_2}\\
= \;& -\left(\frac{1-\xi_T}{\xi_T}\right)\frac{1}{T}\sum_{(j_3,...,j_8)\in IND_4}\sum_{t=1}^{T}\left\{\widetilde{k}_{j_5...j_8,t}-\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{1}{T}\widetilde{I}^{-1}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{k}_{j_5...j_8,t}\right)\right\}\\
&\times\left\{\frac{\nabla_{\delta_{1j_3}}\nabla_{\delta_{1j_4}}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(1)}_{i_1i_2}+\left(\frac{1-\xi_T}{\xi_T}\right)\frac{\nabla_{\delta_{1j_3}}\nabla_{\delta_{1j_4}}\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(2)}_{i_1i_2}-\alpha^{(1)\prime}_{i_1i_2}\frac{\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(1)}_{j_3j_4}\right.\\
&\left.-\alpha^{(1)}_{i_1i_2}\frac{\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(2)}_{j_3j_4}-\alpha^{(2)\prime}_{i_1i_2}\frac{\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}}{\widetilde{f}_t}\alpha^{(1)}_{j_3j_4}\right\}+O_p\left(T^{-1/2}+|\rho_T|+|p_T-q_T|\right).
\end{aligned}$$

where the elements of IND_4 *are as follows: each pair* (j_3,j_4) *corresponds to one of the 15 outcomes when taking two elements without replacement from* $\{i_3,...,i_8\}$ *(the ordering does not matter), and* $j_5,...,j_8$ *correspond to the remaining four elements.*

Proof of Lemma A.19. We divide the components of $\mathcal{M}^{(7)}_{ji_3...i_8}(p,q,\widetilde{\delta})$ into 7 subsets as explained in Remark 8. The proof of Lemma A.18 implies that those in the first, second, third, and sixth subsets are $O_p\left(T^{-1/2}+|\rho_T|+|p_T-q_T|\right)$. Thus, to have $\mathcal{M}^{(7)}_{ji_3...i_8}(p,q,\widetilde{\delta})=0$, the elements in the fourth and seventh subsets must jointly satisfy

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_3}}...\nabla_{\delta_{2i_8}}\widetilde{B}_t}{\widetilde{B}_t}\frac{\widetilde{M}_{jt}}{\widetilde{B}_t} \qquad\qquad (A.127)\\
= \;& -\frac{1}{T}\sum_{(j_3,...,j_8)\in IND_4}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2j_5}}...\nabla_{\delta_{2j_8}}\widetilde{B}_t}{\widetilde{B}_t}\frac{\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\widetilde{M}_{jt}}{\widetilde{B}_t}+O_p\left(T^{-1/2}+|\rho_T|+|(p_T-q_T)|\right),
\end{aligned}$$

where IND_4 is defined in the Lemma.

By (A.121), $(\nabla_{\delta_{2j_5}}...\nabla_{\delta_{2j_8}}\widetilde{B}_t)/\widetilde{f}_t=\widetilde{k}_{j_5...j_8,t}-(\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}/\widetilde{f}_t)\{T^{-1}\widetilde{I}^{-1}\sum_{t=1}^{T}(\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}\widetilde{k}_{j_5...j_8,t})/\widetilde{f}_t\}+\left(T^{-1/2}+|\rho_T|\right)R_t$, where the effect of $\left(T^{-1/2}+|\rho_T|\right)R_t$ on (A.127) is $O_p\left(T^{-1/2}+|\rho_T|\right)$. By (A.102),

$$\begin{aligned}
&\begin{bmatrix}\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\widetilde{M}_{1t}\\ ...\\ \nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\widetilde{M}_{(n_\beta+n_\delta)t}\end{bmatrix}\\
= \;&\begin{bmatrix}\frac{1-\xi_T}{\xi_T}\nabla_{\delta_{1j_3}}\nabla_{\delta_{1j_4}}\nabla_\beta\widetilde{f}_{1t}+\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\hat{\beta}(\widetilde{\delta})+\xi_T\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\hat{\delta}_1(\widetilde{\delta})\\ \frac{(1-\xi_T)^2}{\xi_T}\nabla_{\delta_{1j_3}}\nabla_{\delta_{1j_4}}\nabla_{\delta_1}\widetilde{f}_{1t}+\xi_T\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\hat{\beta}(\widetilde{\delta})+\xi_T\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\hat{\delta}_1(\widetilde{\delta})\end{bmatrix}+R_t,
\end{aligned}$$

where the effect of R_t on (A.127) is $O_p\left(T^{-1/2}+|\rho_T|\right)$. By (A.96), the effect of $\xi_T\nabla_{\delta_1}\nabla_{\delta_1'}\widetilde{f}_{1t}\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\hat{\delta}_1(\widetilde{\delta})$ on (A.127) is $O_p\left(T^{-1/2}+|\rho_T|\right)$. Therefore, this term can be omitted in the subsequent analysis.

Applying these approximations of $(\nabla_{\delta_{2j_5}}...\nabla_{\delta_{2j_8}}\widetilde{B}_t)/\widetilde{f}_t$ and $\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\widetilde{M}_{jt}$ to (A.127), we obtain

$$\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_3}}...\nabla_{\delta_{2i_8}}\widetilde{B}_t}{\widetilde{f}_t^2}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t} \\
= \; & T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\delta_{2i_3}}...\nabla_{\delta_{2i_8}}\widetilde{B}_t}{\widetilde{f}_t^2}\left[\begin{array}{c}\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\widetilde{M}_{1t} \\ ... \\ \frac{1}{\xi_T}\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\widetilde{M}_{(n_\beta+n_\delta)t}\end{array}\right] \\
= \; & -T^{-1}\sum_{(j_3,...,j_8)\in IND_4}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\left\{\widetilde{k}_{j_5...j_8,t}-\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{1}{T}\widetilde{I}^{-1}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{k}_{j_5...j_8,t}\right)\right\} \\
& \times\left[\begin{array}{c}\left(\frac{1-\xi_T}{\xi_T}\right)\nabla_{\delta_{1j_3}}\nabla_{\delta_{1j_4}}\nabla_\beta\widetilde{f}_{1t}+\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\hat{\beta}(\widetilde{\delta})+\xi_T\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\hat{\delta}_1(\widetilde{\delta}) \\ \left(\frac{1-\xi_T}{\xi_T}\right)^2\nabla_{\delta_{1j_3}}\nabla_{\delta_{1j_4}}\nabla_{\delta_1}\widetilde{f}_{1t}+\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}\nabla_{\delta_{2j_3}}\nabla_{\delta_{2j_4}}\hat{\beta}(\widetilde{\delta})\end{array}\right] \\
& +O_p\left(T^{-1/2}+|\rho_T|+|p_T-q_T|\right).
\end{aligned}$$

By Lemma A.13, the right hand side can be further written as

$$\begin{aligned}
& -T^{-1}\sum_{(j_3,...,j_8)\in IND_4}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\left\{\widetilde{k}_{j_5...j_8,t}-\frac{\nabla_{(\beta',\delta_1')}\widetilde{f}_{1t}}{\widetilde{f}_t}\left(\frac{1}{T}\widetilde{I}^{-1}\sum_{t=1}^{T}\frac{\nabla_{(\beta',\delta_1')'}\widetilde{f}_{1t}}{\widetilde{f}_t}\widetilde{k}_{j_5...j_8,t}\right)\right\} \\
& \times\left[\begin{array}{c}\left(\frac{1-\xi_T}{\xi_T}\right)\nabla_{\delta_{1j_3}}\nabla_{\delta_{1j_4}}\nabla_\beta\widetilde{f}_{1t}-\frac{1-\xi_T}{\xi_T}\nabla_\beta\nabla_{\beta'}\widetilde{f}_{1t}\alpha_{j_3j_4}^{(1)}-\frac{1-\xi_T}{\xi_T}\nabla_\beta\nabla_{\delta_1'}\widetilde{f}_{1t}\alpha_{j_3j_4}^{(2)} \\ \left(\frac{1-\xi_T}{\xi_T}\right)^2\nabla_{\delta_{1j_3}}\nabla_{\delta_{1j_4}}\nabla_{\delta_1}\widetilde{f}_{1t}-\frac{1-\xi_T}{\xi_T}\nabla_{\delta_1}\nabla_{\beta'}\widetilde{f}_{1t}\alpha_{j_3j_4}^{(1)}\end{array}\right] \\
& +O_p\left(T^{-1/2}+|\rho_T|+|p_T-q_T|\right).
\end{aligned}$$

The result in the Lemma follows after multiplying this expression by $\alpha_{i_1i_2}'$.■

Proof of Proposition 2. The proofs of the first two results of the Proposition are similar. They are both based on Lemma A.17 and the continuous mapping theorem. Proving the third result is essentially the same as proving that a bootstrap procedure is weakly consistent.

Case 1: SEQ1 with $a < 1/6$. We first establish the convergence rate of the MLE $\hat{\delta}_2$. Then, we apply this result to eliminate the terms in the Taylor expansion that are asymptotically negligible.

By the definition of SEQ1, there exits an $\varepsilon > 0$, such that $|p_T - q_T| > \varepsilon$ for sufficiently large T. As a result, $\min_{x\in R^{n_\delta},\|x^{\otimes 3}\|=1}\left(x^{\otimes 3}\right)'\Omega^{(3)}(p_T,q_T)\left(x^{\otimes 3}\right) > L$ for some $L > 0$. Given this property and the boundedness of $T^{-1}\mathcal{L}_{i_1,...,i_7}^{(7)}(p_T,q_T,\delta)$, $\hat{\delta}_2$ must satisfy

$$\hat{\delta}_2 - \delta_* = O_p(T^{-1/6}); \tag{A.128}$$

otherwise the log likelihood ratio will be negative with probability close to one in large samples. The detailed proof for (A.128) is the essentially same as that of Lemma A.6, thus it is omitted.

Given (A.128), we can restrict our attention to the following set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \leq CT^{-1/6}\}$, where C is a constant that can be made sufficiently large. Uniformly over this set,

$$\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) = \sum_{k=2}^{6} \frac{1}{k!} \sum_{i_1,\ldots,i_k=1}^{n_\delta} \mathcal{L}^{(k)}_{i_1\ldots i_k}(p_T, q_T, \widetilde{\delta}) d_{i_1} \ldots d_{i_k} + o_p(1), \tag{A.129}$$

where $d_{i_j} = \delta_{2i_j} - \widetilde{\delta}_{i_j}$ for $j = 1, .., k$. Now, suppose the convergence rate of $\hat{\delta}_2$ is T^{-b} for some $b \geq 1/6$. Then, $|d_{i_1} \ldots d_{i_k}| = O_p(T^{-kb})$. By Lemma A.17, the terms in (A.129) satisfy

$$\begin{aligned}
T^{-2b}\mathcal{L}^{(2)}_{i_1 i_2}(p_T, q_T, \widetilde{\delta}) &= O_p(T^{\frac{1}{2}-2b}|\rho_T|), \\
T^{-3b}\mathcal{L}^{(3)}_{i_1 i_2 i_3}(p_T, q_T, \widetilde{\delta}) &= O_p(1), \\
T^{-4b}\mathcal{L}^{(4)}_{i_1\ldots i_4}(p_T, q_T, \widetilde{\delta}) &= -(\rho_T^2 T^{1-4b})(\rho_T^{-2}\widetilde{\omega}^{(2)}_{i_1\ldots i_4}(p_T, q_T)) + O_p\left(T^{-1/6}\right), \\
T^{-5b}\mathcal{L}^{(5)}_{i_1\ldots i_5}(p_T, q_T, \widetilde{\delta}) &= O_p\left(\rho_T^2 T^{1-5b}\right) + o_p(1), \\
T^{-6b}\mathcal{L}^{(6)}_{i_1\ldots i_6}(p_T, q_T, \widetilde{\delta}) &= O_p(1).
\end{aligned}$$

As a result, $T^{-4b}\mathcal{L}^{(4)}_{i_1\ldots i_4}(p_T, q_T, \widetilde{\delta})$ diverges and dominates the other terms unless $\rho_T^2 T^{1-4b} = O_p(1)$, or equivalently $T^{-b} = O_p\left(T^{-1/4}|\rho_T|^{-1/2}\right)$. Therefore, we must have $\hat{\delta}_2 - \delta_* = O_p\left(T^{-1/4}|\rho_T|^{-1/2}\right)$; otherwise the terms depending on the matrices $\rho_T^{-2}\widetilde{\Omega}(p_T, q_T)$ and $\widetilde{\Omega}^{(3)}(p_T, q_T)$ will dominate the remaining terms in (A.129), and $\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta})$ will be negative with probability close to one in large samples. This proves that the convergence rate of $\hat{\delta}_2$ is $T^{-1/4}|\rho_T|^{-1/2} = T^{-1/4+a/2}$.

Given this convergence rate, we can further restrict our attention to the set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \leq CT^{-1/4}|\rho_T|^{-1/2}\}$, where C is a constant that can be made sufficiently large. Uniformly over this set, by Lemma A.17,

$$\begin{aligned}
\mathcal{L}^{(2)}_{i_1 i_2}(p_T, q_T, \widetilde{\delta}) d_{i_1} d_{i_2} &= \sum_{t=1}^{T} \widetilde{U}_{i_1 i_2, t}(p_T, q_T) d_{i_1} d_{i_2} + o_p(1), \qquad \text{(A.130)} \\
\mathcal{L}^{(4)}_{i_1\ldots i_4}(p_T, q_T, \widetilde{\delta}) d_{i_1} \ldots d_{i_4} &= -T\widetilde{\omega}^{(2)}_{i_1\ldots i_4}(p_T, q_T) d_{i_1} \ldots d_{i_4} + o_p(1), \\
\mathcal{L}^{(k)}_{i_1\ldots i_k}(p_T, q_T, \widetilde{\delta}) d_{i_1} \ldots d_{i_k} &= o_p(1) \quad \text{for } k = 3, 5, 6,
\end{aligned}$$

where the leading terms in the first two equations are $O_p(1)$ and not $o_p(1)$. This result determines the leading terms in the expansion (A.129), implying

$$\begin{aligned}
\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) &= (1/2)\sum_{t=1}^{T} \sum_{i_1, i_2=1}^{n_\delta} \widetilde{U}_{i_1 i_2, t}(p_T, q_T) d_{i_1} d_{i_2} \\
&\quad -(1/4!)T \sum_{i_1,\ldots,i_4=1}^{n_\delta} \widetilde{\omega}^{(2)}_{i_1\ldots i_4}(p_T, q_T) d_{i_1} \ldots d_{i_4} + o_p(1)
\end{aligned}$$

over $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \leq CT^{-1/4}|\rho_T|^{-1/2}\}$. Let $\eta_i = T^{1/4} d_i$, and apply the definition of $\widetilde{U}^{(2)}_t(p_T, q_T)$, $\widetilde{\Omega}(p_T, q_T)$ and η to the preceding displayed equation, we have

$$\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) = \frac{1}{2}\left(\eta^{\otimes 2}\right)' \left(T^{-1/2}\sum_{t=1}^{T} \widetilde{U}^{(2)}_t(p_T, q_T)\right) - \frac{1}{8}\left(\eta^{\otimes 2}\right)' \widetilde{\Omega}(p_T, q_T)\left(\eta^{\otimes 2}\right) + o_p(1).$$

To obtain the result in the lemma, we multiply both sides by 2 and compute the supremum over η. The effect of the restriction $\|\delta_2 - \delta_*\| \le CT^{-1/4}|\rho_T|^{-1/2}$ on the approximation can be studied in the same way as in Step 2 of the proof of Proposition 1. We omit the details.

Case 2: SEQ1 with $a = 1/6$**.** The results up to (A.129) in the proof of the $a < 1/6$ case still hold. In particular, by (A.128), the convergence rate of $\hat{\delta}_2$ is $T^{-1/6}$. Similarly to (A.130), uniformly over $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \le CT^{-1/6}\}$ with C being sufficiently large, we have

$$\begin{aligned}
\mathcal{L}^{(2)}_{i_1 i_2}(p_T, q_T, \widetilde{\delta}) d_{i_1} d_{i_2} &= \sum_{t=1}^{T} \widetilde{U}_{i_1 i_2, t}(p_T, q_T) d_{i_1} d_{i_2} + o_p(1), \\
\mathcal{L}^{(3)}_{i_1 \dots i_3}(p_T, q_T, \widetilde{\delta}) d_{i_1} \dots d_{i_3} &= \sum_{t=1}^{T} \widetilde{s}_{i_1 \dots i_3, t}(p_T, q_T) d_{i_1} \dots d_{i_3} + o_p(1), \\
\mathcal{L}^{(4)}_{i_1 \dots i_4}(p_T, q_T, \widetilde{\delta}) d_{i_1} \dots d_{i_4} &= -T \widetilde{\omega}^{(2)}_{i_1 \dots i_4}(p_T, q_T) d_{i_1} \dots d_{i_4} + o_p(1), \\
\mathcal{L}^{(6)}_{i_1 \dots i_6}(p_T, q_T, \widetilde{\delta}) d_{i_1} \dots d_{i_6} &= -T \sum_{(j_1, \dots, j_6) \in I_1(i_1, \dots, i_6)} \omega^{(3)}_{j_1 \dots j_6}(p_T, q_T) d_{i_1} \dots d_{i_6} + o_p(1), \\
\mathcal{L}^{(5)}_{i_1 \dots i_5}(p_T, q_T, \widetilde{\delta}) d_{i_1} \dots d_{i_5} &= o_p(1),
\end{aligned}$$

where $I_1(i_1, \dots, i_6)$ equals IND_1 with (j, k, l, m, n, r) replaced by $(i_1, \dots, i_6)$, and the leading terms in the first four equations are $O_p(1)$ and not $o_p(1)$ if $d_{i_1}, \dots, d_{i_k}$ are $O(T^{-1/6})$. These results determine the leading terms in the expansion (A.129), implying

$$\begin{aligned}
& \mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) \\
= & \frac{1}{2} \sum_{t=1}^{T} \sum_{i_1, i_2 = 1}^{n_\delta} \widetilde{U}_{i_1 i_2, t}(p_T, q_T) d_{i_1} d_{i_2} + \frac{1}{3!} \sum_{t=1}^{T} \sum_{i_1, \dots, i_3 = 1}^{n_\delta} \widetilde{s}_{i_1 \dots i_3, t}(p_T, q_T) d_{i_1} \dots d_{i_3} \\
& - \frac{1}{4!} T \sum_{i_1, \dots, i_4 = 1}^{n_\delta} \widetilde{\omega}^{(2)}_{i_1 \dots i_4}(p_T, q_T) d_{i_1} \dots d_{i_4} - \frac{1}{6!} T \sum_{i_1, \dots, i_6 = 1}^{n_\delta} \sum_{(j_1, \dots, j_6) \in I_1(i_1, \dots, i_6)} \omega^{(3)}_{i_1 \dots i_6}(p_T, q_T) d_{i_1} \dots d_{i_6} + o_p(1)
\end{aligned}$$

over the set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \le CT^{-1/6}\}$. Let $\eta_i = T^{1/4} d_i$, and then apply the definitions of $\widetilde{U}^{(2)}_t(p_T, q_T)$, $\widetilde{U}^{(3)}_t(p_T, q_T)$, $\widetilde{\Omega}(p_T, q_T)$, $\widetilde{\Omega}^{(3)}(p_T, q_T)$, and η to the preceding expression. We have

$$\begin{aligned}
& \mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) \\
= & \frac{1}{2} \left(\eta^{\otimes 2}\right)' \left(T^{-1/2} \sum_{t=1}^{T} \widetilde{U}^{(2)}_t(p_T, q_T) \right) + T^{-1/4} \frac{1}{6} \left(\eta^{\otimes 3}\right)' \left(T^{-1/2} \sum_{t=1}^{T} \widetilde{U}^{(3)}_t(p_T, q_T) \right) \\
& - \frac{1}{8} \left(\eta^{\otimes 2}\right)' \widetilde{\Omega}(p_T, q_T) \left(\eta^{\otimes 2}\right) - \frac{1}{72} T^{-1/2} \left(\eta^{\otimes 3}\right)' \widetilde{\Omega}^{(3)}(p_T, q_T) \left(\eta^{\otimes 3}\right) + o_p(1).
\end{aligned}$$

The result follows after multiplying both sides by 2 and computing the supremum over η.

Case 3: SEQ1 with $a > 1/6$**.** The results up to (A.129) in the proof of the $a < 1/6$ case still hold. Further, the second and fourth order derivative terms in (A.129) are now negligible.

In addition, uniformly over the set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \leq CT^{-1/6}\}$, the third and sixth order derivative terms satisfy

$$\begin{aligned} \mathcal{L}^{(3)}_{i_1...i_3}(p_T, q_T, \widetilde{\delta})d_{i_1}...d_{i_3} &= \sum_{t=1}^{T} \widetilde{s}_{i_1...i_3,t}(p_T, q_T)d_{i_1}...d_{i_3} + o_p(1), \\ \mathcal{L}^{(6)}_{i_1,...i_6}(p_T, q_T, \widetilde{\delta})d_{i_1}...d_{i_6} &= -T \sum_{(j_1,...,j_6)\in I_1(i_1,...,i_6)} \omega^{(3)}_{j_1...j_6}(p_T, q_T)d_{i_1}...d_{i_6} + o_p(1). \end{aligned}$$

These results imply

$$\begin{aligned} & \mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) \\ = & \frac{1}{3!}\sum_{t=1}^{T}\sum_{i_1,...,i_3=1}^{n_\delta} \widetilde{s}_{i_1...i_3,t}(p_T, q_T)d_{i_1}...d_{i_3} - \frac{1}{6!}T\sum_{i_1,...,i_6=1}^{n_\delta}\sum_{(j_1,...,j_6)\in I_1(i_1,...,i_6)} \omega^{(3)}_{j_1...j_6}(p_T, q_T)d_{i_1}...d_{i_6} + o_p(1) \end{aligned}$$

uniformly over $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \leq CT^{-1/6}\}$. Let $\eta_i = T^{1/4}d_i$. The result follows from applying the definition of $\widetilde{U}^{(3)}_t(p_T, q_T)$, $\widetilde{\Omega}^{(3)}(p_T, q_T)$ and η, multiplying both sides by 2, and then computing the supremum over η. Note that the convergence rate of $\hat{\delta}_2$ is $T^{-1/6}$, independent of a.

Case 4: SEQ2 with $a < 1/4$. The proof is similar to but more complex than the $a < 1/6$ case because we need to consider an eighth order expansion.

We first establish the convergence rate of $\hat{\delta}_2$. Because $\min_{x\in R^{n_\delta},\|x^{\otimes 4}\|=1} \left(x^{\otimes 4}\right)' \widetilde{\Omega}^{(4)}(p_T, q_T) \left(x^{\otimes 4}\right)$ is positive and $T^{-1}\mathcal{L}^{(9)}_{i_1,...,i_9}(p_T, q_T, \delta)$ is bounded in probability, $\hat{\delta}_2$ must satisfy $\hat{\delta}_2 - \delta_* = O_p(T^{-1/8})$. Thus, we can restrict the attention to the set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \leq CT^{-1/8}\}$, where C can be made sufficiently large. Over this set, the following expansion holds:

$$\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) = \sum_{k=2}^{8}\frac{1}{k!}\sum_{i_1,...,i_k=1}^{n_\delta} \mathcal{L}^{(k)}_{i_1...i_k}(p_T, q_T, \widetilde{\delta})d_{i_1}...d_{i_k} + o_p(1). \quad \text{(A.131)}$$

By Lemma A.17, for any $b \geq 1/8$, the second and fourth order derivatives satisfy

$$T^{-2b}\mathcal{L}^{(2)}_{i_1 i_2}(p_T, q_T, \widetilde{\delta}) = O_p\left(\sqrt{T^{1-4b}\rho_T^2}\right),$$

and

$$\begin{aligned} T^{-4b}\mathcal{L}^{(4)}_{i_1...i_4}(p_T, q_T, \widetilde{\delta}) &= T^{-4b}\sum_{t=1}^{T} \widetilde{k}_{i_1...i_4,t}(p_T, q_T) - \left(T^{1-4b}\rho_T^2\right)(\rho_T^{-2}\widetilde{\omega}^{(2)}_{i_1...i_4}(p_T, q_T)) + o_p(1) \\ &= -\left(T^{1-4b}\rho_T^2\right)(\rho_T^{-2}\widetilde{\omega}^{(2)}_{i_1...i_4}(p_T, q_T)) + O_p(1) = O_p(T^{1-4b}\rho_T^2 + 1). \end{aligned}$$

As a result,$T^{-4b}\mathcal{L}^{(4)}_{i_1...i_4}(p_T, q_T, \widetilde{\delta})$ diverges unless $T^{1-4b}\rho_T^2 = O(1)$, or equivalently $T^{-b} = O_p\left(T^{-1/4}|\rho_T|^{-1/2}\right)$.

Meanwhile, by Lemma A.17, the remaining four terms in (A.131) satisfy

$$
\begin{aligned}
T^{-3b}\mathcal{L}^{(3)}_{i_1\ldots i_3}(p_T,q_T,\widetilde{\delta}) &= O_p\left(\sqrt{T^{1-4b}\rho_T^2}T^{-b}\right) = o_p\left(\sqrt{T^{1-4b}\rho_T^2}\right), \qquad (A.132)\\
T^{-5b}\mathcal{L}^{(5)}_{i_1\ldots i_5}(p_T,q_T,\widetilde{\delta}) &= O_p\left((T^{1-4b}\rho_T^2)T^{-b}\right) = o_p\left(T^{1-4b}\rho_T^2\right),\\
T^{-6b}\mathcal{L}^{(6)}_{i_1\ldots i_6}(p_T,q_T,\widetilde{\delta}) &= O_p\left((T^{1-4b}\rho_T^2)T^{-2b}\right) = o_p\left(T^{1-4b}\rho_T^2\right),\\
T^{-7b}\mathcal{L}^{(7)}_{i_1\ldots i_7}(p_T,q_T,\widetilde{\delta}) &= O_p((T^{1-4b}\rho_T^2)T^{-3b}\left|\rho_T\right|^{-1}) = o_p\left(T^{1-4b}\rho_T^2\right).
\end{aligned}
$$

These four terms are all dominated by the fourth derivative term unless $T^{1-4b}\rho_T^2 = O_p(1)$. Therefore, we must have $\hat{\delta}_2 - \delta_* = O_p\left(T^{-1/4}|\rho_T|^{-1/2}\right)$; otherwise the the matrices $\rho_T^{-2}\widetilde{\Omega}(p_T,q_T)$ and $\widetilde{\Omega}^{(4)}(p_T,q_T)$ will dominate the rest and $\mathcal{L}(p_T,q_T,\delta_2)-\mathcal{L}(p_T,q_T,\widetilde{\delta})$ will be negative with probability close to one in large samples.

Given this convergence rate, the rest of the proof is the same as the $a<1/6$ case, starting with (A.130). This is because only the second and fourth derivative terms are non-negligible when $\hat{\delta}_2 - \delta_* = O_p\left(T^{-1/4}|\rho_T|^{-1/2}\right)$ in light of (A.132).

Case 5: SEQ2 with $a = 1/4$. The proof is similar to the $a = 1/6$ case. We can restrict attention to the set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \le CT^{-1/8}\}$ with C being sufficiently large. The effects of the third, fifth, sixth and seventh order derivative terms on (A.131) are $o_p(1)$ in light of (A.132). For the remaining terms, uniformly over this set,

$$
\begin{aligned}
\mathcal{L}^{(2)}_{i_1 i_2}(p_T,q_T,\widetilde{\delta})d_{i_1}d_{i_2} &= \sum_{t=1}^{T}\widetilde{U}_{i_1 i_2,t}(p_T,q_T)d_{i_1}d_{i_2} + o_p(1),\\
\mathcal{L}^{(4)}_{i_1\ldots i_4}(p_T,q_T,\widetilde{\delta})d_{i_1}\ldots d_{i_4} &= \sum_{t=1}^{T}\widetilde{k}_{i_1\ldots i_4,t}(p_T,q_T)d_{i_1}\ldots d_{i_4} - T\widetilde{\omega}^{(2)}_{i_1\ldots i_4}(p_T,q_T)d_{i_1}\ldots d_{i_4} + o_p(1),\\
\mathcal{L}^{(8)}_{i_1\ldots i_8}(p_T,q_T,\widetilde{\delta})d_{i_1}\ldots d_{i_8} &= -T\sum_{(j_1,\ldots,j_8)\in I_2(i_1,\ldots,i_8)}\omega^{(4)}_{j_1\ldots j_8}(p_T,q_T)d_{i_1}\ldots d_{i_8} + o_p(1),
\end{aligned}
$$

where $I_2(i_1,\ldots,i_8)$ equals IND_2 with (j,k,l,m,n,r,s,u) replaced by $(i_1,\ldots,i_8)$. This result determines the leading terms in the expansion (A.131), implying

$$
\begin{aligned}
&\mathcal{L}(p_T,q_T,\delta_2) - \mathcal{L}(p_T,q_T,\widetilde{\delta})\\
=\ &\frac{1}{2}\sum_{t=1}^{T}\sum_{i_1,i_2=1}^{n_\delta}\widetilde{U}_{i_1 i_2,t}(p_T,q_T)d_{i_1}d_{i_2} + \frac{1}{4!}\sum_{t=1}^{T}\sum_{i_1,\ldots,i_3=1}^{n_\delta}\widetilde{k}_{i_1\ldots i_4,t}(p_T,q_T)d_{i_1}\ldots d_{i_3}\\
&-\frac{1}{4!}T\sum_{i_1,\ldots,i_4=1}^{n_\delta}\widetilde{\omega}^{(2)}_{i_1\ldots i_4}(p_T,q_T)d_{i_1}\ldots d_{i_4} - \frac{1}{8!}T\sum_{i_1,\ldots,i_8=1}^{n_\delta}\sum_{(j_1,\ldots,j_8)\in I_2(i_1,\ldots,i_8)}\omega^{(4)}_{i_1\ldots i_8}(p_T,q_T)d_{i_1}\ldots d_{i_8} + o_p(1).
\end{aligned}
$$

The leading terms in these three equations are $O_p(1)$ and not $o_p(1)$ uniformly over the set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \le CT^{-1/8}\}$. Let $\eta_i = T^{1/4}d_i$. The result follows after applying the definition of $\widetilde{U}_t^{(2)}(p_T,q_T)$, $\widetilde{U}_t^{(4)}(p_T,q_T)$, $\widetilde{\Omega}(p_T,q_T)$, $\widetilde{\Omega}^{(4)}(p_T,q_T)$ and η, multiplying both sides by 2, and then computing the supremum over η.

Case 6: SEQ2 with $a > 1/4$. The proof is very similar to the $a > 1/6$ case. We can restrict attention to the set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \leq CT^{-1/8}\}$ with C being sufficiently large. Within this set, by Lemma A.17, the second, third, fifth, sixth, and seventh order derivative terms in (A.131) are all negligible. The fourth and eighth order derivatives satisfy

$$\begin{aligned}
\mathcal{L}^{(4)}_{i_1 \dots i_4}(p_T, q_T, \widetilde{\delta}) d_{i_1} \dots d_{i_4} &= \sum_{t=1}^{T} \widetilde{k}_{i_1 \dots i_4, t}(p_T, q_T) d_{i_1} \dots d_{i_4} + o_p(1), \\
\mathcal{L}^{(8)}_{i_1 \dots i_8}(p_T, q_T, \widetilde{\delta}) d_{i_1} \dots d_{i_8} &= -T \sum_{(j_1, \dots, j_8) \in I_2(i_1, \dots, i_8)} \omega^{(4)}_{j_1 \dots j_8}(p_T, q_T) d_{i_1} \dots d_{i_8} + o_p(1),
\end{aligned}$$

where the two leading terms are $O_p(1)$ and not $o_p(1)$. Thus,

$$\begin{aligned}
&\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) \\
= \; &\frac{1}{4!} \sum_{t=1}^{T} \sum_{i_1, \dots, i_3 = 1}^{n_\delta} \widetilde{k}_{jklm,t}(p_T, q_T) d_{i_1} \dots d_{i_3} - \frac{1}{8!} T \sum_{i_1, \dots, i_8 = 1}^{n_\delta} \sum_{(j_1, \dots, j_8) \in I_2(i_1, \dots, i_8)} \omega^{(4)}_{j_1 \dots j_8}(p_T, q_T) d_{i_1} \dots d_{i_8} + o_p(1)
\end{aligned}$$

uniformly over $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \leq CT^{-1/8}\}$. Let $\eta_i = T^{1/4} d_i$. The result follows after applying the definition of $\widetilde{U}^{(4)}_t(p_T, q_T), \widetilde{\Omega}^{(4)}(p_T, q_T)$ and η, multiplying both sides by 2, and then computing the supremum over η.

Proof of the weak consistency. We focus on the case with $a < 1/6$ under SEQ1. It will become clear that the other cases can be proved in the same way.

Recall, by definition,

$$\widetilde{\mathcal{W}}^{(2)}(p_T, q_T, \eta) = \left(\eta^{\otimes 2}\right)' \left(T^{-1/2} \sum_{t=1}^{T} \widetilde{U}^{(2)}_t(p_T, q_T) \right) - \frac{1}{4} \left(\eta^{\otimes 2}\right)' \widetilde{\Omega}(p_T, q_T) \left(\eta^{\otimes 2}\right).$$

Let $h = |\rho_T|^{1/2} \eta$. We can write it as

$$\widetilde{\mathcal{W}}^{(2)}(p_T, q_T, h) = \left(h^{\otimes 2}\right)' \left(T^{-1/2} |\rho_T|^{-1} \sum_{t=1}^{T} \widetilde{U}^{(2)}_t(p_T, q_T) \right) - \frac{1}{4} \left(h^{\otimes 2}\right)' \left(\rho_T^{-2} \widetilde{\Omega}(p_T, q_T) \right) \left(h^{\otimes 2}\right).$$

Because $\eta = T^{-1/4}(\delta_2 - \widetilde{\delta}) = O_p\left(|\rho_T|^{-1/2}\right)$, we can focus on the set $\{h \in R^{n_\delta} : (\|h\| \leq C\}$, where C can be made sufficiently large.

The following results hold (uniformly over the set $\{h \in R^{n_\delta} : (\|h\| \leq C\}$ when applicable):

(i) $\widetilde{\mathcal{W}}^{(2)}(p_T, q_T, h)$ converges weakly to a continuous process over the compact set $\{h \in R^{n_\delta} : (\|h\| \leq C\}$. This follows because $T^{-1/2} |\rho_T|^{-1} \sum_{t=1}^{T} \widetilde{U}^{(2)}_t(p_T, q_T)$ converges to a Normal random vector, $\rho_T^{-2} \widetilde{\Omega}(_T, q_T)$ converges in probability to a nonrandom matrix that equals the covariance of this vector, and $\widetilde{\mathcal{W}}^{(2)}(p_T, q_T, h)$ is continuous in h over the compact set.

(ii) The original data sequence $\{(x'_t, y_t)\}_{t=1}^{\infty}$ contains a subsequence, under which $\rho_T^{-2} \widetilde{\Omega}(p_T, q_T)$ converges almost surely to the nonrandom matrix in (i). This follows from the convergence in probability of $\rho_T^{-2} \widetilde{\Omega}(p_T, q_T)$ in (i) and Theorem 20.5(ii) of Billingsley (1986).

(iii) Conditional on any such subsequence in (ii), the simulated process $\mathcal{W}_b^{(2)}(p_T, q_T, h)$ converges weakly to the continuous limiting process in (i). This follows from the definition of $\mathcal{W}_b^{(2)}(p_T, q_T, h)$ and (ii).

(iv) The conditional distribution of the supremum of $\mathcal{W}_b^{(2)}(p_T, q_T, h)$ converges in probability to the supremum of the limiting process in (i), where the conditioning is on the original data sequence $\{(x_t', y_t)\}_{t=1}^{\infty}$. This follows from the continuous mapping theorem, Theorem 20.5(ii) in Billingsley (1986), and (iii).

Note that (iv) implies that the empirical distribution of $S_b(p_T, q_T)$ is a weakly consistent estimator of the limiting distribution of $LR(p_T, q_T)$ under SEQ1 with $a < 1/6$. The arguments in (i)-(iv) are similar to those of Politis and Romano (1994), originally used to prove a bootstrap procedure is weakly consistent. ■

The next result is needed to prove Corollary 2.

Lemma A.20 *Under the Assumptions of Corollary 2, we have* $T^{-1}\mathcal{L}_{jklmnrs}^{(7)}(p_T, q_T, \widetilde{\delta}) = O_p(T^{-1/2} + |\rho_T| + (p_T - q_T)^2)$.

Proof of Lemma A.20. As in Part 6 of the proof of Lemma A.17, we divide the components of $T^{-1}\mathcal{L}_{jklmnrs}^{(7)}(p_T, q_T, \widetilde{\delta}) - ((1-\xi_T)/\xi_T)\,\mathcal{M}_{(n_\beta+j)klmnrs}^{(7)}(p_T, q_T, \widetilde{\delta})$ into 6 subsets. We only need to study the fourth and fifth subsets because Part 6 already shows that the remaining subsets are $O_p(T^{-1/2} + |\rho_T| + (p_T - q_T)^2)$.

For the fourth subset, we apply (A.119) and (A.117). In the expression of $\widetilde{k}_{1,jklm,t}$, the second term in the braces can be rewritten as $((1-\xi_T)/\xi_T)^2\,\nabla_{\delta_{1j}}\nabla_{\delta_{1l}}\nabla_{\delta_{1l}}\nabla_{\delta_{1m}}\widetilde{f}_{1t} + ((2\xi_T - 1)(1-\xi_T)^2/\xi_T^3)\nabla_{\delta_{1j}}\nabla_{\delta_{1l}}\nabla_{\delta_1'}\widetilde{f}_{1t}\alpha_{km}^{(2)}$, and the third and fourth terms can be rewritten in a similar way. From these expressions and the assumption $E(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_3}}f_t(\beta_*, \delta_*)\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f_t(\beta_*, \delta_*)/f_t(\beta_*, \delta_*)^2|\Omega_{t-1}) = 0$ for i_1,...,$i_4 \in \{n_\beta + 1, ..., n_\beta + n_\delta\}$, it follows that the elements of this subset are $O_p(|\rho_T| + (p_T - q_T)^2 + T^{-1/2})$. For the fifth subset, it is sufficient to study

$$((\xi_T - 1)(1 - 2\xi_T)/\xi_T^2)T^{-1}\sum_{t=1}^{T}\nabla_{\delta_{2i_4}}...\nabla_{\delta_{2i_7}}\widetilde{B}_t\nabla_{\delta_{1i_1}}\nabla_{\delta_{1i_2}}\nabla_{\delta_{1i_3}}\widetilde{f}_{1t}/\widetilde{f}_t^2.$$

This is $O_p(|\rho_T| + (p_T - q_T)^2 + T^{-1/2})$ by (A.121). Therefore, all the components are $O_p(|\rho_T| + (p_T - q_T)^2 + T^{-1/2})$.■

Proof of Corollary 2. We partition the sequence (p_T, q_T) into three subsequences and prove that the result in the Corollary holds for each subsequence. The subsequences are constructed based on the convergence rate of ρ_T to zero and the location of (p_T, q_T) relative to $(0.5, 0.5)$. The first subsequence includes all the (p_T, q_T) that satisfy $|\rho_T| \geq T^{-1/8}/\log T$, the second subsequence includes those satisfying $|\rho_T| \leq T^{-1/8}/\log T$ and $|p_T - q_T| \geq T^{-1/8}\log T$, and the third subsequence includes those satisfying $|\rho_T| \leq T^{-1/8}/\log T$ and $|p_T - q_T| \leq T^{-1/8}\log T$. The three subsequences contain all possible realizations of (p_T, q_T). If the result holds for each subsequence, it must also hold for the original sequence.

