Directionally Differentiable Econometric Models

JIN SEO CHO

School of Economics

Yonsei University

50 Yonsei-ro, Seodaemun-gu, Seoul, 120-749, Korea

Email: jinseocho@yonsei.ac.kr

HALBERT WHITE

Department of Economics

University of California, San Diego

9500 Gilman Dr., La Jolla, CA, 92093-0508, U.S.A.

Email: hwhite@weber.ucsd.edu

January 22, 2015

Abstract

The current paper relaxes the differentiability condition for standard econometric models and instead assumes that models are directionally differentiable. The asymptotic distribution of the extremum estimator for directionally differentiable models is represented as a functional of a Gaussian stochastic process indexed by direction, and we treat the conventional analysis for differentiable models as a special case of directionally differentiable model analysis. Furthermore, the standard quasi-likelihood ratio, Wald, and Lagrange multiplier test statistics are redefined in our context so that their asymptotic null behaviors regularly behave.

Key Words: directionally differentiable model, Gaussian stochastic process, quasi-likelihood ratio test, Wald test, and Lagrange multiplier test statistics.

JEL Classification: C12, C13, C22, C32.

Acknowledgements: The second author (Halbert White) passed away while the final version was written. He formed the outline of the paper and also showed an exemplary guide for writing a quality research paper. The authors are benefited from discussions with Yoichi Arai, Seung Chan Ahn, In Choi, Horag Choi, Robert Davies, Graham Elliott, Chirok Han, John Hillas, Jung Hur, Hide Ichimura, Isao Ishida, Yongho Jeon, Estate Khmaladze, Chang-Jin Kim, Chang Sik Kim, Tae-Hwan Kim, Naoto Kunitomo, Taesuk Lee, Bruce Lehmann, Mark Machina, Jaesun Noh, Kosuke Oya, Taeyoung Park, Peter Phillips, Erwann Sbai, Donggyu Sul, Hung-Jen Wang, Yoshihiro Yajima, Byung Sam Yoo and other seminar participants at Sogang University, the University of Auckland, the University of Tokyo, Osaka University, the Econometrics Study Group of the Korean Econometric Society, VUW, Yonsei University, and other conference participants at NZESG (Auckland, 2013).

1 Introduction

Model differentiability is one of the regularity conditions for analyzing standard econometric models. For example, Wald (1943) proposes it as one of the regularity conditions for his classic test statistic. As another example, Chernoff (1954) considers use of the likelihood ratio (LR) test statistic and approximates log-likelihood functions by Taylor's expansion that requires model differentiability.

Many important econometric models are not differentiable (D). For example, the model examined by King and Shively (1993) is non-differentiable. They attempt to resolve the so-called Davies's (1977,1987) identification problem by re-parameterizing the model using the polar coordinates. This model, however, is not D and only directionally differentiable (D-D). As another example, Aigner, Lovell, and Schmidt (1977) specify the stochastic frontier production function model to capture inefficiently produced outputs. Their model inference cannot also be conducted in the conventional way due to its non-differentiability. In addition to these, many models in the literature are not D, and their model analysis should be different from the standard case.

The goal of this paper, therefore, is to extend the analysis tool for D models to the level of D-D models. For this purpose, D-D model analysis is associated with Billingsley's (1999) tight probability measure condition. Each direction around the parameter of interest is regarded as an index indicating a particular value of directional derivatives. The tight probability measure condition here governs the stochastic interrelationship of the directional derivatives in a way to apply the functional central limit theorem (FCLT) and the uniform law of large numbers (ULLN). Using this property, we analyze D-D models as if they are D models with an additional index.

Another goal of the current study is to demonstrate that the D-D model analysis is applicable to numerous examples. For this goal, we revisit important econometric models. King and Shively's (1993) re-parameterized model is reinvestigated. We aim to provide an appropriate tool for analyzing their re-parameterized model under the null. We also apply our D-D model analysis to Aigner, Lovell, and Schmidt's (1977) stochastic frontier production function models and the Box-Cox transformation. The D-D model analysis deepens the meanings of the hypotheses and derives the asymptotic null distribution efficiently and correctly. Finally, we show that the D-D model analysis reduces to the conventional D model analysis if it is applied to standard D models. For this purpose, we reexamine the standard GMM estimation by the tools for D-D model analysis.

The plan of this paper is as follows. In Section 2, D-D models are defined and examined, and D models are investigated as a special case of D-D models. We also provide regularity conditions for D-D models and consider the asymptotic distribution of the extremum estimator under these conditions. Section 3 considers data inferences using D-D models. For this, we redefine the standard quasi-LR (QLR), Wald, and Lagrange multiplier (LM) test statistics that are modified for D-D models and derive their asymptotic null behaviors. Furthermore, a benchmark model is considered under which these three test statistics are asymptotically equivalent under the null. In the same section, we also conduct Monte Carlo experiments. Section 4 offers concluding remarks, and formal mathematical proofs are given in Appendix.

Before moving to the next section, we introduce mathematical notation that is used throughout this paper. For any $x \in \mathbb{R}^r$, $\|x\|$ stands for the Euclidean norm. Furthermore, $\mathbf{1}_{\{\,\cdot\,\}}$ and $\mathrm{cl}(A)$ stand for an indication function and a closure of set A, respectively. The other is standard.

2 Differentiable and Directionally Differentiable Models

To proceed with our discussion in a manageable way, we first introduce the regularity conditions maintained throughout this paper. The following is the data generating process (DGP) condition.

Assumption 1 (DGP)**.** *A sequence of random variables* $\{\mathbf{X}_t \in \mathbb{R}^m\}_{t=1}^n$ $(m \in \mathbb{N})$ *defined on a complete probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is strictly stationary and ergodic.* □

Assumption 1 is standard for stationary time-series data. Many economic data satisfy the given condition. For example, the standard ARMA process, hidden Markov processes, and GARCH processes are typical examples of this DGP. We examine the following model to capture the DGP properties.

Assumption 2 (Model)**.** *A sum of measurable functions* $\{L_n(\boldsymbol{\theta}) := \sum_{t=1}^n \ell_t(\boldsymbol{\theta}; \mathbf{X}^t) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ *is the model for* $\mathbf{X}^n$ *such that for each t,* $\ell_t(\,\cdot\,; \mathbf{X}^t)$ *is Lipschitz continuous on* $\boldsymbol{\Theta}$ *almost surely–*$\mathbb{P}$ *(a.s.–*$\mathbb{P}$*), where for each* $t = 1, 2, \ldots, n$, $\mathbf{X}^t$ *denotes* $(\mathbf{X}_1, \cdots, \mathbf{X}_t)$*, and* $\boldsymbol{\Theta}$ *is a compact and convex set in* $\mathbb{R}^r$ *with* $r \in \mathbb{N}$. □

This model condition is widely used in the literature. For example, the maximum likelihood (ML) and quasi-maximum likelihood (QML) estimations are typical estimators obtained using Assumption 2. We also suppose that the key properties in the DGP are captured by a particular parameter value that is characterized by

Assumption 3 (Existence and Identification)**.** *(i) For each* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $n^{-1}E[L_n(\boldsymbol{\theta})]$ *exists in* $\mathbb{R}$ *and is finite uniformly in* n*;*

(ii) For a unique $\boldsymbol{\theta}_* \in \boldsymbol{\Theta}$, $E[n^{-1}L_n(\cdot)]$ *is maximized at* $\boldsymbol{\theta}_* \in \boldsymbol{\Theta}$ *uniformly in* n. □

Using Assumptions 2 and 3, we define an extreme estimator $\widehat{\boldsymbol{\theta}}_n$ such that

$$L_n(\widehat{\boldsymbol{\theta}}_n) = \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_n(\boldsymbol{\theta}).$$

Several remarks are warranted. First, Assumption 3(i) requires model identification. Even if models are not identified, model analysis using the framework of Davies (1977,1987) can still be made. Nevertheless, accommodating this makes the key aspects of D-D models obscure. We, therefore, highlight the D-D model analysis by focusing on identified models. Second, $\boldsymbol{\theta}_*$ can be on the boundary of $\boldsymbol{\Theta}$ as often ensued by the re-parameterization method of King and Shively (1993). Assumption 3(ii) permits this. Finally, for notational simplicity, $\ell_t(\,\cdot\,; \mathbf{X}^t)$ is abbreviated into $\ell_t(\cdot)$ from now.

Given the assumptions provided so far, the extremum estimator is consistent as formally stated in the following theorem.

Theorem 1. *Given Assumptions 1 to 3, $\widehat{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}_*$ a.s.–$\mathbb{P}$.* □

Theorem 1 is straightforward and well known in the literature (e.g., Andrews, 1999). The desired property is achieved by applying the ULLN to $n^{-1}L_n(\cdot)$, given that $\boldsymbol{\theta}_*$ is unique. We, therefore, do not prove this in the Appendix. One of the implications of Theorem 1 is that the differentiability condition is not necessary for the consistency of the extremum estimator as Wald (1949) points this out under the ML estimation context. On the other hand, the asymptotic distribution of the extremum estimator is dependent upon the differentiability as we discuss in the next subsection.

2.1 Directionally Differentiable Models

The smoothness condition of D-D functions is important for deriving the asymptotic behavior of $\widehat{\boldsymbol{\theta}}_n$. In this subsection, we define D-D models and characterize D models using this.

Definition 1 (D-D Functions)**.** *(i) A function $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is called* directionally differentiable *(*D-D*) at* $\boldsymbol{\theta}$ in the direction of $\boldsymbol{d} \in \Delta(\boldsymbol{\theta})$*, if*

$$Df(\boldsymbol{\theta}; \boldsymbol{d}) := \lim_{h\downarrow 0} \frac{f(\boldsymbol{\theta} + h\boldsymbol{d}) - f(\boldsymbol{\theta})}{h}$$

exists in $\mathbb{R}$, where $\Delta(\boldsymbol{\theta}) := \{\mathbf{x} \in \mathbb{R}^r : \mathbf{x} + \boldsymbol{\theta} \in \mathrm{cl}\{C(\boldsymbol{\theta})\}, \|\mathbf{x}\| = 1\}$, and $C(\boldsymbol{\theta}) := \{\mathbf{x} \in \mathbb{R}^r : \exists \boldsymbol{\theta}' \in \boldsymbol{\Theta}, \mathbf{x} := \boldsymbol{\theta} + \delta\boldsymbol{\theta}', \delta \in \mathbb{R}^+\}$;

(ii) A function $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is said to be D-D on $\Delta(\boldsymbol{\theta})$*, if for all $\boldsymbol{d} \in \Delta(\boldsymbol{\theta})$, $Df(\boldsymbol{\theta}; \boldsymbol{d})$ exists;*

(iii) A function $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is said to be D-D on $\boldsymbol{\Theta}$*, if for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, f is D-D on $\Delta(\boldsymbol{\theta})$.* □

There are several remarks that are relevant to this definition. First, note that the definition of D-D functions is weaker than that of D functions. D-D functions can have different directional derivatives that nonlinearly depend on the levels of direction, and there can be a continuum number of directions if r is greater than one. On the other hand, if $f(\cdot)$ is D, $Df(\boldsymbol{\theta};\boldsymbol{d})$ is represented as a linear combination of r number of different directional derivatives. Second, $Df(\boldsymbol{\theta};\cdot)$ is defined on $\Delta(\boldsymbol{\theta})$. This requirement is adopted to accommodate Chernoff's (1954) device. Chernoff (1954) notes that it is essential to approximate the parameter space by a cone $C(\boldsymbol{\theta})$ and obtains the asymptotic distribution of the extremum estimator. We define $\Delta(\boldsymbol{\theta})$ to collect only directions relevant to $C(\boldsymbol{\theta})$, and it plays the role of domain of a Gaussian stochastic process that is introduced below. Third, even when $\boldsymbol{\theta}$ is on the boundary of $\boldsymbol{\Theta}$, $\Delta(\boldsymbol{\theta})$ can still be defined not to contain the directions of boundary sides. Finally, another norm different from the Euclidean norm can also be used to define $\Delta(\boldsymbol{\theta})$. For example, $\widetilde{\Delta}(\boldsymbol{\theta}) := \{\mathbf{x} \in \mathbb{R}^r : \mathbf{x} + \boldsymbol{\theta} \in \mathrm{cl}\{C(\boldsymbol{\theta})\}, \|\mathbf{x}\|_\infty = 1\}$ can be used, where $\|\cdot\|_\infty$ is the uniform norm, and it captures the same directions as given by $\Delta(\boldsymbol{\theta})$. The current paper proceeds with $\Delta(\boldsymbol{\theta})$.

The following theorem provides sufficient conditions for a D-D function to be D.

Theorem 2 (TROUTMAN 1996, P. 122)**.** *Suppose that (i) a function $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is D-D on $\boldsymbol{\Theta}$; (ii) for each $\boldsymbol{\theta}, \boldsymbol{\theta}'$ and for some $M < \infty$, $|Df(\boldsymbol{\theta}';\boldsymbol{d}) - Df(\boldsymbol{\theta};\boldsymbol{d})| \leq M\|\boldsymbol{\theta}' - \boldsymbol{\theta}\|$ uniformly on $\Delta(\boldsymbol{\theta}) \cap \Delta(\boldsymbol{\theta}')$; and (iii) for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $Df(\boldsymbol{\theta};\boldsymbol{d})$ is continuous and linear in $\boldsymbol{d} \in \Delta(\boldsymbol{\theta})$. Then, $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is D on $\boldsymbol{\Theta}$.* □

The proof of Theorem 2 is available in Troutman (1996). The linearity condition of $Df(\boldsymbol{\theta};\boldsymbol{d})$ in $\boldsymbol{d}$ is the key condition for a D-D function to be D. Without this, any directional derivative cannot be represented as a linear combination of other r different directional derivatives.

Before illustrating D-D model examples, we provide the definition of continuously D-D functions that also plays another key role in our analysis.

Definition 2 (Twice Continuously D-D Functions)**.** *A function $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is called* twice continuously D-D on $\boldsymbol{\Theta}$*, if for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and for all $\boldsymbol{d} \in \Delta(\boldsymbol{\theta})$, $D^2 f(\boldsymbol{\theta};\boldsymbol{d})$ exists, where*

$$D^2 f(\boldsymbol{\theta};\boldsymbol{d}) := \lim_{h\downarrow 0} \frac{Df(\boldsymbol{\theta} + h\boldsymbol{d};\boldsymbol{d}) - Df(\boldsymbol{\theta};\boldsymbol{d})}{h}. \qquad \square$$

Note that the first-order directional differentiability (d-difflity) is necessary for defining twice continuously D-D functions. Furthermore, twice continuously D functions can be obtained from twice continuously D-D functions if a quadratic function condition with respect to $\boldsymbol{d}$ is further imposed. The following Lemma 1 formally states this.

Lemma 1. *If a function $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ that satisfies the conditions in Theorem 2 is further (i) twice continuously D-D on $\boldsymbol{\Theta}$; (ii) for each $\boldsymbol{\theta}, \boldsymbol{\theta}'$ and for some $M < \infty$, $|D^2 f(\boldsymbol{\theta}'; \boldsymbol{d}) - D^2 f(\boldsymbol{\theta}; \boldsymbol{d})| \leq M\|\boldsymbol{\theta}' - \boldsymbol{\theta}\|$ uniformly on $\Delta(\boldsymbol{\theta}) \cap \Delta(\boldsymbol{\theta}')$; and (iii) for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $D^2 f(\boldsymbol{\theta}; \boldsymbol{d})$ is continuous and quadratic in $\boldsymbol{d} \in \Delta(\boldsymbol{\theta})$, $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is twice continuously D on $\boldsymbol{\Theta}$.* □

2.2 Examples

We illustrate four D-D models: the conditional heteroskedasticity model in King and Shively (1993), the stochastic frontier production function model, the Box-Cox transformation, and the standard D model for GMM estimation. The first, second, and third models are only D-D, and the final one is D. As D models are also D-D, the final example can also be analyzed by the D-D model analysis. Some of these models are associated with other non-standard features such as boundary parameter and/or identification problems.

2.2.1 Example 1: Conditional Heteroskedasticity

King and Shively (1993) examine a model for conditional heteroskedasticity. When a set of economic data $\{(Y_t, \mathbf{Q}_t')' := (Y_t, W_t, \mathbf{R}_t')' \in \mathbb{R}^{2+k}\}$ is given, they assume

$$\mathbf{Y}^n = \mathbf{W}^n \alpha_* + \mathbf{R}^n \boldsymbol{\beta}_* + \mathbf{U}^n,$$

$$\mathbf{U}^n | \mathbf{Q}^n \sim N[\mathbf{0}, \sigma_*^2\{\mathbf{I}_n + \kappa_* \boldsymbol{\Omega}^n(\rho_*)\}],$$

where $\mathbf{Y}^n := (Y_1, \cdots, Y_n)'$, $\mathbf{U}^n := (U_1, \cdots, U_n)'$, $\mathbf{W}^n := (W_1, \cdots, W_n)'$, $\mathbf{R}^n$ is an $n \times k$ matrix with $\mathbf{R}_t{}'$ at t-th row, $\mathbf{Q}^n := (\mathbf{W}^n, \mathbf{R}^n)$, and $\boldsymbol{\Omega}^n(\rho_*)$ is an $n \times n$ square matrix with t-th row and t'-th column element

$$\boldsymbol{\Omega}_{tt'}^n(\rho_*) := W_t W_{t'} {\rho_*}^{|t'-t|} / (1 - {\rho_*}^2).$$

Given this, they let $(\boldsymbol{\gamma}_*', \sigma_*^2, \kappa_*, \rho_*) := (\alpha_*, \boldsymbol{\beta}_*', \sigma_*^2, \kappa_*, \rho_*)$ be an unknown parameter in $\boldsymbol{\Gamma} \times [0, \bar{\sigma}^2] \times [0, \bar{\kappa}] \times [0, \bar{\rho}]$, where $\boldsymbol{\Gamma}$ is a compact and convex subset of $\mathbb{R}^{k+1}$, $\bar{\sigma}^2$ and $\bar{\kappa}$ are positive real numbers, and $\bar{\rho}$ is also a positive real number but less than one. For each $(\boldsymbol{\gamma}, \sigma^2, \kappa, \rho)$, its log-likelihood is written as

$$L_n(\boldsymbol{\gamma}, \sigma^2, \kappa, \rho) = -\frac{1}{2} \log\left((2\pi)^n \det\left[\sigma^2\{\mathbf{I}_n + \kappa \boldsymbol{\Omega}^n(\rho)\}\right]\right) - \frac{1}{2\sigma^2} \mathbf{U}^n(\boldsymbol{\gamma})' \left[\mathbf{I}_n + \kappa \boldsymbol{\Omega}^n(\rho)\right]^{-1} \mathbf{U}^n(\boldsymbol{\gamma}),$$

where $\mathbf{U}^n(\gamma) := \mathbf{Y}^n - \mathbf{Q}^n \boldsymbol{\gamma}$, and $\boldsymbol{\gamma} := (\alpha, \boldsymbol{\beta}')'$.

This model is motivated by Rosenberg (1973), who aims to test $\kappa_* = 0$ to examine whether a systematic risk of an asset is time-varying or not. If $\kappa_* \neq 0$, the conditional covariance of $\mathbf{U}^n|\mathbf{Q}^n$ depends on $\mathbf{W}^n$, so that conditional homoskedasticity can be inferred.

Nevertheless, if $\kappa_* = 0$, ρ_* is not identified, and Davies's (1977,1987) identification problem arises. This renders the asymptotic null distributions of standard tests non-standard as they are often represented as functionals of Gaussian stochastic processes. King and Shively (1993) attempt to resolve the unidentified parameter problem by re-parameterizing the original model:

$$\boldsymbol{\theta}_*' := (\theta_{1*}, \theta_{2*}) := (\kappa_* \cos(\rho_* \pi/2), \kappa_* \sin(\rho_* \pi/2))$$

using the polar coordinates, so that the parameter space of $\boldsymbol{\theta}$ is now obtained as $[0, \bar{\kappa}\cos(\bar{\rho}\pi/2)] \times [0, \bar{\kappa}\sin(\bar{\rho}\ \pi/2)]$, and

$$\mathbf{U}^n|\mathbf{Q}^n \sim\ N[0,\ {\sigma_*}^2\{\mathbf{I}_n + ({\boldsymbol{\theta}_*}'\boldsymbol{\theta}_*)^{1/2}\boldsymbol{\Omega}^n(2\tan^{-1}(\theta_{2*}/\theta_{1*})/\pi)\}].$$

Furthermore, the original hypotheses are rephrased into $H_0' : {\boldsymbol{\theta}_*}'\boldsymbol{\theta}_* = 0$ versus $H_1' : {\boldsymbol{\theta}_*}'\boldsymbol{\theta}_* > 0$ by the re-parameterization, and the identification problem does no longer arises under H_0'.

On the other hand, the model obtained in this way is only D-D under H_0', and the null parameter value is on the boundary: for each $(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\theta})$, the log-likelihood is modified into

$$\begin{aligned} L_n(\boldsymbol{\gamma}, \sigma^2, \boldsymbol{\theta}) = &-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left(\det\left[\sigma^2\{\mathbf{I}_n + (\boldsymbol{\theta}'\boldsymbol{\theta})^{1/2}\boldsymbol{\Omega}^n(2\tan^{-1}(\theta_2/\theta_1)/\pi)\}\right]\right) \\ &- \frac{1}{2\sigma^2}\mathbf{U}^n(\boldsymbol{\gamma})'\left[\mathbf{I}_n + (\boldsymbol{\theta}'\boldsymbol{\theta})^{1/2}\boldsymbol{\Omega}^n(2\tan^{-1}(\theta_2/\theta_1)/\pi)\right]^{-1}\mathbf{U}^n(\boldsymbol{\gamma}), \end{aligned}$$

and from this, θ_{2*}/θ_{1*} has the form of $0/0$ under the null. In addition, for each $\boldsymbol{d} := (\boldsymbol{d}_\gamma', d_\sigma, d_1, d_2)'$ such that $\boldsymbol{\theta}_* = \mathbf{0}$ and $\boldsymbol{d}'\boldsymbol{d} = 1$,

$$\lim_{h \downarrow 0} L_n(\boldsymbol{\gamma}_* + \boldsymbol{d}_\gamma h, \sigma_*^2 + d_{\sigma^2}h, \boldsymbol{\theta}_* + \boldsymbol{d}_{\boldsymbol{\theta}}h) = -\frac{n}{2}\log\left(2\pi\det(\sigma_*^2)\right) - \frac{1}{2\sigma_*^2}\mathbf{U}^n(\boldsymbol{\gamma}_*)'\mathbf{U}^n(\boldsymbol{\gamma}_*)$$

that is the null log-likelihood desired by Rosenberg (1973) and King and Shively (1993). We further note that

$$\begin{aligned} DL_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*; \boldsymbol{d}) = &-\frac{nd_{\sigma^2}}{2\sigma_*^2} - \frac{(d_1^2 + d_2^2)^{1/2}}{2}\mathrm{tr}[\boldsymbol{\Omega}^n\left(2\tan^{-1}(d_2/d_1)/\pi\right)] + \frac{d_{\sigma^2}}{2\sigma_*^4}\mathbf{U}^{n\prime}\mathbf{U}^n \\ &+ \frac{1}{\sigma_*^2}(\mathbf{Q}^n\boldsymbol{d}_\gamma)'\mathbf{U}^n + \frac{(d_1^2 + d_2^2)^{1/2}}{2\sigma_*^2}\mathbf{U}^{n\prime}\boldsymbol{\Omega}^n\left(2\tan^{-1}(d_2/d_1)/\pi\right)\mathbf{U}^n \end{aligned} \tag{1}$$

that is not linear with respect to (d_1, d_2), implying that the model is not D as Theorem 2 implies. The second-order directional derivative is also obtained as

$$\begin{aligned}D^2 L_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*; \boldsymbol{d}) =& \frac{n d_{\sigma^2}^2}{2\sigma_*^4} - \frac{d_{\sigma^2}^2}{\sigma_*^6}\mathbf{U}^{n\prime}\mathbf{U}^n - \frac{1}{\sigma_*^2}(\mathbf{Q}^n \boldsymbol{d}_\gamma)'(\mathbf{Q}^n \boldsymbol{d}_\gamma) - \frac{2d_{\sigma^2}}{\sigma_*^4}(\mathbf{Q}^n \boldsymbol{d}_\gamma)'\mathbf{U}^n \\ &+ \frac{(d_1^2+d_2^2)}{2}\mathrm{tr}\left[\boldsymbol{\Omega}^n(2\tan^{-1}(d_2/d_1)/\pi)^2\right] \\ &- \frac{2(d_1^2+d_2^2)^{1/2}}{\sigma_*^2}(\mathbf{Q}^n \boldsymbol{d}_\gamma)'\left[\boldsymbol{\Omega}^n(2\tan^{-1}(d_2/d_1)/\pi)\right]\mathbf{U}^n \\ &- \frac{d_{\sigma^2}}{\sigma_*^4}(d_1^2+d_2^2)^{1/2}\mathbf{U}^{n\prime}\left[\boldsymbol{\Omega}^n(2\tan^{-1}(d_2/d_1)/\pi)\right]\mathbf{U}^n \\ &- \frac{(d_1^2+d_2^2)}{\sigma_*^2}\mathbf{U}^{n\prime}\left[\boldsymbol{\Omega}^n(2\tan^{-1}(d_2/d_1)/\pi)\right]^2\mathbf{U}^n. \qquad (2)\end{aligned}$$

Note that this is not quadratic with respect to $\boldsymbol{d}$, so that the model is not twice continuously D.

We elaborate on the model assumption for an appropriate model analysis. If d_1 is zero, d_2/d_1 is not defined. We avoid this by letting d_2/d_1 have an upper bound. This restriction is equivalent to letting the parameter space of ρ in the original model have an upper bound strictly less than unity. In addition to this, we do not allow that $d_2 = 0$. If this is allowed, the diagonal elements of $\boldsymbol{\Omega}^n(0)$ contain 0^0, so that the model is not again appropriately defined. Furthermore, Rosenberg's (1973) original purpose to test for conditional heteroskedasticity does not allow the null model to have a time-varying variance. We, therefore, let d_2 be strictly positive. Imposing this lower bound condition is equivalent to letting ρ be away from zero in terms of the original model. Consequently, our parameter space for $\boldsymbol{\theta}$ is refined into

$$\boldsymbol{\Theta} := \{\boldsymbol{\theta} \in [0, \bar{\kappa}\cos(\bar{\pi}/2)] \times [0, \bar{\kappa}\sin(\bar{\pi}/2)] : \underline{c} \times \theta_1 \leq \theta_2 \leq \bar{c} \times \theta_1 \ \exists\ \underline{c} \text{ and } \bar{c} > 0\}.$$

By this modification, d_2/d_1 is constrained to $[\underline{c}, \bar{c}]$. □

2.2.2 Example 2: Stochastic Frontier Production Function Models

Another D-D model is found from the theory of stochastic frontier production function models. Stochastic production function models are often specified for identically and independently distributed (IID) data $\{Y_t, \mathbf{X}_t\}$ as $Y_t = \mathbf{X}_t'\boldsymbol{\beta}_* + U_t$, where $Y_t \in \mathbb{R}$ is the output produced using inputs $\mathbf{X}_t \in \mathbb{R}^k$ such that $\boldsymbol{\beta}_*$ is an interior element of $\mathbf{B} \subset \mathbb{R}^k$, $E[U_t^2] < \infty$, $E[X_{t,j}^2] < \infty$ for $j = 1, 2, \ldots, k$, and $E[\mathbf{X}_t\mathbf{X}_t']$ is positive-definite. Here, U_t stands for an error that is independent of $\mathbf{X}_t$. This model is first introduced by Aigner, Lovell, and Schmidt (1977).