Because $\hat{\delta}_2$ satisfies

$$\hat{\delta}_2 - \delta_* = O_p(T^{-1/8}), \tag{A.133}$$

we can restrict the attention to the set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \leq CT^{-1/8}\}$, where C can be made sufficiently large. The following expansion holds uniformly over this set

$$\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) = \sum_{k=2}^{8} \frac{1}{k!} \sum_{i_1,\ldots,i_k=1}^{n_\delta} \mathcal{L}^{(k)}_{i_1\ldots i_k}(p_T, q_T, \widetilde{\delta}) d_{i_1} \ldots d_{i_k} + o_p(1). \tag{A.134}$$

We study this expansion under the three subsequences separately.

To study the first sequence, as in the proof of the $a < 1/6$ case in Proposition 2, we suppose that the convergence rate of $\hat{\delta}_2$ is T^{-b} for some $b > 0$, where b satisfies $1/8 \leq b \leq 1/4$ in light of (A.133). Note that b can depend on $|p_T - q_T|$ and $|\rho_T|$. This dependence is suppressed to simplify the expressions. By Lemma A.17, for any $1/8 \leq b \leq 1/4$, we have

$$\begin{aligned}
T^{-2b}\mathcal{L}^{(2)}_{i_1 i_2}(p_T, q_T, \widetilde{\delta}) &= T^{-2b} \sum_{t=1}^{T} \widetilde{U}_{i_1 i_2, t}(p_T, q_T) + o_p(1), \qquad \text{(A.135)}\\
T^{-3b}\mathcal{L}^{(3)}_{i_1\ldots i_3}(p_T, q_T, \widetilde{\delta}) &= T^{-3b} \sum_{t=1}^{T} \widetilde{s}_{i_1\ldots i_3, t}(p_T, q_T) + o_p(\sqrt{T^{1-4b}\rho_T^2}),\\
T^{-4b}\mathcal{L}^{(4)}_{i_1\ldots i_4}(p_T, q_T, \widetilde{\delta}) &= -(T^{1-4b}\rho_T^2)(\rho_T^{-2}\widetilde{\omega}^{(2)}_{i_1\ldots i_4}(p_T, q_T)) + T^{\frac{1}{2}-4b}(T^{-1/2}\sum_{t=1}^{T} \widetilde{k}_{i_1\ldots i_4, t}(p_T, q_T)) + o_p(1),\\
T^{-6b}\mathcal{L}^{(6)}_{i_1\ldots i_6}(p_T, q_T, \widetilde{\delta}) &= -T^{1-6b} \sum_{(j_1,\ldots,j_6)\in I_1(i_1,\ldots,i_6)} \omega^{(3)}_{j_1\ldots j_6}(p_T, q_T) + O_p(T^{1-6b}|\rho_T|) + o_p(1),
\end{aligned}$$

where $I_1(i_1, \ldots, i_6)$ equals IND_1 with (j, k, l, m, n, r) replaced by $(i_1, \ldots, i_6)$, and

$$\begin{aligned}
T^{-5b}\mathcal{L}^{(5)}_{i_1\ldots i_5}(p_T, q_T, \widetilde{\delta}) &= o_p\left(T^{1-4b}\rho_T^2\right), \; T^{-7b}\mathcal{L}^{(7)}_{i_1\ldots i_7}(p_T, q_T, \widetilde{\delta}) = O_p\left(T^{1-7b}\right), \qquad \text{(A.136)}\\
T^{-8b}\mathcal{L}^{(8)}_{i_1\ldots i_8}(p_T, q_T, \widetilde{\delta}) &= O_p\left(T^{1-8b}\right).
\end{aligned}$$

The first leading term of $T^{-4b}\mathcal{L}^{(4)}_{i_1\ldots i_4}(p_T, q_T, \widetilde{\delta})$ dominates its second leading term and (A.136) because $b \geq 1/8$, $|\rho_T| \geq T^{-1/8}/\log T$, and $\min_{\|x^{\otimes 2}\|=1} \rho_T^{-2}(x^{\otimes 2})'\widetilde{\Omega}(p_T, q_T)(x^{\otimes 2}) > L > 0$ in probability. Furthermore, this leading term diverges unless $T^{1-4b}\rho_T^2 = O(1)$. Therefore, we must have $T^{1-4b}\rho_T^2 = O_p(1)$, or $T^{-b} = O_p(T^{-1/4}|\rho_T|^{-1/2})$; otherwise, the log likelihood ratio will be negative with probability close to one in large samples. We conclude that $\hat{\delta}_2 - \delta_* = O_p\left(T^{-1/4}|\rho_T|^{-1/2}\right)$ under the first subsequence. This shows $b = -1/4 - (1/2)\log|\rho_T|$.

Given this convergence rate, we can further focus on the set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \leq$

$CT^{-1/4}|\rho_T|^{-1/2}\}$, where C can be made sufficiently large. Then, over this set,

$$\begin{aligned}
\mathcal{L}^{(2)}_{i_1 i_2}(p_T, q_T, \widetilde{\delta}) d_{i_1} d_{i_2} &= \sum_{t=1}^{T} \widetilde{U}_{i_1 i_2,t}(p_T, q_T) d_{i_1} d_{i_2} + o_p(1), \\
\mathcal{L}^{(3)}_{i_1 \ldots i_3}(p_T, q_T, \widetilde{\delta}) d_{i_1} d_{i_2} d_{i_3} &= \sum_{t=1}^{T} \widetilde{s}_{i_1 \ldots i_3,t}(p_T, q_T) d_{i_1} d_{i_2} d_{i_3} + o_p(1), \\
\mathcal{L}^{(4)}_{i_1 \ldots i_4}(p_T, q_T, \widetilde{\delta}) d_{i_1} \ldots d_{i_4} &= -T \widetilde{\omega}^{(2)}_{i_1 \ldots i_4}(p_T, q_T) d_{i_1} \ldots d_{i_4} + o_p(1), \\
\mathcal{L}^{(6)}_{i_1 \ldots i_6}(p_T, q_T, \widetilde{\delta}) d_{i_1} \ldots d_{i_6} &= -T \sum_{(j_1, \ldots, j_6) \in I_1(i_1, \ldots, i_6)} \omega^{(3)}_{j_1 \ldots j_6}(p_T, q_T) d_{i_1} \ldots d_{i_6} + o_p\left(1\right), \\
\mathcal{L}^{(k)}_{i_1 \ldots i_k}(p_T, q_T, \widetilde{\delta}) d_{i_1} \ldots d_{i_k} &= o_p(1) \text{ for } k = 5, 7, 8.
\end{aligned}$$

Applying this to (A.134), we have, over $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \le CT^{-1/4}|\rho_T|^{-1/2}\}$:

$$\begin{aligned}
&\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) \qquad \text{(A.137)} \\
= \;& \frac{1}{2} \sum_{t=1}^{T} \sum_{i_1, i_2 = 1}^{n_\delta} \widetilde{U}_{i_1 i_2,t}(p_T, q_T) d_{i_1} d_{i_2} + \frac{1}{3!} \sum_{t=1}^{T} \sum_{i_1, \ldots, i_3 = 1}^{n_\delta} \widetilde{s}_{i_1 \ldots i_3,t}(p_T, q_T) d_{i_1} \ldots d_{i_3} \\
& - \frac{1}{4!} T \sum_{i_1, \ldots, i_4 = 1}^{n_\delta} \widetilde{\omega}^{(2)}_{i_1 \ldots i_4}(p_T, q_T) d_{i_1} \ldots d_{i_4} - \frac{1}{6!} T \sum_{i_1, \ldots, i_6 = 1}^{n_\delta} \sum_{(j_1, \ldots, j_6) \in I_1(i_1, \ldots, i_6)} \omega^{(3)}_{j_1 \ldots j_6}(p_T, q_T) d_{i_1} \ldots d_{i_6} + o_p\left(1\right).
\end{aligned}$$

Let $\eta_i = T^{1/4} d_i$, and then apply the definitions of $\widetilde{U}^{(2)}_t(p_T, q_T)$, $\widetilde{U}^{(3)}_t(p_T, q_T)$, $\widetilde{\Omega}(p_T, q_T)$, $\widetilde{\Omega}^{(3)}(p_T, q_T)$, and η to the preceding expression. We have

$$\begin{aligned}
&\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) \\
= \;& \frac{1}{2} \left(\eta^{\otimes 2}\right)' \left(T^{-1/2} \sum_{t=1}^{T} \widetilde{U}^{(2)}_t(p_T, q_T) \right) + T^{-1/4} \frac{1}{6} \left(\eta^{\otimes 3}\right)' \left(T^{-1/2} \sum_{t=1}^{T} \widetilde{U}^{(3)}_t(p_T, q_T) \right) \\
& - \frac{1}{8} \left(\eta^{\otimes 2}\right)' \widetilde{\Omega}(p_T, q_T) \left(\eta^{\otimes 2}\right) - \frac{1}{72} T^{-1/2} \left(\eta^{\otimes 3}\right)' \widetilde{\Omega}^{(3)}(p_T, q_T) \left(\eta^{\otimes 3}\right) + o_p\left(1\right).
\end{aligned}$$

We multiply both sides by 2 and compute the supremum over $\{\eta \in R^{n_\delta} : \|\eta\| \le C|\rho_T|^{-1/2}\}$. Note that the effect of the restriction $\|\eta\| \le C|\rho_T|^{-1/2}$ on the approximation can be studied in the same way as in Step 2 of the proof of Proposition 1. We omit the details. Therefore, the result of the Corollary holds for this subsequence.

We now study the second subsequence. The expansion (A.134) still holds. Without loss of generality, we assume $|\rho_T| \ne 0$. Otherwise, the outcomes with $|\rho_T| = 0$ can be treated as a separate subsequence. As when studying (A.136) and (A.135), we suppose that the convergence rate of $\hat{\delta}_2$ is T^{-b} for some $b > 0$ that potentially depends on $|p_T - q_T|$ and $|\rho_T|$. For any $1/8 \le b \le 1/4$, we have

$$\begin{aligned}
T^{-2b}\mathcal{L}^{(2)}_{i_1 i_2}(p_T,q_T,\widetilde{\delta}) &= T^{-2b}\sum_{t=1}^{T}\widetilde{U}_{i_1 i_2,t}(p_T,q_T)+o_p(1),\\
T^{-3b}\mathcal{L}^{(3)}_{i_1\ldots i_3}(p_T,q_T,\widetilde{\delta}) &= T^{-3b}\sum_{t=1}^{T}\widetilde{s}_{i_1\ldots i_3,t}(p_T,q_T)+o_p(|p_T-q_T|),\\
T^{-4b}\mathcal{L}^{(4)}_{i_1\ldots i_4}(p_T,q_T,\widetilde{\delta}) &= -\left(T^{1-4b}\rho_T^2\right)(\rho_T^{-2}\widetilde{\omega}^{(2)}_{i_1\ldots i_4}(p_T,q_T))+T^{\frac{1}{2}-4b}(T^{-1/2}\sum_{t=1}^{T}\widetilde{k}_{i_1\ldots i_4,t}(p_T,q_T))+o_p(1),\\
T^{-6b}\mathcal{L}^{(6)}_{i_1\ldots i_6}(p_T,q_T,\widetilde{\delta}) &= -T^{1-6b}\sum_{(j_1,\ldots,j_6)\in I_1(i_1,\ldots,i_6)}\omega^{(3)}_{j_1\ldots j_6}(p_T,q_T)+o_p(T^{1-6b}\left|p_T-q_T\right|^2)+o_p\left(1\right),
\end{aligned}$$

and

$$\begin{aligned}
T^{-5b}\mathcal{L}^{(5)}_{i_1\ldots i_5}(p_T,q_T,\widetilde{\delta}) &= o_p(T^{1-4b}\rho_T^2)+o_p\left(1\right),\ T^{-7b}\mathcal{L}^{(7)}_{i_1\ldots i_7}(p_T,q_T,\widetilde{\delta})=O_p(T^{1-7b}\left|p_T-q_T\right|^2)+o_p\left(1\right),\\
T^{-8b}\mathcal{L}^{(8)}_{i_1\ldots i_8}(p_T,q_T,\widetilde{\delta}) &= -T^{1-8b}\sum_{(j_1,\ldots,j_8)\in IND_2}\omega^{(4)}_{j_1\ldots j_8}(p_T,q_T)+O_p(T^{1-8b}\left|p_T-q_T\right|)+o_p\left(1\right).
\end{aligned}$$

Because $b\geq 1/8$, $|p_T-q_T|\geq T^{-1/8}\log T$ and $\min_{\|x^{\otimes 3}\|=1}\ (x^{\otimes 3})'\widetilde{\Omega}^{(3)}(p_T,q_T)(x^{\otimes 3})/(p_T-q_T)^2>L>0$ in probability, the first term on the right hand side of $T^{-6b}\mathcal{L}^{(6)}_{i_1\ldots i_6}(p_T,q_T,\widetilde{\delta})$ dominates the following ones: $T^{\frac{1}{2}-4b}T^{-1/2}\sum_{t=1}^{T}\widetilde{k}_{i_1\ldots i_4,t}(p_T,q_T)$, the first two terms on the right hand side of $T^{-8b}\mathcal{L}^{(8)}_{i_1\ldots i_8}(p_T,q_T,\widetilde{\delta})$, and the remainder terms $o_p(T^{1-6b}\left|p_T-q_T\right|^2)$ and $O_p(T^{1-7b}\left|p_T-q_T\right|^2)$. At the same time, the first term on the right hand side of $T^{-4b}\mathcal{L}^{(4)}_{i_1\ldots i_4}(p_T,q_T,\widetilde{\delta})$ dominates $T^{-5b}\mathcal{L}^{(5)}_{i_1\ldots i_5}(p_T,q_T,\widetilde{\delta})$. These relations imply $\hat{\delta}_2-\delta_*=O_p(\min(T^{-1/4}|\rho_T|^{-1/2},T^{-1/6}\left|p_T-q_T\right|^{-1/3}))$. Otherwise, the log likelihood ratio will be negative with probability close to one. This shows $b=\min(-1/4-1/2\log|\rho_T|,-1/6-1/3\log|p_T-q_T|)$.

Given this result, we can further focus on the following set $\{\delta_2\in R^{n_\delta}:(\|\delta_2-\delta_*\|\leq C_T\}$, where $C_T=C\min(T^{-1/4}|\rho_T|^{-1/2},T^{-1/6}\left|p_T-q_T\right|^{-1/3})$ and C can be made sufficiently large. Uniformly over this set, we have

$$\begin{aligned}
\mathcal{L}^{(2)}_{i_1 i_2}(p_T,q_T,\widetilde{\delta})d_{i_1}d_{i_2} &= \sum_{t=1}^{T}\widetilde{U}_{i_1 i_2,t}(p_T,q_T)d_{i_1}d_{i_2}+o_p(1),\\
\mathcal{L}^{(3)}_{i_1\ldots i_3}(p_T,q_T,\widetilde{\delta})d_{i_1}\ldots d_{i_3} &= \sum_{t=1}^{T}\widetilde{s}_{i_1\ldots i_3,t}(p_T,q_T)d_{i_1}\ldots d_{i_3}+o_p(1),\\
\mathcal{L}^{(4)}_{i_1\ldots i_4}(p_T,q_T,\widetilde{\delta})d_{i_1}\ldots d_{i_4} &= -T\widetilde{\omega}^{(2)}_{i_1\ldots i_4}(p_T,q_T)d_{i_1}\ldots d_{i_4}+o_p(1),\\
\mathcal{L}^{(6)}_{i_1\ldots i_6}(p_T,q_T,\widetilde{\delta})d_{i_1}\ldots d_{i_6} &= -T\sum_{(j_1,\ldots,j_6)\in I_1(i_1,\ldots,i_6)}\omega^{(3)}_{j_1\ldots j_6}(p_T,q_T)d_{i_1}\ldots d_{i_6}+o_p\left(1\right),\\
\mathcal{L}^{(k)}_{i_1\ldots i_k}(p_T,q_T,\widetilde{\delta})d_{i_1}\ldots d_{i_k} &= o_p(1)\text{ for }k=5,7,8.
\end{aligned}$$

Therefore, (A.137) holds. Consequently, the result in the Corollary holds for this subsequence.

Finally, we study the third subsequence. Without loss of generality, we assume $|\rho_T|\neq 0$. As before, we suppose that the convergence rate of $\hat{\delta}_2$ is T^{-b} for some $b>0$ that potentially depends on $|p_T - q_T|$ and $|\rho_T|$. For any $1/8 \le b \le 1/4$, we have

$$\begin{aligned}
T^{-2b}\mathcal{L}^{(2)}_{i_1 i_2}(p_T, q_T, \widetilde{\delta}) &= T^{-2b}\sum_{t=1}^{T}\widetilde{U}_{i_1 i_2,t}(p_T,q_T) + o_p(1),\\
T^{-3b}\mathcal{L}^{(3)}_{i_1\ldots i_3}(p_T, q_T, \widetilde{\delta}) &= T^{-3b}\sum_{t=1}^{T}\widetilde{s}_{i_1\ldots i_3,t}(p_T,q_T) + o_p\left(\sqrt{T^{1-4b}\rho_T^2}\right),\\
T^{-4b}\mathcal{L}^{(4)}_{i_1\ldots i_4}(p_T, q_T, \widetilde{\delta}) &= -\left(T^{1-4b}\rho_T^2\right)(\rho_T^{-2}\widetilde{\omega}^{(2)}_{i_1\ldots i_4}(p_T,q_T)) + T^{\frac{1}{2}-4b}(T^{-1/2}\sum_{t=1}^{T}\widetilde{k}_{i_1\ldots i_4,t}(p_T,q_T)) + o_p(1),\\
T^{-6b}\mathcal{L}^{(6)}_{i_1\ldots i_6}(p_T, q_T, \widetilde{\delta}) &= -T^{1-6b}\sum_{(j_1,\ldots,j_6)\in I_1(i_1,\ldots,i_6)}\omega^{(3)}_{j_1\ldots j_6}(p_T,q_T) + o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
T^{-5b}\mathcal{L}^{(5)}_{i_1\ldots i_5}(p_T, q_T, \widetilde{\delta}) &= o_p\left(T^{1-4b}\rho_T^2\right) + o_p(1),\ T^{-7b}\mathcal{L}^{(7)}_{i_1\ldots i_7}(p_T, q_T, \widetilde{\delta}) = o_p(1),\\
T^{-8b}\mathcal{L}^{(8)}_{i_1\ldots i_8}(p_T, q_T, \widetilde{\delta}) &= -T^{1-8b}\sum_{(j_1,\ldots,j_8)\in I_2(i_1,,,.i_8)}\omega^{(4)}_{j_1\ldots j_8}(p_T,q_T) + o_p(1),
\end{aligned}$$

where $I_2(i_1,...,i_8)$ equals IND_2 with (j,k,l,m,n,r,s,u) replaced by $(i_1,...,i_8)$. The first leading terms of $T^{-4b}\mathcal{L}^{(4)}_{i_1\ldots i_4}(p_T,q_T,\widetilde{\delta})$ dominates the $o_p\left(T^{1-4b}\rho_T^2\right)$ term in $T^{-5b}\mathcal{L}^{(5)}_{i_1\ldots i_5}(p_T,q_T,\widetilde{\delta})$ and the $o_p(\sqrt{T^{1-4b}\rho_T^2})$ term in $T^{-3b}\mathcal{L}^{(3)}_{i_1\ldots i_3}(p_T,q_T,\widetilde{\delta})$. This leading term diverges unless $T^{1-4b}\rho_T^2 = O_p(1)$. Given this result, we can further focus on the following set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \le C_T\}$, where $C_T = CT^{-1/4}|\rho_T|^{-1/2}$ and C can be made sufficiently large. Uniformly over this set, we have

$$\begin{aligned}
\mathcal{L}^{(2)}_{i_1 i_2}(p_T, q_T, \widetilde{\delta})d_{i_1}d_{i_2} &= \sum_{t=1}^{T}\widetilde{U}_{i_1 i_2,t}(p_T,q_T)d_{i_1}d_{i_2} + o_p(1),\\
\mathcal{L}^{(3)}_{i_1\ldots i_3}(p_T, q_T, \widetilde{\delta})d_{i_1}\ldots d_{i_3} &= \sum_{t=1}^{T}\widetilde{s}_{i_1\ldots i_3,t}(p_T,q_T)d_{i_1}\ldots d_{i_3} + o_p(1),\\
\mathcal{L}^{(4)}_{i_1\ldots i_4}(p_T, q_T, \widetilde{\delta})d_{i_1}\ldots d_{i_4} &= -T\widetilde{\omega}^{(2)}_{i_1\ldots i_4}(p_T,q_T)d_{i_1}\ldots d_{i_4} + \sum_{t=1}^{T}\widetilde{k}_{i_1\ldots i_4,t}(p_T,q_T)d_{i_1}\ldots d_{i_4} + o_p(1),\\
\mathcal{L}^{(6)}_{i_1\ldots i_6}(p_T, q_T, \widetilde{\delta})d_{i_1}\ldots d_{i_6} &= -T\sum_{(j_1,\ldots,j_6)\in I_1(i_1,\ldots,i_6)}\omega^{(3)}_{j_1\ldots j_6}(p_T,q_T)d_{i_1}\ldots d_{i_6} + o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{L}^{(5)}_{i_1\ldots i_5}(p_T, q_T, \widetilde{\delta})d_{i_1}\ldots d_{i_5} &= o_p(1), \mathcal{L}^{(7)}_{i_1\ldots i_7}(p_T, q_T, \widetilde{\delta})d_{i_1}\ldots d_{i_7} = o_p(1),\\
\mathcal{L}^{(8)}_{i_1\ldots i_8}(p_T, q_T, \widetilde{\delta})d_{i_1}\ldots d_{i_8} &= -T\sum_{(j_1,\ldots,j_8)\in I_2(i_1\ldots i_8)}\omega^{(4)}_{j_1\ldots j_8}(p_T,q_T)d_{i_1}\ldots d_{i_8} + o_p(1).
\end{aligned}$$

Let $\eta_i = T^{1/4}d_i$, and then apply the definitions of $\widetilde{U}^{(2)}_t(p_T,q_T)$, $\widetilde{U}^{(3)}_t(p_T,q_T)$, $\widetilde{U}^{(4)}_t(p_T,q_T)$, $\widetilde{\Omega}(p_T,q_T)$,

$\widetilde{\Omega}^{(3)}(p_T, q_T)$, and $\widetilde{\Omega}^{(4)}(p_T, q_T)$, we have

$$
\begin{aligned}
& 2[\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta})] \\
= \; & \left(\eta^{\otimes 2}\right)' \left(T^{-1/2} \sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T) \right) + T^{-1/4} \frac{1}{3} \left(\eta^{\otimes 3}\right)' \left(T^{-1/2} \sum_{t=1}^{T} \widetilde{U}_t^{(3)}(p_T, q_T) \right) \\
& + T^{-1/2} \frac{1}{12} \left(\eta^{\otimes 4}\right)' \left(T^{-1/2} \sum_{t=1}^{T} \widetilde{U}_t^{(4)}(p_T, q_T) \right) - \frac{1}{4} \left(\eta^{\otimes 2}\right)' \widetilde{\Omega}(p_T, q_T) \left(\eta^{\otimes 2}\right) \\
& - \frac{1}{36} T^{-1/2} \left(\eta^{\otimes 3}\right)' \widetilde{\Omega}^{(3)}(p_T, q_T) \left(\eta^{\otimes 3}\right) - \frac{1}{576} T^{-1} \left(\eta^{\otimes 4}\right)' \widetilde{\Omega}^{(4)}(p_T, q_T) \left(\eta^{\otimes 4}\right) + o_p(1) \\
= \; & \widetilde{\mathcal{W}}^{(2)}(p_T, q_T, \eta) + \widetilde{\mathcal{W}}^{(3)}(p_T, q_T, \eta) + \widetilde{\mathcal{W}}^{(4)}(p_T, q_T, \eta) + o_p(1).
\end{aligned}
$$

Take the supremum with respect to η. This shows that the result of the Corollary holds for the third subsequence.

Under the third subsequence, depending on the values of $|\rho_T|$ and $|p_T - q_T|$, one or two terms among $\widetilde{\mathcal{W}}^{(k)}(p_T, q_T, \eta)$ $(k = 2, 3, 4)$ may be asymptotically negligible. It is also possible that all three terms are non-negligible. Suppose $p_T = 0.5 + c_1 T^{-1/4} + c_2 T^{-1/8}$ and $q_T = 0.5 + c_3 T^{-1/4} - c_2 T^{-1/8}$. Then, $\rho_T = (c_1 + c_3) T^{-1/4}$ and $p_T - q_T = 2c_2 T^{-1/8}$. It follows that $T^{-kb} \mathcal{L}^{(k)}_{i_1 \dots i_k}(p_T, q_T, \widetilde{\delta})$ $(k = 2, 3, 4, 6, 8)$ are $O_p(1)$ but not $o_p(1)$ when $b = 1/8$. Therefore, none of $\widetilde{\mathcal{W}}^{(k)}(p_T, q_T, \eta)$ for $k = 2, 3, 4$ can be omitted from the approximation to $LR(p_T, q_T)$. ■

Remark 9 *Lemma A.21 below takes a close look at the assumption*

$$\min_{x \in R^{n_\delta}, \|x^{\otimes 2}\| = 1} (x^{\otimes 2})' \widetilde{\Omega}(p_T, q_T)(x^{\otimes 2}) / \rho_T^2 > L > 0$$

in the context of the following model:

$$y_t = \delta_1 1_{\{s_t = 1\}} + \delta_2 1_{\{s_t = 2\}} + \alpha x_t + u_t, \tag{A.138}$$

where x_t is a weakly exogenous regressor independent of u_t and $u_t \sim i.i.d.N(0, \sigma^2)$. It shows that this assumption holds except in the special case where $x_t = y_{t-1}$ and $\alpha = 0$. In this special case, nevertheless, the refined approximation is still a weakly consistent estimator of the limiting distribution of $LR(p_T, q_T)$ under both SEQ1 and SEQ2 for any $a \geq 0$ and Corollary 2 also holds; see Lemmas A.22, A.23, and A.24.

Lemma A.21 *Suppose that the null hypothesis and Assumptions 1-9 hold for the model A.138. Assume $\epsilon \leq p_T, q_T \leq 1 - \epsilon$ for any T and $\rho_T \to 0$ as $T \to \infty$. Then, $\widetilde{\Omega}(p_T, q_T)/\rho_T^2$ is strictly positive in probability when x_t is not the lagged dependent variable and when $x_t = y_{t-1}$ but $\alpha \neq 0$. If $x_t = y_{t-1}$ and $\alpha = 0$, then $\widetilde{\Omega}(p_T, q_T)/\rho_T^2 \to^p 0$ and*

$$\widetilde{\Omega}(p_T, q_T) = \frac{4}{\sigma_*^4} \left(\frac{1 - p_T}{1 - q_T} \right)^2 \rho_T^4 + O_p\left(T^{-1/2} \rho_T^2 \right) + o_p\left(\rho_T^4 \right).$$

Proof of Lemma A.21. Under (A.138), we have

$$\widetilde{\Omega}(p_T,q_T)=T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T,q_T)^2-T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T,q_T)\widetilde{S}_t'\left(T^{-1}\sum_{t=1}^{T}\widetilde{S}_t\widetilde{S}_t'\right)^{-1}T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T,q_T)\widetilde{S}_t,$$

where $\widetilde{u}_t=y_t-\widetilde{\delta}+\widetilde{\alpha}x_t$,

$$\widetilde{S}_t=\left[\frac{x_t\widetilde{u}_t}{\widetilde{\sigma}^2}\quad\frac{1}{2\widetilde{\sigma}^2}\left(\frac{\widetilde{u}_t^2}{\widetilde{\sigma}^2}-1\right)\quad\frac{\widetilde{u}_t}{\widetilde{\sigma}^2}\right]',$$

and

$$\begin{aligned}&\widetilde{U}_t^{(2)}(p_T,q_T) &&\text{(A.139)}\\ =\;&\frac{2}{\widetilde{\sigma}^2}\left(\frac{1-p_T}{1-q_T}\right)\rho_T\frac{u_{t-1}\widetilde{u}_t}{\widetilde{\sigma}^2}\\ &+\frac{2}{\widetilde{\sigma}^2}\left(\frac{1-p_T}{1-q_T}\right)\left\{\rho_T^2\frac{\widetilde{u}_{t-2}\widetilde{u}_t}{\widetilde{\sigma}^2}+\rho_T^3\sum_{s=3}^{t-1}\rho_T^{s-3}\frac{\widetilde{u}_{t-s}\widetilde{u}_t}{\widetilde{\sigma}^2}+\rho_T\frac{(\widetilde{u}_{t-1}-u_{t-1})\,\widetilde{u}_t}{\widetilde{\sigma}^2}\right\}.\end{aligned}$$

Because the null regression allows for an intercept, without loss of generality, we assume that x_t and y_t both have mean zero. Note that $T\widetilde{\Omega}(p_T,q_T)$ equals the SSR from the projection of $\widetilde{U}_t^{(2)}(p_T,q_T)$ on the space spanned by $\widetilde{S}_t$ for $t=1,...,T$.

First, we consider the special case where $x_t=y_{t-1}$ and $\alpha=0$. We have

$$\widetilde{S}_t=\left[\frac{u_{t-1}\widetilde{u}_t}{\widetilde{\sigma}^2}\quad\frac{1}{2\widetilde{\sigma}^2}\left(\frac{\widetilde{u}_t^2}{\widetilde{\sigma}^2}-1\right)\quad\frac{\widetilde{u}_t}{\widetilde{\sigma}^2}\right]'. \qquad \text{(A.140)}$$

Because $u_{t-1}\widetilde{u}_t/\widetilde{\sigma}^2$ in (A.139) belongs to the column space of $\widetilde{S}_t$, the first term on the right hand side of (A.139) has zero contribution to the SSR. The product of the second term with $\widetilde{S}_t$ satisfies

$$T^{-1}\sum_{t=1}^{T}\left\{\rho_T^2\frac{\widetilde{u}_{t-2}\widetilde{u}_t}{\widetilde{\sigma}^2}+\rho_T^3\sum_{s=3}^{t-1}\rho_T^{s-3}\frac{\widetilde{u}_{t-s}\widetilde{u}_t}{\widetilde{\sigma}^2}+\rho_T\frac{(\widetilde{u}_{t-1}-u_{t-1})\,\widetilde{u}_t}{\widetilde{\sigma}^2}\right\}\widetilde{S}_t=O_p\left(T^{-1/2}\left|\rho_T\right|+\left|\rho_T^3\right|\right),$$

where the equality holds because of the CLT and LLN. Therefore,

$$\begin{aligned}\widetilde{\Omega}(p_T,q_T)&=\frac{4}{\widetilde{\sigma}^4}\left(\frac{1-p_T}{1-q_T}\right)^2\rho_T^4T^{-1}\sum_{t=1}^{T}\frac{\widetilde{u}_{t-2}^2\widetilde{u}_t^2}{\widetilde{\sigma}^4}+O_p\left(T^{-1/2}\rho_T^2+\left|\rho_T^5\right|\right)\\ &=\frac{4}{\sigma_*^4}\left(\frac{1-p_T}{1-q_T}\right)^2\rho_T^4+O_p\left(T^{-1/2}\rho_T^2\right)+o_p\left(\rho_T^4\right),\end{aligned}$$

where the second equality holds because $T^{-1}\sum_{t=1}^{T}\widetilde{u}_{t-2}^2\widetilde{u}_t^2/\widetilde{\sigma}^4=1+O_p\left(T^{-1/2}\right)$.

Next, we consider the situation where $x_t\neq y_{t-1}$ or $x_t=y_{t-1}$ but $\alpha\neq 0$. We have

$$T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T,q_T)^2=\frac{4}{\widetilde{\sigma}^4}\left(\frac{1-p_T}{1-q_T}\right)^2\rho_T^2T^{-1}\sum_{t=1}^{T}\left(\frac{u_{t-1}\widetilde{u}_t}{\widetilde{\sigma}^2}\right)^2+o_p\left(\rho_T^2\right)=\frac{4}{\sigma_*^4}\left(\frac{1-p_T}{1-q_T}\right)^2\rho_T^2+o_p\left(\rho_T^2\right)$$

and

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T,q_T)\widetilde{S}_t \\
= \; & \frac{2}{\widetilde{\sigma}^2}\left(\frac{1-p_T}{1-q_T}\right)\rho_T T^{-1}\sum_{t=1}^{T}\frac{u_{t-1}\widetilde{u}_t}{\widetilde{\sigma}^2}\widetilde{S}_t \\
& +\frac{2}{\widetilde{\sigma}^2}\left(\frac{1-p_T}{1-q_T}\right)T^{-1}\sum_{t=1}^{T}\left\{\rho_T^2\frac{\widetilde{u}_{t-2}\widetilde{u}_t}{\widetilde{\sigma}^2}+\rho_T^3\sum_{s=3}^{t-1}\rho_T^{s-3}\frac{\widetilde{u}_{t-s}\widetilde{u}_t}{\widetilde{\sigma}^2}+\rho_T\frac{(\widetilde{u}_{t-1}-u_{t-1})\,\widetilde{u}_t}{\widetilde{\sigma}^2}\right\}\widetilde{S}_t \\
= \; & \frac{2}{\widetilde{\sigma}^2}\left(\frac{1-p_T}{1-q_T}\right)\rho_T T^{-1}\sum_{t=1}^{T}\frac{u_{t-1}\widetilde{u}_t}{\widetilde{\sigma}^2}\widetilde{S}_t+O_p\left(T^{-1/2}\left|\rho_T\right|+\rho_T^2\right) \\
= \; & \frac{2}{\widetilde{\sigma}^2}\left(\frac{1-p_T}{1-q_T}\right)\rho_T\left[T^{-1}\sum_{t=1}^{T}\frac{x_t u_{t-1}u_t^2}{\sigma_*^4}\quad 0\quad 0\right]+O_p\left(T^{-1/2}\left|\rho_T\right|+\rho_T^2\right).
\end{aligned}
$$

From these two equations, it follows that

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T,q_T)\widetilde{S}_t'\left(T^{-1}\sum_{t=1}^{T}\widetilde{S}_t\widetilde{S}_t'\right)^{-1}T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T,q_T)\widetilde{S}_t \\
= \; & \frac{4}{\widetilde{\sigma}^4}\left(\frac{1-p_T}{1-q_T}\right)^4\rho_T^2\left(T^{-1}\sum_{t=1}^{T}\frac{u_{t-1}u_t x_t u_t}{\sigma_*^4}\right)\left(T^{-1}\sum_{t=1}^{T}\left(\frac{x_t u_t}{\sigma_*^2}\right)^2\right)^{-1}\left(T^{-1}\sum_{t=1}^{T}\frac{u_{t-1}u_t x_t u_t}{\sigma_*^4}\right)+o_p\left(\rho_T^2\right) \\
= \; & \frac{4}{\sigma_*^4}\left(\frac{1-p_T}{1-q_T}\right)^4\rho_T^2\left(\frac{E\left(u_{t-1}x_t\right)}{\sigma_*(\mathrm{var}(x_t))^{1/2}}\right)^2+o_p\left(\rho_T^2\right)=\frac{4}{\sigma_*^4}\left(\frac{1-p_T}{1-q_T}\right)^4\rho_T^2\,\mathrm{Corr}(u_{t-1},x_t)^2+o_p\left(\rho_T^2\right).
\end{aligned}
$$

Therefore,

$$
\widetilde{\Omega}(p_T,q_T)=\frac{4}{\sigma_*^4}\left(\frac{1-p_T}{1-q_T}\right)^2\rho_T^2\left\{1-\mathrm{Corr}(u_{t-1},x_t)^2\right\}+o_p\left(\rho_T^2\right).
$$

Because $x_t\neq y_{t-1}$ or $x_t=y_{t-1}$ but $\alpha\neq 0$, $\mathrm{Corr}(u_{t-1},x_t)^2<1$. Therefore, the Lemma holds.■

Lemma A.22 *Suppose the null hypothesis and Assumptions 1-9 hold for the model A.138 with* $x_t=y_{t-1}$ *and* $\alpha=0$. *Assume* $\epsilon\leq p_T,q_T\leq 1-\epsilon$ *for any* T *and* $\rho_T\to 0$ *as* $T\to\infty$. *Then, in addition to the results in Lemma A.17, we have* $\mathcal{L}^{(2)}(p_T,q_T,\widetilde{\delta})=\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T,q_T)+O_p(\left|\rho_T^3\right|+T^{-1/2}\rho_T^2)$, $\mathcal{L}^{(3)}(p_T,q_T,\widetilde{\delta})=\sum_{t=1}^{T}\widetilde{s}_t(p_T,q_T)+O_p(T^{1/2}\rho_T^2+T^{1/2}|\rho_T(p_T-q_T)|+|\rho_T|)$, $\mathcal{L}^{(5)}(p_T,q_T,\widetilde{\delta})=O_p(T^{1/2}+T\rho_T^4+T\left|\rho_T^3(p_T-q_T)\right|)$, $\mathcal{L}^{(6)}(p_T,q_T,\widetilde{\delta})=-10T\widetilde{\Omega}^{(3)}(p_T,q_T)+O_p(T^{1/2}+T\left|\rho_T^3\right|+T\rho_T^2\left|p_T-q_T\right|)$, *and* $\mathcal{L}^{(7)}(p_T,q_T,\widetilde{\delta})=O_p(T^{1/2}+T\rho_T^2+T\left(p_T-q_T\right)^2+\left|\left(p_T-q_T\right)\rho_T\right|)$.

Proof of Lemma A.22. The results follow from the proof of Lemma A.17 after taking into account the specific structure of the model (A.138).

For the first result, the expression in the parentheses of (A.101) is $O_p(|\rho_T|)$ by Lemma A.11. In addition, $\nabla_{\delta_{2k}}\hat{\theta}_j(\widetilde{\delta})=O_p(T^{-1/2}\rho_T^2+T^{-1}|\rho_T|)$ except when $j\in\{n_\beta+k,n_\beta+n_\delta+k\}$ because of the structure (A.139)-(A.140). Therefore,

$$T^{-1/2}\mathcal{L}^{(2)}(p_T,q_T,\widetilde{\delta}) = \left(\frac{1-\xi_T}{\xi_T^2}\right)T^{-1/2}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\left\{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}+\nabla_{\delta_2}\widetilde{f}_{2t}\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\right\}$$
$$+\frac{1}{\xi_T}T^{-1/2}\sum_{t=1}^{T}\frac{1}{\widetilde{f}_t}\left\{\nabla_{\delta_2}\widetilde{f}_{2t}\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}-\nabla_{\delta_2}\widetilde{f}_{2t}\nabla_{\delta_2}\widetilde{\xi}_{t|t-1}\right\}+O_p(\left|\rho_T^3\right|+T^{-1/2}\rho_T^2).$$

The result follows because $\nabla_{\delta_1}\widetilde{f}_{1t}=\nabla_{\delta_2}\widetilde{f}_{2t}$ and $\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}=-\nabla_{\delta_2}\widetilde{\xi}_{t|t-1}$.