One of the early uses of this specification is in identifying inefficiently produced outputs. Given output levels subject to the production function and inputs, $E[U_t] < 0$ means that outputs are inefficiently produced.

Aigner, Lovell, and Schmidt (1977) capture this inefficiency by decomposing U_t into $U_t \equiv V_t - W_t$, where $V_t \sim N(0, \tau_*^2)$, $W_t := \max[0, Q_t]$, $Q_t \sim N(\mu_*, \sigma_*^2)$, and V_t is independent of W_t. Here, it is assumed that $\tau_* > 0$, $\sigma_* \geq 0$, and $\mu_* \geq 0$, and W_t is employed to capture inefficiently produced outputs. If $\mu_* = 0$ and $\sigma_*^2 = 0$, this model reduces to Zellner, Kmenta, and Drèze's (1966) stochastic production function model, implying that outputs are efficiently produced. The key to identifying the inefficiency is, therefore, in testing $\mu_* = 0$ and $\sigma_*^2 = 0$.

The original model introduced by Aigner, Lovell, and Schmidt (1977) assumes $\mu_* = 0$, so that the mode of W_t is always achieved at zero. Stevenson (1980) suggests to extend the model scope by estimating μ_*. Since then, the model with unknown μ_* has been popularly specified for empirical data analysis (e.g., Dutta, Narasimhan, and Rajiv (1999), Habib and Ljungqvist (2005), and etc.).

Nevertheless, testing methodology for $\mu_* = 0$ and $\sigma_*^2 = 0$ is missing in the literature. This is mainly because the likelihood value is not defined under the null. Note that for each $(\boldsymbol{\beta}, \sigma, \mu, \tau)$, the log-likelihood is given as

$$L_n(\boldsymbol{\beta}, \sigma, \mu, \tau) = \sum_{t=1}^{n} \left\{ \ln \left[\phi \left(\frac{Y_t - \mathbf{X}_t'\boldsymbol{\beta} + \mu}{\sqrt{\sigma^2 + \tau^2}} \right) \right] - \frac{1}{2} \ln(\sigma^2 + \tau^2) - \ln \left[\Phi \left(\frac{\mu}{\sqrt{\sigma^2}} \right) / \Phi \left(\frac{\widetilde{\mu}_t}{\sqrt{\widetilde{\sigma}^2}} \right) \right] \right\},$$

where $\phi(\,\cdot\,)$ and $\Phi(\,\cdot\,)$ are the probability density function (PDF) and cumulative density function (CDF) of a standard normal random variable, respectively, and

$$\widetilde{\mu}_t := \frac{\tau^2 \mu - \sigma^2 (Y_t - \mathbf{X}_t'\boldsymbol{\beta})}{\tau^2 + \sigma^2} \quad \text{and} \quad \widetilde{\sigma}^2 := \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2}.$$

Here, the log-likelihood is not defined if $\boldsymbol{\theta}_* := (\boldsymbol{\beta}_*', \mu_*, \sigma_*, \tau_*)' = (\boldsymbol{\beta}_*', 0, 0, \tau_*)$ because $\mu_*/\sqrt{\sigma_*^2} = 0/0$, so that $\ln[\Phi(\mu_*/\sqrt{\sigma_*^2})]$ is not properly defined. Furthermore, if we let

$$\widetilde{\mu}_{*t} := \frac{\tau_*^2 \mu_* - \sigma_*^2 U_t}{\tau_*^2 + \sigma_*^2} \quad \text{and} \quad \widetilde{\sigma}_*^2 := \frac{\tau_*^2 \sigma_*^2}{\tau_*^2 + \sigma_*^2},$$

$\widetilde{\mu}_{*t}/\sqrt{\widetilde{\sigma}_*^2} = 0/0$, so that $\ln[\Phi(\widetilde{\mu}_{*t}/\sqrt{\widetilde{\sigma}_*^2})]$ is not also defined.

Even further, this model is not D. This can be seen by examining the first-order directional derivatives. Some tedious algebra shows that for a given $\boldsymbol{d} := (\boldsymbol{d}_{\boldsymbol{\beta}}', d_\mu, d_\sigma, d_\tau)'$,

$$\lim_{h \downarrow 0} L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = -\frac{n}{2} \ln(\tau_*^2) + \sum_{t=1}^{n} \ln \left[\phi \left(\frac{Y_t - \mathbf{X}_t'\boldsymbol{\beta}_*}{\sqrt{\tau_*^2}} \right) \right],$$

which is the log-likelihood desired by the null condition. This limit is obtained by particularly using the fact that

$$\lim_{h\downarrow 0}\Phi\left(\frac{hd_\mu}{\sqrt{(hd_\sigma)^2}}\right)=\Phi\left(\frac{d_\mu}{\sqrt{d_\sigma^2}}\right)\quad\text{and}\quad \lim_{h\downarrow 0}\Phi\left(\frac{\widetilde{\mu}_{*t}(h;\boldsymbol{d})}{\sqrt{\widetilde{\sigma}(h;\boldsymbol{d})^2}}\right)=\Phi\left(\frac{d_\mu}{\sqrt{d_\sigma^2}}\right),$$

where

$$\widetilde{\sigma}_*(h;\boldsymbol{d})^2:=\frac{(\tau_*+hd_\tau)^2(hd_\sigma)^2}{(\tau_*+dd_\tau)^2+(hd_\sigma)^2}\quad\text{and}$$

$$\widetilde{\mu}_{*t}(h;\boldsymbol{d}):=\frac{(\tau_*+hd_\tau)^2hd_\mu-(hd_\sigma)^2(Y_t-\mathbf{X}_t'(\boldsymbol{\beta}_*+h\boldsymbol{d}_{\boldsymbol{\beta}}))}{(\tau_*+hd_\tau)^2+(hd_\sigma)^2}.$$

Using this directional limit, the first- and second-order directional derivatives of $L_n(\,\cdot\,)$ at $(\boldsymbol{\beta}_*,0,0,\tau_*)$ are

$$DL_n(\boldsymbol{\theta}_*;\boldsymbol{d})=\sum_{t=1}^{n}\frac{1}{\tau_*^3}\left\{d_\tau(U_t^2-\tau_*^2)+\left[-d_\mu+\mathbf{X}_t'\boldsymbol{d}_{\boldsymbol{\beta}}-\psi\left(d_\mu,d_\sigma\right)\right]\tau_*U_t\right\},$$

and

$$\begin{aligned}D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{d})&=\sum_{t=1}^{n}\frac{1}{\tau_*^4}\left\{d_\sigma^2(U_t^2-\tau_*^2)+d_\tau^2\tau_*^2\right\}\\&-\sum_{t=1}^{n}\frac{1}{\tau_*^4}\left\{[d_\tau U_t-(d_\mu-\mathbf{X}_t'\boldsymbol{d}_{\boldsymbol{\beta}})\tau_*][3d_\tau U_t-(d_\mu-\mathbf{X}_t'\boldsymbol{d}_{\boldsymbol{\beta}})\tau_*]\right\}\\&-\sum_{t=1}^{n}\frac{1}{\tau_*^4}\left\{\psi(d_\mu,d_\sigma)^2U_t^2+\psi(d_\mu,d_\sigma)[d_\mu U_t^2-4d_\tau\tau_*U_t+(d_\mu-2\mathbf{X}_t'\boldsymbol{d}_{\boldsymbol{\beta}})\tau_*^2]\right\},\end{aligned}$$

respectively, where $\psi(d_\mu,d_\sigma):=|d_\sigma|\phi(d_\mu/|d_\sigma|)/\Phi(d_\mu/|d_\sigma|)$. Here, if $\boldsymbol{\theta}_*=(\boldsymbol{\beta}_*',0,0,\tau_*)$, $U_t\sim N(0,\tau_*^2)$. These directional derivatives are neither linear nor quadratic with respect to $\boldsymbol{d}$, respectively, so that $L_n(\cdot)$ is not twice continuously D. Therefore, this model cannot be analyzed as for standard D models. We examine this model by letting the direction space be

$$\Delta(\boldsymbol{\theta}_*):=\left\{\boldsymbol{d}\in\mathbb{R}^{d+3}:\boldsymbol{d}'\boldsymbol{d}=1,d_\mu\geq 0,\text{ and }d_\sigma\geq 0\right\}$$

to accommodate the condition that $\mu_*\geq 0$ and $\sigma_*\geq 0$. □

2.2.3 Example 3: Box-Cox's (1964) Transformation

Applying the directional derivatives makes model analysis more sensible for nonlinear models with irregular properties. Box and Cox's (1964) transformation belongs to this case. We consider the following model:

$$Y_t=\mathbf{Z}_t{}'\boldsymbol{\theta}_0+\frac{\theta_1}{\theta_2}(X_t^{\theta_2}-1)+U_t,\tag{3}$$

where $\{(Y_t, X_t, \mathbf{Z}_t{}') \in \mathbb{R}^{2+k} : t = 1, 2, \cdots\}$ is assumed to be IID, X_t is strictly greater than zero almost surely, and $U_t := Y_t - E[Y_t|\mathbf{Z}_t, X_t]$. Furthermore, $\boldsymbol{\theta} := (\boldsymbol{\theta}_0', \theta_1, \theta_2)' \in \mathbf{\Theta}_0 \times \mathbf{\Theta}_{12}$, $\mathbf{\Theta}_0$ is a convex and compact set in $\mathbb{R}^k$, and

$$\mathbf{\Theta}_{12} := \{(y, z) \in \mathbb{R}^2 : \underline{c}y \leq z \leq \bar{c}y < \infty,\ 0 < \underline{c} < \bar{c} < \infty, \text{ and } z^2 + y^2 \leq \bar{m} < \infty\}.$$

Our interests are in testing whether X_t influences $E[Y_t|\mathbf{Z}_t, X_t]$ or not by testing that the second term of (3) vanishes.

This model is introduced to avoid Davies's (1977,1987) identification problem. If the Box-Cox transformation is specified in the conventional way as in Hansen (1996b), so that

$$Y_t = \mathbf{Z}_t{}'\boldsymbol{\theta}_0 + \beta_1(X_t^{\gamma} - 1) + U_t$$

is assumed, then γ_* is not identified when $\beta_{1*} = 0$, where the subscript '$*$' indicates the limit of the nonlinear least squares (NLS) estimator. We may instead examine another null hypothesis: $\gamma_* = 0$. Note that letting $\gamma_* = 0$ also renders β_{1*} be unidentified. We, therefore, avoid this problem by re-parameterizing the model using θ_1 and θ_2. If $\theta_{2*} = 0$, θ_{1*} has to be zero by the model condition on $\mathbf{\Theta}_{12}$. The identification problem does not arise any longer.

Nevertheless, the re-parameterized model becomes obscure by the null condition: $\theta_{1*} = 0$ and $\theta_{2*} = 0$. If so, the quasi-likelihood is not defined. Note that $\theta_{1*}(X_t^{\theta_{2*}} - 1)/\theta_{2*} = 0 \times 0/0$, implying that the standard tests cannot be applied to this model as for the first and second examples.

On the other hand, the directional limits are well defined, and they can be used to analyze the asymptotic behavior of the quasi-likelihood. For this purpose, we let $\boldsymbol{d} = (\boldsymbol{d}_0', d_1, d_2)'$ and $\boldsymbol{\theta}_* = (\boldsymbol{\theta}_{0*}{}', 0, 0)'$ with $\boldsymbol{\theta}_{0*}$ interior to $\mathbf{\Theta}_0$. The following quasi-likelihood function is obtained from this:

$$L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = -\frac{1}{2}\sum_{t=1}^{n}\left\{Y_t - \mathbf{Z}_t{}'(\boldsymbol{\theta}_{0*} + \boldsymbol{d}_0 h) - \frac{d_1}{d_2}(X_t^{d_2 h} - 1)\right\}^2,$$

which is now D with respect to h at 0. Therefore, for each $\boldsymbol{d}$, $\lim_{h\downarrow 0} L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = -\frac{1}{2}\sum_{t=1}^{n}\{Y_t - \mathbf{Z}_t{}'\boldsymbol{\theta}_{0*}\}^2$. The directional derivatives are also derived as

$$DL_n(\boldsymbol{\theta}_*; \boldsymbol{d}) = \sum_{t=1}^{n} U_t\{\mathbf{Z}_t{}'\boldsymbol{d}_0 + \log(X_t)d_1\}, \quad \text{and} \tag{4}$$

$$D^2 L_n(\boldsymbol{\theta}_*; \boldsymbol{d}) = -\sum_{t=1}^{n} \{\mathbf{Z}_t{}'\boldsymbol{d}_0 + \log(X_t) d_1\}^2 + \sum_{t=1}^{n} U_t \{\log(X_t)\}^2 d_1 d_2, \tag{5}$$

that are linear and quadratic in $(\boldsymbol{d}_0, d_2, d_2)$, respectively. Therefore, the model may be analyzed as if it is D, although the null model is not defined under the null. □

2.2.4 Example 4: Generalized Method of Moments (GMM)

Hansen (1982) examines an estimation method and relevant inferences by generalizing the method of moments that requires model differentiability as one of the regularity conditions. We consider the GMM estimator $\widehat{\boldsymbol{\theta}}_n$ obtained by maximizing $Q_n(\boldsymbol{\theta}) := \mathbf{g}_n(\mathbf{X}^n; \boldsymbol{\theta})' \{-\mathbf{M}_n\}^{-1} \mathbf{g}_n(\mathbf{X}^n; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, where $\{\mathbf{X}_t : t = 1, 2, \cdots\}$ is a sequence of strictly stationary and ergodic random variables, $\mathbf{g}_n(\mathbf{X}^n; \boldsymbol{\theta}) := n^{-1} \sum_{t=1}^{n} \mathbf{q}(\mathbf{X}_t; \boldsymbol{\theta})$ with $\mathbf{q}_t := \mathbf{q}(\mathbf{X}_t; \cdot) : \boldsymbol{\Theta} \mapsto \mathbb{R}^k$ being continuously D a.s.–$\mathbb{P}$ on $\boldsymbol{\Theta}$ given in Assumption 2 ($r \leq k$) such that for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\mathbf{q}(\,\cdot\,; \boldsymbol{\theta})$ is measurable, and $\mathbf{M}_n$ is a symmetric and positive definite random matrix a.s.–$\mathbb{P}$ uniformly in n that converges to a symmetric and positive definite $\mathbf{M}_*$ a.s.–$\mathbb{P}$. Furthermore, for some integrable $m(\mathbf{X}_t)$, $\|\mathbf{q}_t(\,\cdot\,)\|_\infty \leq m(\mathbf{X}_t)$ and $\|\nabla_{\boldsymbol{\theta}} \mathbf{q}_t(\,\cdot\,)\|_\infty \leq m(\mathbf{X}_t)$, and there is a unique $\boldsymbol{\theta}_*$ that maximizes $E[\mathbf{q}_t(\boldsymbol{\theta})]'\{-\mathbf{M}_*\}^{-1} E[\mathbf{q}_t(\boldsymbol{\theta})]$ on the interior part of $\boldsymbol{\Theta}$. We denote the uniform matrix norm by $\|\cdot\|_\infty$. We further suppose that $n^{1/2}\mathbf{g}_n(\mathbf{X}^n; \boldsymbol{\theta}_*) \Rightarrow \mathbf{W} \sim N(\mathbf{0}, \mathbf{S}_*)$ for some positive definite matrix $\mathbf{S}_*$. The GMM estimator is widely applied for empirical data.

The given conditions for $Q_n(\cdot)$ do not exactly satisfy the model conditions in Assumption 2. Even so, our D-D model analysis can be easily adapted to the GMM framework. Directional derivatives play a key role as before. We note that the first-order directional derivative of $\mathbf{g}_n(\,\cdot\,) := \mathbf{g}_n(\mathbf{X}^n; \,\cdot\,)$ is

$$D\mathbf{g}_n(\boldsymbol{\theta}; \boldsymbol{d}) = \nabla_{\boldsymbol{\theta}} \mathbf{g}_n(\mathbf{X}^n; \boldsymbol{\theta})' \boldsymbol{d}, \tag{6}$$

where $\nabla_{\boldsymbol{\theta}} \mathbf{g}_n(\mathbf{X}^n; \boldsymbol{\theta}) := [\nabla_{\theta_1} g_{1,n}(\mathbf{X}^n; \boldsymbol{\theta}), \cdots, \nabla_{\theta_r} g_{k,n}(\mathbf{X}^n; \boldsymbol{\theta})]'$ and $g_{j,n}(\mathbf{X}^n; \boldsymbol{\theta})$ is the j-th element of $\mathbf{g}_n(\mathbf{X}^n; \boldsymbol{\theta})$. As (6) makes it clear, $D\mathbf{g}_n(\boldsymbol{\theta}; \boldsymbol{d})$ is now linear with respect to $\boldsymbol{d}$. Applying the mean-value theorem implies that for each $\boldsymbol{d}$,

$$\mathbf{g}_n(\boldsymbol{\theta}; \boldsymbol{d}) = \mathbf{g}_n(\boldsymbol{\theta}_*; \boldsymbol{d}) + D\mathbf{g}_n(\bar{\boldsymbol{\theta}}; \boldsymbol{d})(\boldsymbol{\theta} - \boldsymbol{\theta}_*). \tag{7}$$

Here, $\bar{\boldsymbol{\theta}} := [\bar{\boldsymbol{\theta}}_1, \cdots, \bar{\boldsymbol{\theta}}_r]$ is the collection of the parameter values between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_*$, and $D\mathbf{g}_n(\bar{\boldsymbol{\theta}}; \boldsymbol{d})$ denotes $[\nabla_{\theta_1} g_{1,n}(\mathbf{X}^n; \bar{\boldsymbol{\theta}}_1), \cdots, \nabla_{\theta_r} g_{k,n}(\mathbf{X}^n; \bar{\boldsymbol{\theta}}_r)]'\boldsymbol{d}$. Furthermore, $DQ_n(\boldsymbol{\theta}; \boldsymbol{d}) = -2\boldsymbol{d}' \nabla_{\boldsymbol{\theta}} \mathbf{g}_n(\boldsymbol{\theta})' \mathbf{M}_n^{-1} \mathbf{g}_n(\boldsymbol{\theta})$. This implies that for each $\boldsymbol{d}$, $n^{1/2} DQ_n(\boldsymbol{\theta}_*; \boldsymbol{d}) \Rightarrow -2\boldsymbol{d}' \mathbf{C}_*' \mathbf{M}_*^{-1} \mathbf{W}$ by the central limit theorem (CLT) and the law of large numbers (LLN): $\nabla_{\boldsymbol{\theta}} \mathbf{g}_n(\boldsymbol{\theta}_*)$ converges to $\mathbf{C}_* := E[\nabla_{\boldsymbol{\theta}} \mathbf{q}_t(\boldsymbol{\theta}_*)]$ a.s.$-\mathbb{P}$ by the fact that

$\|\nabla_{\boldsymbol{\theta}} \mathbf{q}_t(\,\cdot\,)\|_\infty \le m(\mathbf{X}_t)$. We below use these and the tools for D-D model analysis to obtain the asymptotic behavior of the GMM estimator. $\square$

As these four examples illustrate, models can be undefined under the null (Examples 1, 2, and 3) and are non-differentiable although they are D-D (Examples 1 and 2). Despite these irregular model properties, they can be nicely analyzed by D-D model analysis tools.

Indeed, many nonlinear models share the similar features. For example, table 1 of Cheng, Evans, and Iles (1992) collects numerous nonlinear models with parameter instability problems. Many of them can be analyzed using the approach of the current study. Furthermore, D-D model analysis tools simplify dimensional complexities that arise when higher-order approximations are needed for model analysis. Cho, Ishida, and White (2011,2013) and White and Cho (2012) revisit testing neglected nonlinearity using artificial neural networks, and it requires higher-order model approximations. They resolve the relevant issues by applying the D-D analysis to their models.

2.3 Asymptotic Distribution of the Extremum Estimator

As pointed out in the previous subsection, the most significant difference between D-D and D models is in the linearity condition of directional derivatives in $\boldsymbol{d}$. Further regularity conditions are provided for D-D models.

Assumption 4 (D-Difflity)**.** *A model $\ell_t : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is twice continuously D-D on $\boldsymbol{\Theta}$ a.s.$-\mathbb{P}$, and for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{d} \in \Delta(\boldsymbol{\theta})$, $D^2\ell_t(\,\cdot\,;\boldsymbol{d})$ is continuous on $\boldsymbol{\Theta}$ a.s.$-\mathbb{P}$.* $\square$

All of the above examples trivially satisfy Assumption 4. We use Assumption 4 to approximate D-D models by a second-order directional Taylor expansion for each direction. Although second-order directional derivatives do not have to be quadratic with respect to $\boldsymbol{d}$, the following regularity conditions are imposed:

Assumption 5 (Regular D-Difflity)**.** *(i) For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $D\ell_t(\boldsymbol{\theta};\boldsymbol{d})$ and $D^2\ell_t(\boldsymbol{\theta};\boldsymbol{d})$ are continuous with respect to $\boldsymbol{d} \in \Delta(\boldsymbol{\theta})$ a.s.–$\mathbb{P}$;*

(ii) For each $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}$ and $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}) \cap \Delta(\boldsymbol{\theta}')$, $|D\ell_t(\boldsymbol{\theta};\boldsymbol{d}) - D\ell_t(\boldsymbol{\theta}';\boldsymbol{d})| \le M_t\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|$ and $|D^2\ell_t(\boldsymbol{\theta};\boldsymbol{d}) - D^2\ell_t(\boldsymbol{\theta}';\boldsymbol{d})| \le M_t\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|$, where $\{M_t\}$ is a sequence of positive, stationary and ergodic random variables;

(iii) For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and for all $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \Delta(\boldsymbol{\theta})$, there is $\lambda > 0$ such that $|D\ell_t(\boldsymbol{\theta};\boldsymbol{d}_1) - D\ell_t(\boldsymbol{\theta};\boldsymbol{d}_2)| \le M_t\|\boldsymbol{d}_1 - \boldsymbol{d}_2\|^\lambda$ and $|D^2\ell_t(\boldsymbol{\theta};\boldsymbol{d}_1) - D^2\ell_t(\boldsymbol{\theta};\boldsymbol{d}_2)| \le M_t\|\boldsymbol{d}_1 - \boldsymbol{d}_2\|^\lambda$. $\square$

Note that Assumptions 5(*i* and *ii*) are stochastic analogs of the conditions in Theorem 2 and Lemma 1. All of the examples in the current study satisfy these conditions. Assumption 5(*iii*) is assumed to apply the tightness and the ULLN to the first and second-order directional derivatives, respectively. We detail the tightness and the ULLN below, when they are more relevant. If Assumption 5(*iii*) is replaced by the following stronger Assumption 5(*iii*)*, the model is twice continuously D a.s.–$\mathbb{P}$ by Lemma 1.

Assumption 5 (Regular D-Difflity)**.** *(iii)** *For each* $\boldsymbol{\theta}$ *and for all* $\boldsymbol{d} \in \Delta(\boldsymbol{\theta})$*,* $D\ell_t(\boldsymbol{\theta}; \boldsymbol{d})$ *is linear in* $\boldsymbol{d}$ *and* $D^2\ell_t(\boldsymbol{\theta}; \boldsymbol{d})$ *is quadratic in* $\boldsymbol{d}$ *a.s.*$-\mathbb{P}$. □

We let Assumption 5* denote Assumptions 5(*i*, *ii*, and $iii^\star$) going forward when D models are referred. Unless stated otherwise, Assumption 5 stands for Assumptions 5(*i*, *ii*, and *iii*). The models in Examples 3 and 4 satisfy Assumption 5(*iii*)*.

We impose further regularity conditions for the asymptotic behavior of the extremum estimator.

Assumption 6 (CLT)**.** *(i)* $E[D\ell_t(\boldsymbol{\theta}_*; \boldsymbol{d})] = 0$ *uniformly in* $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$ *and* t*;*

(ii) $A_*(\boldsymbol{d}) := E[n^{-1}D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{d})]$ *is strictly negative and finite uniformly in* $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$ *and* n*;*

(iii) $B_*(\boldsymbol{d}, \boldsymbol{d})$ *is strictly positive and finite uniformly in* $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$ *and* n*, where for each* $\boldsymbol{d}, \widetilde{\boldsymbol{d}}$*,*

$$B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) := \mathrm{acov}\{n^{-1/2}DL_n(\boldsymbol{\theta}_*; \boldsymbol{d}), n^{-1/2}DL_n(\boldsymbol{\theta}_*; \widetilde{\boldsymbol{d}})\},$$

and 'acov' *denotes the asymptotic covariance of given arguments;*

(iv) for some $q > (r-1)/(\lambda\gamma)$ *and* $s > q \geq 2$*, and for each* $f_t \in \bar{\mathbb{L}}$*,* $\|f_t - E[f_t|\mathcal{F}_{t-\tau}^{t+\tau}]\|_q \leq \nu_\tau$*, where* $\bar{\mathbb{L}} := \{a_1 f_1 + a_2 f_2 : f_1, f_2 \in \{D\ell_t(\boldsymbol{\theta}_*; \cdot, \boldsymbol{d}) : \boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)\}, a_1, a_2 \in \mathbb{R}\}$*;* ν_τ *is of size* $-1/(1-\gamma)$ *with* $1/2 \leq \gamma < 1$*;* $\mathcal{F}_{t-\tau}^{t+\tau} := \sigma(\mathbf{X}_{t-\tau}, \cdots, \mathbf{X}_{t+\tau})$*; and* $\{\mathbf{X}_t \in \mathbb{R}^k : t = 1, 2, \cdots\}$ *is a strong mixing sequence with size* $-sq/(s-q)$*. Furthermore,* $E[M_t^s] < \infty$ *and* $\sup_{\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)} \sup_{t=1,2,\cdots} \|D\ell_t(\boldsymbol{\theta}_*; \boldsymbol{d})\|_s < \Delta < \infty$*.* □

Assumption 6(*i*) is imposed to apply the CLT. If Assumption 6(*i*) does not hold, the test statistics considered below are degenerate. Assumption 6(*i*) prevents this. Assumption 6(*iii*) is also imposed for the same purpose. For notational simplicity, we let $B_*(\boldsymbol{d})$ denote $B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$ if $\boldsymbol{d} = \widetilde{\boldsymbol{d}}$ from now. Assumption 6(*iv*) is imposed to apply corollary 3.1 of Wooldridge and White (1988) and theorem 4 of Hansen (1996a). It follows that $n^{-1/2}DL_n(\boldsymbol{\theta}_*; \cdot)$ obeys a functional central limit theorem (FCLT) mainly from Assumption 6(*iv*). Wooldridge and White (1988) provide regularity conditions for the CLT of near-epoch processes as a special case of mixingale processes. Hansen (1996a) generalizes this and provides the regularity conditions for the tightness of Lipschitz continuous functions. In essence, Assumption 6 is used to apply both CLT and tightness to $n^{-1/2}DL_n(\boldsymbol{\theta}_*;\,\cdot\,)$.