To prove the second result, as in the proof of Lemma A.17, we consider (A.39) with the scaling factor $T^{-3/4}$ replaced by $T^{-1/2}$. The second and third terms are $O_p\left(\rho_T^2+T^{-1/2}\left|\rho_T\right|\right)$. For the first term (A.104), (M1)-(M6) are all $O_p(T^{-1/2}\rho_T^2+\left|\rho_T\left(p_T-q_T\right)\right|+T^{-1}\left|\rho_T\right|)$ because of the structure (A.139)-(A.140). The remaining terms in (A.104) can be rewritten as

$$T^{-1/2}\sum_{t=1}^{T}\sum_{s=n_\beta+1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\left\{(1-\xi_T)\nabla_{\delta_2}\nabla_{\theta_s}\widetilde{f}_{2t}-(1-\xi_T)\nabla_{\delta_1}\nabla_{\theta_s}\widetilde{f}_{1t}\right\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta})$$
$$-\frac{1}{\xi_T}T^{-1/2}\sum_{t=1}^{T}\sum_{s=1}^{n_\beta+2n_\delta}\frac{1}{\widetilde{f}_t}\left\{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\theta_s}\widetilde{\xi}_{t|t-1}+(\nabla_{\theta_s}\widetilde{f}_{1t}-\nabla_{\theta_s}\widetilde{f}_{2t})\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\right\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\theta}_s(\widetilde{\delta}).$$

The first double summation equals 0 by Assumption 7. The second double summation equals

$$-\frac{1}{\xi_T}T^{-1/2}\sum_{t=1}^{T}\sum_{s=1}^{n_\delta}\frac{1}{\widetilde{f}_t}\left\{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_{1s}}\widetilde{\xi}_{t|t-1}+\nabla_{\delta_{1s}}\widetilde{f}_{1t}\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}\right\}\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\delta}_{1s}(\widetilde{\delta})=O_p\left(\rho_T^2+T^{-1}\left|\rho_T\right|\right),$$

where the last equality holds because of (A.139)-(A.140). These results imply $\mathcal{L}^{(3)}(p_T,q_T,\widetilde{\delta})=\sum_{t=1}^{T}\widetilde{s}_t(p_T,q_T)+O_p(T^{1/2}\rho_T^2+T^{1/2}|\rho_T(p_T-q_T)|+|\rho_T|)$.

The proofs for the remaining three results of the Lemma follow from similar lines of argument. That is, we use the arguments in the proof of Lemma A.17 and apply (A.139)-(A.140) whenever this is relevant. We omit the details. ■

Lemma A.23 *Suppose the null hypothesis and Assumptions 1-9 hold for the model (A.138) with $x_t=y_{t-1}$ and $\alpha=0$. Then: (i) Proposition 2.1 holds after replacing 1/6 by 1/12. (ii) Proposition 2.2 holds after replacing 1/4 and $\sup_{\eta\in R}\{\widetilde{\mathcal{W}}^{(2)}(p_T,q_T,\eta)+\widetilde{\mathcal{W}}^{(4)}(p_T,q_T,\eta)\}$ by 1/8 and $\sup_{\eta\in R}\{\widetilde{\mathcal{W}}^{(2)}(p_T,q_T,\eta)+\widetilde{\mathcal{W}}^{(3)}(p_T,q_T,\eta)+\widetilde{\mathcal{W}}^{(4)}(p_T,q_T,\eta)\}$. (3) Proposition 2.3 holds.*

Proof of Lemma A.23. We focus on the $a<1/12$ case. Its proof is similar to the $a<1/6$ case of Proposition 2. From this proof, it will become clear that the remaining results in (i) and (ii) can be proved in a similar way, by applying the arguments in the proof of Proposition 2.1-2 and Lemma A.22. The proof for the third result of the lemma is the same as that of Proposition 2.3. It is therefore omitted.

As in the proof of Proposition 2, we first establish the convergence rate of the MLE $\hat{\delta}_2$, and then apply this result to study the Taylor expansion. We can restrict our attention to the following set $\{\delta_2\in R^{n_\delta}:(\|\delta_2-\delta_*\|\le CT^{-1/6}\}$, where C can be made sufficiently large. Define $d=$

$\delta_2 - \widetilde{\delta}$. Then, the following expansion holds uniformly over this set: $\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) = \sum_{k=2}^{6} \frac{1}{k!}\mathcal{L}^{(k)}(p_T, q_T, \widetilde{\delta})d^k + o_p(1)$. We suppose that the convergence rate of $\hat{\delta}_2$ is T^{-b} for some $b > 0$ that can potentially depend on $|p_T - q_T|$ and $|\rho_T|$. By Lemmas A.17 and A.22, for any $b \geq 1/6$:

$$
\begin{aligned}
T^{-2b}\mathcal{L}^{(2)}(p_T, q_T, \widetilde{\delta}) &= O_p(T^{\frac{1}{2}-2b}\rho_T^2) + o_p(1), \\
T^{-3b}\mathcal{L}^{(3)}(p_T, q_T, \widetilde{\delta}) &= O_p(1), \\
T^{-4b}\mathcal{L}^{(4)}(p_T, q_T, \widetilde{\delta}) &= -3(\rho_T^4 T^{1-4b})(\rho_T^{-4}\widetilde{\Omega}(p_T, q_T)) + O_p\left(T^{-1/6}\right), \\
T^{-5b}\mathcal{L}^{(5)}(p_T, q_T, \widetilde{\delta}) &= o_p\left(\rho_T^4 T^{1-4b}\right) + o_p(1), T^{-6b}\mathcal{L}^{(6)}(p_T, q_T, \widetilde{\delta}) = O_p(1).
\end{aligned}
$$

As a result, $T^{-4b}\mathcal{L}^{(4)}(p_T, q_T, \widetilde{\delta})$ diverges and dominates the other terms unless $\rho_T^4 T^{1-4b} = O_p(1)$, or $T^{-b} = O_p\left(T^{-1/4}|\rho_T|^{-1}\right)$. Therefore, we must have $\hat{\delta}_2 - \delta_* = O_p\left(T^{-1/4}|\rho_T|^{-1}\right)$; otherwise $\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta})$ will be negative with probability close to one in large samples. This proves $\hat{\delta}_2 - \delta_* = O_p\left(T^{-1/4}|\rho_T|^{-1}\right)$.

Given this, we can further restrict our attention to the set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \leq CT^{-1/4}|\rho_T|^{-1}\}$ with C being a sufficiently large constant. Let $C_T = CT^{-1/4}|\rho_T|^{-1}$. By Lemmas A.17 and A.22, we have $C_T^2\mathcal{L}^{(2)}(p_T, q_T, \widetilde{\delta}) = C_T^2 \sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T) + o_p(1)$, $C_T^4\mathcal{L}^{(4)}(p_T, q_T, \widetilde{\delta}) = -3TC_T^4\widetilde{\Omega}(p_T, q_T) + o_p(1)$, $C_T^k\mathcal{L}^{(k)}(p_T, q_T, \widetilde{\delta}) = o_p(1)$ for $k = 3, 5, 6$, where the leading terms in the first two equations are both $O_p(1)$ but not $o_p(1)$. This implies $\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) = (1/2)\sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T)d^2 - (1/8)T\widetilde{\Omega}(p_T, q_T)d^4 + o_p(1)$ uniformly over this set. Let $\eta = T^{1/4}d$. The result follows after multiplying both sides by 2 and computing the supremum with respect to η. ■

Lemma A.24 *Assume $\rho_T \to 0$ as $T \to \infty$. Suppose that the null hypothesis, Assumptions 1-9, and the conditions in Proposition 2 hold for the model A.138 with $x_t = y_{t-1}$ and $\alpha = 0$. Then, the conclusion of Corollary 2 holds.*

Proof of Lemma A.24. The proof is essentially the same as that Corollary 2. That is, we partition the sequence (p_T, q_T) into three subsequences and show that the result holds for each subsequence. We omit the details. ■

A.4 Proofs for Subsection 6.2

We need to resolve two challenges in order to derive the refined approximation in Proposition 3. First, because the null hypothesis is violated, the approximations in Lemmas 2, 3 and A.17 are no longer applicable. Second, for the model (39), the contiguous alternatives have a complex structure if the transition probability is allowed to converge to the boundary (34) as $T \to 0$. Their orders can vary between $O_p(T^{-1/4})$ and $O_p(T^{-1/8})$ depending on the convergence rate and the converging path of the transition probability. These challenges are resolved in Lemmas A.25 - A.28, where we provide approximations to various terms in an eighth order Taylor expansion of the log likelihood ratio, allowing the alternatives to be of order $O_p(T^{-a})$, where a can be any value between 1/8 and 1/4. It is unnecessary to consider $a > 1/4$ because the likelihood ratio has trivial asymptotic power in that case. The proofs make heavy use of the structure of the DGP (38) and the model (39).

After proving these lemmas, we apply them to derive the refined approximation in Proposition 3. This is done in two steps as in Proposition 1 and Corollary 1. We first study the likelihood ratio when the transition probability is fixed and bounded away from (34). This produces the leading terms in the approximation in Proposition 3. Then, we study the likelihood when the transition probability is local to the boundary (34). This produces the refinement terms of Proposition 3. The final result follows from merging these terms according to how they affect the likelihood expansion.

We focus on the dynamic model (39) in the subsequent analysis. The static case is similar and simpler. In order to allow the transition probability to converge to (34), we consider a more general DGP than (38) in Lemmas A.25-A.28:

$$y_t = \mu_* + A_T 1_{\{s_t=2\}} + e_t, \tag{A.141}$$

where

$$A_T = c_* T^{-a} \text{ with } 1/8 \leq a \leq 1/4, \tag{A.142}$$

and the underlying transition probability satisfies $\epsilon \leq p_T, q_T \leq 1-\epsilon$ for any T. This DGP reduces to (38) when $(p_T, q_T) = (p_*, q_*)$ and $a = 1/4$. To shorten the expressions, we define

$$\widetilde{s}_t = 1_{\{s_t=2\}} - T^{-1}\sum_{t=1}^{T} 1_{\{s_t=2\}}, \quad \widetilde{y}_{t-1} = y_{t-1} - T^{-1}\sum_{t=1}^{T} y_{t-1}, \quad \widetilde{e}_t = e_t - T^{-1}\sum_{t=1}^{T} e_t.$$

The estimated residuals of the null model satisfy

$$\widetilde{u}_t = \widetilde{e}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} + A_T\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right), \tag{A.143}$$

Finally, note that for this model,

$$\theta = \left(\alpha, \sigma^2, \delta_1, \delta_2\right)'.$$

Because only the intercept is allowed to switch, some notations can be simplified accordingly. For example, we write $\mathcal{L}^{(2)}_{jk}(p_T, q_T, \widetilde{\delta})$ and $\nabla_{\delta_{1j}}\nabla_{\delta_{1k}}\widetilde{f}_{1t}$ as $\mathcal{L}^{(2)}(p_T, q_T, \widetilde{\delta})$ and $\nabla^2_{\delta_1}\widetilde{f}_{1t}$, respectively.

Remark 10 *Under the alternative hypothesis, the stochastic orders of* $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\widetilde{f}_{jt}/\widetilde{f}_{jt})$ *and* $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\widetilde{f}_{jt}/\widetilde{f}_{jt})\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\widetilde{\xi}_{t|t-1}$ *depend on* A_T*, as well as* T *and* $|\rho_T|$*. The next two lemmas study their stochastic orders for various* $i, k, j,$ *and* m*.*

Lemma A.25 *Suppose that the data are generated by (A.141)-(A.142). Then, for* $j = 1, 2$ *and* $i_1, ...i_k \in \{1, ..., 4\}$*, we have* $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\widetilde{f}_{jt})/\widetilde{f}_{jt} = O_p(T^{-1/2} + |A_T|^{\alpha(k)})$*, where* $\alpha(k) = k$ *for* $k = 2, 3, 4$ *and* $\alpha(k) = 4$ *for* $k = 5, .., 8$*.*

Proof of Lemma A.25. The proof uses the orthogonality of Hermite polynomials. Without loss of generality, assume $j = 1, \mu_* = 0$, and $c_* > 0$. We begin with the $k = 2$ case.

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\widetilde{f}_{1t}}{\widetilde{f}_{1t}} &= T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}f^*_{1t}}{f^*_{1t}} \\ &+T^{-1}\sum_{j=1}^{4}\sum_{t=1}^{T}\left(\frac{\nabla_{\theta_j}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}f^*_{1t}}{f^*_{1t}} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}f^*_{1t}\nabla_{\theta_j}f^*_{1t}}{f^{*2}_{1t}}\right)(\widetilde{\theta}_j - \theta_j^{*(s_t)}) + O_p\left(A_T^2\right), \end{aligned}$$

where the asterisk denotes the true value, $\theta^{*(1)} = \left(0, \sigma_*^2, 0, 0\right)$, $\theta^{*(2)} = \left(0, \sigma_*^2, A_T, A_T\right)$ (or, equivalently $\theta^{*(2)} = \left(0, \sigma_*^2, A_T, 0\right)$ since f_{1t} does not depend on δ_2), $\theta^{*(s_t)} = \left(0, \sigma_*^2, A_T 1_{\{s_t=2\}}, A_T 1_{\{s_t=2\}}\right)$, and the remainder term is $O_p\left(A_T^2\right)$ because $||\widetilde{\theta} - \theta^{*(s_t)}|| = O_p\left(A_T\right)$. The first term on the right hand side is $O_p\left(T^{-1/2}\right)$ by the CLT. For the second term, we apply the decomposition $\widetilde{\theta}_j - \theta_j^{*(s_t)} = (\theta_j^{*(1)} - \theta_j^{*(s_t)}) + (\widetilde{\theta}_j - \theta_j^{*(1)})$, where $\theta_j^{*(1)} - \theta_j^{*(s_t)}$ captures the effect of the regime switching, and $\widetilde{\theta}_j - \theta_j^{*(1)}$ the effect of the parameter estimation. We have

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\left(\frac{\nabla_{\theta_j}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}} f_{1t}^*}{f_{1t}^*} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}} f_{1t}^* \nabla_{\theta_j} f_{1t}^*}{f_{1t}^{*2}}\right)(\widetilde{\theta}_j - \theta_j^{*(s_t)}) \qquad \text{(A.144)}\\
= \; & T^{-1}\sum_{t=1}^{T}\left\{\frac{\nabla_{\theta_j}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}} f_{1t}^*}{f_{1t}^*} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}} f_{1t}^* \nabla_{\theta_j} f_{1t}^*}{f_{1t}^{*2}}\right\}(\theta_j^{*(1)} - \theta_j^{*(s_t)})\\
& + T^{-1}\sum_{t=1}^{T}\left\{\frac{\nabla_{\theta_j}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}} f_{1t}^*}{f_{1t}^*} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}} f_{1t}^* \nabla_{\theta_j} f_{1t}^*}{f_{1t}^{*2}}\right\}(\widetilde{\theta}_j - \theta_j^{*(1)})\\
= \; & (S1) + (S2).
\end{aligned}
$$

Below, we consider $j = 1, ..., 4$ separately. When $j = 1$, the expression in the braces has mean zero because of the orthogonality of Hermite polynomials. Further, this expression is not affected by the regime switching. Therefore, $(S1)$ and $(S2)$ are both $O_p(T^{-1/2}A_T)$ in this case. When $j = 2$, $(S1) = 0$ because $\theta_2^{*(1)} = \theta_2^{*(s_t)}$, and $(S2) = O_p\left(A_T^2 + T^{-1/2}\right)$ because $\widetilde{\sigma}_*^2 - \sigma_*^2 = O_p(A_T^2 + T^{-1/2})$. When $j = 3$, $(S1)$ and $(S2)$ are $O_p(T^{-1/2}A_T)$ for the same reason as when $j = 1$. When $j = 4$, $(S1) = (S2) = 0$ because f_{1t} does not depend on δ_2. Therefore, the Lemma holds when $k = 2$.

We now consider the $k = 3$ case. Apply a second order Taylor expansion around $\theta^{*(s_t)}$:

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}\widetilde{f}_{1t}}{\widetilde{f}_{1t}}\\
= \; & T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}} f_{1t}^*}{f_{1t}^*}\\
& + T^{-1}\sum_{j=1}^{4}\sum_{t=1}^{T}\left\{\frac{\nabla_{\theta_j}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}} f_{1t}^*}{f_{1t}^*} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}} f_{1t}^* \nabla_{\theta_j} f_{1t}^*}{f_{1t}^{*2}}\right\}(\widetilde{\theta}_j - \theta_j^{*(s_t)})\\
& + \frac{1}{2T}\sum_{j_1,j_2=1}^{4}\sum_{t=1}^{T}\nabla_{\theta_{j_2}}\left(\frac{\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}} f_{1t}^*}{f_{1t}^*} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}} f_{1t}^* \nabla_{\theta_{j_1}} f_{1t}^*}{f_{1t}^{*2}}\right)(\widetilde{\theta}_{j_1} - \theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2} - \theta_{j_2}^{*(s_t)})\\
& + O_p\left(A_T^3\right),
\end{aligned}
$$

where $\nabla_{\theta_{j_2}}(\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}} f_{1t}^*/f_{1t}^*)$ denotes the derivative of $\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}} f_{1t}/f_{1t}$ with respect to θ_{j_2} evaluated at the true parameter value. The first term on the right hand side is $O_p(T^{-1/2})$ by the CLT; the second is $O_p(T^{-1/2}A_T)$ by the orthogonality of Hermite polynomials.

For the third term, we apply $\widetilde{\theta}_j - \theta_j^{*(s_t)} = (\theta_j^{*(1)} - \theta_j^{*(s_t)}) + (\widetilde{\theta}_j - \theta_j^{*(1)})$ as in (A.144):

$$\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\nabla_{\theta_{j_2}}\left(\frac{\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}}{f_{1t}^{*}} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}\nabla_{\theta_j}f_{1t}^{*}}{f_{1t}^{*2}}\right)(\widetilde{\theta}_{j_1} - \theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2} - \theta_{j_2}^{*(s_t)}) \\
= \; & T^{-1}\sum_{t=1}^{T}\nabla_{\theta_{j_2}}\left(\frac{\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}}{f_{1t}^{*}} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}\nabla_{\theta_{j_1}}f_{1t}^{*}}{f_{1t}^{*2}}\right)(\widetilde{\theta}_{j_1} - \theta_{j_1}^{*(1)})(\widetilde{\theta}_{j_2} - \theta_{j_2}^{*(1)}) \\
& + T^{-1}\sum_{t=1}^{T}\nabla_{\theta_{j_2}}\left(\frac{\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}}{f_{1t}^{*}} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}\nabla_{\theta_{j_1}}f_{1t}^{*}}{f_{1t}^{*2}}\right)(\theta_{j_2}^{*(1)} - \theta_{j_2}^{*(s_t)})(\widetilde{\theta}_{j_1} - \theta_{j_1}^{*(1)}) \\
& + T^{-1}\sum_{t=1}^{T}\nabla_{\theta_{j_2}}\left(\frac{\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}}{f_{1t}^{*}} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}\nabla_{\theta_{j_1}}f_{1t}^{*}}{f_{1t}^{*2}}\right)(\theta_{j_1}^{*(1)} - \theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2} - \theta_{j_2}^{*(1)}) \\
& + T^{-1}\sum_{t=1}^{T}\nabla_{\theta_{j_2}}\left(\frac{\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}}{f_{1t}^{*}} - \frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}\nabla_{\theta_{j_1}}f_{t}^{*}}{f_{1t}^{*2}}\right)(\theta_{j_1}^{*(1)} - \theta_{j_1}^{*(s_t)})(\theta_{j_2}^{*(1)} - \theta_{j_2}^{*(s_t)}) \\
= \; & (S3) + (S4) + (S5) + (S6).
\end{aligned}$$

The expression $(S3)$ equals

$$\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\left\{\frac{\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}}{f_{1t}^{*}} - \frac{\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}\nabla_{\theta_{j_2}}f_{1t}^{*}}{f_{1t}^{*2}}\right. \\
& \qquad \left. - \frac{\nabla_{\theta_{j_2}}\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}\nabla_{\theta_{j_1}}f_{1t}^{*}}{f_{1t}^{*2}}\right\}(\widetilde{\theta}_{j_1} - \theta_{j_1}^{*(1)})(\widetilde{\theta}_{j_2} - \theta_{j_2}^{*(1)}) \\
& - T^{-1}\sum_{t=1}^{T}\left\{\frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}f_{1t}^{*}}{f_{1t}^{*2}}\right\}(\widetilde{\theta}_{j_1} - \theta_{j_1}^{*(1)})(\widetilde{\theta}_{j_2} - \theta_{j_2}^{*(1)}) \\
& + 2T^{-1}\sum_{t=1}^{T}\left\{\frac{\nabla_{\theta_{i_1}}\nabla_{\theta_{i_2}}\nabla_{\theta_{i_3}}f_{1t}^{*}\nabla_{\theta_{j_1}}f_{t}^{*}\nabla_{\theta_{j_2}}f_{1t}^{*}}{f_{1t}^{*3}}\right\}(\widetilde{\theta}_{j_1} - \theta_{j_1}^{*(1)})(\widetilde{\theta}_{j_2} - \theta_{j_2}^{*(1)}).
\end{aligned}$$

We study the three sample averages separately. The three terms inside the first braces are mean zero by the orthogonality of Hermite polynomials. In addition, $(\widetilde{\theta}_{j_1} - \theta_{j_1}^{*(1)})(\widetilde{\theta}_{j_2} - \theta_{j_2}^{*(1)}) = O_p(A_T^2)$. Therefore, the first sample average is $O_p(A_T^2 T^{-1/2})$. For the second sample average, when $j_1 \neq 2$ and $j_2 \neq 2$, the expression in the second braces have mean zero, which implies that it is $O_p(T^{-1/2}A_T^2)$. When $j_1 = 2$ or $j_2 = 2$, $(\widetilde{\theta}_{j_1} - \theta_{j_1}^{*(1)})(\widetilde{\theta}_{j_2} - \theta_{j_2}^{*(1)}) = O_p\left(A_T^3\right)$, which implies that it is $O_p(A_T^3)$. The expression in the third braces has a more complex structure in that it equals the product of three Hermite polynomials. By the integral formula of the Hermite polynomial triple product (7.375.2 on p.804 of Gradshteyn and Ryzhik, 2007), this expression has mean zero except when $j_1 = 2$ or $j_2 = 2$. Therefore, this term is also $O_p\left(A_T^3 + T^{-1/2}A_T^2\right)$. The analysis of $(S4)$ is simpler because $\theta_{j_2}^{*(1)} - \theta_{j_2}^{*(s_t)} = 0$ except when $j_2 = 3, 4$. We have $(S4) = O_p\left(A_T^3 + T^{-1/2}\right)$. The terms $(S5)$ and $(S6)$ can be analyzed similarly. They are also $O_p\left(A_T^3 + T^{-1/2}\right)$; we omit the details.

The proof for $k=4$ is similar to $k=3$, except that we apply a third order Taylor expansion:

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}\widetilde{f}_{1t}}{\widetilde{f}_{1t}} \qquad \text{(A.145)}\\
= \; & T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f^*_{1t}}{f^*_{1t}}+T^{-1}\sum_{j=1}^{4}\sum_{t=1}^{T}\left\{\frac{\nabla_{\theta_j}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f^*_{1t}}{f^*_{1t}}-\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f^*_{1t}\nabla_{\theta_j}f^*_{1t}}{f^{*2}_{1t}}\right\}(\widetilde{\theta}_j-\theta_j^{*(s_t)})\\
& +\frac{1}{2}T^{-1}\sum_{j_1,j_2=1}^{4}\sum_{t=1}^{T}\nabla_{\theta_{j_2}}\left(\frac{\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f^*_{1t}}{f^*_{1t}}\right)(\widetilde{\theta}_{j_1}-\theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2}-\theta_{j_2}^{*(s_t)})\\
& -\frac{1}{2}T^{-1}\sum_{j_1,j_2=1}^{4}\sum_{t=1}^{T}\left\{\nabla_{\theta_{j_2}}\left(\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f^*_{1t}\nabla_{\theta_{j_1}}f^*_{1t}}{f^{*2}_{1t}}\right)\right\}(\widetilde{\theta}_{j_1}-\theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2}-\theta_{j_2}^{*(s_t)})\\
& +\frac{1}{3!}T^{-1}\sum_{j_1,j_2,j_3=1}^{4}\sum_{t=1}^{T}\nabla_{\theta_{j_2}}\nabla_{\theta_{j_3}}\left(\frac{\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f^*_{1t}}{f^*_{1t}}\right)(\widetilde{\theta}_{j_1}-\theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2}-\theta_{j_2}^{*(s_t)})(\widetilde{\theta}_{j_3}-\theta_{j_3}^{*(s_t)})\\
& -\frac{1}{3!}T^{-1}\sum_{j_1,j_2,j_3=1}^{4}\sum_{t=1}^{T}\nabla_{\theta_{j_2}}\nabla_{\theta_{j_3}}\left(\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f^*_{1t}\nabla_{\theta_{j_1}}f^*_{1t}}{f^{*2}_{1t}}\right)(\widetilde{\theta}_{j_1}-\theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2}-\theta_{j_2}^{*(s_t)})(\widetilde{\theta}_{j_3}-\theta_{j_3}^{*(s_t)})\\
& +O_p\left(A_T^4\right).
\end{aligned}
$$

The first three terms on the right hand side are $O_p\left(T^{-1/2}\right)$ by the CLT and the orthogonality of Hermite polynomials. The fourth term is $O_p\left(T^{-1/2}+A_T^4\right)$ because the expression in the braces are mean zero except when $j_1=j_2=2$. The fifth term on the right hand side satisfies

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\nabla_{\theta_{j_2}}\nabla_{\theta_{j_3}}\left(\frac{\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f^*_{1t}}{f^*_{1t}}\right)(\widetilde{\theta}_{j_1}-\theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2}-\theta_{j_2}^{*(s_t)})(\widetilde{\theta}_{j_3}-\theta_{j_3}^{*(s_t)})\\
= \; & T^{-1}\sum_{t=1}^{T}\left\{\nabla_{\theta_{j_3}}\left(\frac{\nabla_{\theta_{j_2}}\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f^*_{1t}}{f^*_{1t}}\right)\right\}(\widetilde{\theta}_{j_1}-\theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2}-\theta_{j_2}^{*(s_t)})(\widetilde{\theta}_{j_3}-\theta_{j_3}^{*(s_t)})\\
& -T^{-1}\sum_{t=1}^{T}\left\{\frac{\nabla_{\theta_{j_3}}\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f^*_{1t}\nabla_{\theta_{j_2}}f^*_{1t}+\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f^*_{1t}\nabla_{\theta_{j_2}}\nabla_{\theta_{j_3}}f^*_{1t}}{f^{*2}_{1t}}\right\}\times\\
& \qquad (\widetilde{\theta}_{j_1}-\theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2}-\theta_{j_2}^{*(s_t)})(\widetilde{\theta}_{j_3}-\theta_{j_3}^{*(s_t)})\\
& +2T^{-1}\sum_{t=1}^{T}\left\{\frac{\nabla_{\theta_{j_1}}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f^*_{1t}\nabla_{\theta_{j_2}}f^*_{1t}\nabla_{\theta_{j_3}}f^*_{1t}}{f^{*3}_{1t}}\right\}(\widetilde{\theta}_{j_1}-\theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2}-\theta_{j_2}^{*(s_t)})(\widetilde{\theta}_{j_3}-\theta_{j_3}^{*(s_t)}).
\end{aligned}
$$

The expressions in the three sets of curly brackets all have mean zero. Therefore, this term is

$O_p\left(T^{-1/2}A_T^3\right)$. Finally, consider the sixth term of (A.145):

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\nabla_{\theta_{j_2}}\nabla_{\theta_{j_3}}\left(\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f_{1t}^{*}\nabla_{\theta_{j_1}}f_{1t}^{*}}{f_{1t}^{*2}}\right)(\widetilde{\theta}_{j_1}-\theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2}-\theta_{j_2}^{*(s_t)})(\widetilde{\theta}_{j_3}-\theta_{j_3}^{*(s_t)}) \\
= \; & T^{-1}\sum_{t=1}^{T}\left\{\nabla_{\theta_{j_3}}\left(\frac{\nabla_{\theta_{j_2}}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f_{1t}^{*}\nabla_{\theta_{j_1}}f_{1t}^{*}}{f_{1t}^{*2}}\right)\right\}(\widetilde{\theta}_{j_1}-\theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2}-\theta_{j_2}^{*(s_t)})(\widetilde{\theta}_{j_3}-\theta_{j_3}^{*(s_t)}) \\
& +T^{-1}\sum_{t=1}^{T}\left\{\nabla_{\theta_{j_3}}\left(\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f_{1t}^{*}\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}f_{1t}^{*}}{f_{1t}^{*2}}\right)\right\}(\widetilde{\theta}_{j_1}-\theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2}-\theta_{j_2}^{*(s_t)})(\widetilde{\theta}_{j_3}-\theta_{j_3}^{*(s_t)}) \\
& -2T^{-1}\sum_{t=1}^{T}\left\{\nabla_{\theta_{j_3}}\left(\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f_{1t}^{*}\nabla_{\theta_{j_1}}f_{1t}^{*}\nabla_{\theta_{j_2}}f_{1t}^{*}}{f_{1t}^{*3}}\right)\right\}(\widetilde{\theta}_{j_1}-\theta_{j_1}^{*(s_t)})(\widetilde{\theta}_{j_2}-\theta_{j_2}^{*(s_t)})(\widetilde{\theta}_{j_3}-\theta_{j_3}^{*(s_t)}).
\end{aligned}
$$

The expression in the first braces has mean zero. The expression in the second braces equals

$$
\frac{\nabla_{\theta_{j_3}}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f_{1t}^{*}\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}f_{1t}^{*}}{f_{1t}^{*2}}+\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f_{1t}^{*}\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}\nabla_{\theta_{j_3}}f_{1t}^{*}}{f_{1t}^{*2}}-2\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f_{1t}^{*}\nabla_{\theta_{j_1}}\nabla_{\theta_{j_2}}f_{1t}^{*}\nabla_{\theta_{j_3}}f_{1t}^{*}}{f_{1t}^{*3}},
$$

which has mean zero when none of j_1, j_2 and j_3 equals 2. The expression in the third braces equals

$$
\begin{aligned}
& -2\frac{\nabla_{\theta_{j_3}}\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f_{1t}^{*}\nabla_{\theta_{j_1}}f_{1t}^{*}\nabla_{\theta_{j_2}}f_{1t}^{*}}{f_{1t}^{*3}}-2\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f_{1t}^{*}\nabla_{\theta_{j_1}}\nabla_{\theta_{j_3}}f_{1t}^{*}\nabla_{\theta_{j_2}}f_{1t}^{*}}{f_{1t}^{*3}} \\
& -2\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f_{1t}^{*}\nabla_{\theta_{j_1}}f_{1t}^{*}\nabla_{\theta_{j_2}}\nabla_{\theta_{j_3}}f_{1t}^{*}}{f_{1t}^{*3}}+6\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}f_{1t}^{*}\nabla_{\theta_{j_1}}f_{1t}^{*}\nabla_{\theta_{j_2}}f_{1t}^{*}\nabla_{\theta_{j_3}}f_{1t}^{*}}{f_{1t}^{*4}}.
\end{aligned} \tag{A.146}
$$

Among the four terms, the first has means zero, and the remaining terms have mean zero when none of j_1, j_2 and j_3 is equal to 2. Therefore, (A.145) is $O_p\left(T^{-1/2}+A_T^4\right)$.

The proof for $k=5,...8$ is essentially the same as the $k=4$ case. That is, we apply the third order Taylor expansion and then repeat the analysis (A.145)-(A.146) with $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_4}}$ replaced by $\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}$ for $k=5,...,8$. The details are omitted. ■

Lemma A.26 *Suppose the data are generated by (A.141)-(A.142). Then, for* $j\in\{1,2\}$,

$$
T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\widetilde{f}_{jt}}{\widetilde{f}_{jt}}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\widetilde{\xi}_{t|t-1}=O_p\left(T^{-1/2}|\rho_T|+A_T^2\rho_T^2\right),
$$

where $k,m\in\{1,...,7\}$, $k+m\leq 8$, *and* $i_1,...,i_k,j_1,...,j_m\in\{1,...,4\}$.

Proof of Lemma A.26. The proof is similar to that of Lemma A.25. The main difference is that an extra Taylor expansion is used to study $\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\widetilde{\xi}_{t|t-1}$. Let $\nabla_{\theta}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\bar{\xi}_{t|t-1}^{*}$ denote the derivative of $\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\xi_{t|t-1}$ with respect to θ evaluated at $\theta^{*(1)}=(0,\sigma_*^2,0,0)'$, which satisfies the conditions of Lemma A.10 because $\delta_1=\delta_2=0$. Without loss of generality, we suppose $j=1, \mu_*=0$, and $c_*>0$. Recall $\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\widetilde{\xi}_{t+1|t}=\rho_T\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\widetilde{\xi}_{t|t-1}+\widetilde{\mathcal{E}}_{j_1...j_m,t}$, where $\widetilde{\mathcal{E}}_{j_1...j_m,t}$ denotes $\bar{\mathcal{E}}_{j_1...j_m,t}$ evaluated at the restricted MLE.

When $k=1$, we have

$$T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\widetilde{f}_{1t}}{\widetilde{f}_{1t}}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\widetilde{\xi}_{t|t-1} \tag{A.147}$$
$$= T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\widetilde{f}_{1t}}{\widetilde{f}_{1t}}\widetilde{\mathcal{E}}_{j_1...j_m,t-1}+\rho_T T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\widetilde{f}_{1t}}{\widetilde{f}_{1t}}\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\widetilde{\xi}_{t-1|t-2}.$$

Regarding $\widetilde{\mathcal{E}}_{j_1...j_m,t-1}$, there are two possibilities: (i) It does not depend on the derivatives of $\widetilde{\xi}_{t-1|t-2}$. (ii) It depends on such derivative, (e.g., $\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_{m-1}}}\widetilde{\xi}_{t-1|t-2}$), c.f., the expressions in Lemma 1. Under (i), the first term on the right hand side of (A.147) is $O_p\left(T^{-1/2}|\rho_T|+A_T^2\rho_T^2\right)$ by direct calculations. Under (ii), we study it using a first order Taylor expansion around $\theta^{*(1)}$:

$$T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\widetilde{f}_{1t}}{\widetilde{f}_{1t}}\widetilde{\mathcal{E}}_{j_1...j_m,t-1} = T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\widetilde{f}_{1t}}{\widetilde{f}_{1t}}\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1} \tag{A.148}$$
$$+T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\widetilde{f}_{1t}}{\widetilde{f}_{1t}}\nabla_{\theta'}\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1}(\widetilde{\theta}-\theta^{*(1)})+O_p\left(A_T^2\rho_T^2\right),$$

where $\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1}$ denotes $\widetilde{\mathcal{E}}_{j_1...j_m,t}$ evaluated at $\theta^{*(1)}$. We consider the cases $i_1=1,2,3,4$ separately. When $i_1=1$, by (A.143),

$$\frac{\nabla_{\theta_{i_1}}\widetilde{f}_{1t}}{\widetilde{f}_{1t}}=\frac{y_{t-1}}{\widetilde{\sigma}^2}\left\{\widetilde{e}_t-\widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}+A_T\left(\widetilde{s}_t-\widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)\right\},$$

which implies

$$T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\widetilde{f}_{1t}}{\widetilde{f}_{1t}}\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1}$$
$$= \frac{1}{\widetilde{\sigma}^2}T^{-1}\sum_{t=1}^{T}y_{t-1}\widetilde{e}_t\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1}-\frac{1}{\widetilde{\sigma}^2}T^{-1}\sum_{t=1}^{T}y_{t-1}\widetilde{y}_{t-1}\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}$$
$$+\frac{1}{\widetilde{\sigma}^2}A_T T^{-1}\sum_{t=1}^{T}y_{t-1}\widetilde{s}_t\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1}-\frac{1}{\widetilde{\sigma}^2}A_T T^{-1}\sum_{t=1}^{T}y_{t-1}\widetilde{y}_{t-1}\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}.$$

The first two terms on the right hand side are $O_p(|\rho_T|\,T^{-1/2})$ by the CLT. The third and fourth terms are $O_p(\rho_T^2 A_T\left(T^{-1/2}+A_T\right))$ by the definition of $\widetilde{s}_t$. When $i_1=2$,

$$T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_{i_1}}\widetilde{f}_{1t}}{\widetilde{f}_t}\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1}$$
$$= T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\sigma^2}f^{*}_{1t}}{f^{*}_t}\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1}+\sum_{i=1}^{4}T^{-1}\sum_{t=1}^{T}\frac{\nabla_{\theta_i}\nabla_{\sigma^2}f^{*}_{1t}}{f^{*}_t}\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1}(\widetilde{\theta}_i-\theta_i^{*(s_t)})$$
$$-T^{-1}\sum_{i=1}^{4}\sum_{t=1}^{T}\frac{\nabla_{\sigma^2}f^{*}_{1t}\nabla_{\theta_i}f^{*}_{1t}}{f^{*2}_t}\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1}(\widetilde{\theta}_i-\theta_i^{*(s_t)})+O_p\left(\rho_T^2A_T^2\right),$$

where the first term is $O_p(|\rho_T|\, T^{-1/2})$; the second term can be studied in the same way as (A.144), and it is $O_p(|\rho_T|\, T^{-1/2})$; the third term is $O_p(\rho_T^2(T^{-1/2}+A_T^2))$ because $E(\nabla_{\sigma^2} f_{1t}^* \nabla_{\theta_i} f_{1t}^*)/f_{1t}^*|\Omega_{t-1}) = 0$ except when $i = 2$ for which $||\widetilde{\theta}_2 - \theta_2^{*(s_t)}|| = O_p\left(A_T^2\right)$. When $i_1 = 3$, the proof is the same way as the $i_1 = 1$ case after replacing y_{t-1} with 1. When $i_1 = 4$, $\nabla_{\theta_{i_1}} \widetilde{f}_{1t} = 0$ because f_{1t} does not depend on δ_2. This completes the analysis of the first right hand side term of (A.148). The analysis of the second term on the right hand side is the same as the first term after replacing $\bar{\mathcal{E}}^*_{j_1 \dots j_m, t-1}$ by $\nabla_{\theta'} \bar{\mathcal{E}}^*_{j_1 \dots j_m, t-1}$. It follows that it is $O_p\left(T^{-1/2}|\rho_T| + A_T^2 \rho_T^2\right)$. Finally, $\rho_T T^{-1} \sum_{t=1}^T (\nabla_{\theta_{i_1}} \widetilde{f}_{1t}/\widetilde{f}_{1t}) \nabla_{\theta_{j_1}} \dots \nabla_{\theta_{j_m}} \widetilde{\xi}_{t-1|t-2} = O_p\left(T^{-1/2}|\rho_T| + A_T^2 \rho_T^2\right)$. Therefore, the Lemma holds when $k = 1$.