The limit distribution of $\widehat{\boldsymbol{\theta}}_n$ is obtained by using the regularity conditions provided so far. Our plan is to approximate the model by a second-order directional Taylor expansion for each direction and relate this to other directional Taylor expansions. Specifically, we first derive the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_n$ for a particular direction $\boldsymbol{d}$ and call it *directional extremum estimator* (*d-extremum estimator*). Next, we examine how this is interrelated with another d-extremum estimator in distribution. For this, we first let $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})$ denote the d-extremum estimator. That is,

$$L_n(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})) = \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_*(\boldsymbol{d})} L_n(\boldsymbol{\theta}),$$

where $\boldsymbol{\Theta}_*(\boldsymbol{d}) := \{\boldsymbol{\theta}' \in \boldsymbol{\Theta} :\ \boldsymbol{\theta}' = \boldsymbol{\theta}_* + h\boldsymbol{d},\ h \in \mathbb{R}^+, \boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)\}$. Note that the d-extremum estimator is constrained by $\boldsymbol{d}$. That is, for given $\boldsymbol{d}$, $\boldsymbol{\Theta}_*(\boldsymbol{d})$ is a straight line starting from $\boldsymbol{\theta}_*$ and ending at the boundary of $\boldsymbol{\Theta}$. Therefore, $\boldsymbol{\Theta}_*(\boldsymbol{d}) \subset \boldsymbol{\Theta}$, so that for each $\boldsymbol{d}$, $L_n(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})) \leq L_n(\widehat{\boldsymbol{\theta}}_n)$.

We can also represent the d-extremum estimator $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})$ using the distance between $\boldsymbol{\theta}_*$ and $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})$. By the constraint that $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d}) \in \boldsymbol{\Theta}_*(\boldsymbol{d})$, we let $\widehat{h}_n(\boldsymbol{d})$ be such that $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d}) \equiv \boldsymbol{\theta}_* + \widehat{h}_n(\boldsymbol{d})\boldsymbol{d}$, from which the asymptotic behavior of $\widehat{h}_n(\boldsymbol{d})$ is associated with that of $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})$. We define the space of h as $H_*(\boldsymbol{d}) := \{h \in \mathbb{R}^+ : \boldsymbol{\theta}_* + h\boldsymbol{d} \in \boldsymbol{\Theta}_*(\boldsymbol{d})\}$, so that

$$\max_{h\in H_*(\boldsymbol{d})} L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = L_n(\widehat{\boldsymbol{\theta}}(\boldsymbol{d})).$$

As $\boldsymbol{\Theta}$ is a compact and convex set in $\mathbb{R}^r$, $H_*(\boldsymbol{d})$ has to be a closed and bounded interval in $\mathbb{R}^+$ with its left-end point equal to zero. We next apply the directional second-order Taylor approximation to $L_n(\boldsymbol{\theta}_* + (\cdot)\boldsymbol{d})$ using the first and second-order directional derivatives. The following lemma shows their asymptotic behaviors.

Lemma 2. *Given Assumptions 1 to 6, for each* $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$,

(i) $n^{-1/2}DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}) \Rightarrow \mathcal{Z}(\boldsymbol{d})$, *where* $\mathcal{Z}(\boldsymbol{d}) \sim N(0, B_*(\boldsymbol{d}))$;

(ii) $n^{-1}D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{d})$ *converges to* $A_*(\boldsymbol{d})$ *a.s.–*$\mathbb{P}$;

(iii) $\left\{n^{-1/2}DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}), n^{-1}D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{d})\right\} \Rightarrow \{\mathcal{Z}(\boldsymbol{d}), A_*(\boldsymbol{d})\}$. □

Lemma 2 is not hard to prove, and it further forms a basis for obtaining the asymptotic distribution of $\widehat{h}_n(\boldsymbol{d})$.

For this purpose, we approximate $L_n(\cdot)$ on $\boldsymbol{\Theta}_*(\boldsymbol{d})$ by the mean-value theorem: for some $\bar{\boldsymbol{\theta}}_n(\boldsymbol{d}) \in \boldsymbol{\Theta}(\boldsymbol{d})$,

$$L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = L_n(\boldsymbol{\theta}_*; \boldsymbol{d}) + DL_n(\boldsymbol{\theta}_*; \boldsymbol{d})h + \frac{1}{2}D^2 L_n(\bar{\boldsymbol{\theta}}_n(\boldsymbol{d}); \boldsymbol{d})h^2, \tag{8}$$

where $h \in H_*(\boldsymbol{d})$ and $L_n(\boldsymbol{\theta}_*; \boldsymbol{d}) := \lim_{h\downarrow 0} L_n(\boldsymbol{\theta}_* + h\boldsymbol{d})$. This approximation is carried out on $H_*(\boldsymbol{d})$, so that for each $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$,

$$2\{L_n(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})) - L_n(\boldsymbol{\theta}_*; \boldsymbol{d})\} \Rightarrow \max_{\widetilde{h} \in \mathbb{R}^+} [2\mathcal{Z}(\boldsymbol{d})\widetilde{h} + A_*(\boldsymbol{d})\widetilde{h}^2] \tag{9}$$

by Lemma 2, where $\widetilde{h}$ captures the asymptotic behavior of $\sqrt{n}h$. Here, $\mathcal{Z}(\cdot)$ and $A_*(\cdot)$ are the weak and almost sure limits of $n^{-1/2}DL_n(\boldsymbol{\theta}_*; \cdot)$ and $n^{-1}D^2 L_n(\boldsymbol{\theta}_*; \cdot)$, respectively, and their convergence rates $n^{1/2}$ and n are used to capture the asymptotic behavior of $\sqrt{n}h$. The large sample properties of relevant statistics are stated in the following theorem.

Theorem 3. *Given Assumptions 1 to 6, for each $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$,*

(i) $\sqrt{n}\widehat{h}_n(\boldsymbol{d}) \Rightarrow \max[0, \mathcal{G}(\boldsymbol{d})]$, *where* $\mathcal{G}(\boldsymbol{d}) := \{-A_*(\boldsymbol{d})\}^{-1}\mathcal{Z}(\boldsymbol{d})$;

(ii) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d}) - \boldsymbol{\theta}_*) \Rightarrow \max[0, \mathcal{G}(\boldsymbol{d})]\boldsymbol{d}$;

(iii) $2\{L_n(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \max[0, \mathcal{Y}(\boldsymbol{d})]^2$, *where for each* $\boldsymbol{d}$, $\mathcal{Y}(\boldsymbol{d}) := \{-A_*(\boldsymbol{d})\}^{1/2}\mathcal{G}(\boldsymbol{d})$. □

Some remarks are warranted. The half-normal random variable $\mathcal{G}(\boldsymbol{d})$ in Theorem 3(*i*) is obtained as the weak limit of the argument in the right-hand side (RHS) of (9). That is,

$$\max[0, \mathcal{G}(\boldsymbol{d})] = \underset{\widetilde{h}(\boldsymbol{d}) \in \mathbb{R}^+}{\arg\max} [2\mathcal{Z}(\boldsymbol{d})\widetilde{h}(\boldsymbol{d}) + A_*(\boldsymbol{d})\widetilde{h}(\boldsymbol{d})^2].$$

Here, 'max' operator is involved because $\widetilde{h}(\boldsymbol{d})$ lies on the positive real line $\{\boldsymbol{\theta}_* + \sqrt{n}h\boldsymbol{d} : \ h \in H_*(\boldsymbol{d})\}$, so that if $\mathcal{Z}(\boldsymbol{d})$ is negative, the RHS of (9) is obtained by letting $\widetilde{h}(\boldsymbol{d}) = 0$. In the literature, Chernoff (1954) first approximates a parameter space by a cone, and Self and Liang (1987) and Andrews (1999) develop this to cope with the boundary parameter problem more fundamentally. We apply their methods to D-D models, and the result in Theorem 3(*i*) is the consequence of this. Theorems 3 (*ii* and *iii*) trivially follow from the identity $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d}) \equiv \boldsymbol{\theta}_* + \widehat{h}_n(\boldsymbol{d})\boldsymbol{d}$ and (9), respectively.

The pointwise results (with respect to $\boldsymbol{d}$) in Theorem 3 are not sufficient to yield the asymptotic behavior of $\widehat{\boldsymbol{\theta}}_n$. It is necessary to consider the stochastic interrelationships between d-extremum estimators. For this, we note that

$$L_n(\widehat{\boldsymbol{\theta}}_n) = \sup_{\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)} L_n(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})). \tag{10}$$

That is, if we let $\widehat{\boldsymbol{d}}_n := \arg\max_{\boldsymbol{d}\in\Delta(\boldsymbol{\theta}_*)} L_n(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d}))$, then $L_n(\widehat{\boldsymbol{\theta}}_n) \equiv L_n(\widehat{\boldsymbol{\theta}}_n(\widehat{\boldsymbol{d}}_n))$. The asymptotic behavior of $\widehat{\boldsymbol{\theta}}_n$ is derived by examining how $\widehat{\boldsymbol{\theta}}_n$ is asymptotically associated with $\widehat{\boldsymbol{\theta}}_n(\cdot)$. The following lemma promotes this examination.

Lemma 3. *Given Assumptions 1 to 6,*

(i) for all $\varepsilon > 0$*, there is* $\delta > 0$ *such that*

$$\limsup_{n\to\infty} \mathbb{P}_n \left(\sup_{\|\boldsymbol{d}_1-\boldsymbol{d}_2\|<\delta} n^{-1/2}|DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}_1) - DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}_2)| > \varepsilon \right) < \varepsilon,$$

where $\mathbb{P}_n$ *is empirical probability measure;*

(ii) for all $\varepsilon > 0$*, there is* $n(\varepsilon)$ *a.s.*$-\mathbb{P}$ *such that if* $n > n(\varepsilon)$*,* $\sup_{\boldsymbol{d}\in\Delta(\boldsymbol{\theta}_*)} |n^{-1}D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{d}) - A_*(\boldsymbol{d})| < \varepsilon$*.*

□

Lemma 3(*i*) implies that the first-order directional derivative weakly forms a Gaussian stochastic process indexed by $\boldsymbol{d}$ (e.g., Billingsley, 1999). In our time-series data context, theorem 1 of Hansen (1996a) provides sufficient regularity conditions for this. As Lemma 3(*i*) is needed to show the desired weak convergence with $r > 1$, we suppose that $\Delta(\boldsymbol{\theta}_*)$ has an uncountable number of directions when proving Lemma 3(*i*).

If $L_n(\cdot)$ is D on $\boldsymbol{\Theta}$, it is trivial to show Lemma 3. By Theorem 2, it follows that $DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}) = \nabla_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)\boldsymbol{d}$, so that

$$\sup_{\|\boldsymbol{d}_1-\boldsymbol{d}_2\|<\delta} n^{-1/2}|DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}_1) - DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}_2)| \le \|n^{-1/2}\nabla_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)\|\delta,$$

implying that for any $\varepsilon > 0$,

$$\mathbb{P}_n \left(\sup_{\|\boldsymbol{d}_1-\boldsymbol{d}_2\|<\delta} n^{-1/2} |DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}_1) - DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}_2)| > \varepsilon \right) \le \mathbb{P}_n \left(\left\| n^{-1/2}\nabla_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*) \right\| \delta > \varepsilon \right). \quad (11)$$

Thus, if $n^{-1/2}\nabla_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)$ obeys the CLT, we can choose δ to have the RHS of (11) be less than ε, and this is what Lemma 3(*i*) states. Likewise, we can also apply the ULLN to the second-order derivatives to show Lemma 3(*ii*): for each $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$, $D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{d}) = \boldsymbol{d}'\nabla^2_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)\boldsymbol{d}$, so that for a nontrivial norm, $\|\cdot\|_\infty$ say,

$$\sup_{\boldsymbol{d}} |n^{-1}\{\boldsymbol{d}'\nabla^2_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)\boldsymbol{d} - \boldsymbol{d}'E[\nabla^2_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)]\boldsymbol{d}\}| \le \sup_{\boldsymbol{d}} \boldsymbol{d}'\boldsymbol{d} \left\| n^{-1}\{\nabla^2_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*) - E[\nabla^2_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)\}]\right\|_\infty,$$

where the RHS can be made as small as we wish by applying the LLN.

Lemma 3 extends Theorem 3 to the level of functional space, and from this, the asymptotic distribution of the extremum estimator follows. If $L_n(\boldsymbol{\theta}_*)$ is not defined as in Examples 1, 2, and 4, we let $L_n(\boldsymbol{\theta}_*) := L_n(\boldsymbol{\theta}_*; \widehat{\boldsymbol{d}}_n)$ unless confusion would otherwise arise.

Theorem 4. *Given Assumptions 1 to 6,*

(i) $\{n^{-1/2} DL_n(\boldsymbol{\theta}_*; \cdot), n^{-1} D^2 L_n(\boldsymbol{\theta}_*; \cdot)\} \Rightarrow (\mathcal{Z}(\cdot), A_*(\cdot))$, *where for each* $\boldsymbol{d}$ *and* $\boldsymbol{d}'$, $E[\mathcal{Z}(\boldsymbol{d})\mathcal{Z}(\boldsymbol{d}')] = B_*(\boldsymbol{d}, \boldsymbol{d}')$*;*

(ii) $\sqrt{n}\widehat{h}_n(\cdot) \Rightarrow \max[0, \mathcal{G}(\cdot)]$*;*

(iii) $2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \sup_{\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)} \max[0, \mathcal{Y}(\boldsymbol{d})]^2$*;*

(iv) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \Rightarrow \max[0, \mathcal{G}(\boldsymbol{d}_*)]\boldsymbol{d}_*$, *where* $\boldsymbol{d}_* := \arg\max_{\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)} \max[0, \mathcal{Z}(\boldsymbol{d})]^2 \{-\mathbf{A}_*(\boldsymbol{d})\}^{-1}$. □

Note that the extremum estimator is now represented as a functional of the Gaussian stochastic process defined on $\Delta(\boldsymbol{\theta}_*)$ at the limit. Even when the model is correctly specified, so that $2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\}$ is now the LR test statistic, its asymptotic null distribution is not chi-squared.

Many statistics are known to follow asymptotic distributions characterized by Gaussian stochastic processes. For example, Davies (1977,1987), Andrews (2001), and Cho and White (2007,2010,2011a) examine these statistics: unidentified parameters yield asymptotic distributions characterized by Gaussian processes.

Theorem 4 accommodates the standard D models as a special case of D-D models. For this examination, we impose

Assumption 6 (CLT). *(ii)** *For a symmetric and negative definite matrix* $\mathbf{A}_*$ *and each* $\boldsymbol{d}$, $A_*(\boldsymbol{d}) = \boldsymbol{d}'\mathbf{A}_*\boldsymbol{d}$*;*

(iii)* *For a symmetric and positive definite matrix* $\mathbf{B}_*$ *and each* $\boldsymbol{d}, \widetilde{\boldsymbol{d}}$, $B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) = \boldsymbol{d}'\mathbf{B}_*\widetilde{\boldsymbol{d}}$. □

Assumptions 6(ii and iii)* correspond to assuming that $\mathbf{A}_* := \lim_{n\to\infty} n^{-1} E[\nabla^2_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)]$ and $\mathbf{B}_* := \mathrm{acov}\{n^{-1/2}\nabla_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)\}$ are negative and positive definite, respectively, in the D model context. These assumptions further refine the results in Theorem 4. We let Assumption 6* denote Assumptions 6(*i*, *ii**, *iii**, and *iv*) from now.

Corollary 1. *Given Assumptions 1 to 4,* $5^\star$*, and* $6^\star$*,*

(i) $\mathcal{Z}(\cdot)$ *is linear in* $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$*, so that for each* $\boldsymbol{d}$, $\mathcal{Z}(\boldsymbol{d}) = \mathbf{Z}'\boldsymbol{d}$ *in distribution, where* $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{B}_*)$*;*

(ii) for each $\boldsymbol{d}$, $\mathcal{G}(\boldsymbol{d}) = \mathbf{Z}'\boldsymbol{d}\{-\boldsymbol{d}'\mathbf{A}_*\boldsymbol{d}\}^{-1}$ *in distribution;*

(iii) for each $\boldsymbol{d}$, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d}) - \boldsymbol{\theta}_*) \Rightarrow \max[0, \{\mathbf{Z}'\boldsymbol{d}\{-\boldsymbol{d}'\mathbf{A}_*\boldsymbol{d}\}^{-1}]\boldsymbol{d}$*;*

(iv) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \Rightarrow \max[0, -\mathbf{Z}'\boldsymbol{d}_*\{\boldsymbol{d}_*{}'\mathbf{A}_*\boldsymbol{d}_*\}^{-1}]\boldsymbol{d}_*$ *with* $\boldsymbol{d}_* := \arg\max_{\boldsymbol{d}} \max[0, \mathbf{Z}'\boldsymbol{d}]^2/\boldsymbol{d}'(-\mathbf{A}_*)\boldsymbol{d}$*;*

(v) $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \Rightarrow (-\mathbf{A}_*)^{-1}\mathbf{Z}$*, provided that* $\boldsymbol{\theta}_*$ *is interior to* $\boldsymbol{\Theta}$*;*

(vi) $2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \sup_{\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)} \max[0, \mathbf{Z}'\boldsymbol{d}]^2\{\boldsymbol{d}'(-\mathbf{A}_*)\boldsymbol{d}\}^{-1}$*;*

(vii) $2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z}$*, provided that* $\boldsymbol{\theta}_*$ *is interior to* $\boldsymbol{\Theta}$. □

Corollary 1 is the same consequence as for the standard case if $\boldsymbol{\theta}_*$ is interior to $\boldsymbol{\Theta}$. Our analysis is more primitive because it involves directional derivatives. Cho (2011) exploits the D-D model analysis for D model estimation in the QML estimation context and examines other aspects that are not contained in Corollary 1.

2.4 Examples

In this subsection, we examine the previous examples using Theorems 3 and 4.

2.4.1 Example 1 (continued)

The first-order directional derivative given in (1) can be partitioned into three pieces: $DL_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*; \boldsymbol{d}) = Z_{1,n}(\boldsymbol{d}) + Z_{2,n}(\boldsymbol{d}) + Z_{3,n}(\boldsymbol{d})$, where for each $\boldsymbol{d}$,

$$Z_{1,n}(\boldsymbol{d}) := \frac{\boldsymbol{d}_{\boldsymbol{\gamma}}'}{\sigma_*^2}\sum_{t=1}^{n}\mathbf{Q}_t U_t, \quad Z_{2,n}(\boldsymbol{d}) := \sum_{t=1}^{n}\left[\frac{d_{\sigma^2}}{2\sigma_*^4} + \frac{(d_1^2+d_2^2)^{1/2}W_t^2}{2\sigma_*^2\{1-m(d_2/d_1)^2\}}\right](U_t^2 - \sigma_*^2),$$

$$Z_{3,n}(\boldsymbol{d}) := \frac{(d_1^2+d_2^2)^{1/2}}{\sigma_*^2\{1-m(d_2/d_1)^2\}}\sum_{t=2}^{n} U_t W_t \sum_{t'=1}^{t-1} U_{t'}W_{t'}m(d_2/d_1)^{t-t'},$$

and $m(d_2/d_1) := 2\tan^{-1}(d_2/d_1)/\pi$. Here, $\{\mathbf{Q}_t U_t\}$ and $\{U_t^2 - \sigma_*^2\}$ are sequences of IID random variables, so that Assumption 6(*iv*) holds for these sequences, and the CLT also holds for $Z_{1,n}(\boldsymbol{d})$ and $Z_{2,n}(\boldsymbol{d})$. Furthermore, $\{U_t W_t \sum_{t'=1}^{t-1} U_{t'} W_{t'} m(d_2/d_1)^{t-t'}\}$ is a martingale difference array (MDA). Therefore, we can apply the CLT for an MDA of McLeish (1974, Theorem 2.3): for each $\boldsymbol{d}$ and $i = 1, 2, 3$, $n^{-1/2}Z_i(\boldsymbol{d}) \Rightarrow \mathcal{Z}_i(\boldsymbol{d})$, where $\mathcal{Z}_i(\boldsymbol{d}) \sim N(0, B_*^{(i)}(\boldsymbol{d},\boldsymbol{d}))$ is independent of $\mathcal{Z}_j(\boldsymbol{d}) \sim N(0, B_*^{(j)}(\boldsymbol{d},\boldsymbol{d}))$ $(i \neq j)$, and for each $(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$,

$$B_*^{(1)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) := \frac{1}{\sigma_*^2}\boldsymbol{d}_{\boldsymbol{\gamma}}' E[\mathbf{Q}_t\mathbf{Q}_t']\widetilde{\boldsymbol{d}}_{\boldsymbol{\gamma}},$$

$$B_*^{(2)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) := E\left\{\left[\frac{d_{\sigma^2}}{\sqrt{2}\sigma_*^2} + \frac{(d_1^2+d_2^2)^{1/2}W_t^2}{\sqrt{2}\{1-m(d_2/d_1)^2\}}\right]\left[\frac{\widetilde{d}_{\sigma^2}}{\sqrt{2}\sigma_*^2} + \frac{(\widetilde{d}_1^2+\widetilde{d}_2^2)^{1/2}W_t^2}{\sqrt{2}\{1-m(\widetilde{d}_2/\widetilde{d}_1)^2\}}\right]\right\},$$

and

$$B_*^{(3)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) := \frac{(d_1^2+d_2^2)^{1/2}(\widetilde{d}_1^2+\widetilde{d}_2^2)^{1/2}}{\{1-m(d_2/d_1)^2\}\{1-m(\widetilde{d}_2/\widetilde{d}_1)^2\}}\left[\frac{m(d_2/d_1)m(\widetilde{d}_2/\widetilde{d}_1)E[W_t^2]^2}{1-m(d_2/d_1)m(\widetilde{d}_2/\widetilde{d}_1)}\right].$$

Note that $\mathcal{Z}_i(\boldsymbol{d})$ and $\mathcal{Z}_j(\boldsymbol{d})$ are asymptotically independent, and this fact enables us to apply example 1.4.6 of van der Vaart and Wellner (1996, p. 31). That is, $\{Z_{1,n}(\boldsymbol{d}), Z_{2,n}(\boldsymbol{d}), Z_{3,n}(\boldsymbol{d})\} \Rightarrow \{\mathcal{Z}_1(\boldsymbol{d}), \mathcal{Z}_2(\boldsymbol{d}), \mathcal{Z}_3(\boldsymbol{d})\}$, and $n^{-1/2}DL_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*; \boldsymbol{d}) \Rightarrow \mathcal{Z}(\boldsymbol{d}) := \mathcal{Z}_1(\boldsymbol{d}) + \mathcal{Z}_2(\boldsymbol{d}) + \mathcal{Z}_3(\boldsymbol{d})$ by the continuous mapping theorem (CMT).

It is not hard to generalize the given pointwise weak convergence to the level of functional space. This can be achieved by showing that $n^{-1/2}DL_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*;\,\cdot\,)$ is tight. As $Z_{1,n}(\cdot)$ and $Z_{2,n}(\cdot)$ are virtually linear with respect to $\mathbf{Q}_t U_t$ and $(U_t^2 - \sigma_*^2)$, respectively, their tightness trivially holds. We, therefore, focus on $Z_{3,n}(\cdot)$. For notational simplicity, we let $\varepsilon_t := W_t U_t$, $m := m(d_2/d_1)$, and $\widetilde{m} := m(\widetilde{d}_2/\widetilde{d}_1)$. Then for any $\epsilon > 0$, there is δ such that

$$\limsup_{n\to\infty} \mathbb{P}_n \left(\sup_{|m-\widetilde{m}|<\delta} \left| n^{-1/2} \sum_{t=2}^{n} \varepsilon_t \sum_{t'=1}^{t-1} \varepsilon_{t'} m^{t-t'} - n^{-1/2} \sum_{t=2}^{n} \varepsilon_t \sum_{t'=1}^{t-1} \varepsilon_{t'} \widetilde{m}^{t-t'} \right| > \epsilon \right) < \epsilon.$$

Note that the sequence $\{\varepsilon_t \sum_{t'=1}^{t-1} \varepsilon_{t'} m^{t-t'}, \sigma(\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots)\}$ is an MDA uniformly in m. We apply Hansen (1996a) to show the tightness. First, his theorem 1 trivially holds if $E[W_t^4] < \Delta^4 < \infty$. We next note that Hansen's (1996a) λ and a are identical to 1 in our model assumption,

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E[\varepsilon_t^2 (\sum_{\tau=1}^{t-1} \varepsilon_\tau m^{t-\tau})^2] = (\sigma_* \Delta)^4 \left(\frac{m^2}{1-m^2} \right) < \infty$$

uniformly in m, and the Lipschitz constant $M_t := \sum_{\tau=1}^{t-1} (t-\tau) \ddot{m}^{t-\tau-1} |\varepsilon_t \varepsilon_\tau|$ satisfies the moment condition:

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E[M_t^2] = (\sigma_* \Delta)^4 \left(\frac{1 + 2\ddot{m} - 2\ddot{m}^3 - \ddot{m}^4}{(1-\ddot{m})^5 (1+\ddot{m})^3} \right) < \infty$$

by the standard argument that $|m|$ is uniformly and strictly bounded by 1 and $E[|\varepsilon_t^2 \varepsilon_\tau \varepsilon_{t'}|] < (\sigma_* \Delta)^4 < \infty$, where $\ddot{m} := \max[|m(\underline{c})|, |m(\bar{c})|]$. These facts imply that his theorem 2 holds for our model, and Assumption 5(*iii*) also follows from this: $n^{-1/2}DL_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*;\,\cdot\,) \Rightarrow \mathcal{Z}(\cdot)$, where for each $\boldsymbol{d}$ and $\widetilde{\boldsymbol{d}}$, $E[\mathcal{Z}(\boldsymbol{d})\mathcal{Z}(\widetilde{\boldsymbol{d}})] = B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) := B_*^{(1)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) + B_*^{(2)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) + B_*^{(3)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$.