When $k > 1$, we have

$$
\begin{aligned}
& T^{-1} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} \dots \nabla_{\theta_{i_k}} \widetilde{f}_{1t}}{\widetilde{f}_{1t}} \nabla_{\theta_{j_1}} \dots \nabla_{\theta_{j_m}} \widetilde{\xi}_{t|t-1} \\
= \; & T^{-1} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} \dots \nabla_{\theta_{i_k}} \widetilde{f}_{1t}}{\widetilde{f}_{1t}} \widetilde{\mathcal{E}}_{j_1 \dots j_m, t-1} + \rho_T T^{-1} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} \dots \nabla_{\theta_{i_k}} \widetilde{f}_{1t}}{\widetilde{f}_{1t}} \nabla_{\theta_{j_1}} \dots \nabla_{\theta_{j_m}} \widetilde{\xi}_{t-1|t-2}
\end{aligned}
$$

As in the $k = 1$ case, under (i) the first term on the right hand side is $O_p\left(T^{-1/2}|\rho_T| + A_T^2 \rho_T^2\right)$. Under (ii), we study it using a first order Taylor expansion around $\theta^{*(1)}$:

$$
\begin{aligned}
& T^{-1} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} \dots \nabla_{\theta_{i_k}} \widetilde{f}_{1t}}{\widetilde{f}_{1t}} \widetilde{\mathcal{E}}_{j_1 \dots j_m, t-1} \\
= \; & T^{-1} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} \dots \nabla_{\theta_{i_k}} \widetilde{f}_{1t}}{\widetilde{f}_{1t}} \bar{\mathcal{E}}^*_{j_1 \dots j_m, t-1} + T^{-1} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} \dots \nabla_{\theta_{i_k}} \widetilde{f}_{1t}}{\widetilde{f}_{1t}} \nabla_{\theta'} \bar{\mathcal{E}}^*_{j_1 \dots j_m, t-1} (\widetilde{\theta} - \theta^{*(1)}) + O_p\left(A_T^2 \rho_T^2\right).
\end{aligned}
$$

Next, by a first order Taylor expansion applied to $\nabla_{\theta_{i_1}} \dots \nabla_{\theta_{i_k}} \widetilde{f}_{1t}/\widetilde{f}_{1t}$, the first term on the right hand side can be represented as

$$
\begin{aligned}
& T^{-1} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} \dots \nabla_{\theta_{i_k}} f_{1t}^*}{f_{1t}^*} \bar{\mathcal{E}}^*_{j_1 \dots j_m, t-1} + T^{-1} \sum_{i=1}^{4} \sum_{t=1}^{T} \frac{\nabla_{\theta_i} \nabla_{\theta_{i_1}} \dots \nabla_{\theta_{i_k}} f_{1t}^*}{f_{1t}^*} \bar{\mathcal{E}}^*_{j_1 \dots j_m, t-1} (\widetilde{\theta}_i - \theta_i^{*(s_t)}) \\
& - T^{-1} \sum_{i=1}^{4} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} \dots \nabla_{\theta_{i_k}} f_{1t}^* \nabla_{\theta_i} f_{1t}^*}{f_{1t}^{*2}} \bar{\mathcal{E}}^*_{j_1 \dots j_m, t-1} (\widetilde{\theta}_i - \theta_i^{*(s_t)}) + O_p\left(A_T^2 \rho_T^2\right).
\end{aligned}
$$

Among the three leading terms, the first satisfies a CLT and is $O_p(T^{-1/2}\, |\rho_T|)$, the second is $O_p(T^{-1/2}\, |\rho_T|\, A_T)$ by the analysis of (A.144), and the third can be written as

$$
\begin{aligned}
& T^{-1} \sum_{i=1}^{4} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} \dots \nabla_{\theta_{i_k}} f_{1t}^* \nabla_{\theta_i} f_{1t}^*}{f_{1t}^{*2}} \bar{\mathcal{E}}^*_{j_1 \dots j_m, t-1} (\widetilde{\theta}_i - \theta_i^{*(1)}) \\
& + T^{-1} \sum_{i=1}^{4} \sum_{t=1}^{T} \frac{\nabla_{\theta_{i_1}} \dots \nabla_{\theta_{i_k}} f_{1t}^* \nabla_{\theta_i} f_{1t}^*}{f_{1t}^{*2}} \bar{\mathcal{E}}^*_{j_1 \dots j_m, t-1} (\theta_i^{*(1)} - \theta_i^{*(s_t)}).
\end{aligned}
$$

These two expressions both satisfy a CLT except when $i = 2$ and $k = 2$ and $i_1, i_2 \in \{1, 3\}$ for which $\theta_i^{*(1)} - \theta_i^{*(s_t)} = 0$ and $\widetilde{\theta}_i - \theta_i^{*(1)} = O_p(A_T^2)$. Therefore, they are both $O_p(\rho^2|(T^{-1/2} + A_T^2))$. Finally, by the same line of argument, the terms $T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\widetilde{f}_{1t})/\widetilde{f}_{1t}\nabla_{\theta'}\bar{\mathcal{E}}^{*}_{j_1...j_m,t-1}(\widetilde{\theta}-\theta^{*(1)})$ and $\rho_T T^{-1}\sum_{t=1}^{T}(\nabla_{\theta_{i_1}}...\nabla_{\theta_{i_k}}\widetilde{f}_{1t}/\widetilde{f}_{1t})\nabla_{\theta_{j_1}}...\nabla_{\theta_{j_m}}\widetilde{\xi}_{t-1|t-2}$ are both $O_p\left(T^{-1/2}|\rho_T|+A_T^2\rho_T^2\right)$. Therefore, the result in the Lemma holds when $k>1$.■

Remark 11 *The next lemma establishes results analogous to those in Lemmas A.12-A.16, under the alternative hypothesis for the DGP (A.141)-(A.142).*

Lemma A.27 *Suppose that the DGP is (A.141)-(A.142). Then, the results of Lemmas A.12-A.16 hold after their remainder terms are replaced by the following*

1. *Lemma A.12:* $O_p\left(T^{-2a}\rho_T^4+T^{-1/2}\left|\rho_T\right|\right)$;
2. *Lemma A.13:* $O_p\left(T^{-1/2}\left|\rho_T\right|+T^{-2a}\rho_T^4+T^{-5a}+T^{-1}\right)$;
3. *Lemma A.14:* $O_p\left(T^{-3a}+T^{-1/2}+\left|\rho_T\right|\right)$;
4. *Lemma A.15:* $O_p\left(T^{-3a}+T^{-1/2}+\left|\rho_T\right|\right)$;
5. *Lemma A.16: No change is needed.*

Proof of Lemma A.27. The proof is similar to that of Lemmas A.12-A.16. The main difference is that Lemmas A.25 and A.26 are applied to account for the departure from the null hypothesis.

To prove the first result, note that in the proof of Lemma A.12, the analysis up to (A.78) does not involve any approximation. To proceed further, we apply Lemma A.26 to the two terms on the right hand side of (A.78). This gives (A.79), except that the remainder term is now $O_p\left(T^{-1/2}\rho_T^2+T^{-2a}\rho_T^4+T^{-1}|\rho_T|\right)$ instead of $O_p\left(T^{-1/2}\left|\rho_T\right|\right)$. The rest of the proof is the same as that of Lemma A.12.

To prove the second result, we make the following three changes to the proof of Lemma A.13: (i) replacing its references to Lemma A.12 by references to Lemma A.27.1; (ii) replacing $T^{-1/2}\left|\rho_T\right|$ by $T^{-1/2}\left|\rho_T\right|+T^{-2a}\rho_T^4$ whenever this appears in any expression; (iii) replacing the remaining $T^{-1/2}$ with $T^{-1/2}+T^{-2a}$. After these three changes, the proof goes through, which implies that the remainder term is $O_p\left(T^{-1/2}\left|\rho_T\right|+T^{-2a}\rho_T^4+T^{-4a}\right)$. Finally, we iterate on (A.83) using the expressions of $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})$ and $\nabla_{\delta_{2k}}\nabla_{\delta_{2l}}\hat{\beta}(\widetilde{\delta})$ and $\alpha_{kl}=(0,2,0)$. This shows that the remainder term can be improved to $O_p\left(T^{-1/2}\left|\rho_T\right|+T^{-2a}\rho_T^4+T^{-5a}+T^{-1}\right)$.

For the third result, note that in the proof of Lemma A.14, the analysis up to (A.85) goes through after making the following two changes: (i) replacing Lemmas A.12 and A.13 by Lemmas A.27.1 and A.27.2; (ii) replacing $T^{-1/2}$ and T^{-1} by T^{-2a} and T^{-4a} whenever they appear in any expression. Apply Lemma A.25 to the last two terms in the braces of (A.85). From this, (A.86) follows except that the remainder term is $O_p\left(T^{-3a}+T^{-1/2}+T^{-2a}\left|\rho_T\right|\right)$ instead of $O_p\left(T^{-1/2}\right)$. Consequently, (A.88) holds with a remainder term $O_p\left(T^{-3a}+T^{-2a}\left|\rho_T\right|+T^{-1/2}\right)$ instead of $O_p\left(T^{-1/2}\right)$. The rest of the proof goes through after replacing $T^{-1/2}$ by $\left(T^{-3a}+T^{-1/2}+T^{-2a}\left|\rho_T\right|\right)$ whenever $T^{-1/2}$ appears in an expression. The result follows from the final displayed equation in that proof.

The fourth result follows from the proof of Lemma A.15 after making the following three changes to it: (i) replacing the references to Lemma A.12, A.13, and A.14 by Lemma A.27.1, A.27.2, and

A.27.3, respectively; (ii) replacing $T^{-1/2}|\rho_T|$ by $T^{-2a}|\rho_T|$ whenever it appears in the proof; (iii) replacing the remaining $T^{-1/2}$ by $T^{-3a}+T^{-2a}|\rho_T|$ whenever it appears in the proof.

The fifth result follows from the proof of Lemma A.27.5 without any modification. ■

Remark 12 *The next lemma establishes the stochastic orders of $\mathcal{L}^{(j)}(p_T,q_T,\widetilde{\delta})$ $(j=2,...,8)$ and their leading terms for the DGP (A.141)-(A.142). These results are important for understanding the local asymptotic power properties of $LR(p_T,q_T)$ under various drifting sequences of (p_T,q_T). In particular, they allow us to compare the power when (p_T,q_T) is bounded away from the boundary (34) with when it converges to (34) it as $T\to\infty$; see Lemma A.29.*

Lemma A.28 *Suppose the data are generated by (A.141)-(A.142). Then,*

1. $T^{-1/2}\mathcal{L}^{(2)}(p_T,q_T,\widetilde{\delta})=T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T,q_T)+O_p\left(T^{1/2-4a}\rho_T^4+T^{-2a}\rho_T^2\right)$,
2. $T^{-1/2}\mathcal{L}^{(3)}(p_T,q_T,\widetilde{\delta})=T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(3)}(p_T,q_T)+O_p(T^{1/2-3a}\rho_T^2(|\rho_T|+|p_T-q_T|)+\rho_T^2+|\rho_T(p_T-q_T)|+T^{-2a}|\rho_T|)$,
3. $T^{-1/2}\mathcal{L}^{(4)}(p_T,q_T,\widetilde{\delta})=T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(4)}(p_T,q_T)-3T^{1/2}\widetilde{\Omega}(p_T,q_T)+O_p(|\rho_T|+T^{-a}+(T^{1/2-2a}\rho_T^2+T^{1/2-3a}|\rho_T|)(|\rho_T|+|\rho_T-q_T|))$,
4. $T^{-1}\mathcal{L}^{(5)}(p_T,q_T,\widetilde{\delta})=O_p(T^{-1/2}+\rho_T^4+|\rho_T^3(p_T-q_T)|+T^{-2a}\rho_T^2+T^{-3a}(|p_T-q_T|+|\rho_T|))$,
5. $T^{-1}\mathcal{L}^{(6)}(p_T,q_T,\widetilde{\delta})=-10\widetilde{\Omega}^{(3)}(p_T,q_T)+O_p(T^{-1/2}+|\rho_T^3|+\rho_T^2|p_T-q_T|+T^{-a}\rho_T^2+|\rho_T|(p_T-q_T)^2+T^{-3a})$,
6. $T^{-1}\mathcal{L}^{(7)}(p_T,q_T,\widetilde{\delta})=O_p(\rho_T^2+|(p_T-q_T)\rho_T|+(p_T-q_T)^2+|p_T-q_T|T^{-a}+T^{-2a})$,
7. $T^{-1}\mathcal{L}^{(8)}(p_T,q_T,\widetilde{\delta})=-35\widetilde{\Omega}^{(4)}(p_T,q_T)+O_p(|\rho_T|+|p_T-q_T|)+o_p(1)$,
8. $T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T,q_T)=O_p\left(\rho_T^2+T^{-1/2}|\rho_T|+T^{1/2-2a}\rho_T^4\right)$,
9. $T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(3)}(p_T,q_T)=O_p\left(|p_T-q_T|+T^{1/2-3a}(p_T-q_T)^2\right)+o_p\left(T^{1/2-3a}|p_T-q_T|\right)$,
10. $T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(4)}(p_T,q_T)=O_p\left(1+T^{1/2-4a}\right)$,
11. $\widetilde{\Omega}(p_T,q_T)=O_p\left(\rho_T^4+T^{-1/2}\rho_T^2\right)$, $\widetilde{\Omega}^{(3)}(p_T,q_T)=O_p((p_T-q_T)^2)$, *and* $\widetilde{\Omega}^{(4)}(p_T,q_T)=O_p(1)$.

where, for $k=2,3,$and 4, $\widetilde{U}_t^{(k)}(p_T,q_T)$ are given by (33) with (p,q) replaced by (p_T,q_T), $\widetilde{\Omega}^{(k)}(p_T,q_T)=\widetilde{V}^{(k)}-\widetilde{D}^{(k)\prime}\widetilde{I}^{-1}\widetilde{D}^{(k)}$, $\widetilde{V}^{(k)}=T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(k)}(p_T,q_T)^2$, $\widetilde{D}^{(k)}=T^{-1}\sum_{t=1}^{T}\widetilde{S}_t\widetilde{U}_t^{(k)}(p_T,q_T)$, and $\widetilde{I}=T^{-1}\sum_{t=1}^{T}\widetilde{S}_t\widetilde{S}_t'$, where $\widetilde{S}_t=[y_{t-1}\widetilde{u}_t/\widetilde{\sigma}^2,\left(\widetilde{u}_t^2/\widetilde{\sigma}^2-1\right)/(2\widetilde{\sigma}^2),\widetilde{u}_t/\widetilde{\sigma}^2]'$.

Proof of Lemma A.28. The proof is similar to that of Lemma A.17. The main difference is that Lemmas A.25, A.26, and A.27 are used to account for the departure from the null hypothesis.

Part 1: Proof of Lemma A.28.1. In the proof of Lemma A.17, (A.101) does not rely on any approximation. Its right hand side can be rewritten as

$$2T^{-1/2}\sum_{t=1}^{T}\frac{\nabla_{\delta_1}\widetilde{f}_{1t}\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T^2}+\left\{T^{-1/2}\sum_{t=1}^{T}\left(\frac{(\nabla_{\theta'}\widetilde{f}_{1t}-\nabla_{\theta'}\widetilde{f}_{2t})\nabla_{\delta_1}\widetilde{\xi}_{t|t-1}}{\widetilde{f}_t\xi_T}+\frac{\nabla_{\delta_2}\widetilde{f}_{2t}}{\widetilde{f}_t\xi_T}\nabla_{\theta'}\widetilde{\xi}_{t|t-1}\right)\right\}R_T,$$

where $R_T=O_p\left(T^{-2a}\rho_T^4+T^{-1/2}\left|\rho_T\right|\right)$ by Lemma A.27.1, and the expression in the braces is $O_p((1+T^{1/2-2a})|\rho_T|)$ by Lemma A.26. Thus, their product is $O_p((T^{-2a}\rho_T^4+T^{-1/2}\left|\rho_T\right|)(1+T^{1/2-2a})|\rho_T|)=O_p\left(T^{1/2-4a}\rho_T^4+T^{-2a}\rho_T^2\right)$, where the equality holds because $a\in[1/8,1/4]$.

Part 2: Proof of Lemma A.28.2. As in the proof of Lemma A.17, we consider (A.39) with $T^{-3/4}$ replaced by $T^{-1/2}$. The last two terms equal zero when $\delta_2=\widetilde{\delta}$. The second and third terms are $O_p(T^{-2a}\rho_T^2+T^{1/2-4a}\rho_T^4+T^{-1/2-2a}\left|\rho_T\right|+\left|\rho_T^3\right|)$ by (A.81) and Lemma A.27.1. For the first term in (A.39), we apply (A.102) and (A.103). This leads to (A.104) except that the remainder term is $O_p(\rho_T^4+T^{-2a}\left|\rho_T\right|)$ instead of $O_p\left(\left|\rho_T\right|\right)$. (M1)-(M6) in (A.104) can be studied in the same way as in Part 2 of Lemma A.17, which shows that they are $O_p(\left|\rho_T\left(p_T-q_T\right)\right|+T^{1/2-3a}\left|\rho_T^2\left(p_T-q_T\right)\right|)$. Further, (M0) equals $[\left(1-\xi_T\right)\left(2\xi_T-1\right)/\xi_T^2]T^{-1/2}\sum_{t=1}^{T}\nabla_{\delta_1}^3\widetilde{f}_{1t}/\widetilde{f}_t+O_p(\rho_T^4+T^{-2a}\left|\rho_T\right|))$. Finally, the remaining two terms in (A.104) are $O_p\left(T^{-2a}\left|\rho_T\right|+\rho_T^4\right)$.

Part 3. Proof of Lemma A.28.3. As in Part 3 of Lemma A.17, We divide the components of (A.41) into three subsets. For the first subset, (A.106) holds after replacing its remainder term $O_p(|\rho_T|)$ by $O_p(|\rho_T|+T^{-a}+T^{1/2-2a}\rho_T^2(|\rho_T|+|p_T-q_T|))$. This follows from Lemma A.25, Lemma A.26, and the expressions in Lemma 1. The components in the second subset are $O_p(|\rho_T|+T^{1/2}T^{-2a}\rho_T^4)$ by Lemma A.27.1. For the third subset, the decomposition below (A.107) holds after replacing its remainder term $O_p\left(\left|\rho_T\right|\right)$ by $O_p\left(\left|\rho_T\right|+T^{1/2-2a}\rho_T^4\right)$. Subsequently, the two products with (N1) and (N2) still produce the expressions (A.108) and (A.109). The products with (N3) and (N4) produces the same leading terms as in (A.110) and (A.111) with an overall remainder term of $O_p(|\rho_T|+T^{-a}+T^{1/2-3a}|\rho_T|(|\rho_T|+|p_T-q_T|))$. The Lemma follows by combining these results.

Part 4. Proof of Lemma A.28.4. As in Part 4 of the proof of Lemma A.17, we divide the components of $T^{-1}\mathcal{L}^{(5)}(p_T,q_T,\widetilde{\delta})-((1-\xi_T)/\xi_T)\mathcal{M}^{(5)}(p_T,q_T,\widetilde{\delta})$ into four subsets $(i=1,2,3,4)$. For $i=1$, the components are $O_p(T^{-1/2}+T^{-2a}\left|\rho_T\right|+T^{-3a}|p_T-q_T|)$ because of Lemmas A.25 and A.26. For $i=2$, they are $O_p(T^{-1/2}\left|\rho_T\right|+T^{-2a}\rho_T^4)$ by (A.81) and Lemma A.27.1. For $i=3$, (A.112) holds after replacing $O_p(T^{-1/2}\left|\rho_T\right|)$ with $O_p(T^{-1/2}\left|\rho_T\right|+T^{-2a}\rho_T^4)$. Further, (A.113) holds, where the effect of R_t on $T^{-1}\mathcal{L}^{(5)}(p_T,q_T,\widetilde{\delta})$ is $O_p(T^{-1/2}+\rho_T^4+|\rho_T^3(p_T-q_T)|+T^{-3a}|p_T-q_T|)$. The components in this subset are therefore $O_p(T^{-1/2}+\rho_T^4+|\rho_T^3(p_T-q_T)|+T^{-3a}|p_T-q_T|+T^{-2a}\rho_T^2|p_T-q_T|)$. For $i=4$, (A.114) can be studied as in Part 4 of the proof of Lemma A.17. In particular, $T^{-1}\sum_{t=1}^{T}\nabla_{\delta_2}^3\widetilde{B}_t\widetilde{M}_{it}/\widetilde{B}_t^2$ and $T^{-1}\sum_{t=1}^{T}\nabla_{\delta_2}^3\widetilde{B}_t\nabla_{\theta_s}\widetilde{f}_{1t}\nabla_{\theta_u}\widetilde{\xi}_{t|t-1}/\widetilde{f}_t^2$ are $O_p(T^{-1/2}+\rho_T^4+T^{-3a}|\rho_T|+T^{-2a}\rho_T^2)$ and $O_p(T^{-1/2}+\rho_T^4+\rho_T^3\left|p_T-q_T\right|+T^{-2a}\rho_T^2+T^{-3a}|p_T-q_T|)$, which implies that the components in this subset are $O_p(T^{-1/2}+\rho_T^4+\rho_T^3\left|p_T-q_T\right|+T^{-2a}\rho_T^2+T^{-3a}|\rho_T|+T^{-3a}|p_T-q_T|)$.

Part 5. Proof of Lemma A.28.5. As in Part 5 of Lemma A.17, we divide the components of $T^{-1}\mathcal{L}^{(6)}(p_T,q_T,\widetilde{\delta})-((1-\xi_T)/\xi_T)\mathcal{M}^{(6)}(p_T,q_T,\widetilde{\delta})$ into 5 subsets and study them separately. The components in the first two subsets are $O_p(T^{-1/2}+T^{-2a}\,|\rho_T|+T^{-3a}|p_T-q_T|)$ and $O_p(T^{-1/2}\,|\rho_T|+T^{-2a}\rho_T^4)$, respectively, for the same reason as the $T^{-1}\mathcal{L}^{(5)}(p_T,q_T,\widetilde{\delta})$ case. For the third subset, (A.116) holds after $O_p(T^{-1/2})$ is replaced by $O_p(T^{-1/2}+\rho_T^4+T^{-3a})$. Consequently, the components in this subset are $O_p(|\rho_T^3|+T^{-1/2}+T^{-2a}\rho_T^2)$.

For the fourth subset, the displayed equation below (A.118) holds after $T^{-1/2}$ is replaced by $T^{-1/2}+T^{-2a}\rho_T^4$ because of Lemma A.27.1. The next displayed result holds after $O_p(T^{-1/2}+\rho_T^2+|\rho_T|\,(p_T-q_T)^2)$ is replaced by $O_p(T^{-1/2}+\rho_T^4+|\rho_T|\,(p_T-q_T)^2+\rho_T^2|p_T-q_T|+T^{-a}\rho_T^2+T^{-3a})$. It follows from these two results and (A.119) that (A.118) equals $-10\widetilde{\Omega}^{(3)}(p_T,q_T)+O_p(T^{-1/2}+\rho_T^4+|\rho_T|\,(p_T-q_T)^2+\rho_T^2|p_T-q_T|+T^{-a}\rho_T^2+T^{-3a})$. Finally, for the fifth subset, (A.120) can studied in the same way as in Part 4 of the proof. It follows that these terms are $O_p\left(T^{-1/2}+|\rho_T^3|+\rho_T^2T^{-a}+T^{-3a}\right)$.

Part 6. Proof of Lemma A.28.6. As in Part 6 of the proof of Lemma A.17, we divide the components of $T^{-1}\mathcal{L}^{(7)}(p_T,q_T,\widetilde{\delta})-((1-\xi_T)/\xi_T)\mathcal{M}^{(7)}(p_T,q_T,\widetilde{\delta})$ into 6 subsets. The components in the first two subsets are $O_p(T^{-1/2}+T^{-2a}\,|\rho_T|+T^{-3a}|p_T-q_T|)$ and $O_p(T^{-1/2}\,|\rho_T|+T^{-2a}\rho_T^4)$, respectively, for the same reason as the $T^{-1}\mathcal{L}^{(5)}(p_T,q_T,\widetilde{\delta})$ case. Those in the third subset are $O_p(T^{-1/2}+\rho_T^2)$ by (A.85).

For the fourth subset, (A.117) holds, where the overall effect of R_t is $O_p(\rho_T^2+T^{-1/2}+T^{-2a}+|\rho_T|\,|p_T-q_T|)$, and (A.119) holds after replacing $T^{-1/2}R_{4t}$ by $(T^{-1/2}+T^{-2a})R_{4t}$. Thus, the components in this subset are $O_p(\rho_T^2+T^{-1/2}+T^{-2a}+|\rho_T|\,|p_T-q_T|+T^{-a}|p_T-q_T|)$.

For the fifth subset, (A.123) holds after replacing $O_p(|\rho_T|)$ by $O_p(|p_T-q_T|\,|\rho_T|+\rho_T^2+T^{-1/2})$. Among the two leading terms, the first is $O_p(|(p_T-q_T)\rho_T|+T^{-1/2}+\rho_T^2+T^{-2a}+|p_T-q_T|\,T^{-a})$ and the second is $O_p(T^{-1/2}+\rho_T^2+T^{-2a})$. Therefore, the components in this subset are $O_p(T^{-1/2}+\rho_T^2+|(p_T-q_T)\rho_T|+|p_T-q_T|\,T^{-a}+T^{-2a})$. Finally, the components in the sixth subset are $O_p(T^{-1/2}+\rho_T^2+T^{-2a}+(p_T-q_T)^2)$ implied by the proof of Lemma A.18.

Part 7. Proof of Proof of Lemma A.28.7. The proof is the same as in Part 7 of Lemma A.17.

Part 8. Proof of Lemma A.28.8-11. These four results follow after applying standard calculations to $\widetilde{U}_t^{(j)}(p_T,q_T)$ and $\widetilde{\omega}^{(j)}(p_T,q_T)$ for $j=2,3,4$. We omit the details. ■

Recall, by (37),

$$\begin{aligned}
\widetilde{\mathcal{W}}^{(2)}(p_T,q_T,\eta) &= \left(T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T,q_T)\right)\eta^2-\frac{1}{4}\widetilde{\Omega}(p_T,q_T)\eta^4,\\
\widetilde{\mathcal{W}}^{(3)}(p_T,q_T,\eta) &= T^{-1/4}\frac{1}{3}\left(T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(3)}(p_T,q_T)\right)\eta^3-T^{-1/2}\frac{1}{36}\widetilde{\Omega}^{(3)}(p_T,q_T)\eta^6,\\
\widetilde{\mathcal{W}}^{(4)}(p_T,q_T,\eta) &= T^{-1/2}\frac{1}{12}\left(T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(4)}(p_T,q_T)\right)\eta^4-T^{-1}\frac{1}{576}\widetilde{\Omega}^{(4)}(p_T,q_T)\eta^8,
\end{aligned}$$

where $\widetilde{U}_t^{(k)}(p_T,q_T)$ and $\widetilde{\Omega}^{(k)}(p_T,q_T)$ are defined in Lemma A.28.

Remark 13 *The next lemma presents approximations to $LR(p_T,q_T)$ for seven different cases: (p_T,q_T) is bounded away from (34) and (p_T,q_T) follows one of the six sequences specified in Proposition 2. In the first case, the local alternatives are of order $T^{-1/4}$, and the asymptotic distribution of $LR(p_T,q_T)$ is determined by the second and fourth derivatives of the log likelihood. In the remaining six cases, the order of the local alternatives varies between $T^{-1/4}$ and $T^{-1/8}$, and the asymptotic distributions can depend on higher order derivatives of the log likelihood.*

Lemma A.29 *Suppose that the DGP is (A.141)-(A.142) with $A_T = c_* T^{-\gamma}$, where $\gamma \in [1/4, 1/8]$, and that there exists some $\eta > 0$ such that $\sup_{|\delta-\widetilde{\delta}|<\eta} T^{-1}|\mathcal{L}^{(j)}(p_T,q_T,\delta)| = O_p(1)$ for $j = 5,7,9$.*

1. *If (p_T,q_T) is fixed and belongs to Λ_ϵ, then the contiguous alternatives of $LR(p_T,q_T)$ are of order $T^{-1/4}$. Further, if $\gamma = 1/4$, then $\Pr(LR(p_T,q_T) \le x) - \Pr(\sup_{\eta\in R} \widetilde{\mathcal{W}}^{(2)}(p_T,q_T,\eta) \le x) \to 0$ for any $x \in R$.*

2. *If (p_T,q_T) follows SEQ1 for some $a > 0$, then the contiguous alternatives of $LR(p_T,q_T)$ are of order $T^{\min(-1/4+a,-1/6)}$. Further, if $\gamma = \min(1/4-a, 1/6)$, then $\Pr(LR(p_T,q_T) \le x) - \Pr(\widetilde{S}_a(p_T,q_T) \le x) \to 0$ for $x \in R$, where*

$$\widetilde{S}_a(p_T,q_T) = \begin{cases} \sup_{\eta\in R} \widetilde{\mathcal{W}}^{(2)}(p_T,q_T,\eta) & \text{if } a < 1/12 \\ \sup_{\eta\in R}\{\widetilde{\mathcal{W}}^{(2)}(p_T,q_T,\eta) + \widetilde{\mathcal{W}}^{(3)}(p_T,q_T,\eta)\} & \text{if } a = 1/12 \\ \sup_{\eta\in R} \widetilde{\mathcal{W}}^{(3)}(p_T,q_T,\eta) & \text{if } a > 1/12 \end{cases}.$$

3. *If (p_T,q_T) follows SEQ2 for some $a > 0$, then the contiguous alternatives of $LR(p_T,q_T)$ are of order $T^{\min(-1/4+a,-1/8)}$. Further, if $\gamma = \min(1/4-a, 1/8)$, then $\Pr(LR(p_T,q_T) \le x) - \Pr(\widetilde{S}_a(p_T,q_T) \le x) \to 0$ for $x \in R$, where*

$$\widetilde{S}_a(p_T,q_T) = \begin{cases} \sup_{\eta\in R} \widetilde{\mathcal{W}}^{(2)}(p_T,q_T,\eta) & \text{if } a < 1/8 \\ \sup_{\eta\in R}\{\widetilde{\mathcal{W}}^{(2)}(p_T,q_T,\eta) + \widetilde{\mathcal{W}}^{(3)}(p_T,q_T,\eta) + \widetilde{\mathcal{W}}^{(4)}(p_T,q_T,\eta)\} & \text{if } a = 1/8 \\ \sup_{\eta\in R} \widetilde{\mathcal{W}}^{(4)}(p_T,q_T,\eta) & \text{if } a > 1/8 \end{cases}.$$

Proof of Lemma A.29. The proof is similar to that of Proposition 2. The role of Lemma A.17 in Proposition 2 is now played by Lemma A.28. Given this similarity, we will omit some details.

Proof of Lemma A.29.1. For alternatives that exceed $O_p(T^{-1/4})$, i.e., $A_T = c_* T^{-\gamma}$ with $1/8 \le \gamma < 1/4$, $\mathcal{L}^{(2)}(p_T,q_T,\widetilde{\delta})(\hat{\delta}_2 - \widetilde{\delta})^2$ diverges with positive probability because the orders of $\mathcal{L}^{(2)}(p_T,q_T,\widetilde{\delta})$ and $\hat{\delta}_2 - \widetilde{\delta}$ are higher than or equal to $O_p(T^{1-2\gamma})$ and $O_p(T^{-1/4})$, respectively. This implies that the order of the contiguous alternatives can not exceed $T^{-1/4}$. Thus, we can focus on $\gamma \ge 1/4$ in the subsequent analysis.

Consider the following Taylor expansion around $\widetilde{\delta}$:

$$\mathcal{L}(p_T,q_T,\delta_2) - \mathcal{L}(p_T,q_T,\widetilde{\delta}) = \sum_{k=2}^{4} \frac{1}{k!}\mathcal{L}^{(k)}(p_T,q_T,\widetilde{\delta})(\delta_2 - \widetilde{\delta})^k + \frac{1}{5!}\mathcal{L}^{(5)}(p_T,q_T,\bar{\delta})(\delta_2 - \widetilde{\delta})^5,$$

where $\bar{\delta} = \widetilde{\delta} + c(\delta_2 - \widetilde{\delta})$ with $c \in (0,1)$. Because (p_T,q_T) belongs to Λ_ϵ, $\widetilde{\Omega}(p_T,q_T)$ is strictly positive in probability. Given this property, the consistency of $\hat{\delta}_2$, and the boundedness of $T^{-1}\mathcal{L}^{(5)}(p_T,q_T,\delta)$,

$\hat{\delta}_2$ must satisfy $\hat{\delta}_2 = O_p\left(T^{-1/4}\right)$, otherwise $LR(p_T, q_T)$ will be negative with probability close to one in large samples. Thus, we can restrict our attention to the set $\{\delta_2 \in R : (|\delta_2| \leq CT^{-1/4}\}$, where C can be made sufficiently large. Define $\eta = T^{1/4}\delta_2$. Then, by Lemma A.28, uniformly over this set for any $\gamma \geq 1/4$:

$$\begin{aligned}
\mathcal{L}^{(2)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^2 &= T^{-1/2}\sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T)\,\eta^2 + o_p(1), \\
\mathcal{L}^{(4)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^4 &= -3\widetilde{\Omega}(p_T, q_T)\eta^4 + o_p(1), \\
\mathcal{L}^{(3)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^3 &= o_p(1),\ \mathcal{L}^{(5)}(p_T, q_T, \bar{\delta})(\delta_2 - \widetilde{\delta})^5 = o_p(1),
\end{aligned}$$

where $T^{-1/2}\sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T)$ and $\widetilde{\Omega}(p_T, q_T)$ are both $O_p(1)$ but not $o_p(1)$. Apply these four approximations to the preceding Taylor expansion. We have

$$\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) = \frac{1}{2}\left(T^{-1/2}\sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T)\right)\eta^2 - \frac{1}{8}\widetilde{\Omega}(p_T, q_T)\eta^4 + o_p(1).$$

The result in the Lemma follows after multiplying both sides by 2 and computing the supremum over η. Note that the asymptotic distribution of $\sup_{\eta\in R}\widetilde{\mathcal{W}}^{(2)}(p_T, q_T, \eta)$ is different from the one under the null hypothesis when $\gamma = 1/4$, but it is the same as the latter when $\gamma > 1/4$. This confirms that the contiguous alternatives are indeed of order $T^{-1/4}$.

Proof of Lemma A.29.2 when $a < 1/12$**.** The proof is similar to that of Lemma A.29.1. If $\gamma < 1/4 - a$, then $\mathcal{L}^{(2)}(p_T, q_T, \widetilde{\delta})(\hat{\delta}_2 - \widetilde{\delta})^2$ diverges with positive probability because $\mathcal{L}^{(2)}(p_T, q_T, \widetilde{\delta})$ and $\hat{\delta}_2 - \widetilde{\delta}$ exceed the orders $O_p\left(T^{1-2\gamma-4a}\right)$ and $O_p\left(T^{-\gamma}\right)$, respectively. This implies that the order of the local alternatives can not exceed $T^{-1/4+a}$. Thus, we can focus on $\gamma \geq 1/4 - a$.

Consider the following Taylor expansion around $\widetilde{\delta}$:

$$\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) = \sum_{k=2}^{6}\frac{1}{k!}\mathcal{L}^{(k)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^k + \frac{1}{5!}\mathcal{L}^{(7)}(p_T, q_T, \bar{\delta})(\delta_2 - \widetilde{\delta})^7,$$

where $\bar{\delta} = \widetilde{\delta} + c(\delta_2 - \widetilde{\delta})$ with $c \in (0, 1)$. Because of $\widetilde{\Omega}^{(3)}(p_T, q_T)$ is strictly positive and $T^{-1}\mathcal{L}^{(7)}(p_T, q_T, \delta)$ is bounded in probability, and $\hat{\delta}_2$ is consistent, $\hat{\delta}_2$ must satisfy $(\hat{\delta}_2 - \widetilde{\delta}) = O_p(T^{-1/6})$; otherwise the log likelihood ratio will be negative with probability close to 1 in large samples. Thus, we can restrict our attention to the following set $\{\delta_2 \in R : |\delta_2| \leq CT^{-1/6}\}$, where C can be made sufficiently large. Then, uniformly over this set, the following results hold for any $\gamma \geq 1/4 - a$:

$$\begin{aligned}
\mathcal{L}^{(2)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^2 &= \sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T)\,(\delta_2 - \widetilde{\delta})^2 + o_p(1), \\
\mathcal{L}^{(3)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^3 &= \sum_{t=1}^{T} \widetilde{U}_t^{(3)}(p_T, q_T)\,(\delta_2 - \widetilde{\delta})^3 + o_p(1) = O_p(1), \\
\mathcal{L}^{(4)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^4 &= -3T\widetilde{\Omega}(p_T, q_T)(\delta_2 - \widetilde{\delta})^4 + o_p(1), \\
\mathcal{L}^{(5)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^5 &= T^{1/2}\rho_T^2(\delta_2 - \widetilde{\delta})^2 * o_p(1) + o_p(1), \\
\mathcal{L}^{(6)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^6 &= O_p(1).
\end{aligned}$$

This implies that $\mathcal{L}^{(4)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^4$ diverges and dominates the other terms unless $T^{1/2}\rho_T^2(\delta_2 - \widetilde{\delta})^2 = O_p(1)$, or $\delta_2 - \widetilde{\delta} = O_p\left(T^{-1/4+a}\right)$, otherwise $\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta})$ will be negative with probability close to one in large samples.

Given this result, we can further restrict our attention to the set $\{\delta_2 \in R : (|\delta_2| \leq CT^{-1/4+a}\}$ with C being a sufficiently large constant. Then,

$$\begin{aligned}
\mathcal{L}^{(2)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^2 &= \sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^2 + o_p(1), \\
\mathcal{L}^{(4)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^4 &= -3\widetilde{\Omega}(p_T, q_T)T(\delta_2 - \widetilde{\delta})^4 + o_p(1), \\
\mathcal{L}^{(k)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^k &= o_p(1) \text{ for } k = 3, 5, 6.
\end{aligned}$$

where the leading terms in the first two equations are both $O_p(1)$ but not $o_p(1)$. This result determines the leading terms in the expansion, implying

$$\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) = \frac{1}{2}\left(T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T, q_T)\right)\eta^2 - \frac{1}{8}\widetilde{\Omega}(p_T, q_T)\eta^4 + o_p(1),$$

where $\eta = T^{1/4}(\delta_2 - \widetilde{\delta})$. The result in the Lemma follows after multiplying both sides by 2 and computing the supremum over η. Note that the asymptotic distribution of $\sup_{\eta \in R}\widetilde{\mathcal{W}}^{(2)}(p_T, q_T, \eta)$ is different from the one under the null hypothesis when $\gamma = 1/4 - a$, and it is the same as the latter if $\gamma > 1/4 - a$. This confirms that the contiguous alternatives are indeed of order $T^{-1/4+a}$.

Proof of Lemma A.29.2 when $a = 1/12$**.** In this case, the order of the contiguous alternatives can not exceed $O_p(T^{-1/6})$, and $\hat{\delta}_2$ must satisfy $\hat{\delta}_2 = O_p(T^{-1/6})$ for the same reason as in the $a < 1/12$ case. Thus, we can restrict our attention to the set $\{\delta_2 \in R : (|\delta_2| \leq CT^{-1/6}\}$ with C being a sufficiently large constant. Over this set, we have

$$\begin{aligned}
\mathcal{L}^{(2)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^2 &= \sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^2 + o_p(1), \\
\mathcal{L}^{(3)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^3 &= \sum_{t=1}^{T} \widetilde{U}_t^{(3)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^3 + o_p(1), \\
\mathcal{L}^{(4)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^4 &= -3\widetilde{\Omega}(p_T, q_T)T(\delta_2 - \widetilde{\delta})^4 + o_p(1), \\
\mathcal{L}^{(6)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^6 &= -10\widetilde{\Omega}^{(3)}(p_T, q_T)T(\delta_2 - \widetilde{\delta})^6 + o_p(1), \\
\mathcal{L}^{(5)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^k &= o_p(1),
\end{aligned}$$

where the leading terms in the first four equations are all $O_p(1)$ but not $o_p(1)$. These results pin down the leading terms in the expansion, implying

$$\begin{aligned}
\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) &= \frac{1}{2}\left(T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T, q_T)\right)\eta^2 + T^{-1/4}\frac{1}{6}\left(T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(3)}(p_T, q_T)\right)\eta^3 \\
&\quad -\frac{1}{8}\widetilde{\Omega}(p_T, q_T)\eta^4 - \frac{1}{72}T^{-1/2}\widetilde{\Omega}^{(3)}(p_T, q_T)\eta^6 + o_p(1),
\end{aligned}$$

where $\eta = T^{1/4}(\delta_2 - \widetilde{\delta})$. The Lemma follows after multiplying both sides by 2 and computing the supremum over η. Note that the asymptotic distribution of $\sup_{\eta\in R}\{\widetilde{\mathcal{W}}^{(2)}(p_T, q_T, \eta) + \widetilde{\mathcal{W}}^{(3)}(p_T, q_T, \eta)\}$ is different from the one under the null hypothesis when $\gamma = 1/6$, and it is the same as the latter if $\gamma > 1/6$. This confirms that the contiguous alternatives are of order $T^{-1/6}$.

Proof of Lemma A.29.2 when $a > 1/12$. As in the previous case, the order of the local alternatives can not exceed $T^{-1/6}$, and $\hat{\delta}_2$ must satisfy $\hat{\delta}_2 = O_p(T^{-1/6})$. Thus, we can restrict our attention to the set $\{\delta_2 \in R : (|\delta_2| \leq CT^{-1/6}\}$ with C being sufficiently large. We have

$$\begin{aligned}
\mathcal{L}^{(3)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^3 &= \sum_{t=1}^{T} \widetilde{U}_t^{(3)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^3 + o_p(1), \\
\mathcal{L}^{(6)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^6 &= 10\widetilde{\Omega}^{(3)}(p_T, q_T)T(\delta_2 - \widetilde{\delta})^6 + o_p(1), \\
\mathcal{L}^{(k)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^k &= o_p(1) \text{ for } k = 2, 4, 5.
\end{aligned}$$

These results imply

$$\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) = T^{-1/4}\frac{1}{6}\left(T^{-1/2}\sum_{t=1}^{T} \widetilde{U}_t^{(3)}(p_T, q_T)\right)\eta^3 - \frac{1}{72}T^{-1/2}\widetilde{\Omega}^{(3)}(p_T, q_T)\eta^6 + o_p(1),$$

where $\eta = T^{1/4}(\delta_2 - \widetilde{\delta})$. The expression in the Lemma follows after multiplying both sides by 2 and computing the supremum over η. Note that the asymptotic distribution of $\sup_{\eta\in R}\widetilde{\mathcal{W}}^{(3)}(p_T, q_T, \eta)$ is different from the one under the null hypothesis when $\gamma = 1/6$, and it is the same as the latter if $\gamma > 1/6$. This confirms that the contiguous alternatives are of order $T^{-1/6}$.

Proof of Lemma A.29.3 when $a < 1/8$. Similarly to the $a < 1/12$ case, $\mathcal{L}^{(2)}(p_T, q_T, \widetilde{\delta})(\hat{\delta}_2 - \widetilde{\delta})^2$ diverges with positive probability if $\gamma < 1/4 - a$. This implies that the order of the local alternatives can not exceed $T^{-1/4+a}$. Thus, we can focus on $\gamma \geq 1/4 - a$.

As in the $a < 1/12$ case, we first establish the convergence rate of $\hat{\delta}_2$. Because $\widetilde{\Omega}^{(4)}(p_T, q_T)$ is strictly positive and $T^{-1}\mathcal{L}^{(9)}(p_T, q_T, \delta)$ is bounded in probability, $\hat{\delta}_2$ must satisfy $\hat{\delta}_2 = O_p(T^{-1/8})$. Thus, we can restrict the attention to the set $\{\delta_2 \in R^{n_\delta} : (\|\delta_2 - \delta_*\| \leq CT^{-1/8}\}$, where C can be made sufficiently large. Over this set, the following expansion holds:

$$\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) = \sum_{k=2}^{8} \frac{1}{k!}\mathcal{L}^{(k)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^k + o_p(1). \tag{A.149}$$

Over this set, we have

$$\begin{aligned}
\mathcal{L}^{(2)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^2 &= \sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^2 + o_p(1), \\
\mathcal{L}^{(4)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^4 &= -3T\widetilde{\Omega}(p_T, q_T)(\delta_2 - \widetilde{\delta})^4 + o_p(1),
\end{aligned}$$

where the two leading terms on the right hand side are $(\delta_2 - \widetilde{\delta})^2 * O_p((T\rho_T^4)^{1/2})$ and $(\delta_2 - \widetilde{\delta})^4 * O_p(T\rho_T^4)$, respectively. These two terms diverge unless $(\delta_2 - \widetilde{\delta})^4 T\rho_T^4 = O_p(1)$. Further,

$$\begin{aligned}
\mathcal{L}^{(3)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^3 &= (\delta_2 - \widetilde{\delta})^2 * o_p((T\rho_T^4)^{1/2}), \\
\mathcal{L}^{(k)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^k &= (\delta_2 - \widetilde{\delta})^4 * o_p\left(T\rho_T^4\right) \text{ for } k = 5, 6, 7, \\
\mathcal{L}^{(8)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^8 &= -35\widetilde{\Omega}^{(4)}(p_T, q_T)\, T(\delta_2 - \widetilde{\delta})^8 + o_p(1).
\end{aligned}$$

Therefore, we must have $\hat{\delta}_2 = O_p\left(T^{-1/4}|\rho_T|^{-1}\right)$; otherwise $\mathcal{L}^{(4)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^4$ dominates the other terms, and consequently $\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta})$ will be negative with probability close to one in large samples.

Given this rate of convergence, the rest of the proof is the same as the $a < 1/12$ case, because only the second and fourth derivative terms are non-negligible. The order of the contiguous alternatives is also the same as that case.

Proof of Lemma A.29.3 when $a = 1/8$. The proof is similar to the $a = 1/12$ case. The order of the local alternatives can not exceed $T^{-1/8}$, and $\hat{\delta}_2$ must satisfy $\hat{\delta}_2 = O_p(T^{-1/8})$. Thus, we can restrict our attention to the set $\{\delta_2 \in R : (|\delta_2| \leq CT^{-1/8}\}$ with C being sufficiently large. We have

$$\begin{aligned}
\mathcal{L}^{(2)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^2 &= \sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^2 + o_p(1), \\
\mathcal{L}^{(3)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^3 &= \sum_{t=1}^{T} \widetilde{U}_t^{(3)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^3 + o_p(1), \\
\mathcal{L}^{(4)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^4 &= \sum_{t=1}^{T} \widetilde{U}_t^{(4)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^4 - 3T\widetilde{\Omega}(p_T, q_T)(\delta_2 - \widetilde{\delta})^4 + o_p(1), \\
\mathcal{L}^{(6)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^6 &= 10\widetilde{\Omega}^{(3)}(p_T, q_T)T(\delta_2 - \widetilde{\delta})^6 + o_p(1), \\
\mathcal{L}^{(8)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^8 &= -35T\widetilde{\Omega}^{(4)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^8 + o_p(1), \\
\mathcal{L}^{(k)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^k &= o_p(1) \quad \text{for } k = 5, 7.
\end{aligned}$$

where the three leading terms on the right hand side are all $O_p(1)$ but not $o_p(1)$. This result determines the leading terms in the expansion (A.131), implying

$$\begin{aligned}
&\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) \\
= &\frac{1}{2}\sum_{t=1}^{T} \widetilde{U}_t^{(2)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^2 + \frac{1}{3!}\sum_{t=1}^{T} \widetilde{U}_t^{(3)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^3 + \frac{1}{4!}\sum_{t=1}^{T} \widetilde{U}_t^{(4)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^4 \\
&- \frac{3}{4!}T\widetilde{\Omega}(p_T, q_T)(\delta_2 - \widetilde{\delta})^4 - \frac{10}{6!}T\widetilde{\Omega}^{(3)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^6 - \frac{35}{8!}T\widetilde{\Omega}^{(4)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^8 + o_p(1).
\end{aligned}$$

The result follows after letting $\eta = T^{1/4}(\delta_2 - \widetilde{\delta})$, multiplying both sides by 2, and computing the supremum over η. Note that the resulting asymptotic distribution is different from the one under the null hypothesis when $\gamma = 1/8$, and it is the same as the latter if $\gamma > 1/8$. This confirms that the contiguous alternatives are of order $T^{-1/8}$.

Proof of Lemma A.29.3 when $\mathbf{a > 1/8}$. The proof is very similar to the $a > 1/8$ case. The order of the contiguous alternatives can not exceed $T^{-1/8}$, and $\hat{\delta}_2$ must satisfy $\hat{\delta}_2 = O_p(T^{-1/8})$. Thus, we can restrict our attention to the set $\{\delta_2 \in R : (|\delta_2| \leq CT^{-1/8}\}$ with C being a sufficiently large constant. Over this set, $\mathcal{L}^{(k)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^k = o_p(1)$ for $k = 2, 3, 5, 6, 7$, and

$$\begin{aligned}
\mathcal{L}^{(4)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^4 &= \sum_{t=1}^{T} \widetilde{U}_t^{(4)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^4 + o_p(1), \\
\mathcal{L}^{(8)}(p_T, q_T, \widetilde{\delta})(\delta_2 - \widetilde{\delta})^8 &= -35T\widetilde{\Omega}^{(4)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^8 + o_p(1),
\end{aligned}$$

where the two leading terms on the right hand side are $O_p(1)$ but not $o_p(1)$. Thus,

$$\mathcal{L}(p_T, q_T, \delta_2) - \mathcal{L}(p_T, q_T, \widetilde{\delta}) = \frac{1}{4!}\sum_{t=1}^{T} \widetilde{U}_t^{(4)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^4 - \frac{35}{8!}T\widetilde{\Omega}^{(4)}(p_T, q_T)(\delta_2 - \widetilde{\delta})^8 + o_p(1).$$

The result follows after letting $\eta = T^{1/4}(\delta_2 - \widetilde{\delta})$, multiplying both sides by 2, and computing the supremum over η. Note that the resulting asymptotic distribution is different from the one under the null hypothesis when $\gamma = 1/8$, and it is the same as the latter if $\gamma > 1/8$. This confirms that the contiguous alternatives are of order $T^{-1/8}$. ■

Remark 14 *Lemma A.29 implies that, in order to obtain an adequate approximation to the LR test under the DGP (38)-(39) over a wide range of transition probabilities, it is important to account for the effect of the second, third, fourth, sixth, and eighth order derivatives of the log likelihood ratio. The next lemma studies the leading terms of these derivatives to deliver the terms needed for the refined approximation.*

Lemma A.30 *Suppose that the DGP is (38) and the model (39) is estimated. Then,*

1. $T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_*, q_*) = \frac{2}{\sigma_*^2}\left(\frac{1-p_*}{1-q_*}\right)T^{-1/2}\sum_{t=1}^{T}\{\sum_{s=2}^{t-1}\rho_*^s\frac{\widetilde{e}_t\widetilde{e}_{t-s}}{\sigma_*^2} + (T^{-1/4}c_*)^2\sum_{s=2}^{t-1}\rho_*^s\frac{\widetilde{s}_t\widetilde{s}_{t-s}}{\sigma_*^2}\}$ $+R_T + o_p(T^{-1/2})$,
2. $T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(3)}(p_*, q_*) = \frac{1}{\sigma_*^3}\frac{(1-p_*)(p_*-q_*)}{(1-q_*)^2}T^{-1/2}\sum_{t=1}^{T}\left\{\frac{\widetilde{e}_t^3}{\sigma_*^3} + (T^{-1/4}c_*)^3\frac{\widetilde{s}_t^3}{\sigma_*^3}\right\} + R_T$,
3. $T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(4)}(p_*, q_*) = \frac{1}{\sigma_*^4}\left\{\frac{1-p_*}{2-p_*-q_*}\left(1 + \left(\frac{1-p_*}{1-q_*}\right)^3\right) - 3\left(\frac{1-p_*}{1-q_*}\right)^2\right\}T^{-1/2}\left\{\sum_{t=1}^{T}\left(\frac{\widetilde{e}_t^4}{\sigma_*^4} - 6\frac{\widetilde{e}_t^2}{\sigma_*^2} + 3\right)\right.$ $\left. + \left(T^{-1/4}c_*\right)^4\left(\sum_{t=1}^{T}\frac{\widetilde{s}_t^4}{\sigma_*^4} - 3\left(T^{-1/2}\sum_{t=1}^{T}\frac{\widetilde{s}_t^2}{\sigma_*^2}\right)^2\right)\right\} + R_T$,
4. $\widetilde{\Omega}(p_*, q_*) = \frac{4}{\sigma_*^4}\left(\frac{1-p_*}{1-q_*}\right)^2 T^{-1}\sum_{t=1}^{T}\left(\sum_{s=2}^{t-1}\rho_*^s\frac{\widetilde{e}_{t-s}\widetilde{e}_t}{\sigma_*^2}\right)^2 + R_T + o_p(T^{-1/2})$,
5. $\widetilde{\Omega}^{(3)}(p_*, q_*) = \frac{1}{\sigma_*^6}\frac{(1-p_*)^2(p_*-q_*)^2}{(1-q_*)^4}T^{-1}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{e}_t}{\sigma_*}\right)^3 - 3\frac{\widetilde{e}_t}{\sigma_*}\right\}^2 + R_T$,
6. $\widetilde{\Omega}^{(4)}(p_*, q_*) = \frac{1}{\sigma_*^8}\left\{\frac{1-p_*}{2-p_*-q_*}\left(1 + \left(\frac{1-p_*}{1-q_*}\right)^3\right) - 3\left(\frac{1-p_*}{1-q_*}\right)^2\right\}^2 T^{-1}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{e}_t}{\sigma_*}\right)^4 - 6\left(\frac{\widetilde{e}_t}{\sigma_*}\right)^2 + 3\right\}^2 +$ R_T,

where R_T represents a remainder term that can differ between the cases. Importantly, R_T is dominated by leading term preceding it, not only under the DGP (38), but also when $(p_*, q_*, \rho_*, T^{-1/4})$ in (38) and the above equations are replaced by $(p_T, q_T, \rho_T, T^{-a})$ for any $a \in [1/8, 1/4]$ and any (p_T, q_T) that follows one of the six sequences specified in Proposition 2.

Proof of Lemma A.29. We prove the six results separately.

Proof of Lemma A.30.1 In this result, the leading term on the right hand side is of exact order $O_p(1)$ under (38) and it is $O_p\left(\rho_T^2 + T^{1/2-2a}\rho_T^4\right)$ when $(p_*, q_*, \rho_*, T^{-1/4})$ is replaced by $(p_T, q_T, \rho_T, T^{-a})$. Below, we show that R_T is $O_p(T^{-1/4})$ under (38) and that it is $o_p\left(\rho_T^2 + T^{1/2-2a}\rho_T^4\right)$ when $(p_*, q_*, \rho_*, T^{-1/4})$ is replaced by $(p_T, q_T, \rho_T, T^{-a})$. Recall

$$\widetilde{U}_t^{(2)}(p_*, q_*) = T^{-1/2}\frac{2}{\widetilde{\sigma}^4}\left(\frac{1-p_*}{1-q_*}\right)\sum_{t=1}^{T}\rho_*\widetilde{u}_t\widetilde{u}_{t-1} + T^{-1/2}\frac{2}{\widetilde{\sigma}^4}\left(\frac{1-p_*}{1-q_*}\right)\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{u}_t\widetilde{u}_{t-s}.$$

The first term on the right hand side satisfies $T^{-1/2}\sum_{t=1}^{T}\widetilde{u}_t\widetilde{u}_{t-1} = T^{-1/2}\sum_{t=1}^{T}(\widetilde{y}_{t-1} - \widetilde{\alpha}\widetilde{y}_{t-2})\widetilde{u}_{t-1} = -\widetilde{\alpha}T^{-1/2}\sum_{t=1}^{T}\widetilde{u}_t\widetilde{y}_{t-2} = O_p(T^{-1/4})$, where the second equality holds because $\widetilde{u}_t$ is orthogonal to the regressors $(1, y_{t-1})$ and the third holds because $\widetilde{u}_t$ satisfies (A.143) and $\widetilde{\alpha} = O_p(T^{-1/4})$. Similarly, when $(p_*, q_*, T^{-1/4})$ is replaced by (p_T, q_T, T^{-a}), $T^{-1/2}\sum_{t=1}^{T}\widetilde{u}_t\widetilde{u}_{t-1} = R_T + o_p\left(T^{-1/2}\right)$, where $R_T = o_p\left(\rho_T^2 + T^{1/2-2a}\rho_T^4\right)$. The second term on the right hand side of the preceding display satisfies

$$\begin{aligned}
& T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{u}_t\widetilde{u}_{t-s} \\
= \; & T^{-1/2}(T^{-1/4}c_*)^2\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)\left(\widetilde{s}_{t-s} - \widetilde{y}_{t-s-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right) \\
& + T^{-1/2}(T^{-1/4}c_*)\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)\left(\widetilde{e}_{t-s} - \widetilde{y}_{t-s-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right) \\
& + T^{-1/2}(T^{-1/4}c_*)\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\left(\widetilde{e}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)\left(\widetilde{s}_{t-s} - \widetilde{y}_{t-s-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right) \\
& + T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\left(\widetilde{e}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)\left(\widetilde{e}_{t-s} - \widetilde{y}_{t-s-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right) \\
= \; & P_1 + P_2 + P_3 + P_4.
\end{aligned}$$

We now study P_1-P_4 separately.

$$\begin{aligned}
P_1 \; = \; & c_*^2T^{-1}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{s}_t\widetilde{s}_{t-s} - c_*^2\left(T^{-1}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{y}_{t-1}\widetilde{s}_{t-s}\right)\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} \\
& - c_*^2\left(T^{-1}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{s}_t\widetilde{y}_{t-s-1}\right)\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} + c_*^2\left(T^{-1}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{y}_{t-1}\widetilde{y}_{t-s-1}\right)\left(\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)^2 \\
= \; & c_*^2T^{-1}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{s}_t\widetilde{s}_{t-s} + O_p(T^{-1/4}),
\end{aligned}$$

where the second equality holds because $T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t = T^{-1}\sum_{t=1}^{T}\widetilde{e}_{t-1}\widetilde{s}_t + T^{-5/4}c_*\sum_{t=1}^{T}\widetilde{s}_{t-1}\widetilde{s}_t = O_p\left(T^{-1/4}\right)$ and the last equality holds because of the CLT. Next,

$$\begin{aligned}
P_2 &= c_*T^{-1/4}\left(T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{s}_t\widetilde{e}_{t-s}\right) - c_*T^{-1/4}\left(T^{-1}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{s}_t\widetilde{y}_{t-s-1}\right)\frac{T^{-1/2}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} \\
&\quad -c_*T^{-1/2}\left(T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{e}_{t-s}\widetilde{y}_{t-1}\right)\frac{T^{-3/4}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} \\
&\quad +c_*T^{-1/2}\left(T^{-1}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{y}_{t-1}\widetilde{y}_{t-s-1}\right)\frac{T^{-3/4}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\frac{T^{-1/2}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} = O_p(T^{-1/4}),
\end{aligned}$$

Now,

$$\begin{aligned}
P_3 &= c_*T^{-1/4}\left(T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{e}_t\widetilde{s}_{t-s}\right) - c_*T^{-1/2}\left(T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{e}_t\widetilde{y}_{t-s-1}\right)\frac{T^{-3/4}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} \\
&\quad -c_*T^{-1/2}\left(T^{3/4}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{s}_{t-s}\widetilde{y}_{t-1}\right)\frac{T^{-1/2}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} \\
&\quad +c_*T^{-3/4}\left(T^{-3/4}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{y}_{t-1}\widetilde{y}_{t-s-1}\right)\frac{T^{-1/2}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\frac{T^{-3/4}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}. = O_p\left(T^{-1/4}\right).
\end{aligned}$$

Finally,

$$\begin{aligned}
P_4 &= T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{e}_t\widetilde{e}_{t-s} - T^{-1/2}\left(T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{e}_t\widetilde{y}_{t-s-1}\right)\frac{T^{-1/2}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} \\
&\quad -T^{-1/2}\left(T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{y}_{t-1}\widetilde{e}_{t-s}\right)\frac{T^{-1/2}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} \\
&\quad +T^{-3/4}\left(T^{-3/4}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{y}_{t-1}\widetilde{y}_{t-s-1}\right)\left(\frac{T^{-1/2}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)^2 \\
&= T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\widetilde{e}_t\widetilde{e}_{t-s} + O_p(T^{-1/2}).
\end{aligned}$$

Combining the above results, we obtain

$$T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_*,q_*) = \frac{2}{\sigma_*^4}\left(\frac{1-p_*}{1-q_*}\right)T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\left\{\rho_*^s\widetilde{e}_t\widetilde{e}_{t-s} + (T^{-1/4}c_*)^2\rho_*^s\widetilde{s}_t\widetilde{s}_{t-s}\right\} + O_p(T^{-1/4}),$$

where the equality holds because of the CLT and $\widetilde{\sigma}^2 - \sigma_*^2 = O_p(T^{-1/2})$.

We now consider the situation where $(p_*, q_*, \rho_*, T^{-1/4})$ is replaced by $(p_T, q_T, \rho_T, T^{-a})$. By repeating the same argument as above for P_1-P_4, it can be verified that the above displayed equations still hold, except that the remainders are now $o_p(\rho_T^2 + T^{1/2-2a}\rho_T^4)$. For example, P_1 equals

$$(T^{-a}c_*)^2 T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_T^s\widetilde{s}_t\widetilde{s}_{t-s} - (T^{-a}c_*)^2 T^{-1/2}\left(\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_T^s\widetilde{y}_{t-1}\widetilde{s}_{t-s}\right)\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}$$
$$-(T^{-a}c_*)^2 T^{-1/2}\left(\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_T^s\widetilde{s}_t\widetilde{y}_{t-s-1}\right)\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}$$
$$+(T^{-a}c_*)^2 T^{-1/2}\left(\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_T^s\widetilde{y}_{t-1}\widetilde{y}_{t-s-1}\right)\left(\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^2,$$

where $T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^2 = O_p(1)$, $T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t = T^{-1}\sum_{t=1}^T (T^{-a}c_*\widetilde{s}_{t-1} + \widetilde{e}_{t-1})\widetilde{s}_t = O_p(T^{-a}|\rho_T| + T^{-1/2})$, $T^{-1}\sum_{t=1}^T\sum_{s=2}^{t-1}\rho_T^s\widetilde{y}_{t-1}\widetilde{s}_{t-s} = T^{-1}\sum_{t=1}^T\sum_{s=2}^{t-1}\rho_T^s(T^{-a}c_*\widetilde{s}_{t-1} + \widetilde{e}_{t-1})\widetilde{s}_{t-s} = O_p(T^{-a}|\rho_T^3| + T^{-1/2}\rho_T^2)$, $T^{-1}\sum_{t=1}^T\sum_{s=2}^{t-1}\rho_T^s\widetilde{s}_t\widetilde{y}_{t-s-1} = O_p(T^{-a}|\rho_T^3| + T^{-1/2}\rho_T^2)$, and $T^{-1}\sum_{t=1}^T\sum_{s=2}^{t-1}\rho_T^s\widetilde{y}_{t-1}\widetilde{y}_{t-s-1} = O_p(T^{-2a}\rho_T^4 + T^{-1/2}\rho_T^2)$. As a result, $P_1 = (T^{-a}c_*)^2 T^{-1/2}\sum_{t=1}^T\sum_{s=2}^{t-1}\rho_T^s\widetilde{s}_t\widetilde{s}_{t-s} + o_p\left(\rho_T^2 + T^{1/2-2a}\rho_T^4\right)$. We omit the details for P_2-P_4 because they are similar to that of P_1.

Proof of Lemma A.30.2. Because $T^{-1/2}\sum_{t=1}^T \widetilde{e}_t^3 = O_p(1)$, the leading term on the right hand side of this result is of exact order $O_p(1)$ or greater. Below, we show that R_T is $O_p(T^{-1/4})$ under (38) and that it is $o_p(|p_T - q_T|)$ when $(p_*, q_*, \rho_*, T^{-1/4})$ is replaced by $(p_T, q_T, \rho_T, T^{-a})$. Note that because $T^{-1/2}\sum_{t=1}^T (T^{-a}c_*)^3\widetilde{s}_t^3 = O_p(1)$ but not $o_p(1)$ when $a = 1/6$, this term is not part of R_T. Recall

$$T^{-1/2}\sum_{t=1}^T \widetilde{U}_t^{(3)}(p_*, q_*) = \frac{(1-p_*)(p_*-q_*)}{(1-q_*)^2}\frac{1}{\widetilde{\sigma}^3}T^{-1/2}\sum_{t=1}^T\left\{\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^3 - 3\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right\}.$$

Because $T^{-1/2}\sum_{t=1}^T \widetilde{u}_t = 0$, it suffices to consider

$$\begin{aligned}
T^{-1/2}\sum_{t=1}^T\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^3 &= T^{-5/4}\frac{c_*^3}{\widetilde{\sigma}^3}\sum_{t=1}^T\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^3 + T^{-1/2}\frac{1}{\widetilde{\sigma}^3}\sum_{t=1}^T\left(\widetilde{e}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^3\\
&\quad + T^{-1}\frac{3c_*^2}{\widetilde{\sigma}^3}\sum_{t=1}^T\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^2\left(\widetilde{e}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)\\
&\quad + T^{-3/4}\frac{3c_*}{\widetilde{\sigma}^3}\sum_{t=1}^T\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)\left(\widetilde{e}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^2\\
&= P_5 + P_6 + P_7 + P_8.
\end{aligned}$$

We study P_5-P_8 separately. First,

$$\begin{aligned}
P_5 &= T^{-5/4}\frac{c_*^3}{\widetilde{\sigma}^3}\sum_{t=1}^T \widetilde{s}_t^3 - T^{-1}\frac{c_*^3}{\widetilde{\sigma}^3}\left(T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^3\right)\left(\frac{T^{-3/4}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^3\\
&\quad + T^{-5/4}\frac{3c_*^3}{\widetilde{\sigma}^3}\left(\frac{\sum_{t=1}^T \widetilde{s}_t\widetilde{y}_{t-1}^2}{(\sum_{t=1}^T \widetilde{y}_{t-1}^2)^{1/2}}\right)\left(\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{(\sum_{t=1}^T \widetilde{y}_{t-1}^2)^{3/4}}\right)^2\\
&\quad - T^{-3/4}\frac{3c_*^3}{\widetilde{\sigma}^3}(T^{-3/4}\sum_{t=1}^T \widetilde{s}_t^2\widetilde{y}_{t-1})\frac{T^{-3/4}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^2} = T^{-5/4}\frac{c_*^3}{\widetilde{\sigma}^3}\sum_{t=1}^T \widetilde{s}_t^3 + O_p(T^{-3/4}).
\end{aligned}$$

Next,

$$
\begin{aligned}
P_6 &= T^{-1/2}\frac{1}{\widetilde{\sigma}^3}\sum_{t=1}^{T}\widetilde{e}_t^3 - T^{-1}\frac{1}{\widetilde{\sigma}^3}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}^3}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\frac{(T^{-1/2}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t)^3}{(T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^2)^2} - T^{-1/2}\frac{3}{\widetilde{\sigma}^3}\frac{\sum_{t=1}^T \widetilde{e}_t^2\widetilde{y}_{t-1}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2} \\
&\quad + T^{-1}\frac{3}{\widetilde{\sigma}^3}\left(T^{-1/2}\sum_{t=1}^T \widetilde{e}_t\widetilde{y}_{t-1}^2\right)\frac{(T^{-1/2}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t)^2}{(T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^2)^2} = T^{-1/2}\frac{1}{\widetilde{\sigma}^3}\sum_{t=1}^T \widetilde{e}_t^3 + O_p(T^{-1/2}).
\end{aligned}
$$

In addition,

$$
\begin{aligned}
P_7 &= T^{-1/2}\frac{3c_*^2}{\widetilde{\sigma}^3}\left(T^{-1/2}\sum_{t=1}^T \widetilde{s}_t^2\widetilde{e}_t\right) - T^{-1/2}\frac{3c_*^2}{\widetilde{\sigma}^3}\frac{\sum_{t=1}^T \widetilde{s}_t^2\widetilde{y}_{t-1}}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\left(T^{-1/2}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t\right) \\
&\quad + T^{-1/2}\frac{3c_*^2}{\widetilde{\sigma}^3}\left(T^{-1/2}\sum_{t=1}^T \widetilde{y}_{t-1}^2\widetilde{e}_t\right)\left(\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^2 + T^{-1/2}\frac{6c_*^2}{\widetilde{\sigma}^3}\frac{\sum_{t=1}^T \widetilde{s}_t\widetilde{y}_{t-1}^2}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\left(T^{-1/2}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t\right) \\
&\quad - T^{-1/2}\frac{3c_*^2}{\widetilde{\sigma}^3}\left(\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^2\left(\frac{\sum_{t=1}^T \widetilde{y}_{t-1}^3}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)T^{-1/2}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t \\
&\quad - T^{-1/2}\frac{6c_*^2}{\widetilde{\sigma}^3}\left(T^{-1/2}\sum_{t=1}^T \widetilde{s}_t\widetilde{y}_{t-1}\widetilde{e}_t\right)\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2} = O_p(T^{-1/2})
\end{aligned}
$$

where the last equality holds because of the CLT. Finally,

$$
\begin{aligned}
P_8 &= T^{-1/4}\frac{3c_*}{\widetilde{\sigma}^3}T^{-1/2}\sum_{t=1}^T \widetilde{s}_t\widetilde{e}_t^2 + T^{-3/4}\frac{3c_*}{\widetilde{\sigma}^3}\left(\frac{\sum_{t=1}^T \widetilde{s}_t\widetilde{y}_{t-1}^2}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)\frac{(\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t)^2}{\sum_{t=1}^T \widetilde{y}_{t-1}^2} \\
&\quad - T^{-3/4}\frac{6c_*}{\widetilde{\sigma}^3}\frac{\sum_{t=1}^T \widetilde{s}_t\widetilde{e}_t\widetilde{y}_{t-1}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2} \\
&\quad - T^{-1/2}\frac{3c_*}{\widetilde{\sigma}^3}\left(T^{-1/2}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t^2\right)\left(\frac{T^{-3/4}\sum_{t=1}^T (T^{-1/4}c_*\widetilde{s}_{t-1}+\widetilde{e}_{t-1})\widetilde{s}_t}{T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right) \\
&\quad - T^{-3/4}\frac{3c_*}{\widetilde{\sigma}^3}\left(\frac{\sum_{t=1}^T \widetilde{y}_{t-1}^3}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\frac{(\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t)^2}{\sum_{t=1}^T \widetilde{y}_{t-1}^2} \\
&\quad + T^{-3/4}\frac{6c_*}{\widetilde{\sigma}^3}\left(\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)\frac{\sum_{t=1}^T \widetilde{y}_{t-1}^2\widetilde{e}_t\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2} \\
&= T^{-1/4}\frac{3c_*}{\widetilde{\sigma}^3}T^{-1/2}\sum_{t=1}^T \widetilde{s}_t\widetilde{e}_t^2 + O_p(T^{-1/2}) = O_p(T^{-1/4}),
\end{aligned}
$$

where $T^{-1/2}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t^2 = O_p(1)$ because the autoregressive coefficient equals zero. Therefore.

$$
T^{-1/2}\sum_{t=1}^T \widetilde{U}_t^{(3)}(p_*,q_*) = \frac{(1-p_*)\,(p_*-q_*)}{(1-q_*)^2}\frac{1}{\widetilde{\sigma}^6}\left\{T^{-1/2}\sum_{t=1}^T \widetilde{e}_t^3 + c_*^3 T^{-5/4}\sum_{t=1}^T \widetilde{s}_t^3\right\} + O_p(T^{-1/4}).
$$

The result follows after applying $\widetilde{\sigma}^2 - \sigma_*^2 = O_p(T^{-1/2})$ to the first term on the right hand side.

When $(p_*, q_*, \rho_*, T^{-1/4})$ is replaced by $(p_T, q_T, \rho_T, T^{-a})$ with $p_T + q_T - 1 \to 0$, by repeating the same argument as above for P_5-P_8, but replacing $c_* T^{-1/4}$ with $c_* T^{-a}$, we can verify that the omitted terms these displayed equations are all $o_p(1)$ for any $a \in [1/8, 1/4]$. For example, P_5 equals

$$\frac{(T^{-a}c_*)^3}{\widetilde{\sigma}^3} T^{-1/2} \sum_{t=1}^{T} \widetilde{s}_t^3 - \frac{(T^{-a}c_*)^3}{\widetilde{\sigma}^3} T^{1/2-3a} \left(T^{-1} \sum_{t=1}^{T} \widetilde{y}_{t-1}^3 \right) \left(\frac{T^{-1+a} \sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{s}_t}{T^{-1} \sum_{t=1}^{T} \widetilde{y}_{t-1}^2} \right)^3$$
$$+ \frac{3(T^{-a}c_*)^3}{\widetilde{\sigma}^3} T^{1/2-4a} \left(T^{-1+2a} \sum_{t=1}^{T} \widetilde{s}_t \widetilde{y}_{t-1}^2 \right) \left(\frac{T^{-1+a} \sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{s}_t}{T^{-1} \sum_{t=1}^{T} \widetilde{y}_{t-1}^2} \right)^2$$
$$- \frac{3(T^{-a}c_*)^3}{\widetilde{\sigma}^3} T^{1/2-2a} (T^{-1+a} \sum_{t=1}^{T} \widetilde{s}_t^2 \widetilde{y}_{t-1}) \frac{T^{-1+a} \sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{s}_t}{T^{-1} \sum_{t=1}^{T} \widetilde{y}_{t-1}^2} = \frac{(T^{-a}c_*)^3}{\widetilde{\sigma}^3} T^{-1/2} \sum_{t=1}^{T} \widetilde{s}_t^3 + O_p(T^{-a}),$$

where the last equality holds because $a \geq 1/8$. We omit the details for P_6-P_8.

Proof of Lemma A.30.3. Because $T^{-1/2} \sum_{t=1}^{T} \left(\frac{\widetilde{e}_t^4}{\sigma_*^4} - 6 \frac{\widetilde{e}_t^2}{\sigma_*^2} + 3 \right) = O_p(1)$ but not $o_p(1)$, the leading term on the right hand side is of exact order $O_p(1)$ or greater. Below, we show R_T is $O_p(T^{-1/4})$ under (38) and it is $o_p(1)$ when $(p_*, q_*, \rho_*, T^{-1/4})$ is replaced by $(p_T, q_T, \rho_T, T^{-a})$. Recall

$$T^{-1/2} \sum_{t=1}^{T} \widetilde{U}_t^{(4)}(p_*, q_*) = \left[\frac{1-p_*}{2-p_*-q_*} \left(1 + \left(\frac{1-p_*}{1-q_*} \right)^3 \right) - 3 \left(\frac{1-p_*}{1-q_*} \right)^2 \right] \times \frac{1}{\widetilde{\sigma}^4} T^{-1/2} \sum_{t=1}^{T} \left\{ \left(\frac{\widetilde{u}_t}{\widetilde{\sigma}} \right)^4 - 6 \left(\frac{\widetilde{u}_t}{\widetilde{\sigma}} \right)^2 + 3 \right\}.$$

We study $T^{-1/2} \sum_{t=1}^{T} (\widetilde{u}_t/\widetilde{\sigma})^4$ and $T^{-1/2} \sum_{t=1}^{T} (\widetilde{u}_t/\widetilde{\sigma})^2$ separately. We have

$$\begin{aligned} T^{-1/2} \sum_{t=1}^{T} \left(\frac{\widetilde{u}_t}{\widetilde{\sigma}} \right)^4 &= T^{-1/2} \frac{1}{\widetilde{\sigma}^4} \sum_{t=1}^{T} \left(\widetilde{e}_t - \widetilde{y}_{t-1} \frac{\sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{e}_t}{\sum_{t=1}^{T} \widetilde{y}_{t-1}^2} \right)^4 + T^{-3/2} \frac{c_*^4}{\widetilde{\sigma}^4} \sum_{t=1}^{T} \left(\widetilde{s}_t - \widetilde{y}_{t-1} \frac{\sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{s}_t}{\sum_{t=1}^{T} \widetilde{y}_{t-1}^2} \right)^4 \\ &+ T^{-1} \frac{6c_*^2}{\widetilde{\sigma}^4} \sum_{t=1}^{T} \left(\widetilde{e}_t - \widetilde{y}_{t-1} \frac{\sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{e}_t}{\sum_{t=1}^{T} \widetilde{y}_{t-1}^2} \right)^2 \left(\widetilde{s}_t - \widetilde{y}_{t-1} \frac{\sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{s}_t}{\sum_{t=1}^{T} \widetilde{y}_{t-1}^2} \right)^2 \\ &+ T^{-3/4} \frac{4c_*}{\widetilde{\sigma}^4} \sum_{t=1}^{T} \left(\widetilde{e}_t - \widetilde{y}_{t-1} \frac{\sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{e}_t}{\sum_{t=1}^{T} \widetilde{y}_{t-1}^2} \right)^3 \left(\widetilde{s}_t - \widetilde{y}_{t-1} \frac{\sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{s}_t}{\sum_{t=1}^{T} \widetilde{y}_{t-1}^2} \right) \\ &+ T^{-5/4} \frac{4c_*^3}{\widetilde{\sigma}^4} \sum_{t=1}^{T} \left(\widetilde{e}_t - \widetilde{y}_{t-1} \frac{\sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{e}_t}{\sum_{t=1}^{T} \widetilde{y}_{t-1}^2} \right) \left(\widetilde{s}_t - \widetilde{y}_{t-1} \frac{\sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{s}_t}{\sum_{t=1}^{T} \widetilde{y}_{t-1}^2} \right)^3 \\ &= P_9 + P_{10} + P_{11} + P_{12} + P_{13}. \end{aligned}$$

We study P_9-P_{13} separately. For P_{13} and P_{12},

$$\begin{aligned} P_{13} &= T^{-3/4} \frac{4c_*^3}{\widetilde{\sigma}^4} \left\{ T^{-1/2} \sum_{t=1}^{T} \widetilde{e}_t \left(\widetilde{s}_t - \widetilde{y}_{t-1} \frac{\sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{s}_t}{\sum_{t=1}^{T} \widetilde{y}_{t-1}^2} \right)^3 \right\} \\ &+ T^{-3/4} \frac{4c_*^3}{\widetilde{\sigma}^4} \left\{ T^{-1} \sum_{t=1}^{T} \widetilde{y}_{t-1} \left(\widetilde{s}_t - \widetilde{y}_{t-1} \frac{\sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{s}_t}{\sum_{t=1}^{T} \widetilde{y}_{t-1}^2} \right)^3 \right\} \frac{T^{-1/2} \sum_{t=1}^{T} \widetilde{y}_{t-1} \widetilde{e}_t}{T^{-1} \sum_{t=1}^{T} \widetilde{y}_{t-1}^2} = O_p(T^{-3/4}) \end{aligned}$$

and

$$P_{12} = T^{-1/4}\frac{4c_*}{\widetilde{\sigma}^4}\left\{T^{-1/2}\sum_{t=1}^{T}\left(\widetilde{e}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^3 \widetilde{s}_t\right\}$$
$$-T^{-1/4}\frac{4c_*}{\widetilde{\sigma}^4}\left\{T^{-1/2}\sum_{t=1}^{T}\left(\widetilde{e}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^3 \widetilde{y}_{t-1}\right\}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2} = O_p(T^{-1/4}).$$