The asymptotic behavior of the second-order directional derivative is related to $B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$. By applying the LLN to (2), we obtain

$$\begin{aligned} D^2 L_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*; \boldsymbol{d}) = &- \frac{1}{\sigma_*^2} \boldsymbol{d_\gamma}' \mathbf{Q}^{n\prime} \mathbf{Q}^n \boldsymbol{d_\gamma} - \frac{d_{\sigma^2} (d_1^2 + d_2^2)^{1/2}}{\sigma_*^4} \mathbf{U}^{n\prime} \left[\boldsymbol{\Omega}^n (m(d_2/d_1)) \right] \mathbf{U}^n \\ &- \frac{n d_{\sigma^2}^2}{2\sigma_*^4} + \frac{(d_1^2 + d_2^2)}{2} \left\{ \mathrm{tr} \left[\boldsymbol{\Omega}^n (m(d_2/d_1))^2 \right] - \frac{2}{\sigma_*^2} \mathbf{U}^{n\prime} \mathbf{D}^n (m(d_2/d_1)) \mathbf{U}^n \right\} \\ &- \frac{(d_1^2 + d_2^2)}{\sigma_*^2} \mathbf{U}^{n\prime} \mathbf{O}^n (m(d_2/d_1)) \mathbf{U}^n + o_{\mathbb{P}}(n), \end{aligned}$$

where $\mathbf{D}^n(\cdot)$ is a diagonal matrix with the diagonal elements of $\boldsymbol{\Omega}^n(\cdot)^2$, and $\mathbf{O}^n(\cdot)$ is an off-diagonal elements of $\boldsymbol{\Omega}^n(\cdot)^2$ with zero diagonal elements, so that $\mathbf{D}^n(\cdot) + \mathbf{O}^n(\cdot) \equiv \boldsymbol{\Omega}^n(\cdot)^2$. Applying theorem 3.7.2

of Stout (1974) shows that $n^{-1}D^2L_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*; \boldsymbol{d}) = -B_*(\boldsymbol{d}, \boldsymbol{d}) + o_{\mathbb{P}}(1)$. The ULLN further strengthens this: $\sup_{\boldsymbol{d}} |n^{-1}D^2L_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*; \boldsymbol{d}) + B_*(\boldsymbol{d}, \boldsymbol{d})| = o_{\mathbb{P}}(1)$, that also leads to the information matrix equality. This follows mainly because $D^2L_n\,(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*; \cdot)$ is differentiable on $\Delta(\boldsymbol{\theta}_*)$, so that Assumption 5(*iii*) holds with respect to the second-order directional derivatives. Therefore,

$$2\{L_n(\widehat{\boldsymbol{\gamma}}_n, \widehat{\sigma}_n^2, \widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*)\} \Rightarrow \sup_{\boldsymbol{d}}[0, \mathcal{Y}(\boldsymbol{d})]^2$$

by Theorem 4(*iii*), where $\mathcal{Y}(\boldsymbol{d}) := \{B_*(\boldsymbol{d}, \boldsymbol{d})\}^{-1/2}\mathcal{Z}(\boldsymbol{d})$, and for each $\boldsymbol{d}$ and $\widetilde{\boldsymbol{d}}$,

$$E[\mathcal{Y}(\boldsymbol{d})\mathcal{Y}(\widetilde{\boldsymbol{d}})] = \frac{B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}})}{\{B_*(\boldsymbol{d}, \boldsymbol{d})\}^{1/2}\{B_*(\widetilde{\boldsymbol{d}}, \widetilde{\boldsymbol{d}})\}^{1/2}}. \qquad \square$$

2.4.2 Example 2 (continued)

It is not hard to identify the asymptotic behaviors of the first and second-order directional derivatives. Note that $DL_n(\boldsymbol{\theta}_*; \boldsymbol{d}) = Z_{1,n}(\boldsymbol{d}) + Z_{2,n}(\boldsymbol{d})$, where for each $\boldsymbol{d}$,

$$Z_{1,n}(\boldsymbol{d}) := \frac{d_\tau}{\tau_*^3}\sum_{t=1}^{n}(U_t^2 - \tau_*^2), \quad Z_{2,n}(\boldsymbol{d}) := \frac{1}{\tau_*^2}\sum_{t=1}^{n}\left[\mathbf{X}_t'\boldsymbol{d}_{\boldsymbol{\beta}} + m\left(d_\mu, d_\sigma\right)\right]U_t,$$

and $m(d_\mu, d_\sigma) := -[d_\mu + \psi\left(d_\mu, d_\sigma\right)]$. Note that $\psi\left(\cdot, \cdot\right)$ is Lipschitz continuous, so that Assumption 5(*iii*) holds with respect to the first-order directional derivative. Furthermore, for each $\boldsymbol{d}$, McLeish's (1974, Theorem 2.3) CLT can be applied to $Z_{1,n}(\boldsymbol{d})$ and $Z_{2,n}(\boldsymbol{d})$: for each $\boldsymbol{d}$,

$$n^{-1/2}\begin{bmatrix} Z_{n,1}(\boldsymbol{d}) \\ Z_{n,2}(\boldsymbol{d}) \end{bmatrix} \Rightarrow \begin{bmatrix} \mathcal{Z}_1(\boldsymbol{d}) \\ \mathcal{Z}_2(\boldsymbol{d}) \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{1}{\tau_*^2}\begin{bmatrix} 2d_\tau^2 & 0 \\ 0 & E[(\mathbf{X}_t'\boldsymbol{d}_{\boldsymbol{\beta}} + m\left(d_\mu, d_\sigma\right))^2] \end{bmatrix}\right).$$

It also follows that for each $\boldsymbol{d}$ and $\widetilde{\boldsymbol{d}}$,

$$E[\mathcal{Z}_1(\boldsymbol{d})\mathcal{Z}_1(\widetilde{\boldsymbol{d}})] = 2\frac{d_\tau\widetilde{d}_\tau}{\tau_*^2}, \quad E[\mathcal{Z}_1(\boldsymbol{d})\mathcal{Z}_2(\widetilde{\boldsymbol{d}})] = 0, \quad \text{and}$$

$$E[\mathcal{Z}_2(\boldsymbol{d})\mathcal{Z}_2(\widetilde{\boldsymbol{d}})] = \frac{1}{\tau_*^2}\begin{bmatrix} m\left(d_\mu, d_\sigma\right) \\ \boldsymbol{d}_{\boldsymbol{\beta}} \end{bmatrix}'\begin{bmatrix} 1 & E[\mathbf{X}_t'] \\ E[\mathbf{X}_t] & E[\mathbf{X}_t\mathbf{X}_t'] \end{bmatrix}\begin{bmatrix} m(\widetilde{d}_\mu, \widetilde{d}_\sigma) \\ \widetilde{\boldsymbol{d}}_{\boldsymbol{\beta}} \end{bmatrix}.$$

Here, $Z_{n,1}(\boldsymbol{d})$ and $Z_{n,2}(\boldsymbol{d})$ are linear with respect d_τ and $[m\left(d_\mu, d_\sigma\right), \boldsymbol{d}_{\boldsymbol{\beta}}']'$, respectively. From this fact, their tightness trivially follows, so that $n^{-1/2}DL_n(\boldsymbol{\theta}_*;\,\cdot\,) \Rightarrow \mathcal{Z}(\cdot)$, where $\mathcal{Z}(\cdot)$ is a zero-mean Gaussian stochastic process such that for each $\boldsymbol{d}$ and $\widetilde{\boldsymbol{d}}$, $E[\mathcal{Z}(\boldsymbol{d})\mathcal{Z}(\widetilde{\boldsymbol{d}})] = B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$ and

$$B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) := \frac{1}{\tau_*^2} \begin{bmatrix} \boldsymbol{d}_{\boldsymbol{\beta}} \\ m\left(d_\mu, d_\sigma\right) \\ d_\tau \end{bmatrix}' \begin{bmatrix} E[\mathbf{X}_t\mathbf{X}_t'] & E[\mathbf{X}_t] & 0 \\ E[\mathbf{X}_t'] & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{d}}_{\boldsymbol{\beta}} \\ m(\widetilde{d}_\mu, \widetilde{d}_\sigma) \\ \widetilde{d}_\tau \end{bmatrix}.$$

It is also possible to define $\mathcal{Z}(\cdot)$ as $\mathcal{Z}_1(\cdot) + \mathcal{Z}_2(\cdot)$.

We provide another Gaussian stochastic process with the same covariance structure as that of $\mathcal{Z}(\cdot)$. If we let $\widetilde{\mathcal{Z}}(\boldsymbol{d}) := \boldsymbol{\delta}(\boldsymbol{d})'\boldsymbol{\Omega}_*^{1/2}\mathbf{W}$ such that for each $\boldsymbol{d}$,

$$\boldsymbol{\delta}(\boldsymbol{d}) := \begin{bmatrix} \boldsymbol{d}_{\boldsymbol{\beta}} \\ m\left(d_\mu, d_\sigma\right) \\ d_\tau \end{bmatrix}, \qquad \boldsymbol{\Omega}_* := \frac{1}{\tau_*^2} \begin{bmatrix} E[\mathbf{X}_t\mathbf{X}_t'] & E[\mathbf{X}_t] & 0 \\ E[\mathbf{X}_t'] & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{W} \sim N(\mathbf{0}_{k+2}, \mathbf{I}_{k+2}),$$

it follows that $E[\widetilde{\mathcal{Z}}(\boldsymbol{d})\widetilde{\mathcal{Z}}(\widetilde{\boldsymbol{d}})] = \boldsymbol{\delta}(\boldsymbol{d})'\boldsymbol{\Omega}_*\boldsymbol{\delta}(\widetilde{\boldsymbol{d}})$, that is identical to $B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$, so that $\widetilde{\mathcal{Z}}(\cdot)$ has the same distribution as $\mathcal{Z}(\cdot)$. Furthermore, $\widetilde{\mathcal{Z}}(\cdot)$ is linear with respect to $\mathbf{W}$. This feature make it convenient to analyze the asymptotic distribution of the first-order directional derivative.

The probability limit of the second-order directional derivative can also be similarly found. Note that $D^2L_n(\boldsymbol{\theta}_*; \cdot)$ is Lipschitz continuous on $\Delta(\boldsymbol{\theta}_*)$, so that Assumption 5(*iii*) holds, and we can apply the LLN:

$$\frac{1}{n}\sum_{t=1}^{n} U_t^2 = \tau_* + o_{\mathbb{P}}(1), \quad \frac{1}{n}\sum_{t=1}^{n} U_t\mathbf{X}_t = o_{\mathbb{P}}(1), \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n} \mathbf{X}_t\mathbf{X}_t' = E[\mathbf{X}_t\mathbf{X}_t'] + o_{\mathbb{P}}(1).$$

This implies that

$$n^{-1}D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{d}) \overset{\text{a.s.}}{\to} -\frac{1}{\tau_*^2}\left\{2d_\tau^2 + E[(d_\mu - \mathbf{X}_t'\boldsymbol{d}_{\boldsymbol{\beta}})^2] + \psi(d_\mu, d_\sigma)^2 + 2[d_\mu - E[\mathbf{X}_t]'\boldsymbol{d}_{\boldsymbol{\beta}}]\psi(d_\mu, d_\sigma)\right\},$$

and this is identical to $-B(\boldsymbol{d}, \boldsymbol{d})$. Thus, $2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \sup_{\boldsymbol{d}\in\Delta(\boldsymbol{\theta}_*)}[0, \mathcal{Y}(\boldsymbol{d})]^2$ by Theorem 4(*iii*), where

$$\mathcal{Y}(\boldsymbol{d}) := \frac{\boldsymbol{\delta}(\boldsymbol{d})'\boldsymbol{\Omega}_*^{1/2}\mathbf{W}}{\{\boldsymbol{\delta}(\boldsymbol{d})'\boldsymbol{\Omega}_*\boldsymbol{\delta}(\boldsymbol{d})\}^{1/2}},$$

and for each $\boldsymbol{d}$ and $\widetilde{\boldsymbol{d}}$,

$$E[\mathcal{Y}(\boldsymbol{d})\mathcal{Y}(\widetilde{\boldsymbol{d}})] = \frac{\boldsymbol{\delta}(\boldsymbol{d})'\boldsymbol{\Omega}_*\boldsymbol{\delta}(\widetilde{\boldsymbol{d}})}{\{\boldsymbol{\delta}(\boldsymbol{d})'\boldsymbol{\Omega}_*\boldsymbol{\delta}(\boldsymbol{d})\}^{1/2}\{\boldsymbol{\delta}(\widetilde{\boldsymbol{d}})'\boldsymbol{\Omega}_*\boldsymbol{\delta}(\widetilde{\boldsymbol{d}})\}^{1/2}}.$$

This result shows that the directional limit of the likelihood is well defined under the null, and the asymptotic null distribution can be obtained using this, although the log-likelihood is not defined under the null. $\square$

2.4.3 Example 3 (continued)

Using the first and second-order directional derivatives in (4) and (5),

$$n^{-1/2}DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}) \Rightarrow \ddot{\boldsymbol{d}}'\mathbf{W} \quad \text{and} \quad n^{-1}D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{d}) \to \ddot{\boldsymbol{d}}'\mathbf{A}_*\ddot{\boldsymbol{d}}$$

a.s.$-\mathbb{P}$, where $\ddot{\boldsymbol{d}} \in \ddot{\Delta}(\boldsymbol{\theta}_*) := \{\mathbf{x} \in \mathbb{R}^{k+1} : \|\mathbf{x}\| = 1\}$, $\mathbf{W}$ is a multivariate normal:

$$\begin{bmatrix} n^{-1/2}\sum U_t\mathbf{Z}_t{}' \\ n^{-1/2}\sum U_t\log(X_t) \end{bmatrix} \Rightarrow \mathbf{W} := \begin{bmatrix} \mathbf{W}_0{}' \\ W_1 \end{bmatrix} \sim N(\mathbf{0}, \mathbf{B}_*)$$

with $\mathbf{B}_*$ being a $(k+1)\times(k+1)$ positive definite matrix with a finite maximum eigenvalue, and

$$\mathbf{A}_* := \begin{bmatrix} \mathbf{A}_*^{(0,0)} & \mathbf{A}_*^{(0,1)} \\ \mathbf{A}_*^{(1,0)} & \mathbf{A}_*^{(1,1)} \end{bmatrix} := \begin{bmatrix} -E[\mathbf{Z}_t\mathbf{Z}_t{}'] & -E[\mathbf{Z}_t\log(X_t)] \\ -E[\log(X_t)\mathbf{Z}_t{}'] & -E[\log(X_t)^2] \end{bmatrix}.$$

Here, we assume $E[\log(X_t)^2] < \infty$ and for each j, $E[Z_{t,j}^2] < \infty$ to obtain these limits. We also separate the set of directions into $\ddot{\Delta}(\boldsymbol{\theta}_*)$ and the set for d_2 to derive the asymptotic distribution more efficiently. By this separation, the maximization process can also be separated into a two-step process:

$$\begin{aligned} 2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} &\Rightarrow \sup_{d_2}\ \sup_{\ddot{\boldsymbol{d}}\in\ddot{\Delta}(\boldsymbol{\theta}_*)} \max[0, \mathbf{W}'\ddot{\boldsymbol{d}}]^2\{-\ddot{\boldsymbol{d}}'\mathbf{A}_*\ddot{\boldsymbol{d}}\}^{-1} \\ &= \sup_{\ddot{\boldsymbol{d}}\in\ddot{\Delta}(\boldsymbol{\theta}_*)} \max[0, \mathbf{W}'\ddot{\boldsymbol{d}}]^2\{-\ddot{\boldsymbol{d}}'\mathbf{A}_*\ddot{\boldsymbol{d}}\}^{-1} = \mathbf{W}'(-\mathbf{A}_*)^{-1}\mathbf{W} \end{aligned}$$

by Theorem 4(*iii*), where $\widehat{\boldsymbol{\theta}}_n$ is the NLS estimator, and applying the proof of Corollary 1(*vii*) obtains the last equality. Note that maximizing the limit with respect to d_2 is an innocuous process to obtaining the asymptotic null distribution because d_2 vanishes at the limit. We also note that the limit result is the same as obtained when identified models are D. $\square$

2.4.4 Example 4 (continued)

Given (6), it is trivial to show that $\{n^{1/2}DQ_n(\boldsymbol{\theta}_*;\,\cdot)\}$ is tight by the fact that it is linear with respect to $\boldsymbol{d}$. Next, we obtain that for some $\bar{\boldsymbol{\theta}}$ between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_*$,

$$n\{Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}_*)\} = -2\boldsymbol{d}'\nabla_{\boldsymbol{\theta}}\mathbf{g}_n(\bar{\boldsymbol{\theta}})'\mathbf{M}_n^{-1}\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_*)\sqrt{n}h - \boldsymbol{d}'\nabla_{\boldsymbol{\theta}}\mathbf{g}_n(\bar{\boldsymbol{\theta}})'\mathbf{M}_n^{-1}\nabla_{\boldsymbol{\theta}}\mathbf{g}_n(\bar{\boldsymbol{\theta}})\boldsymbol{d}(\sqrt{n}h)^2$$

by substituting $\mathbf{g}_n$ in (7) into $Q_n(\cdot)$, and so

$$\{Q_n(\widehat{\boldsymbol{\theta}}_n) - Q_n(\boldsymbol{\theta}_*)\} \Rightarrow \sup_{\boldsymbol{d}}\sup_{h} -2\boldsymbol{d}'\mathbf{C}_*{}'\mathbf{M}_*^{-1}\mathbf{W}h - \boldsymbol{d}'\mathbf{C}_*{}'\mathbf{M}_*^{-1}\mathbf{C}_*\boldsymbol{d}h^2.$$

We may transform this into (9) by letting $\mathcal{Z}(\boldsymbol{d}) := -\boldsymbol{d}'\mathbf{C}_*{}'\mathbf{M}_*^{-1}\mathbf{W}$ and $A_*(\boldsymbol{d}) := -\boldsymbol{d}'\mathbf{C}_*{}'\mathbf{M}_*^{-1}\ \mathbf{C}_*\boldsymbol{d}$. Note that these derivatives are linear and quadratic in $\boldsymbol{d}$, respectively. Therefore,

$$\{Q_n(\widehat{\boldsymbol{\theta}}_n) - Q_n(\boldsymbol{\theta}_*)\} \Rightarrow \mathbf{W}'\mathbf{M}_*^{-1}\mathbf{C}_*\{-\mathbf{C}_*{}'\mathbf{M}_*^{-1}\mathbf{C}_*\}^{-1}\mathbf{C}_*{}'\mathbf{M}_*^{-1}\mathbf{W}$$

by Corollary 1(*vii*). Furthermore, Corollary 1(*v*) implies that

$$\begin{aligned}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) &\Rightarrow -\{\mathbf{C}_*{}'\mathbf{M}_*^{-1}\mathbf{C}_*\}^{-1}\mathbf{C}_*{}'\mathbf{M}_*^{-1}\mathbf{W}\\ &\qquad\sim N(\mathbf{0}, \{\mathbf{C}_*{}'\mathbf{M}_*^{-1}\mathbf{C}_*\}^{-1}\{\mathbf{C}_*{}'\mathbf{M}_*^{-1}\mathbf{S}_*\mathbf{M}_*^{-1}\mathbf{C}_*\}\{\mathbf{C}_*{}'\mathbf{M}_*^{-1}\mathbf{C}_*\}^{-1}).\end{aligned}$$

These are the same results as in the standard GMM literature (e.g., Newey and West, 1987). □

3 Testing Hypotheses with D-D Models

In this section, we examine data inferences using D-D models. To this end, the standard QLR, Wald, and LM test statistics are modified to accommodate d-difflity.

It is efficient to specify first the roles of each parameter. We partition $\boldsymbol{\theta}$ into $(\boldsymbol{\pi}', \boldsymbol{\tau}')' = (\boldsymbol{\lambda}', \boldsymbol{\upsilon}', \boldsymbol{\tau}')'$ such that the directional derivatives of $L_n(\cdot)$ with respect to $\boldsymbol{\lambda}$ ($\in \mathbb{R}^{r_\lambda}$) and $\boldsymbol{\upsilon}$ ($\in \mathbb{R}^{r_\upsilon}$) are linear and non-linear with respect to $\boldsymbol{d}_{\boldsymbol{\lambda}}$ and $\boldsymbol{d}_{\boldsymbol{\upsilon}}$, respectively. The parameter $\boldsymbol{\tau}$ ($\in \mathbb{R}^{r_\tau}$) consists of other nuisance parameters that are asymptotically orthogonal to $\boldsymbol{\pi} := (\boldsymbol{\lambda}', \boldsymbol{\upsilon}')'$ ($\in \mathbb{R}^{r_\pi}$) in terms of the second-order directional derivative. More specifically, we suppose that for each $\boldsymbol{d}$, $DL_n(\boldsymbol{\theta}_*; \boldsymbol{d})$ can be written as

$$DL_n(\boldsymbol{\theta}_*; \boldsymbol{d}) = \boldsymbol{d}_{\boldsymbol{\lambda}}{}'DL_n^{(\boldsymbol{\lambda})} + DL_n^{(\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) + DL_n^{(\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}})$$

such that for each $(\boldsymbol{d_\lambda}', \boldsymbol{d_\upsilon}', \boldsymbol{d_\tau}')'$,

$$\frac{1}{\sqrt{n}}\begin{bmatrix} DL_n^{(\boldsymbol{\pi})}(\boldsymbol{d_\pi}) \\ DL_n^{(\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \end{bmatrix} := \frac{1}{\sqrt{n}}\begin{bmatrix} \boldsymbol{d_\lambda}' DL_n^{(\boldsymbol{\lambda})} \\ DL_n^{(\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon}) \\ DL_n^{(\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \end{bmatrix} \Rightarrow \begin{bmatrix} \boldsymbol{\mathcal{Z}}^{(\boldsymbol{\pi})}(\boldsymbol{d_\pi}) \\ \mathcal{Z}^{(\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \end{bmatrix} := \begin{bmatrix} \boldsymbol{d_\lambda}' \mathbf{Z}^{(\boldsymbol{\lambda})} \\ \mathcal{Z}^{(\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon}) \\ \mathcal{Z}^{(\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \end{bmatrix} \sim N(\mathbf{0}, \mathbf{B}_*(\boldsymbol{d})),$$

and $n^{-1/2}(DL_n^{(\boldsymbol{\pi})}(\cdot), DL_n^{(\boldsymbol{\tau})}(\cdot)) \Rightarrow (\boldsymbol{\mathcal{Z}}^{(\boldsymbol{\pi})}(\cdot), \mathcal{Z}^{(\boldsymbol{\tau})}(\cdot))$, where for each $\boldsymbol{d}, \widetilde{\boldsymbol{d}} \in \Delta(\boldsymbol{\theta}_*)$,

$$\begin{aligned} \mathbf{B}_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}})_{(3\times 3)} &:= \begin{bmatrix} \mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d_\pi}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}}) & \mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d_\pi}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \\ \mathbf{B}_*^{(\boldsymbol{\tau},\boldsymbol{\pi})}(\boldsymbol{d_\tau}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}})' & B_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{d_\tau}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \end{bmatrix} \\ &:= \begin{bmatrix} \boldsymbol{d_\lambda}' \mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})} \widetilde{\boldsymbol{d}}_{\boldsymbol{\lambda}} & \boldsymbol{d_\lambda}' \mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\widetilde{\boldsymbol{d}}_{\boldsymbol{\upsilon}}) & \boldsymbol{d_\lambda}' \mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \\ \mathbf{B}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{d_\upsilon})' \widetilde{\boldsymbol{d}}_{\boldsymbol{\lambda}} & B_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\upsilon}}) & B_*^{(\boldsymbol{\upsilon},\boldsymbol{\tau})}(\boldsymbol{d_\upsilon}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \\ \mathbf{B}_*^{(\boldsymbol{\tau},\boldsymbol{\lambda})}(\boldsymbol{d_\tau})' \widetilde{\boldsymbol{d}}_{\boldsymbol{\lambda}} & B_*^{(\boldsymbol{\tau},\boldsymbol{\upsilon})}(\boldsymbol{d_\tau}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\upsilon}}) & B_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{d_\tau}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \end{bmatrix}, \end{aligned} \tag{12}$$

$DL_n^{(\boldsymbol{\lambda})} \in \mathbb{R}^{r_\lambda}$, $DL_n^{(\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon}) \in \mathbb{R}$, $DL_n^{(\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \in \mathbb{R}$, $\mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})} \in \mathbb{R}^{r_\lambda} \times \mathbb{R}^{r_\lambda}$, $\mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon}) \in \mathbb{R}^{r_\lambda}$, $\mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \in \mathbb{R}^{r_\lambda}$, $\mathbf{B}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{d_\upsilon}) = \mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon})$, and $\mathbf{B}_*^{(\boldsymbol{\tau},\boldsymbol{\lambda})}(\boldsymbol{d_\tau}) = \mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d_\tau})$. Thus, it also follows that

$$\text{acov}\left\{n^{-1/2}DL_n(\boldsymbol{\theta}_*; \boldsymbol{d}), n^{-1/2}DL_n(\boldsymbol{\theta}_*; \widetilde{\boldsymbol{d}})\right\} = \boldsymbol{\iota}_3' \mathbf{B}_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}})\boldsymbol{\iota}_3,$$

where $\boldsymbol{\iota}_\ell$ is the $\ell \times 1$ vector of ones. In the same way, we also suppose that $A_*(\boldsymbol{d}) = \boldsymbol{\iota}_3' \mathbf{A}_*(\boldsymbol{d})\boldsymbol{\iota}_3$, where

$$\begin{aligned} \mathbf{A}_*(\boldsymbol{d})_{(3\times 3)} &:= \begin{bmatrix} \mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d_\pi}) & \mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d_\pi}, \boldsymbol{d_\tau}) \\ \mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\pi})}(\boldsymbol{d_\tau}, \boldsymbol{d_\pi})' & A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \end{bmatrix} \\ &:= \begin{bmatrix} \boldsymbol{d_\lambda}' \mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})} \boldsymbol{d_\lambda} & \boldsymbol{d_\lambda}' \mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon}) & \boldsymbol{d_\lambda}' \mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \\ \mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{d_\upsilon})' \boldsymbol{d_\lambda} & A_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon}) & A_*^{(\boldsymbol{\upsilon},\boldsymbol{\tau})}(\boldsymbol{d_\upsilon}, \boldsymbol{d_\tau}) \\ \mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\lambda})}(\boldsymbol{d_\tau})' \boldsymbol{d_\lambda} & A_*^{(\boldsymbol{\tau},\boldsymbol{\upsilon})}(\boldsymbol{d_\tau}, \boldsymbol{d_\upsilon}) & A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \end{bmatrix}, \end{aligned} \tag{13}$$