Next, P_{11} is equal to

$$T^{-1}\frac{6c_*^2}{\widetilde{\sigma}^4}\sum_{t=1}^T \widetilde{e}_t^2\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^2$$
$$+T^{-1}\frac{6c_*^2}{\widetilde{\sigma}^4}\left\{T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^2\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^2\right\}\left(\frac{T^{-1/2}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^2$$
$$-T^{-1}\frac{12c_*^2}{\widetilde{\sigma}^4}\left\{T^{-1/2}\sum_{t=1}^T\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^2 \widetilde{y}_{t-1}\widetilde{e}_t\right\}\frac{T^{-1/2}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^2}$$
$$= T^{-1}\frac{6c_*^2}{\widetilde{\sigma}^4}\sum_{t=1}^T \widetilde{e}_t^2\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^2 + O_p(T^{-1}),$$

whose first term is equal to

$$T^{-1}\frac{6c_*^2}{\widetilde{\sigma}^4}\sum_{t=1}^T \widetilde{e}_t^2\widetilde{s}_t^2 + T^{-1}\frac{6c_*^2}{\widetilde{\sigma}^4}\sum_{t=1}^T \widetilde{e}_t^2\widetilde{y}_{t-1}^2\left(\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^2 - T^{-1}\frac{12c_*^2}{\widetilde{\sigma}^4}\sum_{t=1}^T \widetilde{e}_t^2\widetilde{s}_t\widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}$$
$$= T^{-1}\frac{6c_*^2}{\widetilde{\sigma}^4}\sum_{t=1}^T \widetilde{e}_t^2\widetilde{s}_t^2 + O_p(T^{-1/2}).$$

Therefore, $P_{11} = T^{-1}6c_*^2\widetilde{\sigma}^{-4}\sum_{t=1}^T \widetilde{e}_t^2\widetilde{s}_t^2 + O_p(T^{-1/2})$. For P_9, we have

$$P_9 = T^{-1/2}\frac{1}{\widetilde{\sigma}^4}\sum_{t=1}^T \widetilde{e}_t^4 + T^{-3/2}\frac{1}{\widetilde{\sigma}^4}\left(T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^4\right)\left(\frac{T^{-1/2}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^4$$
$$+6T^{-1/2}\frac{1}{\widetilde{\sigma}^4}\frac{\sum_{t=1}^T \widetilde{e}_t^2\widetilde{y}_{t-1}^2}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\frac{(\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t)^2}{\sum_{t=1}^T \widetilde{y}_{t-1}^2} - 4T^{-1/2}\frac{1}{\widetilde{\sigma}^4}\frac{\sum_{t=1}^T \widetilde{e}_t^3\widetilde{y}_{t-1}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}$$
$$-4T^{-3/2}\frac{1}{\widetilde{\sigma}^4}\left(T^{-1/2}\sum_{t=1}^T \widetilde{e}_t\widetilde{y}_{t-1}^3\right)\left(\frac{T^{-1/2}\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{e}_t}{T^{-1}\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^3 = T^{-1/2}\frac{1}{\widetilde{\sigma}^4}\sum_{t=1}^T \widetilde{e}_t^4 + O_p(T^{-1/2}).$$

For P_{10}, we have

$$P_{10} = T^{-3/2}\frac{c_*^4}{\widetilde{\sigma}^4}\sum_{t=1}^T\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^T \widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^T \widetilde{y}_{t-1}^2}\right)^4 = T^{-3/2}\frac{c_*^4}{\widetilde{\sigma}^4}\sum_{t=1}^T \widetilde{s}_t^4 + O_p\left(T^{-1/4}\right).$$

Combining the results for P_9-P_{13}, we obtain

$$T^{-1/2}\sum_{t=1}^{T}\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^4 = T^{-3/2}\frac{c_*^4}{\widetilde{\sigma}^4}\sum_{t=1}^{T}\widetilde{s}_t^4 + T^{-1/2}\frac{1}{\widetilde{\sigma}^4}\sum_{t=1}^{T}\widetilde{e}_t^4 + T^{-1}\frac{6c_*^2}{\widetilde{\sigma}^4}\sum_{t=1}^{T}\widetilde{e}_t^2\widetilde{s}_t^2 + O_p\left(T^{-1/4}\right). \quad \text{(A.150)}$$

Now, we consider

$$\begin{aligned}
&-6T^{-1/2}\sum_{t=1}^{T}\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^2 \\
= &-\frac{6}{\widetilde{\sigma}^2}T^{-1/2}\sum_{t=1}^{T}\left(\widetilde{e}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)^2 - \frac{6c_*^2}{\widetilde{\sigma}^2}T^{-1}\sum_{t=1}^{T}\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)^2 \\
&-\frac{12c_*}{\widetilde{\sigma}^2}T^{-3/4}\sum_{t=1}^{T}\left(\widetilde{e}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right) = P_{14} + P_{15} + P_{16}.
\end{aligned}$$

For P_{14}, we have

$$P_{14} = -\frac{6}{\widetilde{\sigma}^2}T^{-1/2}\sum_{t=1}^{T}\widetilde{e}_t^2 + \frac{6}{\widetilde{\sigma}^2}T^{-1/2}\frac{(\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t)^2}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} = -\frac{6}{\widetilde{\sigma}^2}T^{-1/2}\sum_{t=1}^{T}\widetilde{e}_t^2 + O_p\left(T^{-1/2}\right).$$

For P_{15}, we have

$$P_{15} = -\frac{6c_*^2}{\widetilde{\sigma}^2}T^{-1}\sum_{t=1}^{T}\widetilde{s}_t^2 + \frac{6c_*^2}{\widetilde{\sigma}^2}T^{-1/2}\frac{(T^{-3/4}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t)^2}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} = -\frac{6c_*^2}{\widetilde{\sigma}^2}T^{-1}\sum_{t=1}^{T}\widetilde{s}_t^2 + O_p\left(T^{-1/2}\right).$$

For P_{16}, we have

$$\begin{aligned}
P_{16} &= -\frac{12c_*}{\widetilde{\sigma}^2}T^{-1/4}\left(T^{-1/2}\sum_{t=1}^{T}\widetilde{e}_t\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)\right) \\
&\quad -\frac{12c_*}{\widetilde{\sigma}^2}T^{-1/4}\left\{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}\left(\widetilde{s}_t - \widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)\right\}\frac{T^{-1/2}\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{T^{-1}\sum_{t=1}^{T}\widetilde{y}_{t-1}^2} = O_p\left(T^{-1/4}\right).
\end{aligned}$$

Therefore,

$$-6T^{-1/2}\sum_{t=1}^{T}\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^2 = -\frac{6}{\widetilde{\sigma}^2}T^{-1/2}\sum_{t=1}^{T}\widetilde{e}_t^2 - \frac{6c_*^2}{\widetilde{\sigma}^2}T^{-1}\sum_{t=1}^{T}\widetilde{s}_t^2 + O_p\left(T^{-1/4}\right). \quad \text{(A.151)}$$

Combining (A.150) and (A.151), we have

$$\begin{aligned}
&T^{-1/2}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^4 - 6\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^2 + 3\right\} \\
= &\; T^{-1/2}\sum_{t=1}^{T}\left\{\frac{\widetilde{e}_t^4}{\widetilde{\sigma}^4} - 6\frac{\widetilde{e}_t^2}{\widetilde{\sigma}^2} + 3\right\} + T^{-3/2}\frac{c_*^4}{\widetilde{\sigma}^4}\sum_{t=1}^{T}\widetilde{s}_t^4 + T^{-1}\frac{6c_*^2}{\widetilde{\sigma}^4}\sum_{t=1}^{T}\left(\widetilde{e}_t^2 - \widetilde{\sigma}^2\right)\widetilde{s}_t^2 + O_p\left(T^{-1/4}\right).
\end{aligned}$$

Further,

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T}\left(\widetilde{e}_t^2-\widetilde{\sigma}^2\right)\widetilde{s}_t^2 &= \left(T^{-1}\sum_{t=1}^{T}\widetilde{s}_t^2\right)\left(\sigma_*^2-\widetilde{\sigma}^2\right)+T^{-1}\sum_{t=1}^{T}\widetilde{s}_t^2\left(\widetilde{e}_t^2-\sigma_*^2\right)\\
&= \left(T^{-1}\sum_{t=1}^{T}\widetilde{s}_t^2\right)\left(\sigma_*^2-\widetilde{\sigma}^2\right)+O_p\left(T^{-1/2}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
&\sigma_*^2-\widetilde{\sigma}^2 \qquad\qquad \text{(A.152)}\\
=\ & \sigma_*^2-T^{-1}\left\{\sum_{t=1}^{T}\widetilde{e}_t-\widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{e}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}+c_*T^{-1/4}\left(\widetilde{s}_t-\widetilde{y}_{t-1}\frac{\sum_{t=1}^{T}\widetilde{y}_{t-1}\widetilde{s}_t}{\sum_{t=1}^{T}\widetilde{y}_{t-1}^2}\right)\right\}^2\\
=\ & -T^{-3/2}c_*^2\sum_{t=1}^{T}\widetilde{s}_t^2+\left(\sigma_*^2-T^{-1}\sum_{t=1}^{T}\widetilde{e}_t^2\right)+O_p(T^{-3/4})=-T^{-3/2}c_*^2\sum_{t=1}^{T}\widetilde{s}_t^2+O_p(T^{-1/2}).
\end{aligned}
$$

Therefore, by rearrange the terms, we obtain

$$
\begin{aligned}
& T^{-1/2}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^4-6\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^2+3\right\}\\
=\ & T^{-1/2}\sum_{t=1}^{T}\left\{\frac{\widetilde{e}_t^4}{\widetilde{\sigma}^4}-6\frac{\widetilde{e}_t^2}{\widetilde{\sigma}^2}+3\right\}+T^{-3/2}\frac{c_*^4}{\widetilde{\sigma}^4}\sum_{t=1}^{T}\widetilde{s}_t^4-T^{-1/2}\frac{6c_*^4}{\widetilde{\sigma}^4}\left(T^{-1}\sum_{t=1}^{T}\widetilde{s}_t^2\right)^2+O_p\left(T^{-1/4}\right).
\end{aligned}
$$

Finally, we study the first summation on the right hand side of the preceding display using a second order Taylor expansion:

$$
\begin{aligned}
& T^{-1/2}\sum_{t=1}^{T}\left(\frac{1}{\widetilde{\sigma}^4}\widetilde{e}_t^4-6\frac{1}{\widetilde{\sigma}^2}\widetilde{e}_t^2+3\right)\\
=\ & T^{-1/2}\sum_{t=1}^{T}\left(\sigma_*^{-4}\widetilde{e}_t^4-6\sigma_*^{-2}\widetilde{e}_t^2+3\right)+T^{-1/2}\sum_{t=1}^{T}\left(-2\sigma_*^{-6}\widetilde{e}_t^4+6\sigma_*^{-4}\widetilde{e}_t^2\right)\left(\widetilde{\sigma}^2-\sigma_*^2\right)\\
& +\frac{1}{2}T^{-1/2}\sum_{t=1}^{T}\left(6\sigma_*^{-8}\widetilde{e}_t^4-12\sigma_*^{-6}\widetilde{e}_t^2\right)\left(\widetilde{\sigma}^2-\sigma_*^2\right)^2+O_p\left(T^{1/2}\left(\widetilde{\sigma}^2-\sigma_*^2\right)^3\right)\\
=\ & T^{-1/2}\sum_{t=1}^{T}\left(\sigma_*^{-4}\widetilde{e}_t^4-6\sigma_*^{-2}\widetilde{e}_t^2+3\right)+T^{-1/2}\sum_{t=1}^{T}\left(3\sigma_*^{-8}\widetilde{e}_t^4-6\sigma_*^{-6}\widetilde{e}_t^2\right)\left(\widetilde{\sigma}^2-\sigma_*^2\right)^2+O_p\left(T^{-1/2}\right)\\
=\ & T^{-1/2}\sum_{t=1}^{T}\left(\sigma_*^{-4}\widetilde{e}_t^4-6\sigma_*^{-2}\widetilde{e}_t^2+3\right)+3\sigma_*^{-4}c_*^4T^{-1/2}\left(T^{-1}\sum_{t=1}^{T}\widetilde{s}_t^2\right)^2+O_p\left(T^{-1/2}\right),
\end{aligned}
$$

where the second equality holds because of (A.152) and

$$
T^{-1/2}\sum_{t=1}^{T}\left(-2\sigma_*^{-6}\widetilde{e}_t^4+6\sigma_*^{-4}\widetilde{e}_t^2\right)=-2\sigma_*^{-2}T^{-1/2}\sum_{t=1}^{T}\left(\sigma_*^{-4}\widetilde{e}_t^4-3\sigma_*^{-2}\widetilde{e}_t^2\right)=O_p\left(1\right),
$$

and the last equality holds because

$$
\begin{aligned}
& T^{-1/2}\sum_{t=1}^{T}\left(3\sigma_*^{-8}\widetilde{e}_t^4 - 6\sigma_*^{-6}\widetilde{e}_t^2\right)\left(\widetilde{\sigma}^2-\sigma_*^2\right)^2 \\
= \;& 3\sigma_*^{-4}\left(T^{-1}\sum_{t=1}^{T}\left(\sigma_*^{-4}\widetilde{e}_t^4 - 2\sigma_*^{-2}\widetilde{e}_t^2\right)\right)T^{-1/2}\left(\widetilde{\sigma}^2-\sigma_*^2\right)^2 \\
= \;& 3\sigma_*^{-4}\left(T^{-1}\sum_{t=1}^{T}\left(\sigma_*^{-4}\widetilde{e}_t^4 - 2\sigma_*^{-2}\widetilde{e}_t^2\right)\right)c_*^4T^{-1/2}\left(T^{-1}\sum_{t=1}^{T}\widetilde{s}_t^2\right)^2 + O_p\left(T^{-1/2}\right) \\
= \;& 3\sigma_*^{-4}c_*^4T^{-1/2}\left(T^{-1}\sum_{t=1}^{T}\widetilde{s}_t^2\right)^2 + O_p\left(T^{-1/2}\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& T^{-1/2}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^4 - 6\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^2 + 3\right\} \\
= \;& T^{-1/2}\sum_{t=1}^{T}\left\{\frac{\widetilde{e}_t^4}{\sigma_*^4} - 6\frac{\widetilde{e}_t^2}{\sigma_*^2} + 3\right\} + T^{-1/2}\frac{c_*^4}{\sigma_*^4}\left(T^{-1}\sum_{t=1}^{T}\widetilde{s}_t^4 - 3\left(T^{-1}\sum_{t=1}^{T}\widetilde{s}_t^2\right)^2\right) + O_p\left(T^{-1/4}\right).
\end{aligned}
$$

When $(p_*, q_*, \rho_*, T^{-1/4})$ is replaced by $(p_T, q_T, \rho_T, T^{-a})$ with $p_T + q_T - 1 \to 0$, by repeating the same argument as the above, but replacing $c_*T^{-1/4}$ by c_*T^{-a}, we can show that the omitted terms from the above displayed equations are all $o_p(1)$ for any $a \in [1/8, 1/4]$. We omit the details.

Proof of Lemma A.30.4. The leading term on the right hand side of the result is $O_p(1)$ under the DGP (38), and it is $O_p\left(\rho_T^4\right)$ when $(p_*, q_*, \rho_*, T^{-1/4})$ is replaced by $(p_T, q_T, \rho_T, T^{-a})$. Below, we show that R_T is $O_p(T^{-1/4})$ under (38) and that it is $o_p\left(\rho_T^4\right)$ when $(p_*, q_*, \rho_*, T^{-1/4})$ is replaced by $(p_T, q_T, \rho_T, T^{-a})$.

Under (A.138), we have

$$
\widetilde{\Omega}(p_*, q_*) = T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_*, q_*)^2 - T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_*, q_*)\widetilde{S}_t'\left(T^{-1}\sum_{t=1}^{T}\widetilde{S}_t\widetilde{S}_t'\right)^{-1}T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_*, q_*)\widetilde{S}_t,
$$

where

$$
\widetilde{u}_t = y_t - \widetilde{\delta} + \widetilde{\alpha}y_{t-1},\; \widetilde{S}_t = \begin{bmatrix}\frac{y_{t-1}\widetilde{u}_t}{\widetilde{\sigma}^2} & \frac{1}{2\widetilde{\sigma}^2}\left(\frac{\widetilde{u}_t^2}{\widetilde{\sigma}^2} - 1\right) & \frac{\widetilde{u}_t}{\widetilde{\sigma}^2}\end{bmatrix}',
$$

and

$$
\begin{aligned}
\widetilde{U}_t^{(2)}(p_*, q_*) \;=\; & \frac{2}{\widetilde{\sigma}^2}\left(\frac{1-p_*}{1-q_*}\right)\rho_*\frac{y_{t-1}\widetilde{u}_t}{\widetilde{\sigma}^2} + \\
& \frac{2}{\widetilde{\sigma}^2}\left(\frac{1-p_*}{1-q_*}\right)\left\{\rho_*^2\sum_{s=2}^{t-1}\rho_*^{s-2}\frac{\widetilde{u}_{t-s}\widetilde{u}_t}{\widetilde{\sigma}^2} + \rho_*\frac{\left(\widetilde{u}_{t-1} - y_{t-1}\right)\widetilde{u}_t}{\widetilde{\sigma}^2}\right\}.
\end{aligned} \tag{A.153}
$$

Note that $T\widetilde{\Omega}(p_*, q_*)$ equals the SSR from the projection of $\widetilde{U}_t^{(2)}(p_*, q_*)$ on the space spanned by $\widetilde{S}_t$ for $t = 1, ..., T$. Because $y_{t-1}\widetilde{u}_t/\widetilde{\sigma}^2$ in (A.153) belongs to the column space of $\widetilde{S}_t$, the first term on the right hand side of (A.153) has no contribution to the SSR. This is the same situation as under the null hypothesis. Further, the product of the second term with $\widetilde{S}_t$ satisfies

$$T^{-1}\sum_{t=1}^{T}\left\{\sum_{s=2}^{t-1}\rho_*^s\frac{\widetilde{u}_{t-s}\widetilde{u}_t}{\widetilde{\sigma}^2}+\rho_*\frac{(\widetilde{u}_{t-1}-y_{t-1})\,\widetilde{u}_t}{\widetilde{\sigma}^2}\right\}\widetilde{S}_t = O_p\left(T^{-1/2}\right).$$

Therefore,

$$\begin{aligned}\widetilde{\Omega}(p_*, q_*) &= \frac{4}{\widetilde{\sigma}^4}\left(\frac{1-p_*}{1-q_*}\right)^2 T^{-1}\sum_{t=1}^{T}\left(\sum_{s=2}^{t-1}\rho_*^s\frac{\widetilde{u}_{t-s}\widetilde{u}_t}{\widetilde{\sigma}^2}\right)^2 + O_p\left(T^{-1/4}\right)\\ &= \frac{4}{\widetilde{\sigma}^4}\left(\frac{1-p_*}{1-q_*}\right)^2 T^{-1}\sum_{t=1}^{T}\left(\sum_{s=2}^{t-1}\rho_*^s\frac{\widetilde{e}_{t-s}\widetilde{e}_t}{\widetilde{\sigma}^2}\right)^2 + O_p\left(T^{-1/4}\right),\end{aligned}$$

where the second equation holds because of (A.143).

When $(p_*, q_*, \rho_*, T^{-1/4})$ is replaced by $(p_T, q_T, \rho_T, T^{-a})$ with $p_T + q_T - 1 \to 0$,

$$T^{-1}\sum_{t=1}^{T}\left\{\sum_{s=2}^{t-1}\rho_T^s\frac{\widetilde{u}_{t-s}\widetilde{u}_t}{\widetilde{\sigma}^2}+\rho_T\frac{(\widetilde{u}_{t-1}-y_{t-1})\,\widetilde{u}_t}{\widetilde{\sigma}^2}\right\}\widetilde{S}_t = O_p\left(T^{-1/2}\rho_T^2+\rho_T^4T^{-2a}\right).$$

As a result,

$$\begin{aligned}\widetilde{\Omega}(p_T, q_T) &= \frac{4}{\widetilde{\sigma}^4}\left(\frac{1-p_T}{1-q_T}\right)^2 T^{-1}\sum_{t=1}^{T}\left(\sum_{s=2}^{t-1}\rho_T^s\frac{\widetilde{e}_{t-s}\widetilde{e}_t}{\widetilde{\sigma}^2}\right)^2 + O_p\left(T^{-1/2}\rho_T^2+T^{-a}\rho_T^4\right)\\ &= \frac{4}{\widetilde{\sigma}^4}\left(\frac{1-p_T}{1-q_T}\right)^2 T^{-1}\sum_{t=1}^{T}\left(\sum_{s=2}^{t-1}\rho_T^s\frac{\widetilde{e}_{t-s}\widetilde{e}_t}{\widetilde{\sigma}^2}\right)^2 + R_T + o_p\left(T^{-1/2}\right),\end{aligned}$$

where $R_T = O_p(T^{-a}\rho_T^4) = o_p(\rho_T^4)$. Thus, it is dominated by the leading term.

Proof of Lemma A.30.5. The leading term on the right hand side of the result is $O_p(1)$ under the DGP (38) and $O_p\left((p_T-q_T)^2\right)$ when $(p_*, q_*, \rho_*, T^{-1/4})$ is replaced by $(p_T, q_T, \rho_T, T^{-a})$. Below, we show that R_T is $O_p(T^{-1/2})$ under (38) and that it is $O_p\left((p_T-q_T)^2T^{-2a}\right)$ in the latter case.

Under (A.138), we have

$$\widetilde{\Omega}^{(3)}(p_*, q_*) = T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(3)}(p_*, q_*)^2 - T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(3)}(p_*, q_*)\widetilde{S}_t'\left(T^{-1}\sum_{t=1}^{T}\widetilde{S}_t\widetilde{S}_t'\right)^{-1}T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(3)}(p_*, q_*)\widetilde{S}_t,$$

where

$$\widetilde{U}_t^{(3)}(p_*, q_*) = \frac{(1-p_*)\,(p_*-q_*)}{(1-q_*)^2}\frac{1}{\widetilde{\sigma}^3}T^{-1/2}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^3-3\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right\}$$

We have

$$T^{-1}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^3-3\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right\}\widetilde{S}_t = O_p(T^{-1/2})$$

because the orthogonality of the Hermite polynomials, and

$$T^{-1}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^3-3\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right\}^2=T^{-1}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{e}_t}{\sigma_*}\right)^3-3\frac{\widetilde{e}_t}{\sigma_*}\right\}^2+O_p(T^{-1/2})$$

by (A.143) and a first order Taylor expansion. When $(p_*,q_*,\rho_*,T^{-1/4})$ is replaced by (p_T,q_T,ρ_T,T^{-a}), the preceding two displayed equations still hold after replacing their remainder terms by $O_p(T^{-a})$ and $O_p(T^{-2a})$ respectively. Because $\widetilde{U}_t^{(3)}(p_T,q_T)$ depends on (p_T-q_T), it follows that $\widetilde{\Omega}^{(3)}(p_*,q_*)=T^{-1}\sum_{t=1}^{T}\left\{(\widetilde{e}_t/\sigma_*)^3-3\widetilde{e}_t/\sigma_*\right\}^2+O_p((p_T-q_T)^2\,T^{-2a})$.

Proof of Lemma A.30.6. The proof is similar to that of Lemma A.30.5. We have

$$\widetilde{\Omega}^{(4)}(p_*,q_*)=T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(4)}(p_*,q_*)^2-T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(4)}(p_*,q_*)\widetilde{S}_t'\left(T^{-1}\sum_{t=1}^{T}\widetilde{S}_t\widetilde{S}_t'\right)^{-1}T^{-1}\sum_{t=1}^{T}\widetilde{U}_t^{(4)}(p_*,q_*)\widetilde{S}_t,$$

where

$$\widetilde{U}_t^{(4)}(p_*,q_*)=\left[\frac{1-p_*}{2-p_*-q_*}\left(1+\left(\frac{1-p_*}{1-q_*}\right)^3\right)-3\left(\frac{1-p_*}{1-q_*}\right)^2\right]\frac{1}{\widetilde{\sigma}^4}\left\{\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^4-6\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^2+3\right\}.$$

We have

$$T^{-1}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^4-6\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^2+3\right\}\widetilde{S}_t=O_p(T^{-1/2})$$

because of the orthogonality of the Hermite polynomials, and

$$T^{-1}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^4-6\left(\frac{\widetilde{u}_t}{\widetilde{\sigma}}\right)^2+3\right\}^2=T^{-1}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{e}_t}{\sigma_*}\right)^4-6\left(\frac{\widetilde{e}_t}{\sigma_*}\right)^2+3\right\}^2+O_p(T^{-1/2})$$

because of (A.143). When $(p_*,q_*,\rho_*,T^{-1/4})$ is replaced by (p_T,q_T,ρ_T,T^{-a}), the preceding two displayed equations still hold after replacing their remainder terms by $O_p(T^{-a})$ and $O_p(T^{-2a})$ respectively. It follows that $\widetilde{\Omega}^{(4)}(p_*,q_*)=T^{-1}\sum_{t=1}^{T}\left\{(\widetilde{e}_t/\sigma_*)^4-6(\widetilde{e}_t/\sigma_*)^3+3\right\}^2+O_p(T^{-2a})$. ■

Remark 15 *In the next proof, we derive the refined approximation to $LR(p_*,q_*)$ under DGP (38) in two steps by building on Lemmas A.29 and A.30. First, we apply Lemmas A.29.1 and A.30.1 and A.30.4 to obtain the leading terms of this approximation. Then, we apply Lemmas A.28 and A.30 to obtain the refinement terms.*

Proof of Proposition 3. By Lemmas A.29.1, under DGP (38) the local alternatives and the convergence rate of $\hat{\delta}_2$ are both $T^{-1/4}$. Let $\eta=T^{1/4}(\delta_2-\widetilde{\delta})$. Then, for any $\varepsilon>0$, there exists an $M<\infty$ such that $P(\|\eta\|\leq M)\geq 1-\varepsilon$ for sufficiently large T. Over this set,

$$\begin{aligned}&\sup_{\|\delta_2-\widetilde{\delta}\|\leq T^{-1/4}M}2\left[\mathcal{L}(p_*,q_*,\delta_2)-\mathcal{L}(p_*,q_*,\widetilde{\delta})\right] \qquad (A.154)\\ =\;&\sup_{\|\eta\|\leq M}\left[\left(T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_*,q_*)\right)\eta^2-\frac{1}{4}\widetilde{\Omega}(p_*,q_*)\eta^4\right]+o_p(1).\end{aligned}$$

Because of the first and fourth results of Lemma A.30, the right hand side can be written as

$$\sup_{\|\eta\|\le M}\left[\left(\frac{2}{\sigma_*^2}\left(\frac{1-p_*}{1-q_*}\right)T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\frac{\widetilde{e}_t\widetilde{e}_{t-s}}{\sigma_*^2}\right)\eta^2+\left(\frac{2c_*^2}{\sigma_*^2}\left(\frac{1-p_*}{1-q_*}\right)T^{-1}\sum_{s=2}^{t-1}\rho_*^s\frac{\widetilde{s}_t\widetilde{s}_{t-s}}{\sigma_*^2}\right)\eta^2\right.$$
$$\left.-\frac{1}{4}\left(\frac{4}{\sigma_*^4}\left(\frac{1-p_*}{1-q_*}\right)^2T^{-1}\sum_{t=1}^{T}\left(\sum_{s=2}^{t-1}\rho_*^s\frac{\widetilde{e}_{t-s}\widetilde{e}_t}{\sigma_*^2}\right)^2\right)\eta^4\right]+o_p(1).$$

The three summations satisfy the LLN or CLT:

$$\frac{2}{\sigma_*^2}\left(\frac{1-p_*}{1-q_*}\right)T^{-1/2}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_*^s\frac{\widetilde{e}_t\widetilde{e}_{t-s}}{\sigma_*^2} \Rightarrow G(p_*,q_*), \tag{A.155}$$
$$\frac{2c_*^2}{\sigma_*^2}\left(\frac{1-p_*}{1-q_*}\right)T^{-1}\sum_{s=2}^{t-1}\rho_*^s\frac{\widetilde{s}_t\widetilde{s}_{t-s}}{\sigma_*^2} \to^p A_{2d}(p_*,q_*),$$
$$\frac{4}{\sigma_*^4}\left(\frac{1-p_*}{1-q_*}\right)^2T^{-1}\sum_{t=1}^{T}\left(\sum_{s=2}^{t-1}\rho_*^s\frac{\widetilde{e}_{t-s}\widetilde{e}_t}{\sigma_*^2}\right)^2 \to^p \Omega(p_*,q_*)>0,$$

where $G(p_*,q_*)\sim N(0,\Omega(p_*,q_*))$. Applying these limiting results to (A.154), we have

$$\sup_{\|\delta_2-\widetilde{\delta}\|\le T^{-1/4}M}2\left[\mathcal{L}(p_*,q_*,\delta_2)-\mathcal{L}(p_*,q_*,\widetilde{\delta})\right]\Rightarrow\sup_{\|\eta\|\le M}\left[\mathcal{W}_b^{(2)}(p_*,q_*,\eta)+A_{2d}(p_*,q_*)\eta^2\right], \tag{A.156}$$

where $\mathcal{W}^{(2)}(p_*,q_*,\eta)=G(p_*,q_*)\eta^2-(1/4)\Omega(p_*,q_*)\eta^4$. This result provides the leading terms for the refined approximation to $LR(p_*,q_*)$.

Next, we incorporate the refinement terms into this approximation. Apply the LLN and CLT to the leading terms in the second and fifth results of Lemma A.30, we have

$$\frac{1}{\sigma_*^3}\frac{(1-p_*)(p_*-q_*)}{(1-q_*)^2}T^{-1/2}\sum_{t=1}^{T}\frac{\widetilde{e}_t^3}{\sigma_*^3} \Rightarrow G^{(3)}(p_*,q_*),$$
$$\frac{c_*^3}{\sigma_*^3}\frac{(1-p_*)(p_*-q_*)}{(1-q_*)^2}T^{-5/4}\sum_{t=1}^{T}\frac{\widetilde{s}_t^3}{\sigma_*^3} = \frac{1}{3T^{1/4}}A_3(p_*,q_*)+O_p(T^{-3/4}),$$
$$\frac{1}{\sigma_*^6}\frac{(1-p_*)^2(p_*-q_*)^2}{(1-q_*)^4}T^{-1}\sum_{t=1}^{T}\left\{\left(\frac{\widetilde{e}_t}{\sigma_*}\right)^3-3\frac{\widetilde{e}_t}{\sigma_*}\right\}^2 \to^p \Omega^{(3)}(p_*,q_*)>0,$$

where $G^{(3)}(p_*,q_*)\sim N(0,\Omega^{(3)}(p_*,q_*))$. Let

$$C(p_*,q_*)=\frac{1}{\sigma_*^4}\left\{\frac{1-p_*}{2-p_*-q_*}\left(1+\left(\frac{1-p_*}{1-q_*}\right)^3\right)-3\left(\frac{1-p_*}{1-q_*}\right)^2\right\}.$$

Apply the LLN and CLT to the leading terms in the third and sixth results of Lemma A.30:

$$C(p_*, q_*)T^{-1/2} \sum_{t=1}^{T} \left(\frac{\widetilde{e}_t^4}{\sigma_*^4} - 6\frac{\widetilde{e}_t^2}{\sigma_*^2} + 3 \right) \Rightarrow G^{(4)}(p_*, q_*),$$

$$c_*^4 C(p_*, q_*)T^{-1/2} \left(T^{-1} \sum_{t=1}^{T} \frac{\widetilde{s}_t^4}{\sigma_*^4} - 3 \left(T^{-1} \sum_{t=1}^{T} \frac{\widetilde{s}_t^2}{\sigma_*^2} \right)^2 \right) = \frac{1}{12T^{1/2}} A_4(p_*, q_*) + O_p(T^{-1}),$$

$$C(p_*, q_*)^2 T^{-1} \sum_{t=1}^{T} \left\{ \left(\frac{\widetilde{e}_t}{\sigma_*} \right)^4 - 6 \left(\frac{\widetilde{e}_t}{\sigma_*} \right)^2 + 3 \right\}^2 \to^p \Omega^{(4)}(p_*, q_*) > 0,$$

where $G^{(4)}(p_*, q_*) \sim N(0, \Omega^{(4)}(p_*, q_*))$. These limits can be expressed using the the notation of (27):

$$\begin{aligned} \mathcal{W}^{(3)}(p_*, q_*, \eta) &= T^{-1/4}\frac{1}{3}G^{(3)}(p_*, q_*)\eta^3 - T^{-1/2}\frac{1}{36}\Omega^{(3)}(p_*, q_*)\eta^6, \\ \mathcal{W}^{(4)}(p_*, q_*, \eta) &= T^{-1/2}\frac{1}{12}G^{(4)}(p_*, q_*)\eta^4 - T^{-1}\frac{1}{576}\Omega^{(4)}(p_*, q_*)\eta^8. \end{aligned}$$

Incorporating $\mathcal{W}^{(3)}(p_*, q_*, \eta), \mathcal{W}^{(4)}(p_*, q_*, \eta), A_3(p_*, q_*)$, and $A_4(p_*, q_*)$ into (A.156), we obtain the following approximation to $\sup_{\|\delta_2 - \widetilde{\delta}\| \le T^{-1/4}M} 2[\mathcal{L}(p_*, q_*, \delta_2) - \mathcal{L}(p_*, q_*, \widetilde{\delta})]$:

$$\sup_{\|\eta\| \le M} \{ \sum_{j=2}^{4} \mathcal{W}^{(j)}(p_*, q_*, \eta) + A_{2d}(p_*, q_*)\eta^2 + T^{-1/2}A_3(p_*, q_*)\eta^3 + T^{-1}A_4(p_*, q_*)\eta^4 \}.$$

Because $\mathcal{W}^{(3)}(p_*, q_*, \eta), \mathcal{W}^{(4)}(p_*, q_*, \eta), A_3(p_*, q_*)$, and $A_4(p_*, q_*)$ only have high order effects, this approximation is asymptotically equivalent to (A.156). Because ε can be made arbitrarily small, $\|\eta\| \le M$ can be dropped (c.f. the argument in Step 2 of the proof of Proposition 1), and we obtain the following asymptotically valid approximation to $LR(p_*, q_*)$:

$$\mathcal{S}(p_*, q_*) = \sup_{\eta \in R} \{ \sum_{j=2}^{4} \mathcal{W}^{(j)}(p_*, q_*, \eta) + A_{2d}(p_*, q_*)\eta^2 + T^{-1/2}A_3(p_*, q_*)\eta^3 + T^{-1}A_4(p_*, q_*)\eta^4 \}. \quad \blacksquare$$

Remark 16 *The proof of Proposition 3 implies that the distribution of $\mathcal{S}(p_*, q_*)$ can be consistently estimated using the simulation algorithm Section 5.3 in the paper, provided that the constants $A_{2d}(p_*, q_*)$, $A_3(p_*, q_*)$, and $A_4(p_*, q_*)$ are added to the Gaussian random vector $[G_b(p_*, q_*)$, $G_b^{(3)}(p_*, q_*)$, $G_b^{(4)}(p_*, q_*)]$ in Step 3 before the optimization in Step 4. That is, we simulate*

$$\mathcal{S}_b(p_*, q_*) = \sup_{\eta \in R} \{ \sum_{j=2}^{4} \mathcal{W}_b^{(j)}(p_*, q_*, \eta) + A_{2d}(p_*, q_*)\eta^2 + T^{-1/2}A_3(p_*, q_*)\eta^3 + T^{-1}A_4(p_*, q_*)\eta^4 \},$$

where $\mathcal{W}_b^{(j)}(p_*, q_*, \eta)$ is defined by (32). The next result shows that the empirical distribution of $\mathcal{S}_b(p_T, q_T)$ is a weakly consistent estimator of the limiting distribution of $LR(p_T, q_T)$ under the contiguous alternatives when (p_T, q_T) follows SEQ1 *and* SEQ2 *for any $a > 0$. This provides a further justification for this refined approximation under the alternative hypothesis.*

Corollary A.1 *Suppose that the data are generated by (A.141)-(A.142) with* $A_T = c_* T^{-\gamma}$ *for some* $c_* \neq 0$*, and that there exists some* $\eta > 0$ *such that* $\sup_{|\delta-\widetilde{\delta}|<\eta} T^{-1}|\mathcal{L}^{(j)}(p_T, q_T, \delta)| = O_p(1)$ *holds for* $j = 5, 7, 9$*. Further, suppose* $\gamma = \min(1/4 - a, 1/6)$ *if* (p_T, q_T) *follows* Seq1 *and* $\gamma = \min(1/4 - a, 1/8)$ *if it follows* Seq2 *. Then, the empirical distribution of* $\mathcal{S}_b(p_T, q_T)$ *is a weakly consistent estimator of the limiting distribution of* $LR(p_T, q_T)$ *under* Seq1 *and* Seq2 *for any* $a > 0$*.*

Proof of Corollary A.1. The results follow from Lemmas A.29 and A.30.