$\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})} \in \mathbb{R}^{r_\lambda} \times \mathbb{R}^{r_\lambda}$, $\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon}) \in \mathbb{R}^{r_\lambda}$, $\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \in \mathbb{R}^{r_\lambda}$, $\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{d_\upsilon}) = \mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon})$, and $\mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\lambda})}(\boldsymbol{d_\tau}) = \mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d_\tau})$. We also let $\boldsymbol{\pi}$ be orthogonal to $\boldsymbol{\tau}$: for each $\boldsymbol{d}$, $\mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\pi})}(\boldsymbol{d_\tau}, \boldsymbol{d_\pi}) = \mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d_\pi}, \boldsymbol{d_\tau}) = \mathbf{0}$. This assumption is useful in eliminating the nuisance parameters that are asymptotically irrelevant to testing hypotheses. We also permit that r_υ, r_λ, and r_τ can be zero, so that $\boldsymbol{\lambda}$, $\boldsymbol{\upsilon}$, or $\boldsymbol{\tau}$ may be absent in the model. If r_υ and r_τ are zero, the model is twice continuously D. We formally collect these conditions into

Assumption 7 (D-Derivatives)**.** *(i) For each* $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$, $DL_n(\boldsymbol{\theta}_*; \boldsymbol{d}) = DL_n^{(\boldsymbol{\pi})}(\boldsymbol{d_\pi}) + DL_n^{(\boldsymbol{\tau})}(\boldsymbol{d_\tau})$, *and*

$n^{-1/2}(DL_n^{(\boldsymbol{\pi})}(\cdot), DL_n^{(\boldsymbol{\tau})}(\cdot)) \Rightarrow (\boldsymbol{\mathcal{Z}}^{(\boldsymbol{\pi})}(\cdot), \mathcal{Z}^{(\boldsymbol{\tau})}(\cdot))$;

(ii) For each $\boldsymbol{d}, \widetilde{\boldsymbol{d}} \in \Delta(\boldsymbol{\theta}_*)$, $B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) = \boldsymbol{\iota}_3{}'\mathbf{B}_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}})\boldsymbol{\iota}_3$, *where for each* $\boldsymbol{d}$, $\mathbf{B}_*(\boldsymbol{d}, \boldsymbol{d})$ *is symmetric and positive definite;*

(iii) For each $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$, $A_*(\boldsymbol{d}) = \boldsymbol{\iota}_3{}'\mathbf{A}_*(\boldsymbol{d})\boldsymbol{\iota}_3$, *where for each* $\boldsymbol{d}$, $\mathbf{A}_*(\boldsymbol{d})$ *is symmetric and negative definite;*

(iv) $\mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\tau}}, \boldsymbol{d}_{\boldsymbol{\pi}}) = \mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \boldsymbol{d}_{\boldsymbol{\tau}}) = \mathbf{0}$ *uniformly in* $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$*;*

(v) $\boldsymbol{\Theta} = \boldsymbol{\Pi} \times \mathbf{T}$ *and* $C(\boldsymbol{\theta}_*) = C(\boldsymbol{\pi}_*) \times C(\boldsymbol{\tau}_*)$, *where* $C(\boldsymbol{\pi}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\pi} : \exists \boldsymbol{\pi}' \in \boldsymbol{\Pi}, \mathbf{x} := \boldsymbol{\pi}_* + \delta\boldsymbol{\pi}', \delta \in \mathbb{R}^+\}$ *and* $C(\boldsymbol{\tau}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\tau} : \exists \boldsymbol{\tau}' \in \mathbf{T}, \mathbf{x} := \boldsymbol{\tau}_* + \delta\boldsymbol{\tau}', \delta \in \mathbb{R}^+\}$. □

Assumption 7(*v*) lets the parameter space $\boldsymbol{\Theta}$ be the Cartesian product of two separate parameter spaces $\boldsymbol{\Pi}$ and $\mathbf{T}$. We use this property to represent $L_n(\cdot)$ as the sum of two independent functions as discussed in Theorem 5 below.

We further partition $\boldsymbol{\lambda}$ and $\boldsymbol{\upsilon}$ into $(\boldsymbol{\delta}', \boldsymbol{\phi}')'$ $(\in \mathbb{R}^{r_\delta + r_\phi})$ and $(\boldsymbol{\xi}', \boldsymbol{\psi}')'$ $(\in \mathbb{R}^{r_\xi + r_\psi})$, respectively, so that $\mathbf{Z}^{(\boldsymbol{\lambda})}$, $\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}$, and $\mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}$ can be accordingly represented as $\mathbf{Z}^{(\boldsymbol{\lambda})} = (\mathbf{Z}^{(\boldsymbol{\delta})\prime}, \mathbf{Z}^{(\boldsymbol{\phi})\prime})'$,

$$\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})} = \begin{bmatrix} \mathbf{A}_*^{(\boldsymbol{\delta},\boldsymbol{\delta})} & \mathbf{A}_*^{(\boldsymbol{\delta},\boldsymbol{\phi})} \\ \mathbf{A}_*^{(\boldsymbol{\phi},\boldsymbol{\delta})} & \mathbf{A}_*^{(\boldsymbol{\phi},\boldsymbol{\phi})} \end{bmatrix} \text{ and } \mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})} = \begin{bmatrix} \mathbf{B}_*^{(\boldsymbol{\delta},\boldsymbol{\delta})} & \mathbf{B}_*^{(\boldsymbol{\delta},\boldsymbol{\phi})} \\ \mathbf{B}_*^{(\boldsymbol{\phi},\boldsymbol{\delta})} & \mathbf{B}_*^{(\boldsymbol{\phi},\boldsymbol{\phi})} \end{bmatrix}.$$

Similarly, for each $\boldsymbol{d}_{\boldsymbol{\upsilon}}$, we let $\mathcal{Z}^{(\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) = \mathcal{Z}^{(\boldsymbol{\xi})}(\boldsymbol{d}_{\boldsymbol{\xi}}) + \mathcal{Z}^{(\boldsymbol{\psi})}(\boldsymbol{d}_{\boldsymbol{\psi}})$, where for each $\boldsymbol{d}_{\boldsymbol{\xi}}$ and $\boldsymbol{d}_{\boldsymbol{\psi}}$, $\boldsymbol{\mathcal{Z}}^{(\upsilon)}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) := (\mathcal{Z}^{(\boldsymbol{\xi})}(\boldsymbol{d}_{\boldsymbol{\xi}}), \mathcal{Z}^{(\boldsymbol{\psi})}(\boldsymbol{d}_{\boldsymbol{\psi}}))' \sim N(\mathbf{0}, \mathbf{B}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}, \boldsymbol{d}_{\boldsymbol{\upsilon}}))$, and

$$\mathbf{B}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\upsilon}}) := \begin{bmatrix} B_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{d}_{\boldsymbol{\xi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\xi}}) & B_*^{(\boldsymbol{\xi},\boldsymbol{\psi})}(\boldsymbol{d}_{\boldsymbol{\xi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\psi}}) \\ B_*^{(\boldsymbol{\psi},\boldsymbol{\xi})}(\boldsymbol{d}_{\boldsymbol{\psi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\xi}}) & B_*^{(\boldsymbol{\psi},\boldsymbol{\psi})}(\boldsymbol{d}_{\boldsymbol{\psi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\psi}}) \end{bmatrix}.$$

This implies that $B_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}, \boldsymbol{d}_{\boldsymbol{\upsilon}}) = B_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{d}_{\boldsymbol{\xi}}, \boldsymbol{d}_{\boldsymbol{\xi}}) + 2B_*^{(\boldsymbol{\xi},\boldsymbol{\psi})}(\boldsymbol{d}_{\boldsymbol{\xi}}, \boldsymbol{d}_{\boldsymbol{\psi}}) + B_*^{(\boldsymbol{\psi},\boldsymbol{\psi})}(\boldsymbol{d}_{\boldsymbol{\psi}}, \boldsymbol{d}_{\boldsymbol{\psi}})$. For a consistent presentation, we also assume that $A_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) = \boldsymbol{\iota}_2{}'\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}})\boldsymbol{\iota}_2$ by letting

$$\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) := \begin{bmatrix} A_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{d}_{\boldsymbol{\xi}}) & A_*^{(\boldsymbol{\xi},\boldsymbol{\psi})}(\boldsymbol{d}_{\boldsymbol{\xi}}, \boldsymbol{d}_{\boldsymbol{\psi}}) \\ A_*^{(\boldsymbol{\psi},\boldsymbol{\xi})}(\boldsymbol{d}_{\boldsymbol{\psi}}, \boldsymbol{d}_{\boldsymbol{\xi}}) & A_*^{(\boldsymbol{\psi},\boldsymbol{\psi})}(\boldsymbol{d}_{\boldsymbol{\psi}}) \end{bmatrix}$$

be negative definite. We formally collect these conditions into

Assumption 8 (Inference)**.** *(i) For each* $\boldsymbol{d}_{\boldsymbol{\upsilon}}$ *and for the symmetric and positive definite* $\mathbf{B}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}$ $(\boldsymbol{d}_{\boldsymbol{\upsilon}}, \boldsymbol{d}_{\boldsymbol{\upsilon}})$ *given in (3),* $\mathcal{Z}^{(\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) = \boldsymbol{\mathcal{Z}}^{(\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}})'\boldsymbol{\iota}_2$ *such that* $\boldsymbol{\mathcal{Z}}^{(\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) \sim N(\mathbf{0}, \mathbf{B}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}, \boldsymbol{d}_{\boldsymbol{\upsilon}}))$*;*

(ii) For each $\boldsymbol{d}_{\boldsymbol{\upsilon}}$ *and for the symmetric and negative definite* $\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}})$ *given in (3),* $A_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) =$

$\boldsymbol{\iota}_2{}'\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}})\boldsymbol{\iota}_2$. $\square$

Assumption 8 does not impose any sign restriction to $\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}$ and $\mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}$ because Assumption 7 already implies that both $-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}$ and $\mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}$ are symmetric and positive definite.

Given the structure of the parameters, we let $\boldsymbol{\delta}$ or $\boldsymbol{\xi}$ be the parameter of interest, and the hypotheses of interest are given as

$$\left\{ \begin{array}{l} H_0' : \boldsymbol{\delta}_* = \boldsymbol{\delta}_0; \\ H_1' : \boldsymbol{\delta}_* \neq \boldsymbol{\delta}_0 \end{array} \right\} \quad \text{or} \quad \left\{ \begin{array}{l} H_0'' : \boldsymbol{\xi}_* = \boldsymbol{\xi}_0; \\ H_1'' : \boldsymbol{\xi}_* \neq \boldsymbol{\xi}_0 \end{array} \right\}.$$

The role of the parameters in the first hypotheses (H_0' versus H_1') is different from that in the second hypotheses (H_0'' versus H_1''). The directional derivative with respect to $\boldsymbol{d}_{\boldsymbol{\delta}}$ is linear, whereas it is nonlinear with respect to $\boldsymbol{d}_{\boldsymbol{\xi}}$. For future reference, we let $\boldsymbol{\Theta}_0$ be the parameter space constrained by the null hypotheses. That is, $\boldsymbol{\Theta}_0$ is either $\boldsymbol{\Theta}_0^{(\boldsymbol{\delta})} := \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \boldsymbol{\delta} = \boldsymbol{\delta}_0\}$ or $\boldsymbol{\Theta}_0^{(\boldsymbol{\xi})} := \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \boldsymbol{\xi} = \boldsymbol{\xi}_0\}$.

We handle these two hypotheses more efficiently by introducing generic notation. The parameter $\boldsymbol{\pi}$ is rewritten as $\boldsymbol{\eta} := (\boldsymbol{\mu}', \boldsymbol{\omega}')'$ such that $\boldsymbol{\mu}$ is either $\boldsymbol{\delta}$ or $\boldsymbol{\xi}$, and $\boldsymbol{\omega}$ is the rest of $\boldsymbol{\pi}$. Thus, $\boldsymbol{\eta}$ is either $(\boldsymbol{\delta}', \boldsymbol{\omega}')' = (\boldsymbol{\delta}', \boldsymbol{\phi}', \boldsymbol{\xi}', \boldsymbol{\psi}')'$ or $(\boldsymbol{\xi}', \boldsymbol{\omega}')' = (\boldsymbol{\xi}', \boldsymbol{\delta}', \boldsymbol{\phi}', \boldsymbol{\psi}')'$. We also let $\mathbf{H}$, $\mathbf{M}$, and $\boldsymbol{\Omega}$ be the parameter spaces of $\boldsymbol{\eta}$, $\boldsymbol{\mu}$, and $\boldsymbol{\omega}$, respectively, so that the two hypotheses given as H_0' versus H_1' and H_0'' versus H_1'' are generically written as

$$H_0 : \ \boldsymbol{\mu}_* = \boldsymbol{\mu}_0 \quad \text{versus} \quad H_1 : \ \boldsymbol{\mu}_* \neq \boldsymbol{\mu}_0,$$

where $\boldsymbol{\mu}_0$ is either $\boldsymbol{\delta}_0$ or $\boldsymbol{\xi}_0$. We also let $\boldsymbol{d}_{\boldsymbol{\eta}} := (\boldsymbol{d}_{\boldsymbol{\mu}}', \boldsymbol{d}_{\boldsymbol{\omega}}')'$ and $\boldsymbol{\mathcal{Z}}^{(\boldsymbol{\eta})}(\boldsymbol{d}_{\boldsymbol{\eta}}) := (\mathcal{Z}^{(\boldsymbol{\mu})}(\boldsymbol{d}_{\boldsymbol{\mu}}), \mathcal{Z}^{(\boldsymbol{\omega})}(\boldsymbol{d}_{\boldsymbol{\omega}}))'$, respectively, and also reformulate $\mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}})$ and $\mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}})$ into

$$\mathbf{A}_*^{(\boldsymbol{\eta},\boldsymbol{\eta})}(\boldsymbol{d}_{\boldsymbol{\mu}}, \boldsymbol{d}_{\boldsymbol{\omega}}) := \begin{bmatrix} A_*^{(\boldsymbol{\mu},\boldsymbol{\mu})}(\boldsymbol{d}_{\boldsymbol{\mu}}) & A_*^{(\boldsymbol{\mu},\boldsymbol{\omega})}(\boldsymbol{d}_{\boldsymbol{\mu}}, \boldsymbol{d}_{\boldsymbol{\omega}}) \\ A_*^{(\boldsymbol{\omega},\boldsymbol{\mu})}(\boldsymbol{d}_{\boldsymbol{\omega}}, \boldsymbol{d}_{\boldsymbol{\mu}}) & A_*^{(\boldsymbol{\omega},\boldsymbol{\omega})}(\boldsymbol{d}_{\boldsymbol{\omega}}) \end{bmatrix}$$

and

$$\mathbf{B}_*^{(\boldsymbol{\eta},\boldsymbol{\eta})}((\boldsymbol{d}_{\boldsymbol{\mu}}, \boldsymbol{d}_{\boldsymbol{\omega}}), (\widetilde{\boldsymbol{d}}_{\boldsymbol{\mu}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\omega}})) := \begin{bmatrix} B_*^{(\boldsymbol{\mu},\boldsymbol{\mu})}(\boldsymbol{d}_{\boldsymbol{\mu}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\mu}}) & B_*^{(\boldsymbol{\mu},\boldsymbol{\omega})}(\boldsymbol{d}_{\boldsymbol{\mu}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\omega}}) \\ B_*^{(\boldsymbol{\omega},\boldsymbol{\mu})}(\boldsymbol{d}_{\boldsymbol{\omega}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\mu}}) & B_*^{(\boldsymbol{\omega},\boldsymbol{\omega})}(\boldsymbol{d}_{\boldsymbol{\omega}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\omega}}) \end{bmatrix},$$

respectively. By these reformulations, Assumption 7(v) can be restructured as

Assumption 8 (Inference)**.** *(iii)* $\mathbf{H} = \mathbf{M} \times \boldsymbol{\Omega}$ *and* $C(\boldsymbol{\pi}_*) = C(\boldsymbol{\mu}_*) \times C(\boldsymbol{\omega}_*)$, *where* $C(\boldsymbol{\mu}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\mu} : \exists \boldsymbol{\mu}' \in \mathbf{M}, \mathbf{x} := \boldsymbol{\mu}_* + \delta \boldsymbol{\mu}', \delta \in \mathbb{R}^+\}$ *and* $C(\boldsymbol{\omega}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\omega} : \exists \boldsymbol{\omega}' \in \boldsymbol{\Omega}, \mathbf{x} := \boldsymbol{\omega}_* + \delta \boldsymbol{\omega}', \delta \in \mathbb{R}^+\}$. $\square$

3.1 Qausi-Likelihood Ratio Test Statistic

The standard QLR test statistic for D models can be used for D-D models. We formally define the QLR test statistic as

$$\mathcal{LR}_n := 2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\ddot{\boldsymbol{\theta}}_n)\},$$

where $\ddot{\boldsymbol{\theta}}_n$ is such that $L_n(\ddot{\boldsymbol{\theta}}_n) := \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} L_n(\boldsymbol{\theta})$.

For the analysis of the QLR test statistic, we split $\mathcal{LR}_n$ into $\mathcal{LR}_n^{(1)}$ and $\mathcal{LR}_n^{(2)}$ such that $\mathcal{LR}_n^{(1)} := 2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\}$ and $\mathcal{LR}_n^{(2)} := 2\{L_n(\ddot{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\}$. Theorem 4(*iii*) already provides the asymptotic distribution of $\mathcal{LR}_n^{(1)}$. For examining the roles of Assumption 7, we first let $\Delta(\boldsymbol{\pi}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\pi} : \boldsymbol{\pi}_* + \mathbf{x} \in \mathrm{cl}\{C(\boldsymbol{\pi}_*)\}, \|\mathbf{x}\| = 1\ \}$, $\Delta(\boldsymbol{\tau}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\tau} : \boldsymbol{\tau}_* + \mathbf{x} \in \mathrm{cl}\{C(\boldsymbol{\tau}_*)\}, \|\mathbf{x}\| = 1\ \}$ and denote their representative components by $\boldsymbol{s_\pi} (= (\boldsymbol{s}_{\boldsymbol{\lambda}}', \boldsymbol{s}_{\boldsymbol{\upsilon}}')')$ and $\boldsymbol{s_\tau}$, respectively. Here, $\boldsymbol{s}$ is used to distinguish its role from that of $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$. Note that $\Delta(\boldsymbol{\pi}_*)$ and $\Delta(\boldsymbol{\tau}_*)$ are subsets of $\Delta(\boldsymbol{\theta}_*)$. The following theorem provides the asymptotic behavior of $\mathcal{LR}_n^{(1)}$.

Theorem 5. *Given Assumptions 1 to 7,*

$$2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \mathcal{H}_1 + \mathcal{H}_2 := \sup_{\boldsymbol{s_\pi} \in \Delta(\boldsymbol{\pi}_*)} \max[0, \mathcal{Y}^{(\boldsymbol{\pi})}(\boldsymbol{s_\pi})]^2 + \sup_{\boldsymbol{s_\tau} \in \Delta(\boldsymbol{\tau}_*)} \max[0, \mathcal{Y}^{(\boldsymbol{\tau})}(\boldsymbol{s_\tau})]^2,$$

where for each $\boldsymbol{s_\pi}$ *and* $\boldsymbol{s_\tau}$,

$$\mathcal{Y}^{(\boldsymbol{\pi})}(\boldsymbol{s_\pi}) := \{-\boldsymbol{\iota}_2{}' \mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{s_\pi})\boldsymbol{\iota}_2\}^{-1/2} \mathcal{Z}^{(\boldsymbol{\pi})}(\boldsymbol{s_\pi}), \quad \mathcal{Y}^{(\boldsymbol{\tau})}(\boldsymbol{s_\tau}) := \{-A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{s_\tau})\}^{-1/2} \mathcal{Z}^{(\boldsymbol{\tau})}(\boldsymbol{s_\tau}),$$

respectively, and $\mathcal{Z}^{(\boldsymbol{\pi})}(\boldsymbol{s_\pi}) := \boldsymbol{s_\lambda}{}' \boldsymbol{\mathcal{Z}}^{(\boldsymbol{\lambda})} + \mathcal{Z}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})$. □

The orthogonality condition in Assumption 7(*iv*) and the parameter space condition in Assumption 7(*v*) asymptotically separate $\mathcal{LR}_n^{(1)}$ into the sum of $\mathcal{H}_1$ and $\mathcal{H}_2$ as given in Theorem 5. This also implies that we can ignore the effects of $\boldsymbol{\tau}$ (resp. $\boldsymbol{\pi}$) when testing the null hypothesis that is associated with only $\boldsymbol{\pi}$ (resp. $\boldsymbol{\tau}$).

Theorem 5 is also related to the asymptotic distribution of $\mathcal{LR}_n^{(2)}$ because its weak limit contains the same component as $\mathcal{LR}_n^{(1)}$. The following theorem shows the asymptotic behavior of $\ddot{\boldsymbol{\theta}}_n$ and $\mathcal{LR}_n^{(2)}$.

Theorem 6. *(i) Given Assumptions 1 to 3 and* H_0, $\ddot{\boldsymbol{\theta}}_n$ *converges to* $\boldsymbol{\theta}_*$ a.s.–$\mathbb{P}$;

(ii) Given Assumptions 1 to 8 and H_0, $\mathcal{LR}_n^{(2)} \Rightarrow \mathcal{H}_0 + \mathcal{H}_2 := \sup_{\boldsymbol{s_\omega} \in \Delta(\boldsymbol{\omega}_*)} \max[0, \mathcal{Y}^{(\boldsymbol{\omega})}(\boldsymbol{s_\omega})]^2 + \mathcal{H}_2$, *where for each* $\boldsymbol{s_\omega}$, $\mathcal{Y}^{(\boldsymbol{\omega})}(\boldsymbol{s_\omega}) := \{-A_*^{(\boldsymbol{\omega},\boldsymbol{\omega})}(\boldsymbol{s_\omega})\}^{-1/2} \mathcal{Z}^{(\boldsymbol{\omega})}(\boldsymbol{s_\omega})$ and $\Delta(\boldsymbol{\omega}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\omega} : \boldsymbol{\omega}_* + \mathbf{x} \in \mathrm{cl}\{C(\boldsymbol{\omega}_*)\}, \|\mathbf{x}\| = 1\ \}$. □

Note that $\mathcal{H}_2$ in Theorem 6(*ii*) is the same random variable as that in Theorem 5(*ii*).

Several remarks are in order. First, the consequence in Theorem 6(*ii*) can be understood as a corollary of Theorem 5. Note that if $\boldsymbol{\mu} = \boldsymbol{\delta}$ and $\boldsymbol{\omega} = (\boldsymbol{\phi}', \boldsymbol{v}')'$,

$$\mathcal{LR}_n^{(2)} \Rightarrow \sup_{(\boldsymbol{s_\phi}, \boldsymbol{s_v}) \in \Delta(\boldsymbol{\omega}_*)} \left\{ \frac{\max[0, \boldsymbol{s_\phi}'\mathbf{Z}^{(\boldsymbol{\phi})} + \mathcal{Z}^{(\boldsymbol{v})}(\boldsymbol{s_v})]^2}{\boldsymbol{s_\phi}'(-\mathbf{A}_*^{(\boldsymbol{\phi},\boldsymbol{\phi})})\boldsymbol{s_\phi} - A_*^{(\boldsymbol{v},\boldsymbol{v})}(\boldsymbol{s_v}) - 2\boldsymbol{s_\phi}'\mathbf{A}_*^{(\boldsymbol{\phi},\boldsymbol{v})}(\boldsymbol{s_v})} \right\} + \mathcal{H}_2$$

by Theorem 6. In a similar manner, if $\boldsymbol{\mu} = \boldsymbol{\xi}$ and $\boldsymbol{\omega} = (\boldsymbol{\lambda}', \boldsymbol{\psi}')'$,

$$\mathcal{LR}_n^{(2)} \Rightarrow \sup_{(\boldsymbol{s_\lambda}, \boldsymbol{s_\psi}) \in \Delta(\boldsymbol{\omega}_*)} \left\{ \frac{\max[0, \boldsymbol{s_\lambda}'\mathbf{Z}^{(\boldsymbol{\lambda})} + \mathcal{Z}^{(\boldsymbol{\psi})}(\boldsymbol{s_\psi})]^2}{\boldsymbol{s_\lambda}'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})\boldsymbol{s_\lambda} - A_*^{(\boldsymbol{\psi},\boldsymbol{\psi})}(\boldsymbol{s_\psi}) - 2\boldsymbol{s_\lambda}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\psi})}(\boldsymbol{s_\psi})} \right\} + \mathcal{H}_2. \tag{14}$$

Second, the weak limit in Theorem 6(*ii*) is jointly achieved with that of $\mathcal{LR}_n^{(1)}$ because all of these are obtained by applying the CMT to Theorem 4(*i*). Furthermore, $\mathcal{H}_2$ in $\mathcal{LR}_n^{(2)}$ is identical to that of $\mathcal{LR}_n^{(1)}$. From this fact, the asymptotic null distribution of the QLR test statistic is immediately obtained.

Theorem 7. *Given Assumptions 1 to 8, and H_0, $\mathcal{LR}_n \Rightarrow \mathcal{H}_1 - \mathcal{H}_0$.* □

The asymptotic null distribution in Theorem 7 can have various forms depending on the model properties. For example, if $\boldsymbol{\mu} = \boldsymbol{\delta}$ and $r_v = 0$, then $\boldsymbol{\pi} = \boldsymbol{\lambda}$, $\boldsymbol{\phi} = \boldsymbol{\omega}$, and

$$\mathcal{LR}_n \Rightarrow \sup_{\boldsymbol{s_\lambda} \in \Delta(\boldsymbol{\lambda}_*)} \left\{ \frac{\max[0, \boldsymbol{s_\lambda}'\mathbf{Z}^{(\boldsymbol{\lambda})}]^2}{\boldsymbol{s_\lambda}'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})\boldsymbol{s_\lambda}} \right\} - \sup_{\boldsymbol{s_\phi} \in \Delta(\boldsymbol{\phi}_*)} \left\{ \frac{\max[0, \boldsymbol{s_\phi}'\mathbf{Z}^{(\boldsymbol{\phi})}]^2}{\boldsymbol{s_\phi}'(-\mathbf{A}_*^{(\boldsymbol{\phi},\boldsymbol{\phi})})\boldsymbol{s_\phi}} \right\}$$

under H_0'. Furthermore, if $\boldsymbol{\lambda}_*$ is an interior element, applying the proof of Corollary 1(*vii*) shows that

$$\mathcal{LR}_n \Rightarrow (\widetilde{\mathbf{Z}}^{(\boldsymbol{\delta})})'(-\widetilde{\mathbf{A}}_*^{(\boldsymbol{\delta},\boldsymbol{\delta})})^{-1}(\widetilde{\mathbf{Z}}^{(\boldsymbol{\delta})}), \tag{15}$$

where $\widetilde{\mathbf{Z}}^{(\boldsymbol{\delta})} := \mathbf{Z}^{(\boldsymbol{\delta})} - (\mathbf{A}_*^{(\boldsymbol{\delta},\boldsymbol{\phi})})(\mathbf{A}_*^{(\boldsymbol{\phi},\boldsymbol{\phi})})^{-1}\mathbf{Z}^{(\boldsymbol{\phi})}$ and $\widetilde{\mathbf{A}}_*^{(\boldsymbol{\delta},\boldsymbol{\delta})} := \mathbf{A}_*^{(\boldsymbol{\delta},\boldsymbol{\delta})} - (\mathbf{A}_*^{(\boldsymbol{\delta},\boldsymbol{\phi})})(\mathbf{A}_*^{(\boldsymbol{\phi},\boldsymbol{\phi})})^{-1}(\mathbf{A}_*^{(\boldsymbol{\phi},\boldsymbol{\delta})})'$. Therefore, the same result is obtained as for the standard D model analysis. If $\boldsymbol{\lambda}_*$ is a boundary parameter, the QLR test statistic has an asymptotic null distribution that depends on the property of the parameter space. For example, if $\boldsymbol{\delta} \in [\delta_*, \bar{\delta}] \subset \mathbb{R}$ and $\boldsymbol{\phi}_*$ is an interior element of its parameter space, so that $\widetilde{\mathbf{A}}_*^{(\boldsymbol{\delta},\boldsymbol{\delta})}$ and $\widetilde{\mathbf{Z}}^{(\delta)}$ are now scalars, and

$$\sup_{\boldsymbol{s_\lambda} \in \Delta(\boldsymbol{\lambda}_*)} \left\{ \frac{\max[0, \boldsymbol{s_\lambda}'\mathbf{Z}^{(\boldsymbol{\lambda})}]^2}{\boldsymbol{s_\lambda}'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})\boldsymbol{s_\lambda}} \right\} = (\widetilde{\mathbf{Z}}^{(\delta)})'(-\widetilde{\mathbf{A}}_*^{(\boldsymbol{\delta},\boldsymbol{\delta})})^{-1}(\widetilde{\mathbf{Z}}^{(\delta)})\mathbf{1}_{\{\widehat{s}_{\boldsymbol{\delta}} > 0\}} + (\mathbf{Z}^{(\boldsymbol{\phi})})'(-\mathbf{A}_*^{(\boldsymbol{\phi},\boldsymbol{\phi})})^{-1}(\mathbf{Z}^{(\boldsymbol{\phi})}),$$

where $\widehat{\boldsymbol{s}}_{\boldsymbol{\lambda}} := (\widehat{\boldsymbol{s}}_{\boldsymbol{\delta}}', \widehat{\boldsymbol{s}}_{\boldsymbol{\phi}}')'$ solves the LHS. Furthermore, it is a standard exercise to show that $\widetilde{\mathbf{Z}}^{(\delta)} < 0$ if and only if $\widehat{s}_{\boldsymbol{\delta}} \leq 0$. Thus, it follows that $\mathcal{LR}_n \Rightarrow (-\widetilde{\mathbf{A}}_*^{(\boldsymbol{\delta},\boldsymbol{\delta})})^{-1} \max[0, \widetilde{\mathbf{Z}}^{(\delta)}]^2$ under H_0' because

$$\sup_{\boldsymbol{s}_{\boldsymbol{\phi}} \in \Delta(\boldsymbol{\phi}_*)} \left\{ \frac{\max[0, \boldsymbol{s}_{\boldsymbol{\phi}}{}' \mathbf{Z}^{(\boldsymbol{\phi})}]^2}{\boldsymbol{s}_{\boldsymbol{\phi}}{}'(-\mathbf{A}_*^{(\boldsymbol{\phi},\boldsymbol{\phi})})\boldsymbol{s}_{\boldsymbol{\phi}}} \right\} = (\mathbf{Z}^{(\boldsymbol{\phi})})'(-\mathbf{A}_*^{(\boldsymbol{\phi},\boldsymbol{\phi})})^{-1}(\mathbf{Z}^{(\boldsymbol{\phi})})$$

by the proof of Corollary 1(*vii*). As another example, if $\boldsymbol{\mu} = \boldsymbol{\xi}$ and $r_\psi = 0$, a different asymptotic null distribution is obtained. In this case, $\boldsymbol{\lambda} = \boldsymbol{\omega}$ and $\boldsymbol{\pi} = (\boldsymbol{\lambda}', \boldsymbol{\xi}')'$, so that

$$\mathcal{H}_0 = \sup_{\boldsymbol{s}_{\boldsymbol{\lambda}} \in \Delta(\boldsymbol{\lambda}_*)} \left\{ \frac{\max[0, \boldsymbol{s}_{\boldsymbol{\lambda}}{}' \mathbf{Z}^{(\boldsymbol{\lambda})}]^2}{\boldsymbol{s}_{\boldsymbol{\lambda}}{}'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})\boldsymbol{s}_{\boldsymbol{\lambda}}} \right\}, \tag{16}$$

and

$$\mathcal{H}_1 = \sup_{(\boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\xi}}) \in \Delta(\boldsymbol{\pi}_*)} \left\{ \frac{\max[0, \mathcal{Z}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) + \boldsymbol{s}_{\boldsymbol{\lambda}}{}' \mathbf{Z}^{(\boldsymbol{\lambda})}]^2}{-A_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) + 2\boldsymbol{s}_{\boldsymbol{\lambda}}{}'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})) + \boldsymbol{s}_{\boldsymbol{\lambda}}{}'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})\boldsymbol{s}_{\boldsymbol{\lambda}}} \right\} \tag{17}$$

by applying (14). Therefore, $\mathcal{H}_1 - \mathcal{H}_0$ becomes the asymptotic null distribution of the QLR test statistic.

Regularized asymptotic null distribution is obtained if the parameter space condition is imposed as

Assumption 9 (Benchmark)**.** *(i)* $r_\psi = 0$*;*

(ii) $\boldsymbol{\lambda}_*$ *is an interior element of* $\boldsymbol{\Lambda}$*;*

(iii) $\boldsymbol{\Pi} = \boldsymbol{\Lambda} \times \boldsymbol{\Xi} \subset \mathbb{R}^{r_\lambda} \times \mathbb{R}^{r_\xi}$ *such that* $\boldsymbol{\Lambda}$ *and* $\boldsymbol{\Xi}$ *are parameter spaces for* $\boldsymbol{\lambda}$ *and* $\boldsymbol{\xi}$*, respectively;*

(iv) $C(\boldsymbol{\pi}_*) = \mathbb{R}^{r_\lambda} \times C(\boldsymbol{\xi}_*)$*, where* $C(\boldsymbol{\xi}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\lambda} : \exists \boldsymbol{\xi}' \in \boldsymbol{\Xi}, \mathbf{x} := \boldsymbol{\xi}_* + \delta\boldsymbol{\xi}', \delta \in \mathbb{R}^+\}$. □

The null model is D due to Assumption 9(*i*) and does not contain a boundary parameter by Assumption 9(*ii*). Assumption 9(*iv*) ensures that $C(\boldsymbol{\pi}_*)$ is a Cartesian product of two cones, and one of them is $\mathbb{R}^{r_\lambda}$, which follows from the interiority condition in Assumption 9(*ii*). This parameter space condition is popular for D-D models. For example, Examples 1, 2, and 3 satisfy this condition. The following corollary is obtained by adding Assumption 9 to the set of previous assumptions.

Corollary 2. *Given Assumptions 1 to 9,*

(i) $\mathcal{LR}_n^{(1)} \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\xi}} \in \Delta(\boldsymbol{\xi}_*)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})]^2 + (\mathbf{Z}^{(\boldsymbol{\lambda})})'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})^{-1}(\mathbf{Z}^{(\boldsymbol{\lambda})}) + \mathcal{H}_2$*, where for each* $\boldsymbol{s}_{\boldsymbol{\xi}} \in \Delta(\boldsymbol{\xi}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\xi} : \boldsymbol{\xi}_* + \mathbf{x} \in \mathrm{cl}\{C(\boldsymbol{\xi}_*)\}, \|\mathbf{x}\| = 1\}$, $\widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) := (-\widetilde{A}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}))^{-1/2}\widetilde{\mathcal{Z}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})$, $\widetilde{A}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) := A_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) - \mathbf{A}_*^{(\boldsymbol{\xi},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}})'(\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})^{-1}\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})$*, and* $\widetilde{\mathcal{Z}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) := \mathcal{Z}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) - \mathbf{A}_*^{(\boldsymbol{\xi},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}})(\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})^{-1}\mathbf{Z}^{(\boldsymbol{\lambda})}$*;*

(ii) if further H_0'' *holds,* $\mathcal{LR}_n \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\xi}} \in \Delta(\boldsymbol{\xi}_0)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})]^2$. □

Note that $\widetilde{\mathcal{Z}}^{(\boldsymbol{\xi})}(\cdot)$ is obtained by projecting $\mathcal{Z}^{(\boldsymbol{\xi})}(\cdot)$ on $\mathbf{Z}^{(\boldsymbol{\lambda})}$. This projection is needed because the QLR test statistic is constructed by minimizing the impact of the parameter estimation error that arises when estimating the unknown nuisance parameter $\boldsymbol{\lambda}_*$. Furthermore, Corollary 2(*ii*) follows by combining Theorem 7, (16), (17), and Corollary 2. Here, we also note that $\mathcal{H}_0 = (\mathbf{Z}^{(\boldsymbol{\lambda})})'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})^{-1}(\mathbf{Z}^{(\boldsymbol{\lambda})})$ by (16) and Assumption 9(*ii*).

3.2 Wald Test Statistic

Before defining Wald test statistic, we first examine the asymptotic null distribution of the distance between $\widehat{\boldsymbol{\mu}}_n$ and $\boldsymbol{\mu}_0$. Note that the distance between $\widehat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_*$ captured by $\widehat{h}_n(\cdot)$ cannot be used to test the null hypothesis because the inference on $\boldsymbol{\mu}_*$ is mixed with that of the other nuisance parameters $\boldsymbol{\omega}_*$ and $\boldsymbol{\tau}_*$. The distance $\widehat{h}_n(\cdot)$ needs to be broken into pieces corresponding to $\boldsymbol{\mu}$, $\boldsymbol{\omega}$, and $\boldsymbol{\tau}$, and this process is achieved by separating the set of directions into the sets for the directions of $\boldsymbol{\mu}$, $\boldsymbol{\omega}$, and $\boldsymbol{\tau}$. Specifically, for any $h\boldsymbol{d}$ and $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$, there are $h^{(\boldsymbol{\mu})}$, $h^{(\boldsymbol{\omega})}$, $h^{(\boldsymbol{\tau})}$, and $(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) \in \Delta(\boldsymbol{\mu}_*) \times \Delta(\boldsymbol{\omega}_*) \times \Delta(\boldsymbol{\tau}_*)$ such that $h\boldsymbol{d} = (h^{(\boldsymbol{\mu})}\boldsymbol{s}_{\boldsymbol{\mu}}', h^{(\boldsymbol{\omega})}\boldsymbol{s}_{\boldsymbol{\omega}}', h^{(\boldsymbol{\tau})}\boldsymbol{s}_{\boldsymbol{\tau}}')'$ if each parameter space of $\boldsymbol{\mu}$, $\boldsymbol{\omega}$, and $\boldsymbol{\tau}$ is approximated by a cone and the parameter space of $\boldsymbol{\theta}$ is also approximated by the Cartesian product of these cones as assumed in Assumptions 7 and 8, respectively. Therefore, the following equality holds:

$$\sup_{\boldsymbol{d}}\sup_{h} L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = \sup_{\{\boldsymbol{s}_{\boldsymbol{\mu}},\, \boldsymbol{s}_{\boldsymbol{\omega}},\, \boldsymbol{s}_{\boldsymbol{\tau}}\}}\ \sup_{\{h^{(\boldsymbol{\mu})},\, h^{(\boldsymbol{\omega})},\, h^{(\boldsymbol{\tau})}\}} L_n(\boldsymbol{\theta}_* + (h^{(\boldsymbol{\mu})}\boldsymbol{s}_{\boldsymbol{\mu}}', h^{(\boldsymbol{\omega})}\boldsymbol{s}_{\boldsymbol{\omega}}', h^{(\boldsymbol{\tau})}\boldsymbol{s}_{\boldsymbol{\tau}}')'), \tag{18}$$

and if we suppose that $(\widehat{\boldsymbol{d}}_n, \widehat{h}_n(\cdot))$ and $(\widehat{\boldsymbol{s}}_{\boldsymbol{\mu},n}', \widehat{\boldsymbol{s}}_{\boldsymbol{\omega},n}', \widehat{\boldsymbol{s}}_{\boldsymbol{\tau},n}', \widehat{h}_n^{(\boldsymbol{\mu})}(\cdot), \widehat{h}_n^{(\boldsymbol{\omega})}(\cdot), \widehat{h}_n^{(\boldsymbol{\tau})}(\cdot))'$ solve the LHS and RHS of (18), respectively, then it also follows that

$$\widehat{h}_n(\widehat{\boldsymbol{d}}_n)\widehat{\boldsymbol{d}}_n = (\widehat{h}_n^{(\boldsymbol{\mu})}(\widehat{\boldsymbol{s}}_{\boldsymbol{\mu},n}, \widehat{\boldsymbol{s}}_{\boldsymbol{\omega},n}, \widehat{\boldsymbol{s}}_{\boldsymbol{\tau},n})\widehat{\boldsymbol{s}}_{\boldsymbol{\mu},n}', \widehat{h}_n^{(\boldsymbol{\omega})}(\widehat{\boldsymbol{s}}_{\boldsymbol{\mu},n}, \widehat{\boldsymbol{s}}_{\boldsymbol{\omega},n}, \widehat{\boldsymbol{s}}_{\boldsymbol{\tau},n})\widehat{\boldsymbol{s}}_{\boldsymbol{\omega},n}', \widehat{h}_n^{(\boldsymbol{\tau})}(\widehat{\boldsymbol{s}}_{\boldsymbol{\mu},n}, \widehat{\boldsymbol{s}}_{\boldsymbol{\omega},n}, \widehat{\boldsymbol{s}}_{\boldsymbol{\tau},n})\widehat{\boldsymbol{s}}_{\boldsymbol{\tau},n}')',$$

and the Wald testing principle can be applied to the asymptotic behavior of $\widehat{h}_n^{(\boldsymbol{\mu})}(\widehat{\boldsymbol{s}}_{\boldsymbol{\mu},n}, \widehat{\boldsymbol{s}}_{\boldsymbol{\omega},n}, \widehat{\boldsymbol{s}}_{\boldsymbol{\tau},n})$.

For this purpose, we first examine the asymptotic distribution of $\widehat{\boldsymbol{h}}_n(\cdot) := (\widehat{h}_n^{(\boldsymbol{\mu})}(\cdot), \widehat{h}_n^{(\boldsymbol{\omega})}(\cdot), \widehat{h}_n^{(\boldsymbol{\tau})}(\cdot))'$. First, for each $(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) \in \Delta(\boldsymbol{\mu}_*) \times \Delta(\boldsymbol{\omega}_*) \times \Delta(\boldsymbol{\tau}_*)$, we let

$$\begin{bmatrix} \mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}) \\ \mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}) \\ \mathcal{G}^{(\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}}) \end{bmatrix} := \begin{bmatrix} \boldsymbol{\mathcal{G}}^{(\boldsymbol{\eta})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}) \\ \mathcal{G}^{(\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}}) \end{bmatrix} := \begin{bmatrix} \{-\mathbf{A}_*^{(\boldsymbol{\eta},\boldsymbol{\eta})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}})\}^{-1}\boldsymbol{\mathcal{Z}}^{(\boldsymbol{\eta})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}) \\ \{-A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}})\}^{-1}\mathcal{Z}^{(\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}}) \end{bmatrix},$$

where for each $(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}})$, $\boldsymbol{\mathcal{Z}}^{(\boldsymbol{\eta})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}) := (\mathcal{Z}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}), \mathcal{Z}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\omega}}))'$. Next, for each $(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}) \in \Delta(\boldsymbol{\mu}_*) \times$

$\Delta(\boldsymbol{\omega}_*)$, we also let

$$\left[\begin{array}{c} \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}) \\ \dot{\mathcal{G}}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\omega}}) \end{array}\right] := \left[\begin{array}{c} \{-A_*^{(\boldsymbol{\mu},\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}})\}^{-1}\mathcal{Z}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}) \\ \{-A_*^{(\boldsymbol{\omega},\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\omega}})\}^{-1}\mathcal{Z}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\omega}}) \end{array}\right].$$

These constitute the asymptotic behavior of $\widehat{\boldsymbol{h}}_n(\cdot)$ as established in

Lemma 4. *Given Assumptions 1 to 8,*

$$\sqrt{n}\widehat{\boldsymbol{h}}_n(\cdot) \Rightarrow \left[\begin{array}{c} \mathcal{G}^{(\boldsymbol{\mu})}(\cdot) \\ \mathcal{G}^{(\boldsymbol{\omega})}(\cdot) \\ 0 \end{array}\right] \times \mathbf{1}_{\{\min[\mathcal{G}^{(\boldsymbol{\mu})}(\cdot),\mathcal{G}^{(\boldsymbol{\omega})}(\cdot)]>0\}} + \left[\begin{array}{c} \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\cdot)] \times \mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\mu})}(\cdot)\geq 0>\mathcal{G}^{(\boldsymbol{\omega})}(\cdot)\}} \\ \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\omega})}(\cdot)] \times \mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\omega})}(\cdot)\geq 0>\mathcal{G}^{(\boldsymbol{\mu})}(\cdot)\}} \\ \max[0, \mathcal{G}^{(\boldsymbol{\tau})}(\cdot)] \end{array}\right]. \quad (19)$$

□

Lemma 4 can be heuristically described as follows. First, note that both $(\widehat{h}_n^{(\boldsymbol{\mu})}, \widehat{h}_n^{(\boldsymbol{\omega})})'$ and $\widehat{h}_n^{(\boldsymbol{\tau})}$ are initially defined on $\Delta(\boldsymbol{\mu}_*) \times \Delta(\boldsymbol{\omega}_*) \times \Delta(\boldsymbol{\tau}_*)$, but their asymptotic counterparts in the RHS of (19) are now separated into $\Delta(\boldsymbol{\mu}_*) \times \Delta(\boldsymbol{\omega}_*)$ and $\Delta(\boldsymbol{\tau}_*)$. This is mainly due to Assumption 7(*iv* and *v*). By supposing that $A_*^{(\boldsymbol{\eta},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\eta}}, \boldsymbol{d}_{\boldsymbol{\tau}}) = 0$, the maximization process in the RHS of (18) is asymptotically separated into two independent maximization procedures, and this results in the separated domains at the limit. Second, $\widehat{h}_n^{(\boldsymbol{\mu})}(\cdot)$ and $\widehat{h}_n^{(\boldsymbol{\omega})}(\cdot)$ cannot be less than zero. Thus, for each $(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}})$, one of following four different events arises: (*i*) $\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) > 0$ and $\widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) > 0$; (*ii*) $\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) > 0$, $\widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$; (*iii*) $\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$, $\widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) > 0$; and (*iv*) $\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$, $\widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$. These four different events are determined by the sign of $\boldsymbol{\mathcal{G}}^{(\boldsymbol{\eta})}(\cdot)$, and Lemma 4 distinguishes them using the separate indication functions. In particular, if $\widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$ and $\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) > 0$, estimating the nuisance parameter $\boldsymbol{\omega}_*$ does not affect the asymptotic distribution of $\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}})$, so that it follows that $\sqrt{n}\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) \Rightarrow \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}})$. Note that this limit is not influenced by the projecting procedure that minimizes the impact of the parameter estimation error, because $\widehat{\boldsymbol{\omega}}_n(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = \boldsymbol{\omega}_*$ by the fact that $\widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$. That is, the parameter estimation error for $\boldsymbol{\omega}_*$ does not arise. For the cases of (*iii*) and (*iv*), the same interpretation applies. Finally, if both $\boldsymbol{\omega}_*$ and $\boldsymbol{\mu}_*$ are interior elements, it is impossible to avoid the parameter estimation error for $\boldsymbol{\mu}_*$ and $\boldsymbol{\omega}_*$, so that the corresponding second term in the RHS of (19) can be virtually ignored.

We now define the Wald test statistic as

$$\mathcal{W}_n := \sup_{\boldsymbol{s}_{\boldsymbol{\mu}} \in \Delta(\boldsymbol{\mu}_0)} n\{\widetilde{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}})\}\{\widehat{W}_n(\boldsymbol{s}_{\boldsymbol{\mu}})\}\{\widetilde{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}})\},$$

where $\widetilde{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu})$ is such that for each $\boldsymbol{s_\mu} \in \Delta(\boldsymbol{\mu}_0)$,

$$L_n(\boldsymbol{\mu}_0 + \widetilde{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu})\boldsymbol{s_\mu}, \widetilde{\boldsymbol{\omega}}_n(\boldsymbol{s_\mu}), \widetilde{\boldsymbol{\tau}}_n(\boldsymbol{s_\mu})) = \sup_{\{h^{(\boldsymbol{\mu})}, \boldsymbol{\omega}, \boldsymbol{\tau}\}} L_n(\boldsymbol{\mu}_0 + h^{(\boldsymbol{\mu})}\boldsymbol{s_\mu}, \boldsymbol{\omega}, \boldsymbol{\tau}),$$

and $\widehat{W}_n(\cdot)$ is a weight function, which can estimate a non-random positive function $W_*(\cdot)$ uniformly on $\Delta(\boldsymbol{\mu}_0)$. Note that $\widehat{h}_n^{(\boldsymbol{\mu})}(\cdot)$ is equivalent to $\widetilde{h}_n^{(\boldsymbol{\mu})}(\cdot)$ under the null from the fact that $\sup_{h^{(\boldsymbol{\mu})}, \boldsymbol{\omega}, \boldsymbol{\tau}} L_n(\boldsymbol{\mu}_0 + h^{(\boldsymbol{\mu})}\boldsymbol{s_\mu}, \boldsymbol{\omega}, \boldsymbol{\tau})$ is equivalent to $\sup_{\{\boldsymbol{s_\omega}, \boldsymbol{s_\tau}\}} \sup_{\{h^{(\boldsymbol{\mu})}, h^{(\boldsymbol{\omega})}, h^{(\boldsymbol{\tau})}\}} L_n(\boldsymbol{\mu}_0 + h^{(\boldsymbol{\mu})}\boldsymbol{s_\mu}, \boldsymbol{\omega}_* + h^{(\boldsymbol{\omega})}\boldsymbol{s_\omega}, \boldsymbol{\tau}_* + h^{(\boldsymbol{\tau})}\boldsymbol{s_\tau})$. On the other hand, this equivalency does not hold under the alternative. As the weight function is an important component of the Wald test statistic, we formally state its condition.

Assumption 10 (Weight Function I). *$\widehat{W}_n(\cdot)$, which is strictly positive uniformly on $\Delta(\boldsymbol{\mu}_0)$ and n, converges to $W_*(\cdot)$ a.s.–$\mathbb{P}$ uniformly on $\Delta(\boldsymbol{\mu}_0)$ as n tends to infinity, where $W_*(\cdot)$ is strictly positive and bounded from above uniformly on $\Delta(\boldsymbol{\mu}_0)$. That is, $\sup_{\boldsymbol{s_\mu}} |\widehat{W}_n(\boldsymbol{s_\mu}) - W_*(\boldsymbol{s_\mu})| \to 0$ a.s.–$\mathbb{P}$.* □

Under the Wald testing principle, the weight function $W_*(\cdot)$ is typically the asymptotic variance function of $\sqrt{n}\widetilde{h}_n^{(\boldsymbol{\mu})}(\cdot)$. If the parameter of interest is on the boundary, the weight function needs to be carefully chosen because the asymptotic variance function of $\sqrt{n}\widetilde{h}_n^{(\boldsymbol{\mu})}(\cdot)$ is different from the interior parameter case.

The asymptotic null distribution of the Wald test statistic is obtained as follows.

Theorem 8. *Given Assumptions 1 to 8, 10, and H_0,*

$$\begin{aligned}\mathcal{W}_n \Rightarrow \sup_{\boldsymbol{s_\mu} \in \Delta(\boldsymbol{\mu}_0)} & \mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu}))W_*(\boldsymbol{s_\mu})\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu}))\mathbf{1}_{\{\min[\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu})), \mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu}))]>0\}} \\ & + \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu})]W_*(\boldsymbol{s_\mu}) \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu})]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu}, \boldsymbol{s_\omega}\widetilde{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu}))\geq 0 > \mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu}, \widetilde{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu}))\}},\end{aligned}$$

where for each $\boldsymbol{s_\mu}$, $\bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu})$ and $\widetilde{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu})$ are such that

$$\begin{aligned}&\sup_{\boldsymbol{s_\omega} \in \Delta(\omega_*)} \boldsymbol{\mathcal{G}}^{(\boldsymbol{\eta})}(\boldsymbol{s_\mu}, \boldsymbol{s_\omega})'(-\mathbf{A}_*^{(\boldsymbol{\eta}, \boldsymbol{\eta})}(\boldsymbol{s_\mu}, \boldsymbol{s_\omega}))\boldsymbol{\mathcal{G}}^{(\boldsymbol{\eta})}(\boldsymbol{s_\mu}, \boldsymbol{s_\omega})\mathbf{1}_{\{\min[\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu}, \boldsymbol{s_\omega}), \mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu}, \boldsymbol{s_\omega})]>0\}} \\ &= \boldsymbol{\mathcal{G}}^{(\boldsymbol{\eta})}(\boldsymbol{s_\mu}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu}))'(-\mathbf{A}_*^{(\boldsymbol{\eta}, \boldsymbol{\eta})}(\boldsymbol{s_\mu}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu})))\boldsymbol{\mathcal{G}}^{(\boldsymbol{\eta})}(\boldsymbol{s_\mu}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu}))\mathbf{1}_{\{\min[\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu})), \mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu}))]>0\}}\end{aligned} \tag{20}$$

and

$$\begin{aligned}&\sup_{\boldsymbol{s_\omega} \in \Delta(\boldsymbol{\omega}_*)} \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu})](-A_*^{(\boldsymbol{\mu}, \boldsymbol{\mu})}(\boldsymbol{s_\mu})) \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu})]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu}, \boldsymbol{s_\omega})\geq 0 > \mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu}, \boldsymbol{s_\omega})\}} \\ &= \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu})](-A_*^{(\boldsymbol{\mu}, \boldsymbol{\mu})}(\boldsymbol{s_\mu})) \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu})]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu}, \widetilde{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu}))\geq 0 > \mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu}, \widetilde{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s_\mu}))\}},\end{aligned} \tag{21}$$

respectively. $\square$

Theorem 8 is proved by noting that for each $\boldsymbol{s}_{\boldsymbol{\mu}}$, maximizing $2\{L_n(\boldsymbol{\mu}_0 + h^{(\boldsymbol{\mu})}\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{\omega}, \boldsymbol{\tau}) - L_n(\boldsymbol{\mu}_0, \boldsymbol{\omega}_*, \boldsymbol{\tau}_*)\}$ with respect to $h^{(\boldsymbol{\mu})}$, $\boldsymbol{\omega}$, and $\boldsymbol{\tau}$ is equivalent to

$$\sup_{\{\boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}\}} \sup_{\{h^{(\boldsymbol{\mu})}, h^{(\boldsymbol{\omega})}, h^{(\boldsymbol{\tau})}\}} 2\{L_n(\boldsymbol{\mu}_0 + h^{(\boldsymbol{\mu})}\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{\omega}_* + h^{(\boldsymbol{\omega})}\boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{\tau}_* + h^{(\boldsymbol{\tau})}\boldsymbol{s}_{\boldsymbol{\tau}}) - L_n(\boldsymbol{\mu}_0, \boldsymbol{\omega}_*, \boldsymbol{\tau}_*)\}. \tag{22}$$

For each $(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}})$, (22) is approximated by a quadratic function of $(h^{(\boldsymbol{\mu})}, h^{(\boldsymbol{\omega})}, h^{(\boldsymbol{\tau})})$, and the signs of $\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}})$ and $\widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}})$ result in different approximations as pointed out by Lemma 4. As a result of this, the asymptotic null distribution of the Wald test statistic is determined by the signs. Theorem 8 derives this using the indication functions of the factors that determine the signs at the limit. Specific functional forms of the signs are given in (20) and (21). Here, (20) considers when both $\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}})$ and $\widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}})$ are strictly greater than zero at the limit, whereas (21) considers when $\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) > 0$ but $\widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$ at the limit. If the asymptotic sign of $\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{s}_{\boldsymbol{\tau}})$ is zero, the Wald test statistic converges to zero in probability.