Consider Seq1 with $a < 1/12$. By the second result of Lemma A.29,

$$\Pr\left(LR(p_T, q_T) \le x\right) - \Pr\left(\sup_{\eta \in R}\left\{T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T, q_T)\eta^2 - \frac{1}{4}\widetilde{\Omega}(p_T, q_T)\eta^4\right\} \le x\right) \to 0$$

for any $x \in R$. Because η is unrestricted, we can redefine it as $T^a\eta$. Then,

$$\Pr\left(LR(p_T, q_T) \le x\right) - \Pr\left(\sup_{\eta \in R}\left\{T^{-1/2+2a}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T, q_T)\eta^2 - \frac{1}{4}T^{4a}\widetilde{\Omega}(p_T, q_T)\eta^4\right\} \le x\right) \to 0.$$

We now study the two terms inside the braces. Lemma A.30 implies

$$T^{4a}\widetilde{\Omega}(p_T, q_T) = \frac{4}{\sigma_*^4}\left(\frac{1-p_T}{1-q_T}\right)^2 T^{-1+4a}\sum_{t=1}^{T}\left(\sum_{s=2}^{t-1}\rho_T^s\frac{\widetilde{e}_{t-s}\widetilde{e}_t}{\sigma_*^2}\right)^2 + T^{4a}R_T + o_p(T^{-1/2+4a})$$

and

$$\begin{aligned}
& T^{-1/2+2a}\sum_{t=1}^{T}\widetilde{U}_t^{(2)}(p_T, q_T) \\
= \; & \frac{2}{\sigma_*^2}\left(\frac{1-p_T}{1-q_T}\right)T^{-1/2+2a}\sum_{t=1}^{T}\{\sum_{s=2}^{t-1}\rho_T^s\frac{\widetilde{e}_t\widetilde{e}_{t-s}}{\sigma_*^2} + (T^{-1/4+a}c_*)^2\sum_{s=2}^{t-1}\rho_T^s\frac{\widetilde{s}_t\widetilde{s}_{t-s}}{\sigma_*^2}\} + T^{2a}R_T + o_p(T^{-1/2+2a}),
\end{aligned}$$

where R_T is dominated by the term preceding it. Further, by the LLN,

$$\frac{4}{\sigma_*^4}\left(\frac{1-p_T}{1-q_T}\right)^2 T^{-1+4a}\sum_{t=1}^{T}\left(\sum_{s=2}^{t-1}\rho_T^s\frac{\widetilde{e}_{t-s}\widetilde{e}_t}{\sigma_*^2}\right)^2 \to^p \bar{\Omega}(p, q),$$

where (p, q) denotes the limit of (p_T, q_T) and $\bar{\Omega}(p, q) > 0$. In addition,

$$\begin{aligned}
\frac{2}{\sigma_*^2}\left(\frac{1-p_T}{1-q_T}\right)T^{-1/2+2a}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_T^s\frac{\widetilde{e}_t\widetilde{e}_{t-s}}{\sigma_*^2} &= \frac{2}{\sigma_*^2}\left(\frac{1-p_T}{1-q_T}\right)T^{-1/2}\sum_{t=1}^{T}\frac{\widetilde{e}_t\widetilde{e}_{t-2}}{\sigma_*^2} + o_p(1) \\
&\Rightarrow \bar{G}^{(2)}(p, q), \\
\frac{2c_*^2}{\sigma_*^2}\left(\frac{1-p_T}{1-q_T}\right)T^{-1+4a}\sum_{t=1}^{T}\sum_{s=2}^{t-1}\rho_T^s\frac{\widetilde{s}_t\widetilde{s}_{t-s}}{\sigma_*^2} &= \frac{2c_*^2}{\sigma_*^2}\left(\frac{1-p_T}{1-q_T}\right)T^{-1}\sum_{t=1}^{T}\rho_T^{-2}\frac{\widetilde{s}_t\widetilde{s}_{t-2}}{\sigma_*^2} + o_p(1) \\
&\to {}^p\bar{A}_{2d}(p, q),
\end{aligned}$$

where the second equality holds because $Cov(s_t s_{t-s}) = O_p(|\rho_T|^s)$, $\bar{G}^{(2)}(p,q) \sim N(0, \bar{\Omega}(p,q))$, and $\bar{A}_{2d}(p,q)$ is finite. Applying these results, we have

$$LR(p_T, q_T) \Rightarrow \sup_{\eta \in R} \left\{ \bar{G}^{(2)}(p,q)\eta^2 + \bar{A}_{2d}(p,q)\eta^2 - \frac{1}{4}\bar{\Omega}(p,q)\eta^4 \right\}. \quad \text{(A.157)}$$

We now compare this limit with $\mathcal{S}_b(p_T, q_T)$. Because $a < 1/12$, the effect of $\mathcal{W}_b^{(3)}(p_*, q_*, \eta)$, $\mathcal{W}_b^{(4)}(p_*, q_*, \eta)$, $T^{-1/2}A_3(p_*, q_*)\eta^3$, and $T^{-1}A_4(p_*, q_*)\eta^4$ on $\mathcal{S}_b(p_T, q_T)$ is asymptotically negligible. Thus, the distribution of $\mathcal{S}_b(p_T, q_T)$ is asymptotically equivalent to that of $\sup_{\eta \in R}\{G_b^{(2)}(p_T, q_T)\,\eta^2 + A_{2d}(p_T, q_T)\eta^2 - \frac{1}{4}\widetilde{\Omega}(p_T, q_T)\eta^4\}$. Applying a change of variable, we can rewrite the latter as

$$\sup_{\eta \in R} \left\{ T^{2a} G_b^{(2)}(p_T, q_T)\,\eta^2 + T^{2a} A_{2d}(p_T, q_T)\eta^2 - \frac{1}{4} T^{4a}\widetilde{\Omega}(p_T, q_T)\eta^4 \right\}.$$

There are two differences between this expression and the limit of $LR(p_T, q_T)$ in (A.157). First, $T^{2a}A_{2d}(p_T, q_T)$ and $T^{4a}\widetilde{\Omega}(p_T, q_T)$ are replaced by their limits $\bar{A}_{2d}(p,q)$ and $\bar{\Omega}(p,q)$. Second, we have $T^{2a}G_b^{(2)}(p_T, q_T) \sim N(0, T^{4a}\widetilde{\Omega}(p_T, q_T))$ while $\bar{G}^{(2)}(p,q) \sim N(0, \text{plim}_{T\to\infty} T^{4a}\widetilde{\Omega}(p_T, q_T))$. These differences are asymptotically negligible, in the sense that the distribution of $\mathcal{S}_b(p_T, q_T)$ is a weakly consistent estimator of the limiting distribution of $LR(p_T, q_T)$ in (A.157) by the same argument as in the proof of Proposition 2.3. This completes the proof for the $a < 1/12$ case.

The proofs for the remaining cases under SEQ1 and SEQ2 are similar. That is, we obtain the weak limit of $LR(p_T, q_T)$, compare it with $\mathcal{S}_b(p_T, q_T)$, and then apply the argument in the proof of the third result of Proposition 2. The details are omitted. ■

Remark 17 *The next result presents the analytical expressions of various moments of* $1_{\{s_t=2\}}$. *They are used to compute* $A_{2s}(p_*, q_*), A_{2d}(p_*, q_*), A_3(p_*, q_*)$,*and* $A_4(p_*, q_*)$.

Lemma A.31 *Suppose the data are generated by DGP (38). Then,*

$$\begin{aligned}
E\left(1_{\{s_t=2\}} - E1_{\{s_t=2\}}\right)^2 &= \frac{(1-p_*)(1-q_*)}{(2-p_*-q_*)^2}, \\
\sum_{j=1}^{\infty} \rho_*^j \operatorname{Cov}\left(1_{\{s_t=2\}}, 1_{\{s_{t-j}=2\}}\right) &= \frac{\rho_*^2}{1-\rho_*^2}\frac{(1-p_*)(1-q_*)}{(2-p_*-q_*)^2}, \\
\sum_{j=2}^{\infty} \rho_*^j \operatorname{Cov}\left(1_{\{s_t=2\}}, 1_{\{s_{t-j}=2\}}\right) &= \frac{\rho_*^4}{1-\rho_*^2}\frac{(1-p_*)(1-q_*)}{(2-p_*-q_*)^2}, \\
E\left(1_{\{s_t=2\}} - E1_{\{s_t=2\}}\right)^3 &= \frac{(1-p_*)(1-q_*)(p_*-q_*)}{(2-p_*-q_*)^3}, \\
E\left(1_{\{s_t=2\}} - E1_{\{s_t=2\}}\right)^4 &= \frac{\left(p_*^2+q_*^2-p_*q_*-p_*-q_*+1\right)(1-q_*)(1-p_*)}{(2-p_*-q_*)^4}.
\end{aligned}$$

Proof of Lemma A.31. The first result holds because

$$E\left(1_{\{s_t=2\}} - E1_{\{s_t=2\}}\right)^2 = E1_{\{s_t=2\}} - \left(E1_{\{s_t=2\}}\right)^2 = \frac{1-p_*}{2-p_*-q_*} - \left(\frac{1-p_*}{2-p_*-q_*}\right)^2.$$

The second result holds because $1_{\{s_t=2\}} - E1_{\{s_t=2\}}$ is a Markov process and its second order properties are the same as those of an AR(1) process with autoregressive coefficient ρ_*. The third result holds for the same reason. The fourth result holds because

$$
\begin{aligned}
& E\left(1_{\{s_t=2\}} - E1_{\{s_t=2\}}\right)^3 \\
= \;& E\left(1_{\{s_t=2\}} + 3 * 1_{\{s_t=2\}}(E1_{\{s_t=2\}})^2 - 3 * 1_{\{s_t=2\}}E(1_{\{s_t=2\}}) - \left(E1_{\{s_t=2\}}\right)^3\right) \\
= \;& E\left\{\frac{1-p_*}{2-p_*-q_*} + 3\left(\frac{1-p_*}{2-p_*-q_*}\right)^3 - 3\left(\frac{1-p_*}{2-p_*-q_*}\right)^2 - \left(\frac{1-p_*}{2-p_*-q_*}\right)^3\right\}.
\end{aligned}
$$

For the fifth result,

$$
\begin{aligned}
& E\left(1_{\{s_t=2\}} - E1_{\{s_t=2\}}\right)^4 \\
= \;& E\left(1_{\{s_t=2\}} + \left(E1_{\{s_t=2\}}\right)^4 + 6 * 1_{\{s_t=2\}}\left(E1_{\{s_t=2\}}\right)^2 - 4 * 1_{\{s_t=2\}}\left(E1_{\{s_t=2\}}\right) - 4 * 1_{\{s_t=2\}}\left(E1_{\{s_t=2\}}\right)^3\right) \\
= \;& \frac{1-p_*}{2-p_*-q_*} + \left(\frac{1-p_*}{2-p_*-q_*}\right)^4 + 6\left(\frac{1-p_*}{2-p_*-q_*}\right)^3 - 4\left(\frac{1-p_*}{2-p_*-q_*}\right)^2 - 4\left(\frac{1-p_*}{2-p_*-q_*}\right)^4.
\end{aligned}
$$

The result follows after rearranging the terms on the right hand side. ■

References

Billingsley, P. (1986). *Probability and Measure* (Second ed.). New York: Wiley.

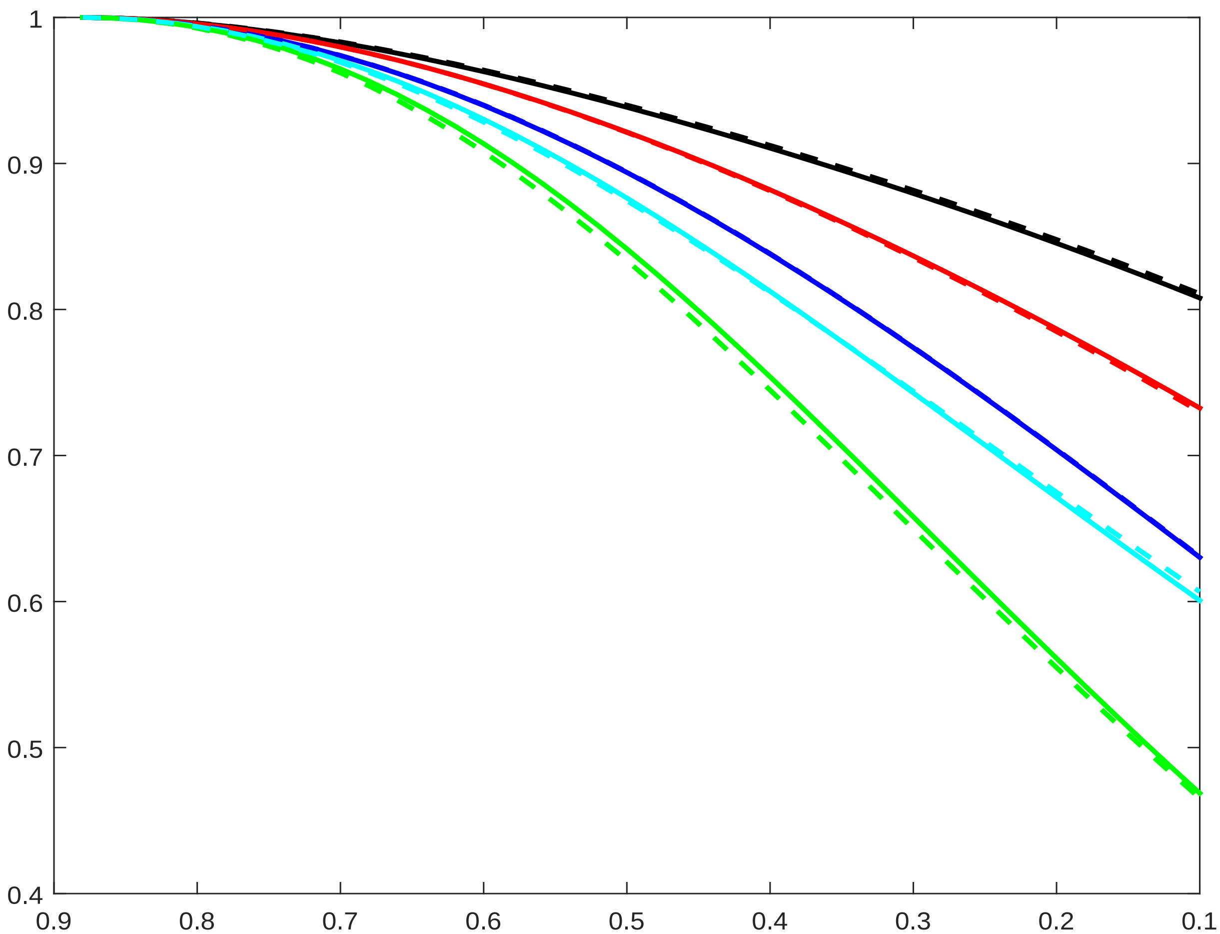
—— (1999): *Convergence of Probability Measures*, John Wiley & Sons.

Gradshteyn, I.S. and I.M. Ryzhik (2007): *Table of Integrals, Series, and Products (7th ed)*, edited by A. Jeffrey and D. Zwillinger, Academic Press.

Politis, D.N. and J.P. Romano (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.

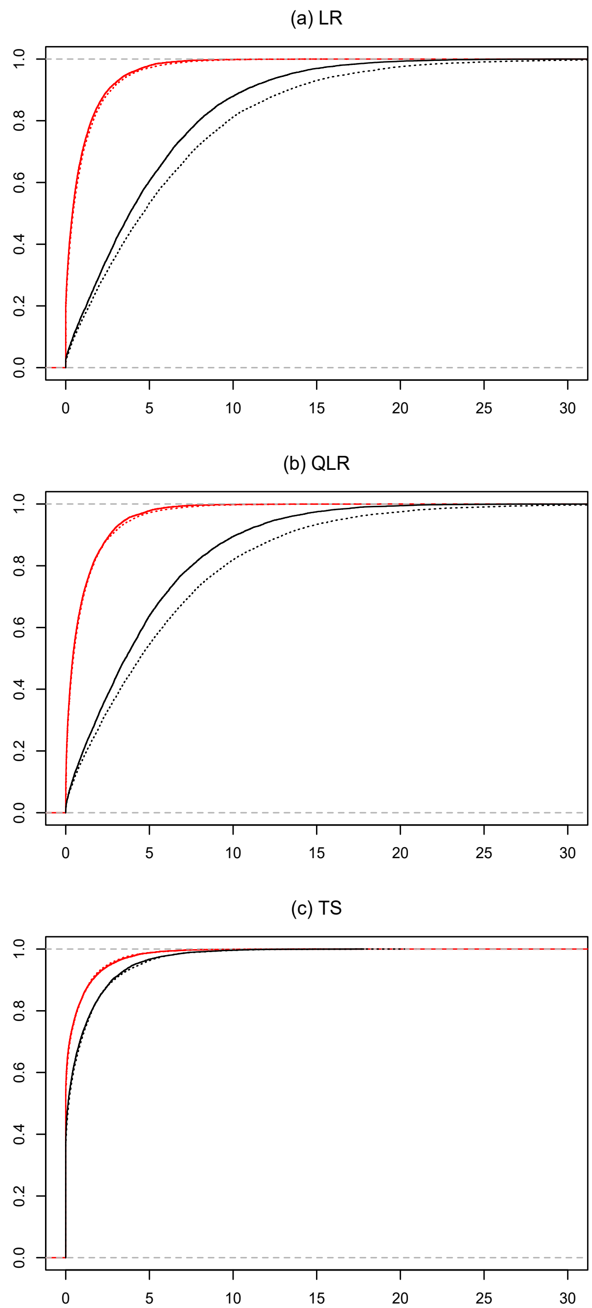
White, H. (2001): *Asymptotic Theory for Econometricians*, Academic Press.

Figure S1. Correlation functions



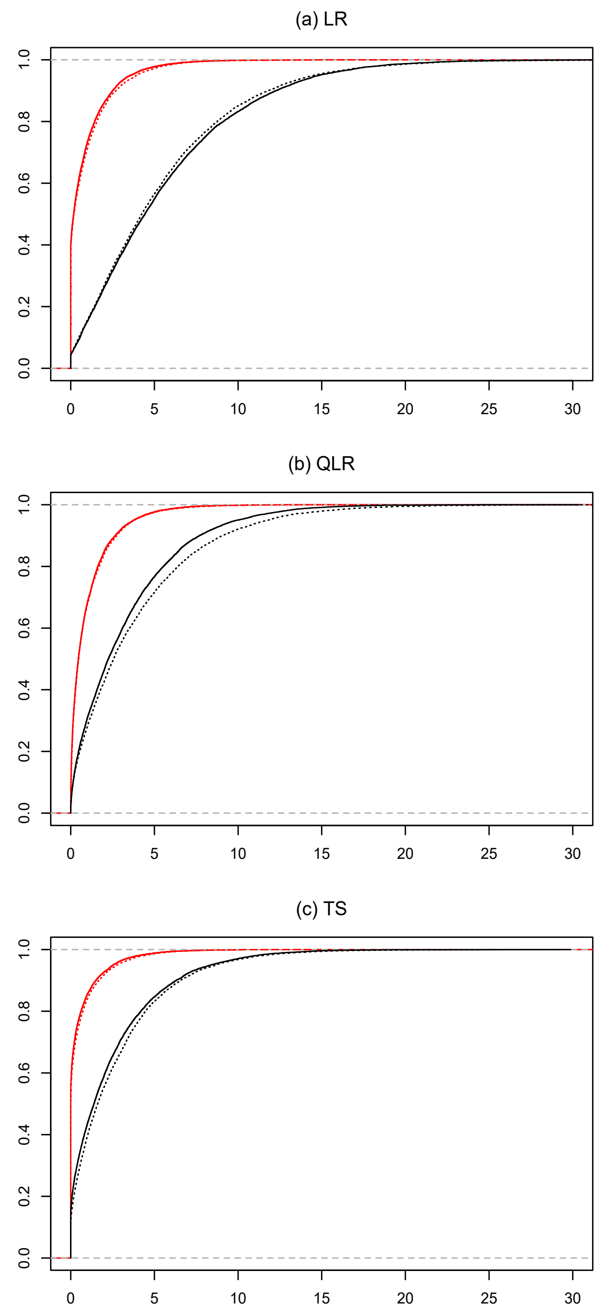
Note. The figure shows correlations between $G(p_r, q_r)$ and $G(p_s, q_s)$ with $(p_r, q_r) = (0.6, 0.9)$ and $(p_s, q_s) = (0.6, x)$, where x varies between 0.1 and 0.9. The solid lines starting from the top correspond to expressions in displays (A.56), (A.54), (A.52), (A.55) and (A.53) in the appendix. The dashed lines are correlations computed using simulations with $T = 250$.

Figure S2. Distributions under a simple DGP: static model, q_{*} =0.2



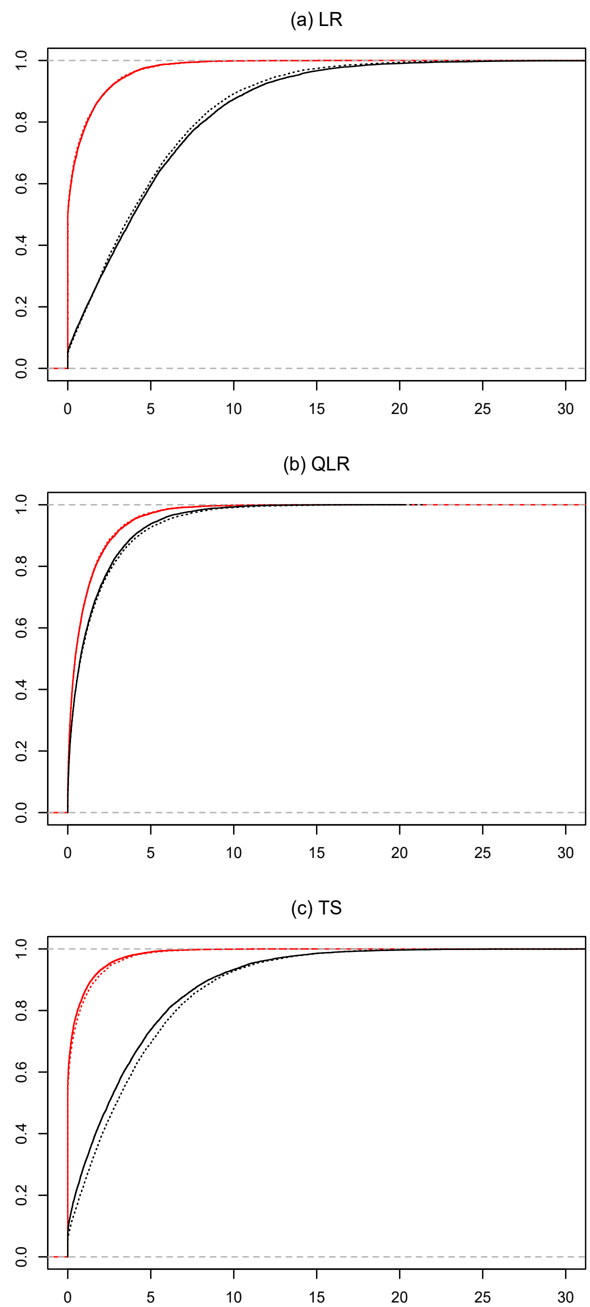
Note. LR: likelihood ratio; QLR: Cho and White (2007); TS: Carrasco, Hu, and Ploberger (2014). CDFs of finite sample distributions and asymptotic approximations are reported. Red dotted line: approximation under the null; red solid line: finite sample distribution under the null; black dotted line: approximation under the alternative; black solid line: finite sample distribution under the alternative.

Figure S3. Distributions under a simple DGP: static model, $q_* = 0.4$



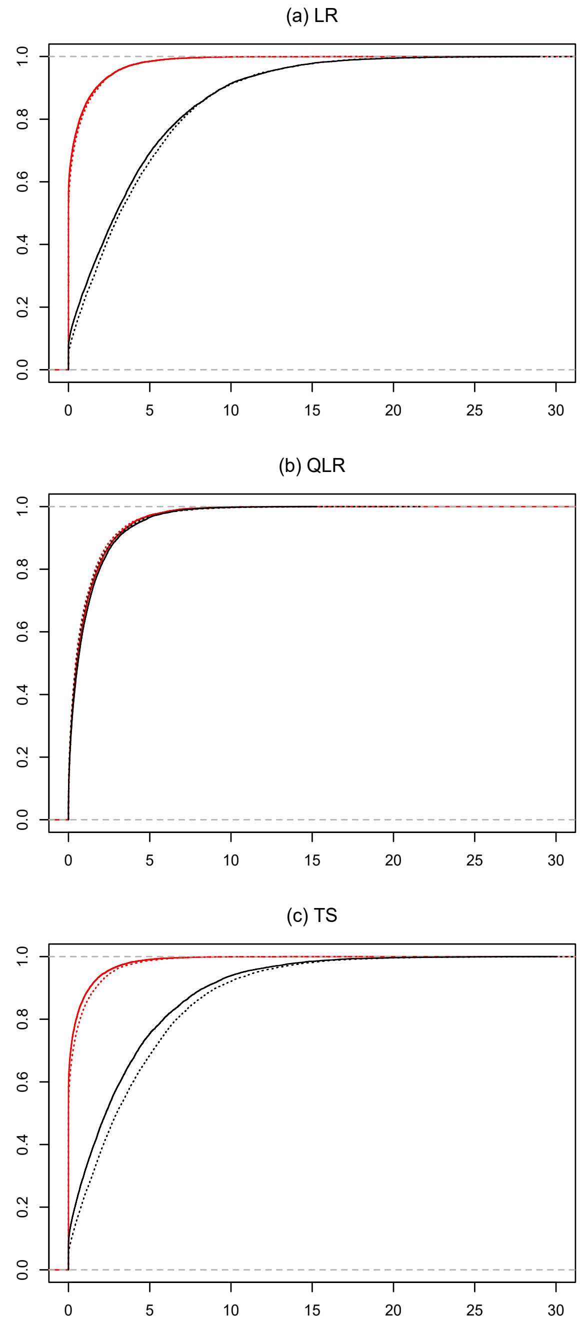
Note. See Figure S2.

Figure S4. Distributions under a simple DGP: static model, $q_* = 0.6$



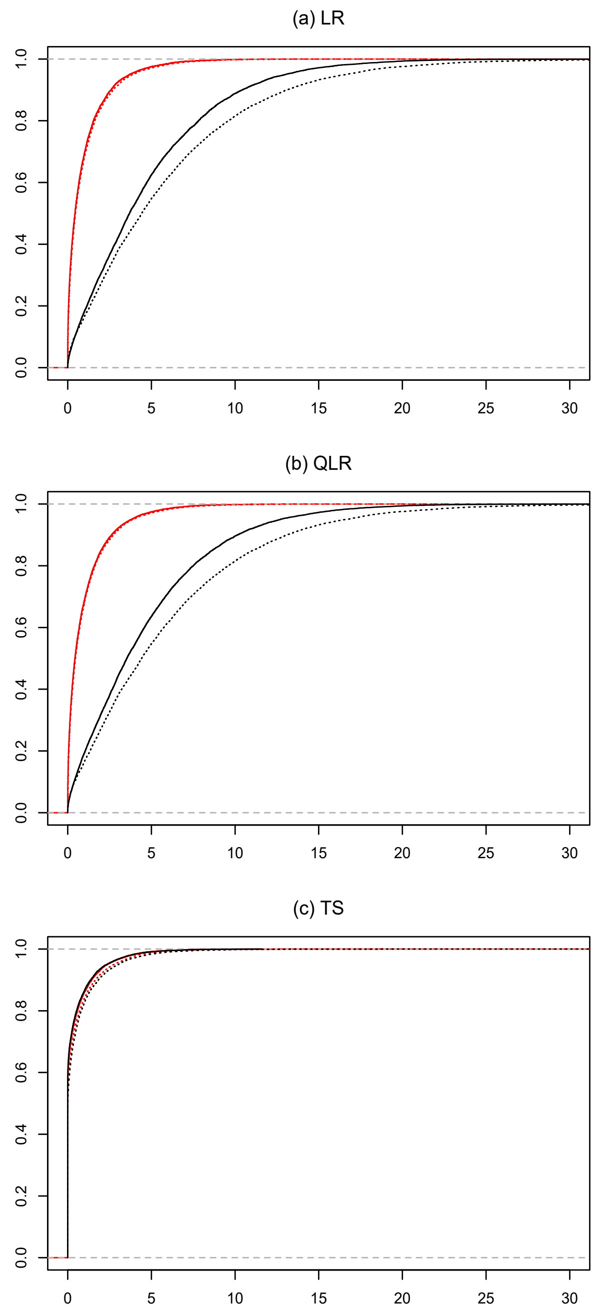
Note. See Figure S2.

Figure S5. Distributions under a simple DGP: static model, q_{*} =0.8



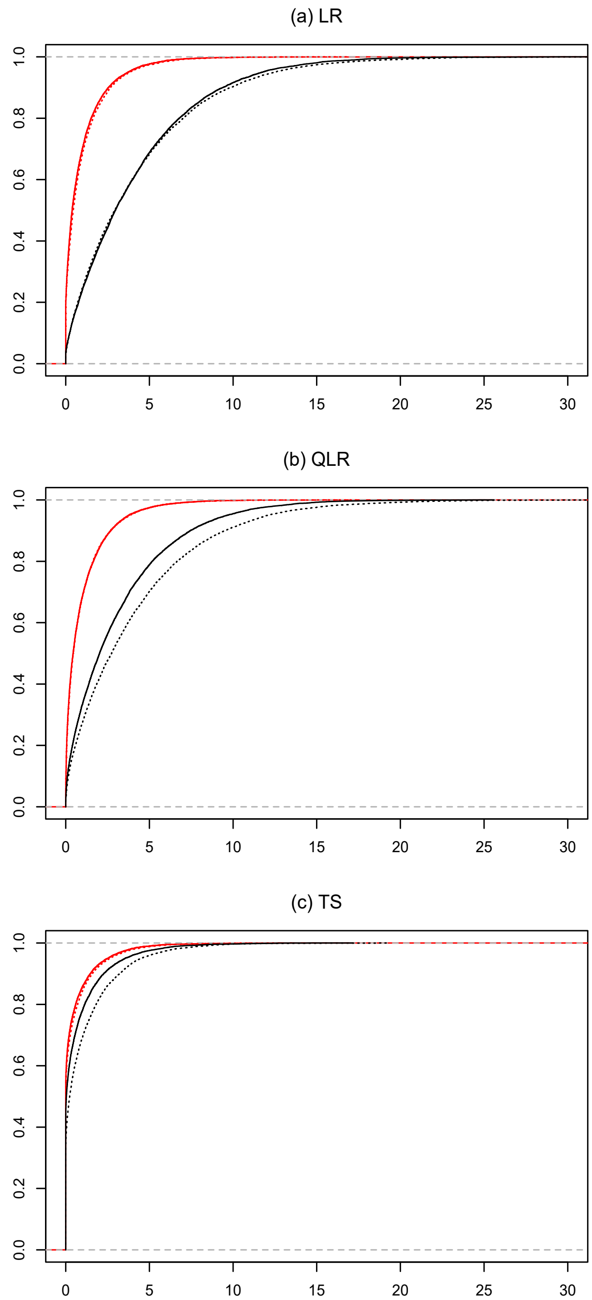
Note. See Figure S2.

Figure S6. Distributions under a simple DGP: dynamic model, q_{*} =0.2



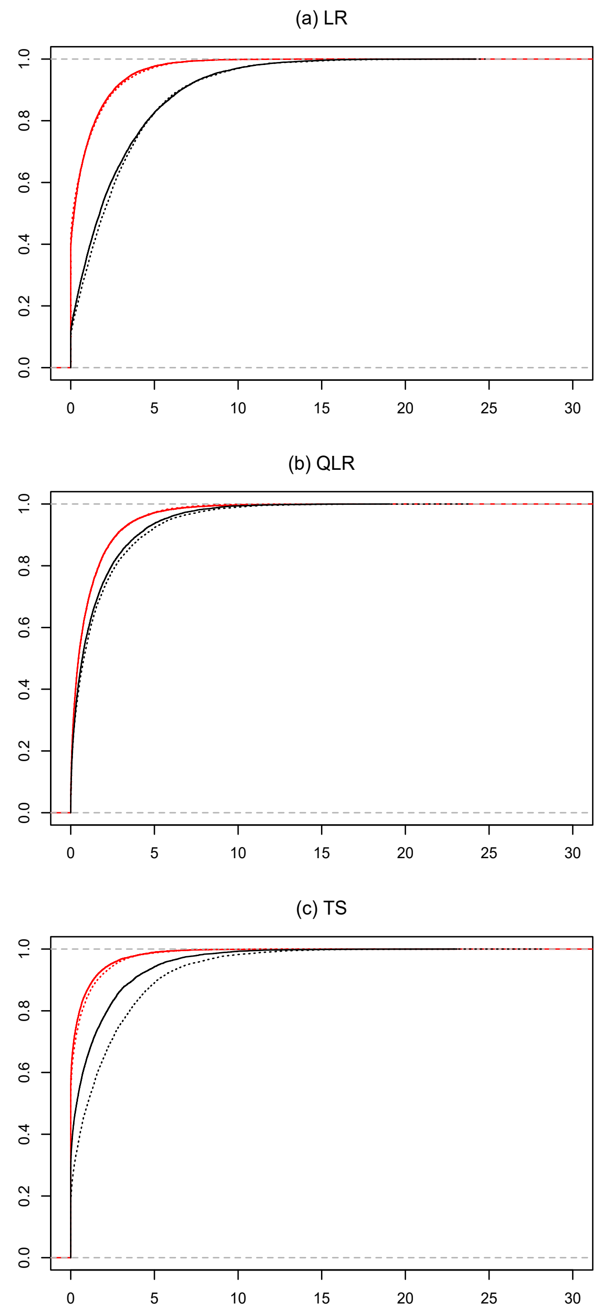
Note. See Figure S2.

Figure S7. Distributions under a simple DGP: dynamic model, q_{*} =0.4



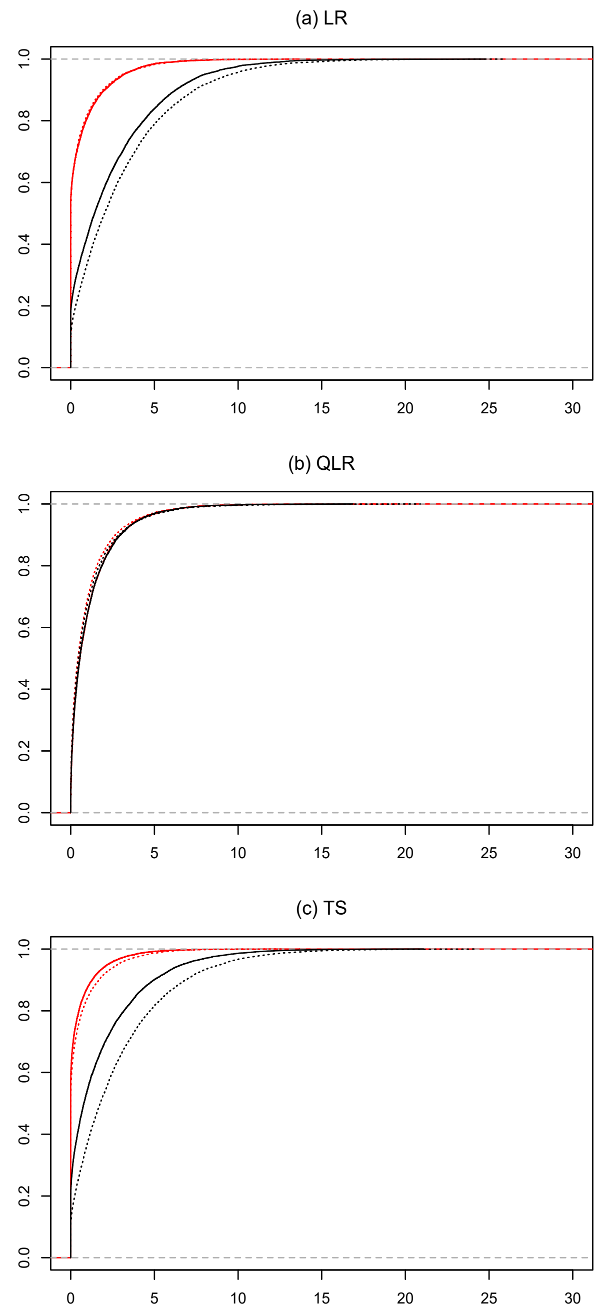
Note. See Figure S2.

Figure S8. Distributions under a simple DGP: dynamic model, q_* =0.6



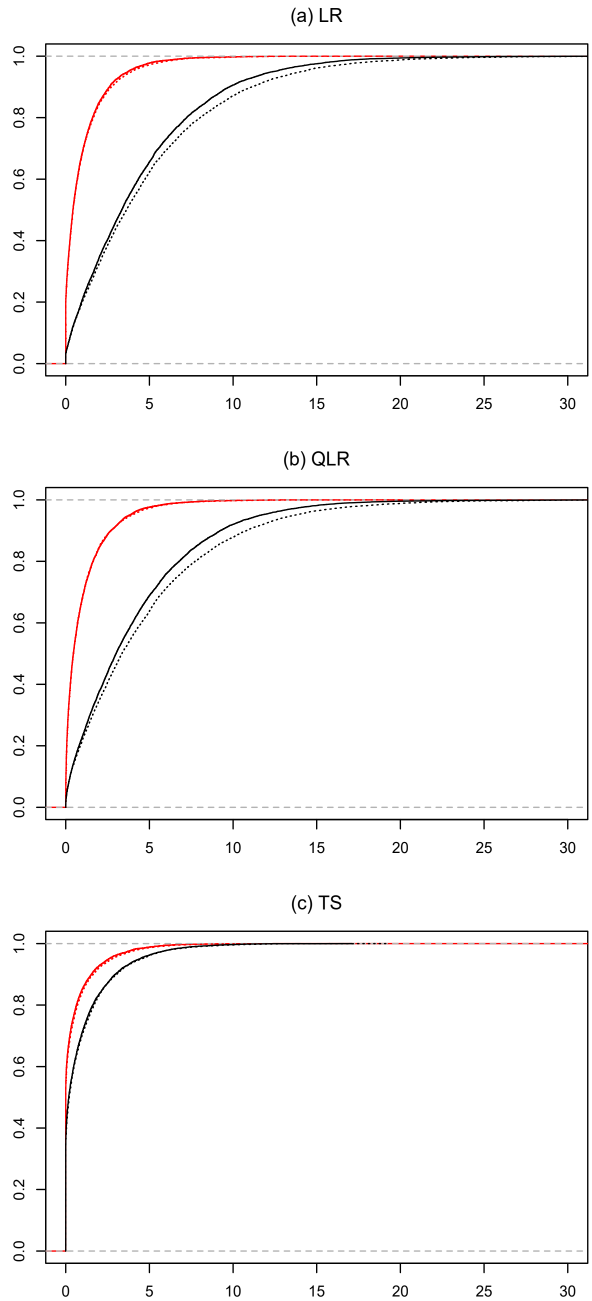
Note. See Figure S2.

Figure S9. Distributions under a simple DGP: dynamic model, $q_* = 0.8$



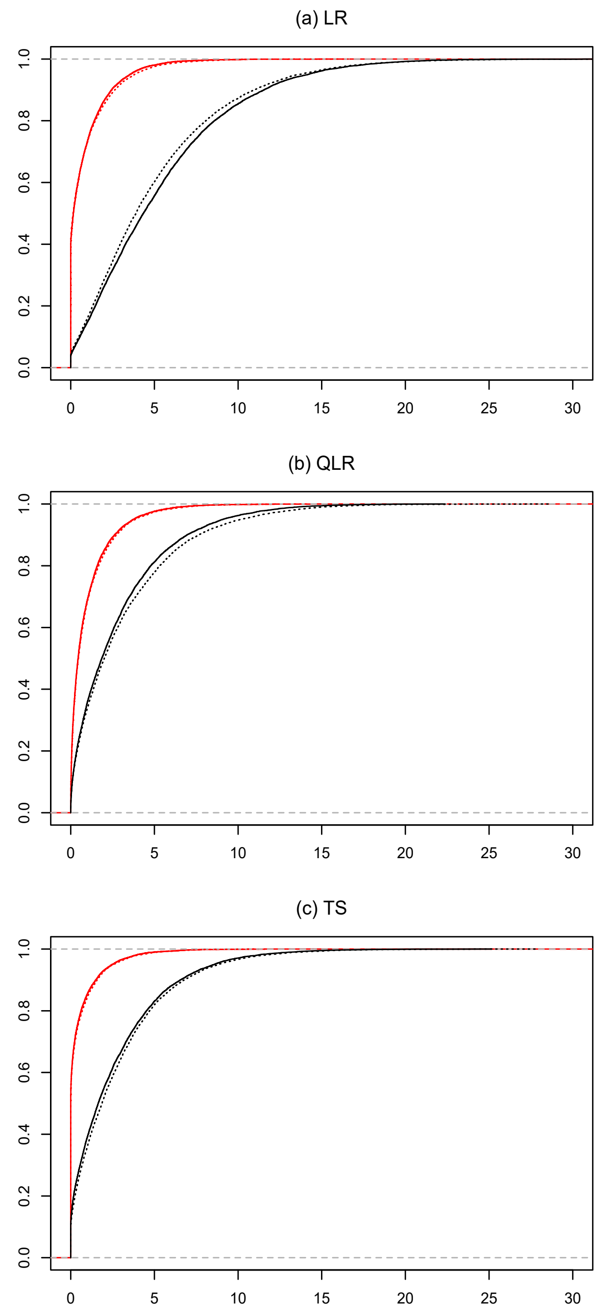
Note. See Figure S2.