Theorem 8 is further simplified if the benchmark condition in Assumption 9 is also imposed.

Corollary 3. *Given Assumptions 1 to 9 and H_0'', if* $\sup_{\boldsymbol{s}_{\boldsymbol{\xi}} \in \Delta(\boldsymbol{\xi}_0)} |\widehat{W}_n(\boldsymbol{s}_{\boldsymbol{\xi}}) + \widetilde{A}_*^{(\boldsymbol{\xi}, \boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})| \to 0$ *a.s.–*$\mathbb{P}$*,* $\mathcal{W}_n \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\xi}} \in \Delta(\boldsymbol{\xi}_0)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})]^2$. $\square$

Note that the weak limit of the Wald test statistic in Corollary 3 is identical to that given in Corollary 2.

3.3 Lagrange Multiplier Test Statistic

The LM test statistic can be appropriately defined for D-D models to test whether the slopes of models are distributed around zero under the null. We let the LM test statistic be

$$\mathcal{LM}_n := \sup_{(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}) \in \Delta(\boldsymbol{\mu}_0) \times \Delta(\ddot{\boldsymbol{\omega}}_n)} n\widetilde{W}_n(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}) \max\left[0, \frac{-DL_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\mu}})}{\widetilde{D}^2 L_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}})}\right]^2,$$

where for each $(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}})$, $\Delta(\ddot{\boldsymbol{\omega}}_n) := \{\mathbf{x} \in \mathbb{R}^{r_\omega} : \mathbf{x} + \ddot{\boldsymbol{\omega}}_n \in \mathrm{cl}\{C(\ddot{\boldsymbol{\omega}}_n)\}, \|\mathbf{x}\| = 1\}$,

$$\widetilde{D}^2 L_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}) := D^2 L_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\mu}}) - DL_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\mu}}; \boldsymbol{s}_{\boldsymbol{\omega}})(D^2 L_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\omega}}))^{-1} DL_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\omega}}; \boldsymbol{s}_{\boldsymbol{\mu}}),$$

$$DL_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\mu}}; \boldsymbol{s}_{\boldsymbol{\omega}}) := \lim_{h \downarrow 0} h^{-1}\{DL_n(\boldsymbol{\mu}_0, \ddot{\boldsymbol{\omega}}_n + h\boldsymbol{s}_{\boldsymbol{\omega}}, \ddot{\boldsymbol{\tau}}_n; \boldsymbol{s}_{\boldsymbol{\mu}}) - DL_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\mu}})\},$$

$$DL_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\omega}}; \boldsymbol{s}_{\boldsymbol{\mu}}) := \lim_{h \downarrow 0} h^{-1}\{DL_n(\boldsymbol{\mu}_0 + h\boldsymbol{s}_{\boldsymbol{\mu}}, \ddot{\boldsymbol{\omega}}_n, \ddot{\boldsymbol{\tau}}_n; \boldsymbol{s}_{\boldsymbol{\omega}}) - DL_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\omega}})\},$$

and $\widetilde{W}_n(\cdot)$ is a weight function that satisfies

Assumption 11 (Weight Function II)**.** *(i) The unknown nuisance parameter $\boldsymbol{\omega}_*$ is interior to $\boldsymbol{\Omega}$;*

(ii) A weight function $\widetilde{W}_n(\cdot)$, which is strictly positive uniformly on $\Delta(\boldsymbol{\mu}_0)\times\Delta(\ddot{\boldsymbol{\omega}}_n)$ and n, converges to $\widetilde{W}(\cdot)$ a.s.$-\mathbb{P}$ uniformly on $\Delta(\boldsymbol{\mu}_0)\times\Delta(\ddot{\boldsymbol{\omega}}_n)$ as n tends to infinity, where $\widetilde{W}(\cdot)$ is strictly positive and bounded from above uniformly on $\Delta(\boldsymbol{\mu}_0)\times\Delta(\boldsymbol{\omega}_*)$. That is, $\sup_{(\boldsymbol{s_\mu},\boldsymbol{s_\omega})\in\Delta(\boldsymbol{\mu}_0)\times\Delta(\ddot{\boldsymbol{\omega}}_n)}|\widetilde{W}_n(\boldsymbol{s_\mu},\boldsymbol{s_\omega})-\widetilde{W}(\boldsymbol{s_\mu},\boldsymbol{s_\omega})|\to 0$ a.s.–$\mathbb{P}$. □

There are several remarks relevant to the definition of the LM test statistic. First, the LM test statistic has a structure that yields the same asymptotic null distribution as those of the QLR and Wald tests. That is, the LM test statistic is defined using the first and second-order directional derivatives of $L_n(\boldsymbol{\mu}_0+h^{(\boldsymbol{\mu})}\boldsymbol{s_\mu},\ddot{\boldsymbol{\omega}}_n+h^{(\boldsymbol{\omega})}\boldsymbol{s_\omega},\ddot{\boldsymbol{\tau}}_n+h^{(\boldsymbol{\tau})}\boldsymbol{s_\tau})$ with respect to $\boldsymbol{s_\mu}$ and $\boldsymbol{s_\omega}$, where $({\boldsymbol{\mu}_0}',\ddot{\boldsymbol{\omega}}_n',\ddot{\boldsymbol{\tau}}_n')'=\ddot{\boldsymbol{\theta}}_n$, and 'max' operator is used to capture the non-negativity of $\sqrt{n}\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega},\boldsymbol{s_\tau})$. Second, the LM test statistic is asymptotically the supremum of a squared random score function with respect to $(\boldsymbol{s_\mu},\boldsymbol{s_\omega})$, if $\widetilde{W}_n(\cdot)$ is asymptotically equivalent to $-n^{-1}\widetilde{D}^2L_n(\ddot{\boldsymbol{\theta}}_n;\cdot)$. Third, $\widetilde{W}_n(\cdot)$ is defined on $\Delta(\boldsymbol{\mu}_0)\times\Delta(\ddot{\boldsymbol{\omega}}_n)$, and $\boldsymbol{\omega}_*$ is an interior element of $\boldsymbol{\Omega}$. Note that the domain $\Delta(\ddot{\boldsymbol{\omega}}_n)$ estimates $\Delta(\boldsymbol{\omega}_*)$. The interiority condition lets $\Delta(\ddot{\boldsymbol{\omega}}_n)$ converge to $\Delta(\boldsymbol{\omega}_*)$ asymptotically. If $\boldsymbol{\omega}_*$ is on the boundary, $\Delta(\ddot{\boldsymbol{\omega}}_n)$ can be different from $\Delta(\boldsymbol{\omega}_*)$, and the asymptotic null distribution of the LM test statistic is affected by this. Assumption 11(*i*) precludes this possibility.

The asymptotic null distribution of the LM test statistic is straightforwardly obtained by the CMT.

Theorem 9. *Given Assumptions 1 to 8 11, and H_0,*

$$\mathcal{LM}_n\Rightarrow\sup_{(\boldsymbol{s_\mu},\boldsymbol{s_\omega})\in\Delta(\boldsymbol{\mu}_0)\times\Delta(\boldsymbol{\omega}_*)}\max[0,\ddot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega};\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})]\widetilde{W}(\boldsymbol{s_\mu},\boldsymbol{s_\omega})\max[0,\ddot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega};\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})],$$

where for each $(\boldsymbol{s_\mu},\boldsymbol{s_\omega})\in\Delta(\boldsymbol{\mu}_*)\times\Delta(\boldsymbol{\omega}_*)$, $\ddot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega};\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}):=\{-\widetilde{A}_*^{(\boldsymbol{\mu},\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega})\}^{-1}\ddot{\mathcal{Z}}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu};\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})$, $\ddot{\mathcal{Z}}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu};\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}):=\mathcal{Z}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu})-(-A_*^{(\boldsymbol{\mu},\boldsymbol{\omega})}(\boldsymbol{s_\mu},\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}))(-A_*^{(\boldsymbol{\omega},\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}))^{-1}\mathcal{Z}^{(\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})$, $\widetilde{A}_*^{(\boldsymbol{\mu},\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega}):=A_*^{(\boldsymbol{\mu},\boldsymbol{\mu})}(\boldsymbol{s_\mu})-A_*^{(\boldsymbol{\mu},\boldsymbol{\omega})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega})(A_*^{(\boldsymbol{\omega},\boldsymbol{\omega})}(\boldsymbol{s_\omega}))^{-1}A_*^{(\boldsymbol{\omega},\boldsymbol{\mu})}(\boldsymbol{s_\omega},\boldsymbol{s_\mu})$, and we let $\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}$ be such that $\max[0,\mathcal{Y}^{(\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})]^2=\sup_{\boldsymbol{s_\omega}\in\Delta(\boldsymbol{\omega}_*)}\max[0,\mathcal{Y}^{(\boldsymbol{\omega})}(\boldsymbol{s_\omega})]^2$. □

Note that the asymptotic null distribution of the LM test statistic depends upon that of $\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}$. This is mainly because the weak limit of $n^{-1/2}DL_n(\ddot{\boldsymbol{\theta}}_n;\,\cdot\,)$ is captured by $\ddot{\mathcal{Z}}^{(\boldsymbol{\mu})}(\,\cdot\,;\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})$, and $\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}$ has its own asymptotic distribution. We accommodate this influence by approximating $n^{-1/2}DL_n(\ddot{\boldsymbol{\theta}}_n;\,\cdot\,)$ via Taylor's expansion.

Theorem 9 is further simplified if the benchmark model assumption is also imposed.

Corollary 4. *Given Assumptions 1 to 9 and H_0'', if $\sup_{(\boldsymbol{s_\xi},\boldsymbol{s_\lambda})\in\Delta(\boldsymbol{\xi}_0)\times\Delta(\ddot{\boldsymbol{\lambda}}_n)}|\widetilde{W}_n(\boldsymbol{s_\xi},\boldsymbol{s_\lambda})+\widetilde{A}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s_\xi})|\to 0$ a.s.–$\mathbb{P}$, $\mathcal{LM}_n\Rightarrow\sup_{\boldsymbol{s_\xi}\in\Delta(\boldsymbol{\xi}_0)}\max[0,\widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s_\xi})]^2$.* □

Therefore, the QLR, Wald, and LM test statistics are asymptotically equivalent under the null for the benchmark model.

3.4 Examples

3.4.1 Example 1 (continued)

The main interests of King and Shively (1993) can be analyzed by the three test statistics. First, we reconcile the parameters in the model with the parameters defined in this section. Specifically, we let $\boldsymbol{\xi} = (\theta_1, \theta_2)'$, $\boldsymbol{\lambda} = \sigma^2$, $\boldsymbol{\tau} = \boldsymbol{\gamma}$, and $\boldsymbol{\pi} = (\sigma^2, \theta_1, \theta_2)'$. Then, for each $\boldsymbol{d}$ and $\widetilde{\boldsymbol{d}}$,

$$\mathbf{B}_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) = \begin{bmatrix} \mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}}) & \mathbf{0}' \\ \mathbf{0} & \frac{1}{\sigma_*^2}\boldsymbol{d}_{\boldsymbol{\gamma}}' E[\mathbf{Q}_t\mathbf{Q}_t']\widetilde{\boldsymbol{d}}_{\boldsymbol{\gamma}} \end{bmatrix},$$

and

$$\begin{aligned} &\mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}}) \\ &= \begin{bmatrix} \frac{1}{2\sigma_*^4} d_{\sigma^2}\widetilde{d}_{\sigma^2} & \frac{1}{2\sigma_*^2} d_{\sigma^2} h(\widetilde{d}_1, \widetilde{d}_2) E[W_t^2] \\ \frac{1}{2\sigma_*^2}\widetilde{d}_{\sigma^2} h(d_1, d_2) E[W_t^2] & h(d_1, d_2)h(\widetilde{d}_1, \widetilde{d}_2)\left[\frac{1}{2}E[W_t^4] + k(d_2/d_1, \widetilde{d}_2/\widetilde{d}_1)E[W_t^2]^2\right] \end{bmatrix}, \end{aligned}$$

where for each (d_1, d_2) and $(\widetilde{d}_1, \widetilde{d}_2)$,

$$h(d_1, d_2) := \frac{(d_1^2 + d_2^2)^{1/2}}{1 - m(d_2/d_1)^2}, \quad \text{and} \quad k(d_2/d_1, \widetilde{d}_2/\widetilde{d}_1) := \frac{m(d_2/d_1)m(\widetilde{d}_2/\widetilde{d}_1)}{1 - m(d_2/d_1)m(\widetilde{d}_2/\widetilde{d}_1)}.$$

Because of the information matrix equality and the fact that $\mathbf{B}_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$ is block diagonal, the asymptotic null distribution associated with each block matrix can be separately treated. Furthermore, $\sigma_*^{-2}\boldsymbol{d}_{\boldsymbol{\gamma}}' E[\mathbf{Q}_t\mathbf{Q}_t']\widetilde{\boldsymbol{d}}_{\boldsymbol{\gamma}}$ is associated only with $\boldsymbol{\gamma}$, so that it can be ignored, when deriving the asymptotic null distributions of the test statistics. We further note that $\boldsymbol{\iota}_3'\mathbf{B}_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}})\boldsymbol{\iota}_3 = B_*^{(1)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) + B_*^{(2)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) + B_*^{(3)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$, where each $B_*^{(i)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$ $(i = 1, 2, 3)$ is the covariance constituting the independent Gaussian stochastic processes that we have already derived above.

The asymptotic null distributions of the test statistics are more easily obtained by the special features of $\mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}})$: first, let

$$\widetilde{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}}) := \begin{bmatrix} \ddot{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}}) & \mathbf{0} \\ \mathbf{0}' & q(d_1, d_2, \widetilde{d}_1, \widetilde{d}_2)E[W_t^2]^2 \end{bmatrix},$$

where

$$\ddot{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}}) := \frac{1}{2}\left[\begin{array}{cc} h(d_1,d_2)h(\widetilde{d}_1,\widetilde{d}_2)E[W_t^4] & \frac{1}{\sigma_*^2}\widetilde{d}_{\sigma^2}h(d_1,d_2)E[W_t^2] \\ \frac{1}{\sigma_*^2}d_{\sigma^2}h(\widetilde{d}_1,\widetilde{d}_2)E[W_t^2] & \frac{1}{\sigma_*^4}d_{\sigma^2}\widetilde{d}_{\sigma^2} \end{array}\right],$$

and

$$q(d_1,d_2,\widetilde{d}_1,\widetilde{d}_2) := h(d_1,d_2)h(\widetilde{d}_1,\widetilde{d}_2)k(d_2/d_1,\widetilde{d}_2/\widetilde{d}_1),$$

and note that $\boldsymbol{\iota}_3{}'\widetilde{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}})\boldsymbol{\iota}_3 = B_*^{(2)}(\boldsymbol{d},\widetilde{\boldsymbol{d}}) + B_*^{(3)}(\boldsymbol{d},\widetilde{\boldsymbol{d}})$ and $\boldsymbol{\iota}_2{}'\ddot{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}})\boldsymbol{\iota}_2 = B_*^{(2)}(\boldsymbol{d},\widetilde{\boldsymbol{d}})$. Here, the Gaussian stochastic process associated with $\ddot{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}$ $(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}})$ is independent of that associated with $q(d_1,d_2,\widetilde{d}_1,\widetilde{d}_2)E[W_t^2]^2$ because $\widetilde{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}})$ is block diagonal. Furthermore, $\ddot{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}$ $(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}})$ is bilinear with respect to $h(d_1,d_2)$ and d_{σ^2}. By these facts, it becomes straightforward to derive the asymptotic null distributions of the test statistics. First, the asymptotic null distribution of the QLR test statistic is obtained as $\mathcal{LR}_n^{(1)} \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\pi}}\in\Delta(\boldsymbol{\pi}_*)}\max[0,\mathcal{Y}^{(\boldsymbol{\pi})}(\boldsymbol{s}_{\boldsymbol{\pi}})]^2 + \mathcal{H}_2$ by Theorem 5, where $\mathcal{Y}^{(\boldsymbol{\pi})}(\cdot)$ is a standard Gaussian stochastic process with

$$E[\mathcal{Y}^{(\boldsymbol{\pi})}(\boldsymbol{s}_{\boldsymbol{\pi}})\mathcal{Y}^{(\boldsymbol{\pi})}(\widetilde{\boldsymbol{s}}_{\boldsymbol{\pi}})] = \frac{\boldsymbol{\iota}_3{}'\widetilde{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{s}_{\boldsymbol{\pi}},\widetilde{\boldsymbol{s}}_{\boldsymbol{\pi}})\boldsymbol{\iota}_3}{\{\boldsymbol{\iota}_3{}'\widetilde{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{s}_{\boldsymbol{\pi}},\boldsymbol{s}_{\boldsymbol{\pi}})\boldsymbol{\iota}_3\}^{1/2}\{\boldsymbol{\iota}_3{}'\widetilde{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\widetilde{\boldsymbol{s}}_{\boldsymbol{\pi}},\widetilde{\boldsymbol{s}}_{\boldsymbol{\pi}})\boldsymbol{\iota}_3\}^{1/2}},$$

and $\mathcal{H}_2$ is a chi-squared random variable with degrees of freedom $k+1$. Second, $\mathcal{LR}_n^{(2)} \Rightarrow \sup_{s_{\sigma^2}\in\{-1,1\}}$ $\max[0,\mathcal{Y}^{(\sigma^2)}(s_{\sigma^2})]^2 + \mathcal{H}_2$ by Theorem 6, where $\mathcal{Y}^{(\sigma^2)}(\cdot)$ is a standard Gaussian stochastic process with $E[\mathcal{Y}^{(\sigma^2)}(s_{\sigma^2})\mathcal{Y}^{(\sigma^2)}(\widetilde{s}_{\sigma^2})] = 1$. This implies that it is free of the direction s_{σ^2}, so that $\sup_{s_{\sigma^2}\in\{-1,1\}}\max[0,$ $\mathcal{Y}^{(\sigma^2)}(s_{\sigma^2})]^2$ is a chi-squared random variable with one degree of freedom by the proof of Corollary 1(*vii*). Thus, $\mathcal{LR}_n \Rightarrow \sup_{s_2/s_1\in[\underline{c},\bar{c}]}\max[0,\widetilde{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1,s_2)]^2$ by Corollary 2(*ii*), where $\widetilde{\mathcal{Y}}^{(\boldsymbol{\theta})}$ is a standard Gaussian stochastic process with covariance structure

$$\frac{c(s_2/s_1,\widetilde{s}_2/\widetilde{s}_1)}{\{c(s_2/s_1,s_2/s_1)\}^{1/2}\{c(\widetilde{s}_2/\widetilde{s}_1,\widetilde{s}_2/\widetilde{s}_1)\}^{1/2}}$$

and for each $(s_2/s_1,\widetilde{s}_2/\widetilde{s}_1)$, $c(s_2/s_1,\widetilde{s}_2/\widetilde{s}_1) := \frac{1}{2}\mathrm{var}(W_t^2) + k(s_2/s_1,\widetilde{s}_2/\widetilde{s}_1)E[W_t^2]^2$. This structure is homogenous of degree zero with respect to s_1 and s_2, so that $\widetilde{\mathcal{Y}}^{(\boldsymbol{\theta})}(\cdot)$ can be equivalently stated as a function of s_2/s_1.

Second, we apply the Wald test statistic to this model. By the requirement of Corollary 3, we let the weight function be

$$\widehat{W}_n(\widetilde{s}_2/\widetilde{s}_1,s_2/s_1) := \frac{1}{(1-m(\widetilde{s}_2/\widetilde{s}_1)^2)(1-m(s_2/s_1)^2)}\left[\frac{\widehat{\mathrm{var}}_n(W_t^2)}{2} + k(\widetilde{s}_2/\widetilde{s}_1,s_2/s_1)\widehat{E}_n[W_t^2]^2\right],$$

where $\widehat{E}_n[W_t^2] := n^{-1}\sum_{t=1}^n W_t^2$ and $\widehat{\mathrm{var}}_n[W_t^2] := n^{-1}\sum_{t=1}^n W_t^4 - (n^{-1}\sum_{t=1}^n W_t^2)^2$. This statistic satisfies Assumption 10, the Wald test statistic is accordingly defined as

$$\mathcal{W}_n := n\{\widetilde{h}_n^{(\boldsymbol{\theta})}(s_2/s_1)\}\widehat{W}_n(s_2/s_1, s_2/s_1)\{\widetilde{h}_n^{(\boldsymbol{\theta})}(s_2/s_1)\},$$

and $\mathcal{W}_n \Rightarrow \sup_{s_2/s_1\in[\underline{c},\bar{c}]}\max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1,s_2)]^2$ under the null by Corollary 3, where $\widetilde{h}_n^{(\boldsymbol{\theta})}(s_2/s_1)$ is such that

$$L_n(\widetilde{\gamma}_n, \widetilde{\sigma}_n^2, \widetilde{h}_n^{(\boldsymbol{\theta})}(s_2/s_1)s_1, \widetilde{h}_n^{(\boldsymbol{\theta})}(s_2/s_1)s_2) = \sup_{(h^{(\boldsymbol{\theta})},\boldsymbol{\gamma},\sigma^2)} L_n(\boldsymbol{\gamma}, \sigma^2, h^{(\boldsymbol{\theta})}s_1, h^{(\boldsymbol{\theta})}s_2)$$

and $s_1^2 + s_2^2 = 1$.

Finally, we apply the LM test statistic to this model. Following the definition of the LM test statistic, we let

$$\mathcal{LM}_n := \sup_{s_2/s_1\in[\underline{c},\bar{c}]} n\left\{\frac{\max\left[0, DL_n(\ddot{\gamma}_n, \ddot{\sigma}_n^2, \mathbf{0}; s_1, s_2)\right]}{-\widetilde{D}^2L_n(\ddot{\gamma}_n, \ddot{\sigma}_n^2, \mathbf{0}; s_1, s_2, s_{\sigma^2})}\right\}^2 \widehat{W}_n(s_2/s_1, s_2/s_1),$$

where

$$DL_n(\ddot{\gamma}_n, \ddot{\sigma}_n^2, \mathbf{0}; s_1, s_2) = \{2\ddot{\sigma}_n^2\}^{-1}\{\mathbf{U}^n(\ddot{\gamma}_n)'\boldsymbol{\Omega}^n\left(m(s_2/s_1)\right)\mathbf{U}^n(\ddot{\gamma}_n) - \ddot{\sigma}_n^2\mathrm{tr}[\boldsymbol{\Omega}^n\left(m(s_2/s_1)\right)]\},$$

$$\begin{aligned}\widetilde{D}^2L_n(\ddot{\gamma}_n, \ddot{\sigma}_n^2, \mathbf{0}; s_1, s_2, s_{\sigma^2}) :=& \frac{1}{2}\left\{\mathrm{tr}(\boldsymbol{\Omega}^n(m(s_2/s_1)^2)) - \frac{2}{\ddot{\sigma}_n^2}\mathbf{U}^n(\ddot{\gamma}_n)'\boldsymbol{\Omega}^n(m(s_2/s_1)^2)\mathbf{U}^n(\ddot{\gamma}_n)\right\}\\ &- \mathbf{U}^n(\ddot{\gamma}_n)'\boldsymbol{\Omega}^n\left(m(s_2/s_1)\right)\mathbf{U}^n(\ddot{\gamma}_n)\left[\frac{n}{2} - \frac{1}{\ddot{\sigma}_n^2}\mathbf{U}^n(\ddot{\gamma}_n)'\mathbf{U}^n(\ddot{\gamma}_n)\right]^{-1}\\ &\times \mathbf{U}^n(\ddot{\gamma}_n)'\boldsymbol{\Omega}^n\left(m(s_2/s_1)\right)\mathbf{U}^n(\ddot{\gamma}_n),\end{aligned}$$

$(\ddot{\gamma}_n, \ddot{\sigma}_n^2)$ is such that $L_n(\ddot{\gamma}_n, \ddot{\sigma}_n^2, \mathbf{0}) = \sup_{(\boldsymbol{\gamma},\sigma^2)} L_n(\boldsymbol{\gamma}, \sigma^2, \mathbf{0})$, and the same weight matrix is used as for the Wald test statistic. Here, $\widetilde{D}^2L_n(\ddot{\gamma}_n, \ddot{\sigma}_n^2, \mathbf{0}; s_1, s_2, s_{\sigma^2})$ is virtually indexed only by (s_1, s_2) because s_{σ^2} disappears by its construction. Corollary 4 now implies that $\mathcal{LM}_n \Rightarrow \sup_{s_2/s_1\in[\underline{c},\bar{c}]}\max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1,s_2)]^2$, so that all three test statistics are asymptotically equivalent.