Figure S10. Distributions under a simple DGP: static model, q_* =0.2, T =1000



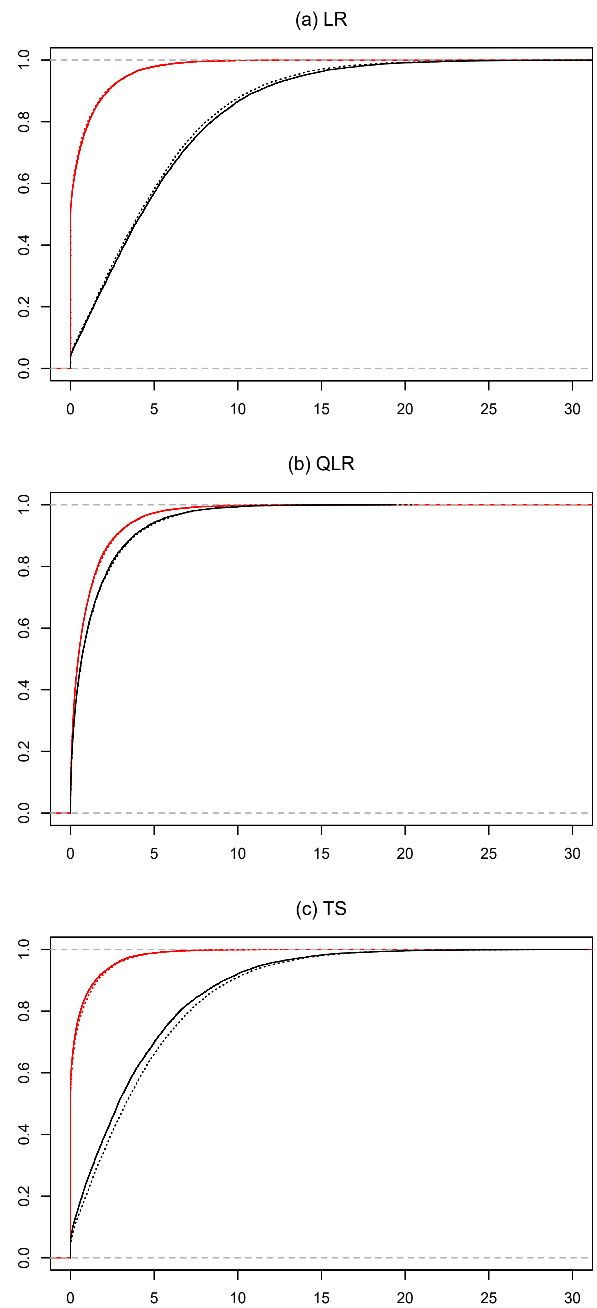
Note. See Figure S2. $T = 1000$.

Figure S11. Distributions under a simple DGP: static model, q_* =0.4, T =1000



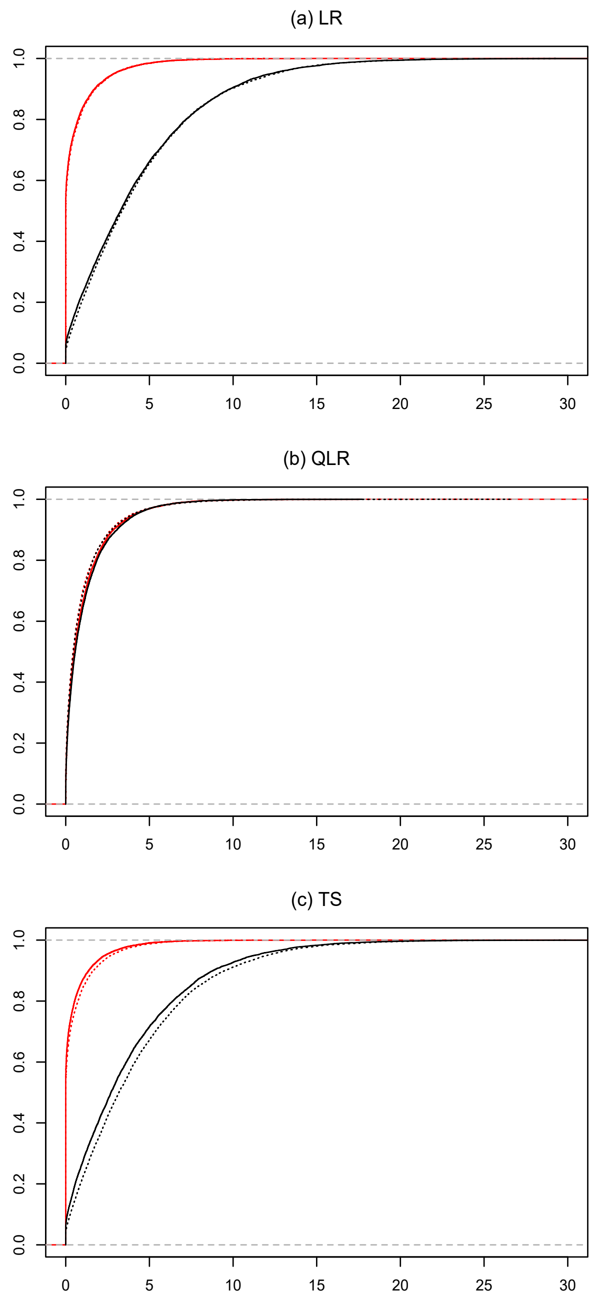
Note. See Figure S2.

Figure S12. Distributions under a simple DGP: static model, q_{*} =0.6, T =1000



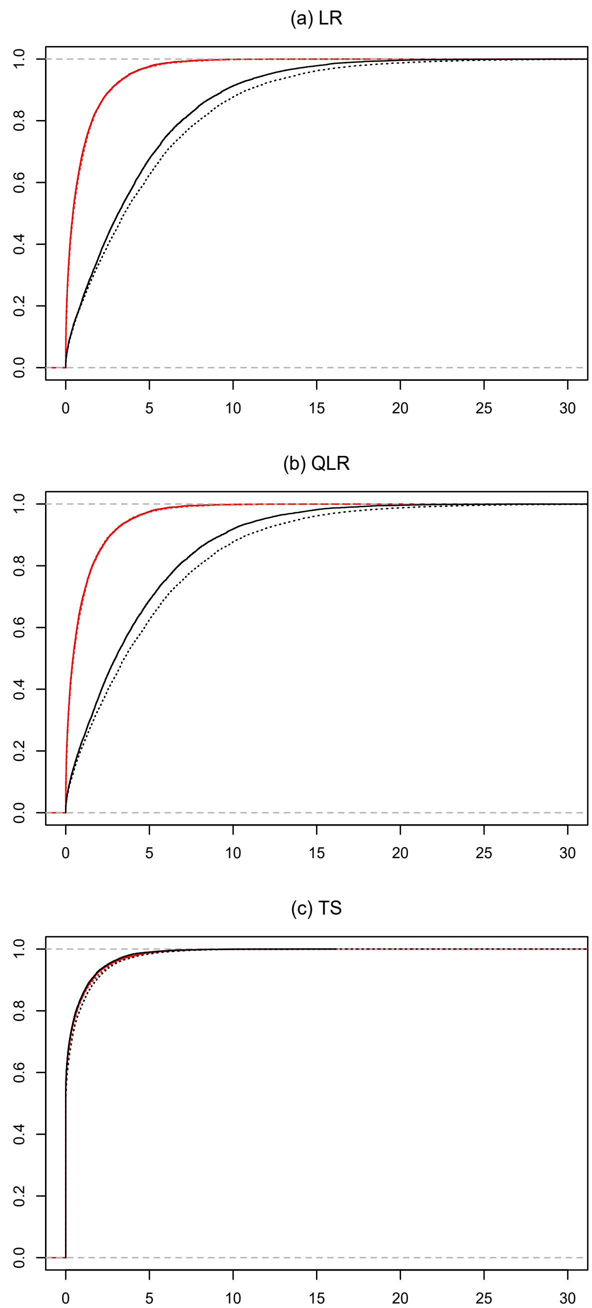
Note. See Figure S2.

Figure S13. Distributions under a simple DGP: static model, q_* =0.8, T =1000



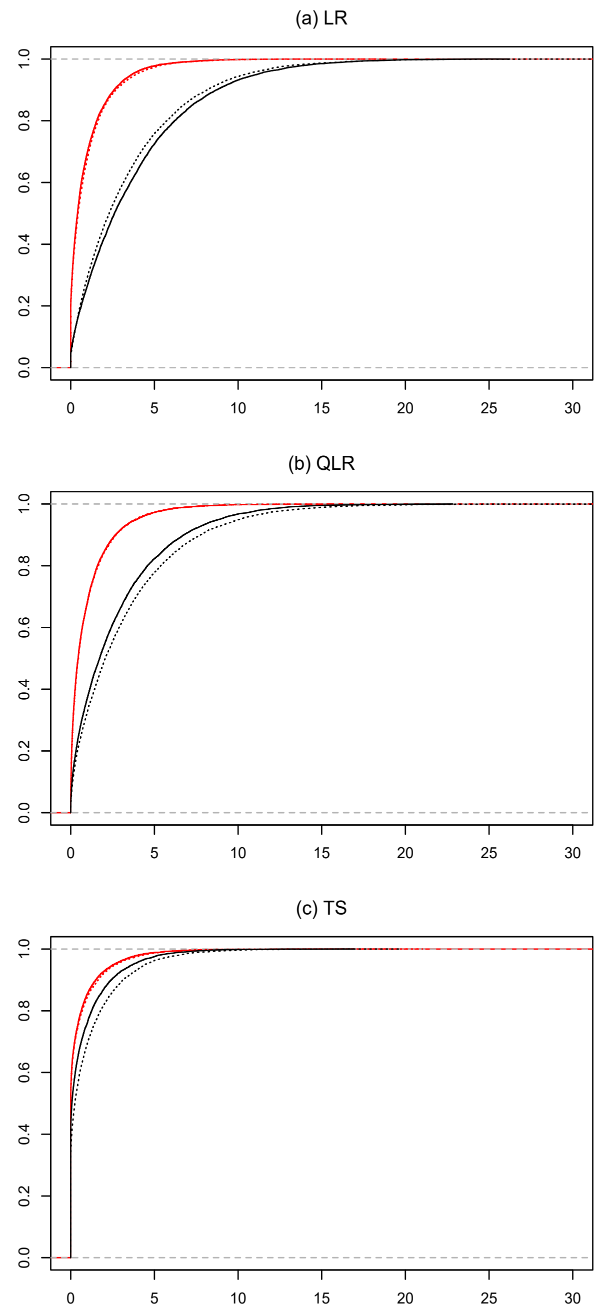
Note. See Figure S2.

Figure S14. Distributions under a simple DGP: dynamic model, q_* =0.2, T =1000



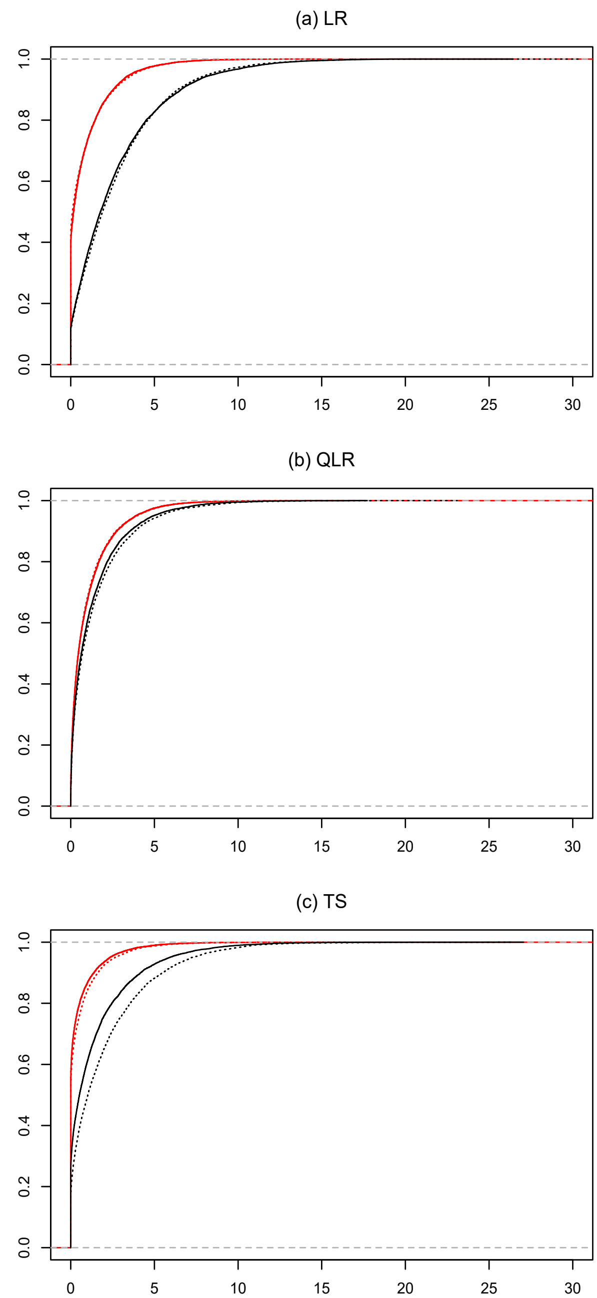
Note. See Figure S2.

Figure S15. Distributions under a simple DGP: dynamic model, q_* =0.4, T =1000



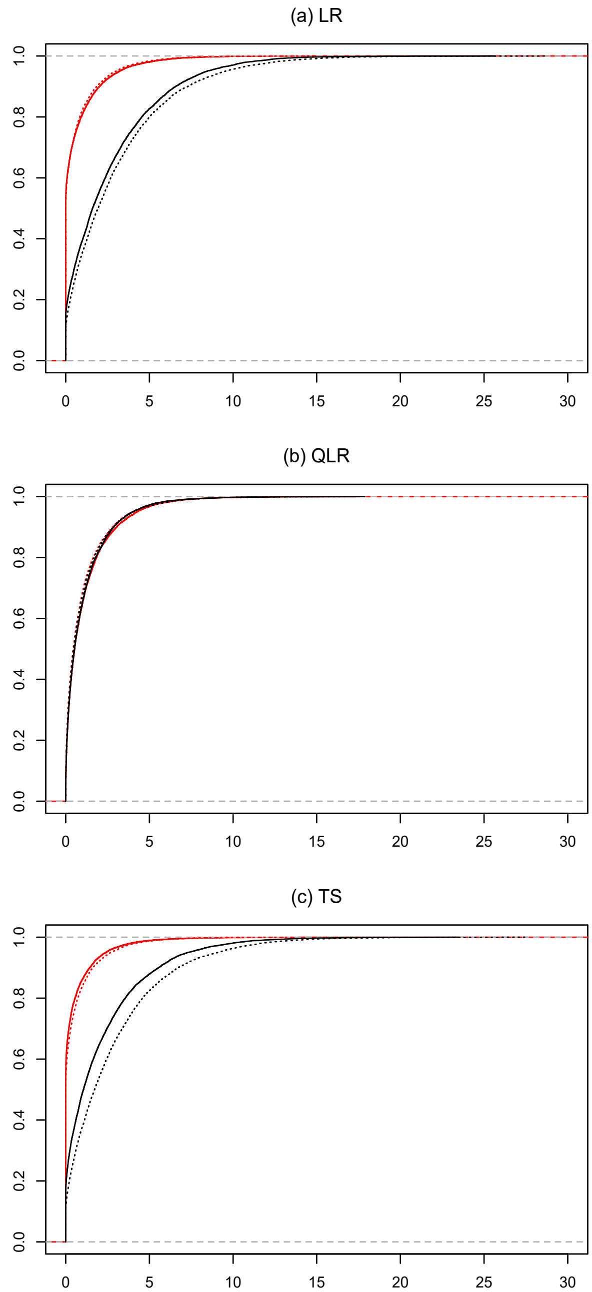
Note. See Figure S2.

Figure S16. Distributions under a simple DGP: dynamic model, q_* =0.6, T =1000



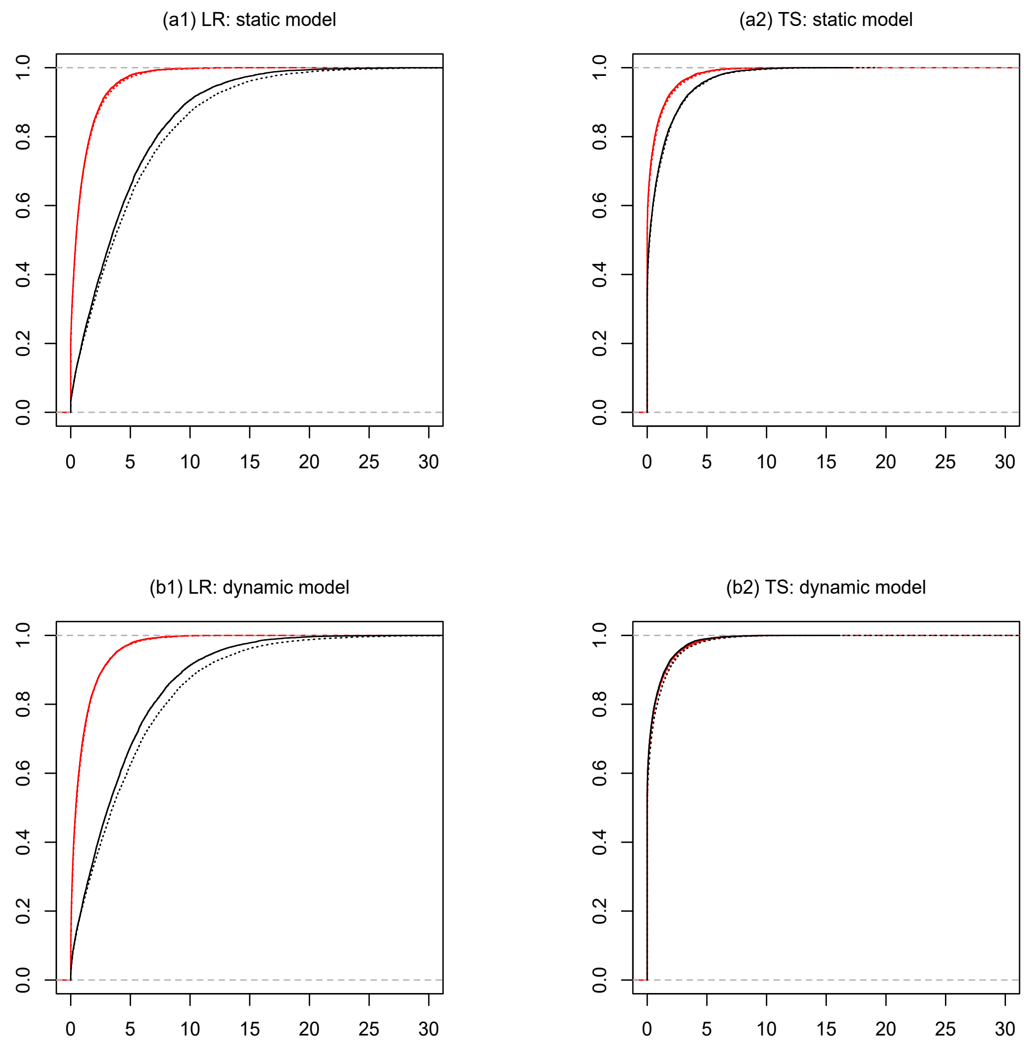
Note. See Figure S2.

Figure S17. Distributions under a simple DGP: dynamic model, $q_* = 0.8$, $T = 1000$



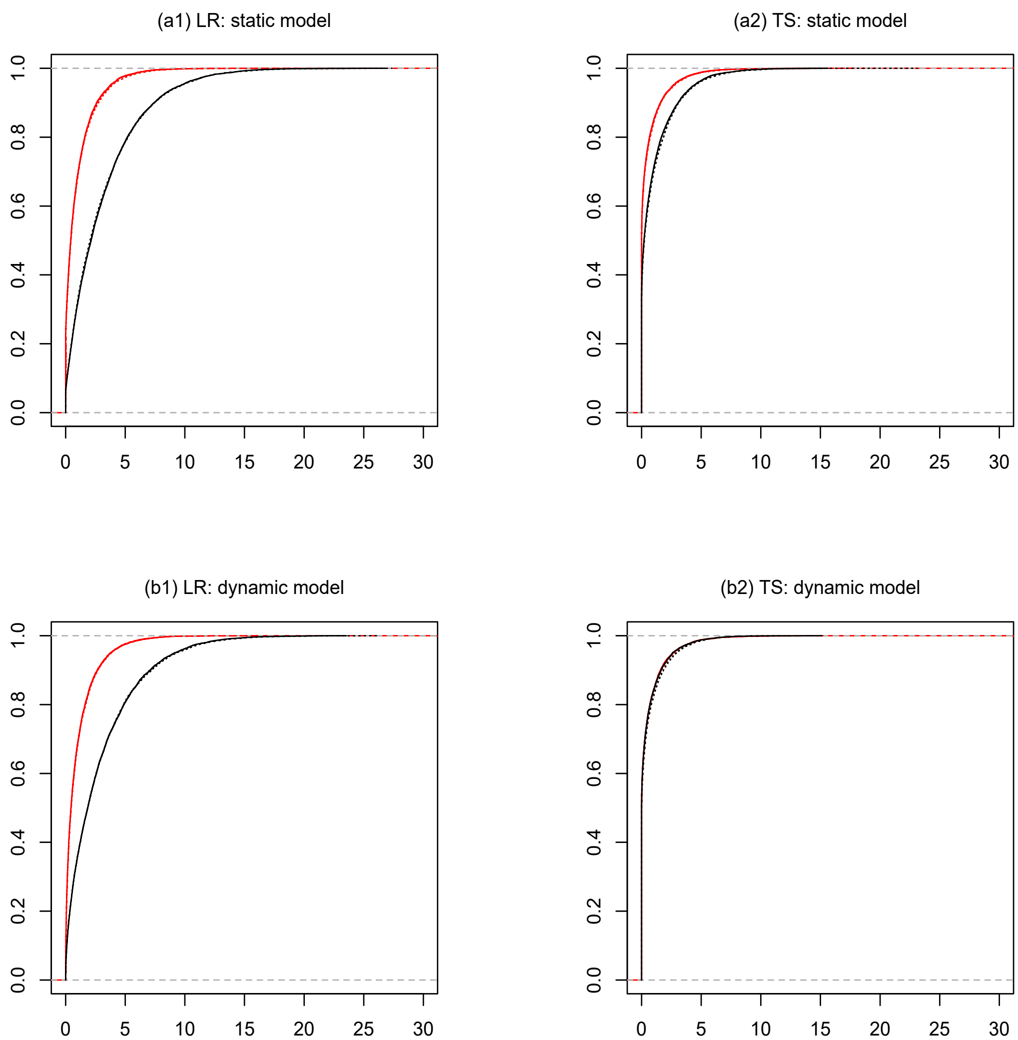
Note. See Figure S2.

Figure S18. The refined approximations and their finite sample counterparts: T =1000



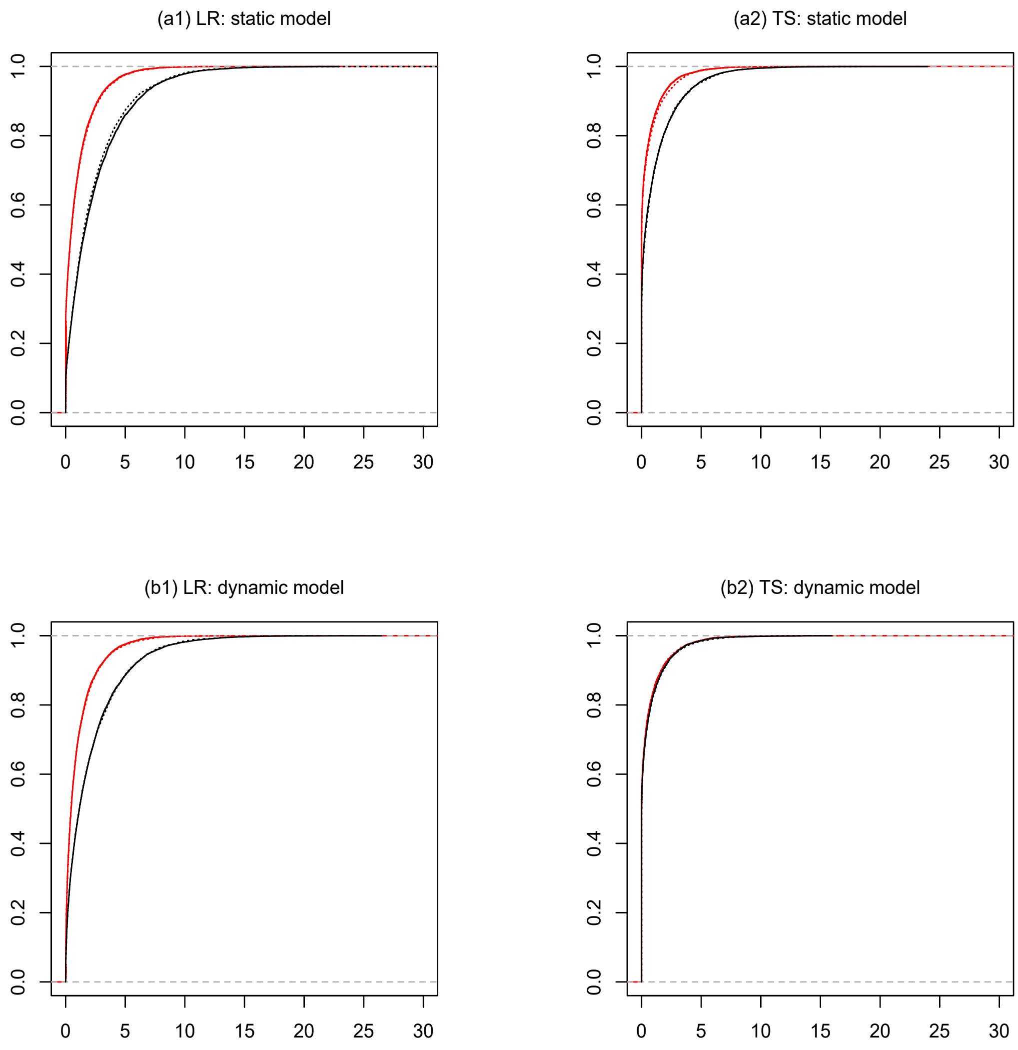
Note. LR: likelihood ratio; TS: Carrasco, Hu, and Ploberger (2014). CDFs of finite sample distributions and asymptotic approximations are reported. Red dotted line: approximation under the null; red solid line: finite sample distribution under the null; black dotted line: approximation under the alternative; black solid line: finite sample distribution under the alternative. (a1) and (a2) are computed using the same data; so are (b1) and (b2).

Figure S19. The refined approximations and their finite sample counterparts: T =5000



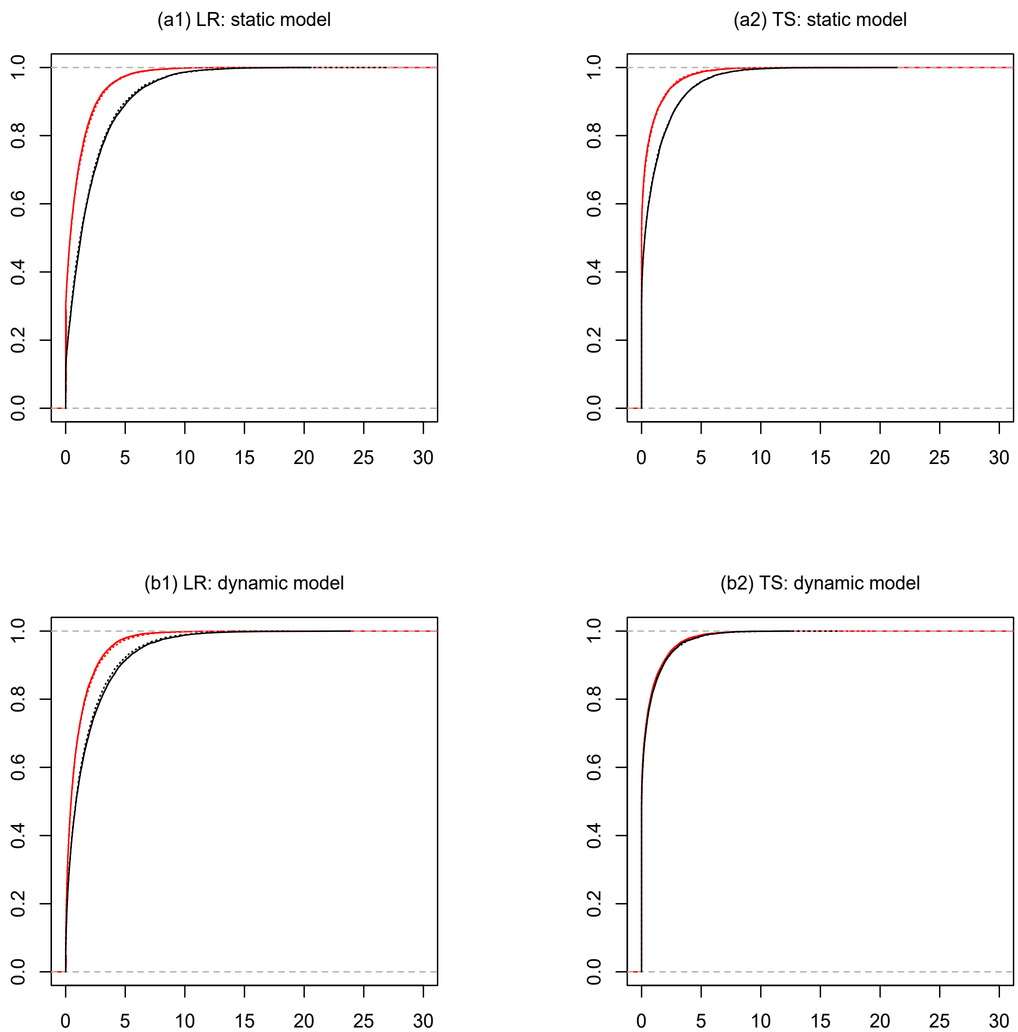
Note. LR: likelihood ratio; TS: Carrasco, Hu, and Ploberger (2014). CDFs of finite sample distributions and asymptotic approximations are reported. Red dotted line: approximation under the null; red solid line: finite sample distribution under the null; black dotted line: approximation under the alternative; black solid line: finite sample distribution under the alternative. (a1) and (a2) are computed using the same data; so are (b1) and (b2).

Figure S20. The refined approximations and their finite sample counterparts: T =20000



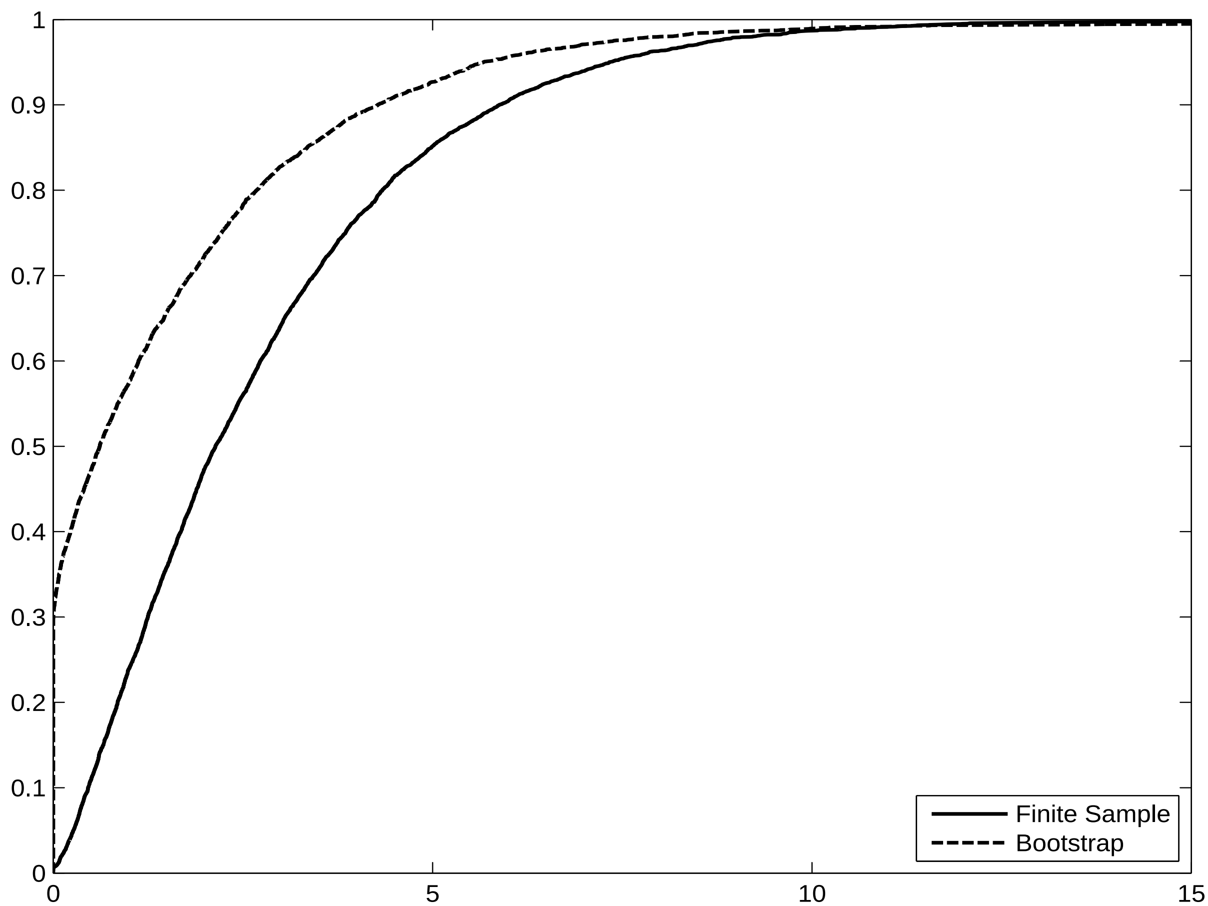
Note. LR: likelihood ratio; TS: Carrasco, Hu, and Ploberger (2014). CDFs of finite sample distributions and asymptotic approximations are reported. Red dotted line: approximation under the null; red solid line: finite sample distribution under the null; black dotted line: approximation under the alternative; black solid line: finite sample distribution under the alternative. (a1) and (a2) are computed using the same data; so are (b1) and (b2).

Figure S21. The refined approximations and their finite sample counterparts: T =50000



Note. LR: likelihood ratio; TS: Carrasco, Hu, and Ploberger (2014). CDFs of finite sample distributions and asymptotic approximations are reported. Red dotted line: approximation under the null; red solid line: finite sample distribution under the null; black dotted line: approximation under the alternative; black solid line: finite sample distribution under the alternative. (a1) and (a2) are computed using the same data; so are (b1) and (b2).

FIGURE S22. A bootstrap procedure applied to an AR(1) model



Note. The model under the null hypothesis is $y_t = \mu + \alpha y_{t-1} + u_t$ with $u_t \sim i.i.d.N(0, \sigma^2)$. The figure shows the finite sample distribution when testing for regime switching in the intercept (the solid line) and the bootstrapped distribution obtained by keeping the regressor fixed (the dashed line). $T = 250$. The true parameter values are $\mu = 0, \alpha = 0.5, \sigma = 1$.

Figure S23: Recession probabilities based on hours worked (1960-2014)

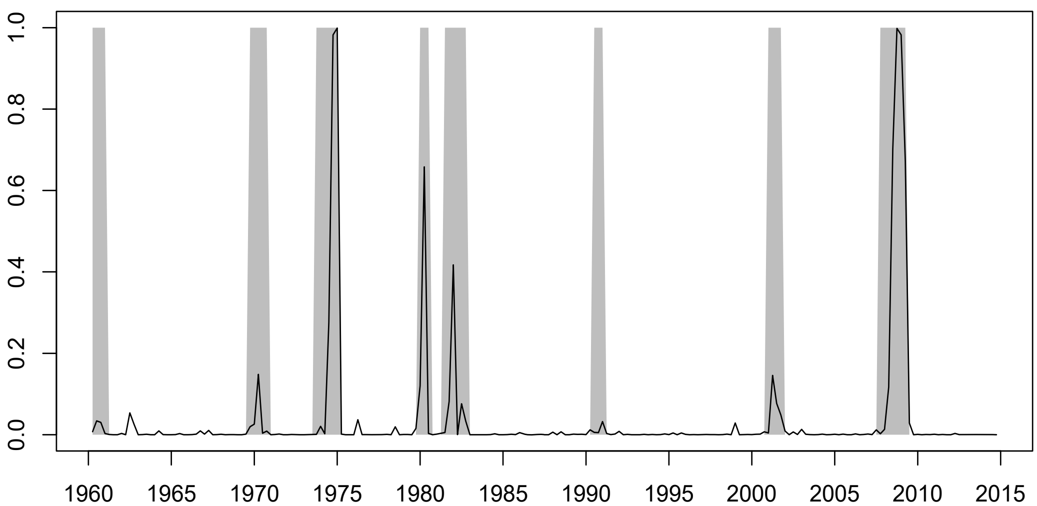


Figure S24: Recession probabilities based on capacity utilization (1967-2014)

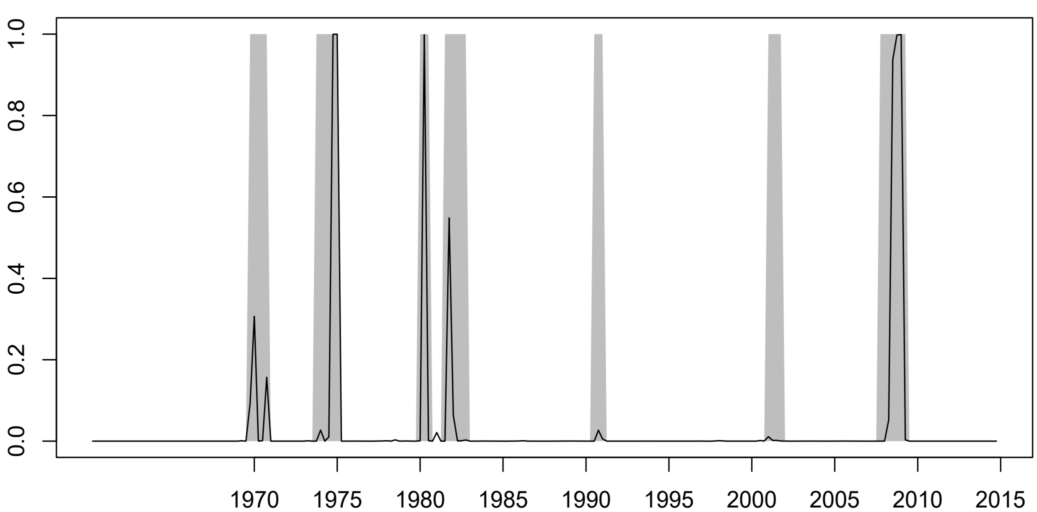


Figure S25: Recession probabilities based on unemployment (1960-2014)

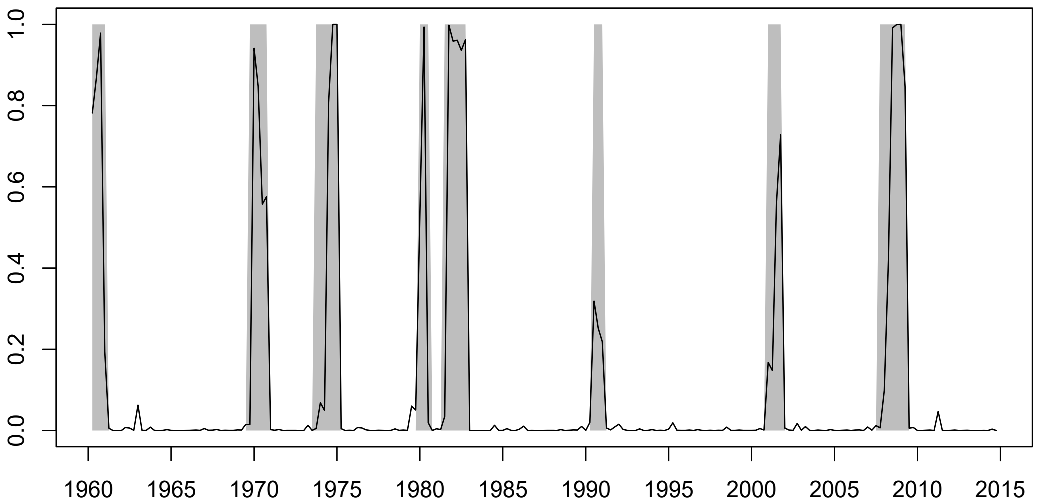


Figure S26: Recession probabilities based on consumption (1960-2014)