The distribution of the Gaussian stochastic process $\widetilde{\mathcal{Y}}^{(\boldsymbol{\theta})}(\cdot)$ can be uncovered by several simulation methods. First, the Monte Carlo method proposed by Dufour (2006) can be used as the model is correctly specified. Although the null likelihood is not defined, the directional limits obtained under the null can be used to form the QLR test. Second, Hansen's (1996b) weighted bootstrap can also be used to estimate the asymptotic p-values. Even when models are misspecified, the weighted bootstrap is applicable. Finally, we note that the covariance structure of $\widetilde{\mathcal{Y}}^{(\boldsymbol{\theta})}(\cdot)$ is the same as that of

$$\ddot{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2) := \frac{1}{c(s_2/s_1, s_2/s_1)^{1/2}} \left[\left\{ \frac{\mathrm{var}(W_t^2)}{2} \right\}^{1/2} Z_0 + E[W_t^2] \sum_{j=1}^{\infty} m(s_2/s_1)^j Z_j \right],$$

where $Z_j \sim$ IID $N(0,1)$. Due to this IID condition, it is not hard to simulate $\ddot{\mathcal{Y}}^{(\boldsymbol{\theta})}(\cdot)$ by a simulation method. When simulating $\ddot{\mathcal{Y}}^{(\boldsymbol{\theta})}(\cdot)$, there are a couple cautions. First, $\mathrm{var}(W_t^2)$ and $E[W_t^2]$ are unknown, and we should instead use their consistent estimators. Second, the running index j in the definition of $\ddot{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2)$ must be truncated at a moderately large level so that it does not significantly affect the null distribution.

We implement Monte Carlo simulations by the last simulation method. The DGP for $Y_t = U_t \sim$ IID $N(0,1)$ and $W_t \sim$ IID $N(0,1)$, which is independent of U_t. We assume that the parameters other than α_*, σ_*^2, θ_{1*}, and θ_{2*} are known and also let $\underline{c} = 0.5$, $\bar{c} = 1.5$. The total number of replications is 2,000, and the sample size is 500. Figure 1 shows the Q-Q plot between $\sup_{(s_1,s_2)} \max[0, \ddot{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2)]^2$ and $\sup_{(s_1,s_2)} \max[0, \widehat{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2)]^2$, where for each (s_1, s_2),

$$\widehat{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2) := \frac{1}{\widehat{c}_n(s_2/s_1, s_2/s_1)^{1/2}} \left[\left\{ \frac{\widehat{\mathrm{var}}_n(W_t^2)}{2} \right\}^{1/2} Z_0 + \widehat{E}_n[W_t^2] \sum_{j=1}^{150} m(s_2/s_1)^j Z_j \right],$$

and $\widehat{\mathrm{var}}_n(W_t^2)$ and $\widehat{E}_n[W_t^2]$ are the method of moments estimators for $\mathrm{var}(W_t^2)$ and $E[W_t^2]$, respectively. Note that the Q-Q line in Figure 1 is almost identical to the 45-degree line, which implies that estimating $\mathrm{var}(W_t^2)$ and $E[W_t^2]$ does not modify the asymptotic distribution. Figure 2 shows the empirical distributions of the QLR test statistic for various sample sizes and the asymptotic null distribution obtained by the simulation method. The empirical distribution of the QLR test statistic approaches the asymptotic distribution as n increases. $\square$

3.4.2 Example 2 (continued)

The efficient production hypothesis can be tested by the QLR, Wald, and LM test statistics. For this examination, we let $\boldsymbol{\xi} = (\mu, \sigma)'$, $\boldsymbol{\lambda} = \boldsymbol{\phi} = \boldsymbol{\beta}$, $\boldsymbol{\tau} = \tau$, and $\boldsymbol{\pi} = (\boldsymbol{\beta}', \boldsymbol{\xi}')' = (\boldsymbol{\beta}', \mu, \sigma)'$. The hypotheses of interest here are

$$H_0''' : \boldsymbol{\xi}_* = \mathbf{0} \quad \text{vs.} \quad H_1''' : \boldsymbol{\xi}_* \neq \mathbf{0}.$$

Then, for each $\boldsymbol{d}$ and $\widetilde{\boldsymbol{d}}$,

$$\mathbf{B}_*(\boldsymbol{d},\widetilde{\boldsymbol{d}}) = \begin{bmatrix} \mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}},\widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}}) & \mathbf{0}' \\ \mathbf{0} & \frac{2}{\tau_*^2} d_\tau{}'\widetilde{d}_\tau \end{bmatrix},$$

and

$$\mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}},\widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}}) = \frac{1}{\tau_*^2}\begin{bmatrix} \boldsymbol{d}_{\boldsymbol{\beta}}' E[\mathbf{X}_t\mathbf{X}_t']\widetilde{\boldsymbol{d}}_{\boldsymbol{\beta}} & \boldsymbol{d}_{\boldsymbol{\beta}}' E[\mathbf{X}_t'] m(d_\mu, d_\sigma) \\ m(\widetilde{d}_\mu,\widetilde{d}_\sigma)E[\mathbf{X}_t]\widetilde{\boldsymbol{d}}_{\boldsymbol{\beta}} & m(d_\mu,d_\sigma)m(\widetilde{d}_\mu,\widetilde{d}_\sigma) \end{bmatrix}.$$

By the information matrix equality, for each $\boldsymbol{d}$, $\mathbf{B}_*(\boldsymbol{d})$ is identical to $-\mathbf{A}_*(\boldsymbol{d})$.

The asymptotic null distributions of the test statistics are identified by the theorems of this section. First, we apply the QLR test. Applying Theorem 5 shows that

$$\mathcal{LR}_n^{(1)} := 2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\pi}}\in\Delta(\boldsymbol{\pi}_*)} \max[0, \mathcal{Y}^{(\boldsymbol{\pi})}(\boldsymbol{s}_{\boldsymbol{\pi}})]^2 + \mathcal{H}_2,$$

where for each $\boldsymbol{s}_{\boldsymbol{\pi}} \in \Delta(\boldsymbol{\pi}_*) := \{(\boldsymbol{s}_{\boldsymbol{\beta}}', s_\mu, s_\sigma)' \in \mathbb{R}^{k+2} : \boldsymbol{s}_{\boldsymbol{\beta}}'\boldsymbol{s}_{\boldsymbol{\beta}} + s_\mu^2 + s_\sigma^2 = 1, s_\mu > 0, \text{ and } s_\sigma > 0\}$,

$$\mathcal{Y}^{(\boldsymbol{\pi})}(\boldsymbol{s}_{\boldsymbol{\pi}}) := \{E[(\boldsymbol{s}_{\boldsymbol{\beta}}'\mathbf{X}_t + m(s_\mu,s_\sigma))^2]\}^{-1/2}\mathcal{Z}^{(\boldsymbol{\pi})}(\boldsymbol{s}_{\boldsymbol{\pi}}),$$

$\mathcal{Z}^{(\boldsymbol{\pi})}(\boldsymbol{s}_{\boldsymbol{\pi}}) := \boldsymbol{s}_{\boldsymbol{\beta}}'\mathbf{Z}^{(\boldsymbol{\beta})} + m(s_\mu,s_\sigma)Z^{(\boldsymbol{\xi})}$, and

$$\begin{bmatrix} \mathbf{Z}^{(\boldsymbol{\beta})} \\ Z^{(\boldsymbol{\xi})} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} E[\mathbf{X}_t\mathbf{X}_t'] & E[\mathbf{X}_t] \\ E[\mathbf{X}_t'] & 1 \end{bmatrix}\right).$$

We can see that $[\mathbf{Z}^{(\boldsymbol{\beta})'}, Z^{(\boldsymbol{\xi})}]'$ is the weak limit of $n^{-1/2}\tau_*^{-1}\sum_{t=1}^n [U_t\mathbf{X}_t', U_t]'$. Corollary 2($i$) also implies that $\mathcal{LR}_n^{(1)} := 2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\xi}}\in\Delta(\boldsymbol{\xi}_*)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})]^2 + \mathbf{Z}^{(\boldsymbol{\beta})'}E[\mathbf{X}_t\mathbf{X}_t']^{-1}\mathbf{Z}^{(\boldsymbol{\beta})} + \mathcal{H}_2$, where for each $\boldsymbol{s}_{\boldsymbol{\xi}_*} \in \Delta(\boldsymbol{\xi}_*) := \{(s_\mu, s_\sigma)' \in \mathbb{R}^2 : s_\mu^2 + s_\sigma^2 = 1, s_\mu > 0, \text{ and } s_\sigma > 0\}$,

$$\widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) := (\widetilde{B}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}))^{-1/2}\widetilde{\mathcal{Z}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}),$$

$$\widetilde{B}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) := m(s_\mu,s_\sigma)^2\{1 - E[\mathbf{X}_t]'E[\mathbf{X}_t\mathbf{X}_t']^{-1}E[\mathbf{X}_t]\},$$

and also $\widetilde{\mathcal{Z}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) := m(s_\mu,s_\sigma)\{Z^{(\boldsymbol{\xi})} - E[\mathbf{X}_t]'E[\mathbf{X}_t\mathbf{X}_t']^{-1}\mathbf{Z}^{(\boldsymbol{\beta})}\}$. Furthermore, Theorem 6 shows that

$$\mathcal{LR}_n^{(2)} := 2\{L_n(\ddot{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\beta}}\in\Delta(\boldsymbol{\beta}_*)} \max[0, \mathcal{Y}^{(\boldsymbol{\beta})}(\boldsymbol{s}_{\boldsymbol{\beta}})]^2 + \mathcal{H}_2,$$

where for each $\boldsymbol{s}_{\boldsymbol{\beta}} \in \Delta(\boldsymbol{\beta}_*) := \{\boldsymbol{s}_{\boldsymbol{\beta}} \in \mathbb{R}^k : \boldsymbol{s}_{\boldsymbol{\beta}}'\boldsymbol{s}_{\boldsymbol{\beta}} = 1\}$, $\mathcal{Y}^{(\boldsymbol{\beta})}(\boldsymbol{s}_{\boldsymbol{\beta}}) := \{\boldsymbol{s}_{\boldsymbol{\beta}}'E[\mathbf{X}_t\mathbf{X}_t']\boldsymbol{s}_{\boldsymbol{\beta}}\}^{-1/2}\mathbf{Z}^{(\boldsymbol{\beta})'}\boldsymbol{s}_{\boldsymbol{\beta}}$, and applying Corollary 2(*i*) implies that $\mathcal{LR}_n^{(2)} := 2\{L_n(\ddot{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \mathbf{Z}^{(\boldsymbol{\beta})\prime} E[\mathbf{X}_t \mathbf{X}_t']^{-1} \mathbf{Z}^{(\boldsymbol{\beta})} + \mathcal{H}_2$. Therefore, Corollary 2(*ii*) now yields that

$$\mathcal{LR}_n \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\xi}} \in \Delta(\boldsymbol{\xi}_0)} \max \left[0, \frac{m(s_\mu, s_\sigma)}{|m(s_\mu, s_\sigma)|} Z\right]^2$$

under H_0''', where $Z := \{1 - E[\mathbf{X}_t]' E[\mathbf{X}_t \mathbf{X}_t']^{-1} E[\mathbf{X}_t]\}^{-1/2} \{Z^{(\boldsymbol{\xi})} - E[\mathbf{X}_t]' E[\mathbf{X}_t \mathbf{X}_t']^{-1} \mathbf{Z}^{(\boldsymbol{\beta})}\} \sim N(0, 1)$. If we let $r(x) := \phi(x)/[x\Phi(x)]$,

$$\frac{m(s_\mu, s_\sigma)}{|m(s_\mu, s_\sigma)|} = -\frac{s_\mu}{|s_\mu|} \left(\frac{1 + r(s_\mu/|s_\sigma|)}{|1 + r(s_\mu/|s_\sigma|)|} \right),$$

which is -1 uniformly on $\Delta(\boldsymbol{\xi}_0)$. Thus, the asymptotic null distribution reduces to $\max[0, -Z]^2$, and this implies that $\mathcal{LR}_n \overset{\text{A}}{\sim} \max[0, -Z]^2$ under H_0'''.

We conduct simulations to verify this. We let $(\mathbf{X}_t', U_t)' \sim$ IID $N(\mathbf{0}_2, \mathbf{I}_2)$ and obtain the null distribution of the QLR test statistic by repeating the same independent experiments 2,000 times for $n =$ 50, 100, and 200. Simulation results are summarized in Figure 3. Note that the null distributions of the QLR test statistics exactly overlap with that of $\max[0, -Z]^2$.

Second, we apply the Wald test. For this, if we let

$$\widehat{W}_n(s_\mu, s_\sigma) := m(s_\mu, s_\sigma)^2 \left\{ 1 - n^{-1} \sum_{t=1}^n \mathbf{X}_t' (n^{-1} \sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t')^{-1} n^{-1} \sum_{t=1}^n \mathbf{X}_t \right\},$$

the LLN implies that $\sup_{s_\mu, s_\sigma} |\widehat{W}_n(s_\mu, s_\sigma) - \widetilde{B}_*^{(\boldsymbol{\xi}, \boldsymbol{\xi})}(s_\mu, s_\sigma)| \to 0$ a.s.$-\mathbb{P}$. In particular, $m(\,\cdot\,,\,\cdot\,)^2$ is bounded by 1 and $2/\pi$ from above and below, respectively. Using $\widehat{W}_n(s_\mu, s_\sigma)$, we let the Wald test statistic be

$$\mathcal{W}_n := \sup_{s_\mu, s_\sigma} n\{\widetilde{h}_n^{(\boldsymbol{\xi})}(s_\mu, s_\sigma)\}\{\widehat{W}_n(s_\mu, s_\sigma)\}\{\widetilde{h}_n^{(\boldsymbol{\xi})}(s_\mu, s_\sigma)\},$$

where $\widetilde{h}_n^{(\boldsymbol{\xi})}(s_\mu, s_\sigma)$ is such that for each (s_μ, s_σ),

$$\begin{aligned} L_n(\widetilde{h}_n^{(\boldsymbol{\xi})}(s_\mu, s_\sigma) s_\mu, \widetilde{h}_n^{(\boldsymbol{\xi})}(s_\mu, s_\sigma) s_\sigma, \widetilde{\boldsymbol{\beta}}_n(s_\mu, s_\sigma), \widetilde{\tau}_n(s_\mu, s_\sigma)) \\ = \sup_{\{h^{(\boldsymbol{\xi})}, \boldsymbol{\beta}, \tau\}} L_n(h^{(\boldsymbol{\xi})}(s_\mu, s_\sigma) s_\mu, h^{(\boldsymbol{\xi})}(s_\mu, s_\sigma) s_\sigma, \boldsymbol{\beta}, \tau). \end{aligned}$$

Corollary 3 now implies that $\mathcal{W}_n \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\xi}} \in \Delta(\boldsymbol{\xi}_0)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})]^2$, and this is the weak limit identical to that of the QLR test statistic. Thus, $\mathcal{W}_n \overset{\text{A}}{\sim} \max[0, -Z]^2$ under H_0'''.

Finally, we investigate the LM test statistic. We let

$$\mathcal{LM}_n := \sup_{(s_\mu, s_\sigma, \boldsymbol{s_\beta}) \in \Delta(\boldsymbol{\xi}_0) \times \Delta(\ddot{\beta}_n)} n\widetilde{W}_n(s_\mu, s_\sigma, \boldsymbol{s_\beta}) \max\left[0, \frac{-DL_n(\ddot{\boldsymbol{\theta}}_n; s_\mu, s_\sigma)}{\widetilde{D}^2 L_n(\ddot{\boldsymbol{\theta}}_n; s_\mu, s_\sigma, \boldsymbol{s_\beta})}\right]^2,$$

where $\ddot{\boldsymbol{\theta}}_n = (\ddot{\boldsymbol{\beta}}_n, 0, 0, \ddot{\tau}_n)$ with $\ddot{\boldsymbol{\beta}}_n = (\sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t')^{-1} \sum_{t=1}^n \mathbf{X}_t Y_t$, $\ddot{\tau}_n = (n^{-1} \sum_{t=1}^2 \ddot{U}_t^2)^{1/2}$, $\ddot{U}_t := Y_t - \mathbf{X}_t' \ddot{\boldsymbol{\beta}}_n$, $\Delta(\ddot{\boldsymbol{\beta}}_n) := \{\mathbf{x} \in \mathbb{R}^k : \mathbf{x}'\mathbf{x} = 1\}$, $DL_n(\ddot{\boldsymbol{\theta}}_n; s_\mu, s_\sigma) = \{m(s_\mu, s_\sigma)/\ddot{\tau}_n^2\} \sum_{t=1}^n \ddot{U}_t$, and

$$\begin{aligned} -\widetilde{D}^2 L_n(\ddot{\boldsymbol{\theta}}_n; s_\mu, s_\sigma, \boldsymbol{s_\beta}) =& \frac{1}{\ddot{\tau}_n^4} \sum_{t=1}^n \{s_\sigma^2(\ddot{\tau}_n^2 - \ddot{U}_t^2) + \psi(s_\mu, s_\sigma)^2 \ddot{U}_t^2 + \psi(s_\mu, s_\sigma) s_\mu (\ddot{U}_t^2 + \ddot{\tau}_n^2) + s_\mu^2 \ddot{\tau}_n^2\} \\ & - \frac{m(s_\mu, s_\sigma)^2}{\ddot{\tau}_n^2} \sum_{t=1}^n \boldsymbol{s_\beta}' \mathbf{X}_t \left(\boldsymbol{s_\beta}' \sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t' \boldsymbol{s_\beta}\right)^{-1} \sum_{t=1}^n \mathbf{X}_t' \boldsymbol{s_\beta}. \end{aligned}$$

In particular, applying the LLN implies that for each (s_μ, s_σ),

$$-\frac{1}{n}\widetilde{D}^2 L_n(\ddot{\boldsymbol{\theta}}_n; s_\mu, s_\sigma, \boldsymbol{s_\beta}) = \frac{m(s_\mu, s_\sigma)^2}{\tau_*^2}\{1 - \boldsymbol{s_\beta}' E[\mathbf{X}_t](\boldsymbol{s_\beta}' E[\mathbf{X}_t \mathbf{X}_t'] \boldsymbol{s_\beta})^{-1} E[\mathbf{X}_t'] \boldsymbol{s_\beta}\} + o_{\mathbb{P}}(1).$$

This LLN also holds uniformly on $\Delta(\boldsymbol{\xi}_0) \times \Delta(\ddot{\beta}_n)$. Thus, for each $(s_\mu, s_\sigma, \boldsymbol{s_\beta})$, we may let

$$\widetilde{W}_n(s_\mu, s_\sigma, \boldsymbol{s_\beta}) := \frac{m(s_\mu, s_\sigma)^2}{\tau_*^2}\left\{1 - n^{-1} \sum_{t=1}^n \boldsymbol{s_\beta}' \mathbf{X}_t \left(\boldsymbol{s_\beta}' n^{-1} \sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t' \boldsymbol{s_\beta}\right)^{-1} n^{-1} \sum_{t=1}^n \mathbf{X}_t' \boldsymbol{s_\beta}\right\}.$$

Here, applying the proof of Corollary 1(*vii*) implies that

$$\begin{aligned} &\sup_{\boldsymbol{s_\beta} \in \Delta(\ddot{\beta}_n)} n\widetilde{W}_n(s_\mu, s_\sigma, \boldsymbol{s_\beta}) \max\left[0, \frac{-DL_n(\ddot{\boldsymbol{\theta}}_n; s_\mu, s_\sigma)}{\widetilde{D}^2 L_n(\ddot{\boldsymbol{\theta}}_n; s_\mu, s_\sigma, \boldsymbol{s_\beta})}\right]^2 \\ &\qquad = \max\left[0, \frac{m(s_\mu, s_\sigma)}{|m(s_\mu, s_\sigma)|} \frac{n^{-1/2} \sum_{t=1}^n \ddot{U}_t}{\{\tau_*^2(1 - E[\mathbf{X}_t]' E[\mathbf{X}_t \mathbf{X}_t']^{-1} E[\mathbf{X}_t])\}^{1/2}}\right]^2 + o_{\mathbb{P}}(1) \end{aligned}$$

by optimizing the objective function with respect to $\boldsymbol{s_\beta}$, so that

$$\mathcal{LM}_n = \sup_{(s_\mu, s_\sigma) \in \Delta(\boldsymbol{\xi}_0)} \max\left[0, \frac{m(s_\mu, s_\sigma)}{|m(s_\mu, s_\sigma)|} \frac{n^{-1/2} \sum_{t=1}^n \ddot{U}_t}{\{\tau_*^2(1 - E[\mathbf{X}_t]' E[\mathbf{X}_t \mathbf{X}_t']^{-1} E[\mathbf{X}_t])\}^{1/2}}\right]^2 + o_{\mathbb{P}}(1)$$

under H_0'''. Therefore, $\mathcal{LM}_n \overset{\text{A}}{\sim} \max[0, -Z]^2$ by noting that $\frac{m(\cdot,\cdot)}{|m(\cdot,\cdot)|} = -1$ on $\Delta(\boldsymbol{\xi}_0)$ and $n^{-1/2} \sum_{t=1}^n \ddot{U}_t \sim N[0, \tau_*^2(1 - E[\mathbf{X}_t]' E[\mathbf{X}_t \mathbf{X}_t']^{-1} E[\mathbf{X}_t])]$. This is exactly what Corollary 4 asserts. $\square$

3.4.3 Example 3 (continued)

For this model examination, we let $\boldsymbol{\pi} = (\boldsymbol{\phi}', \boldsymbol{\xi}')'$ such that $\boldsymbol{\omega} = \boldsymbol{\phi} = \boldsymbol{\theta}_0$ and $\boldsymbol{\mu} = \boldsymbol{\xi} = \theta_2$, so that $\boldsymbol{\Omega} = \boldsymbol{\Theta}_0$, and $\mathbf{M}$ is a closed interval with zero as an interior element. Note that $\theta_{1*} = 0$ if and only if $\theta_{2*} = 0$ from the model assumption. Using these conditions, Corollary 2(*ii*) can be applied. That is,

$$\mathcal{LR}_n \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\xi}} \in \Delta(\boldsymbol{\xi}_0)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})]^2,$$

where $\boldsymbol{s}_{\boldsymbol{\xi}} := s_1$, $\Delta(\boldsymbol{\xi}_0) := \{-1, 1\}$, and

$$\widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) := \frac{s_1 \widetilde{\mathcal{Z}}^{(\boldsymbol{\xi})}}{|s_1|(\widetilde{A}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})})^{1/2}} := \frac{s_1(W_1 - (-\mathbf{A}_*^{(0,1)})(-\mathbf{A}_*^{(0,0)})^{-1}\mathbf{W}_0)}{|s_1|\{(-A_*^{(1,1)}) - (-A_*^{(1,0)})(-\mathbf{A}_*^{(0,0)})^{-1}(-A_*^{(0,1)})\}^{1/2}}.$$

Note that $s_1/|s_1| = \pm 1$, and from this, $\mathcal{LR}_n \Rightarrow \widetilde{\mathcal{Z}}^{(\boldsymbol{\xi})}(\widetilde{A}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})})^{-1}\widetilde{\mathcal{Z}}^{(\boldsymbol{\xi})}$.

In a similar way, we can apply Corollary 3 for the Wald test statistic. Note that $\sqrt{n}\widetilde{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\xi}}) \Rightarrow (\widetilde{A}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})})^{-1}\max[0, s_1\widetilde{\mathcal{Z}}^{(\boldsymbol{\xi})}]$, and select $\widehat{W}_n$ to be a consistent estimator for $(\widetilde{A}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})})^{-1}$. For example, if we let

$$\widehat{W}_n := \{(n^{-1}\sum \log(X_t)^2) - (n^{-1}\sum \log(X_t)\mathbf{Z}_t')(n^{-1}\sum \mathbf{Z}_t\mathbf{Z}_t')^{-1}(n^{-1}\sum \mathbf{Z}_t \log(X_t))\},$$

$\mathcal{W}_n := n\{\widetilde{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\xi}})\}\{\widehat{W}_n\}\{\widetilde{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\xi}})\} \Rightarrow \widetilde{\mathcal{Z}}^{(\boldsymbol{\xi})}(\widetilde{A}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})})^{-1}\widetilde{\mathcal{Z}}^{(\boldsymbol{\xi})}$ by Corollary 3.

Finally, Corollary 4 obtains the same asymptotic null distribution for the LM test statistic using the same weight function. $\square$

3.4.4 Example 4 (continued)

As the model is D, $r_{\boldsymbol{\upsilon}} = 0$. Furthermore, $r_{\boldsymbol{\tau}} = 0$ does not reduce the applicability of Theorem 5. Thus, we simply let $\boldsymbol{\theta} = \boldsymbol{\pi} = \boldsymbol{\lambda} = (\boldsymbol{\delta}', \boldsymbol{\phi}')' = (\boldsymbol{\mu}', \boldsymbol{\omega}')'$.

The objective function $Q_n(\cdot)$ does not satisfy the model condition in Assumption 2. Therefore, the definition of the QLR test statistic cannot be exactly applied to this model. Nevertheless, a QLR test-like test statistic can be defined. We let

$$\mathcal{QLR}_n := \{\sup_{\boldsymbol{\delta},\boldsymbol{\phi}} Q_n(\boldsymbol{\delta}, \boldsymbol{\phi}) - \sup_{\boldsymbol{\phi}} Q_n(\boldsymbol{\delta}_0, \boldsymbol{\phi})\}$$

and also let $\mathbf{C}_*'\{-\mathbf{M}_*\}^{-1}\mathbf{W}$ and $\mathbf{C}_*'\{-\mathbf{M}_*\}^{-1}\mathbf{C}_*$ be $\mathbf{Z}^{(\boldsymbol{\lambda})} = (\mathbf{Z}^{(\delta)'}, \mathbf{Z}^{(\phi)'})'$ and $\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}$ of this section, respectively. The asymptotic null distribution of the QLR test statistic is already given in (15). We can also define Wald test statistic using the GMM estimator and derive its asymptotic null distribution in the same way:

$$\mathcal{QW}_n := \sup_{s_{\delta} \in \Delta(\delta_0)} n\{\widetilde{h}_n^{(\delta)}(s_{\delta})\}\{\widehat{W}_n(s_{\delta})\}\{\widetilde{h}_n^{(\delta)}(s_{\delta})\},$$

where $\widetilde{h}_n^{(\delta)}(s_{\delta})$ is such that for each $s_{\delta} \in \Delta(\delta_0)$,

$$Q_n(\delta_0 + \widetilde{h}_n^{(\delta)}(s_{\delta})s_{\delta}, \widetilde{\phi}_n(s_{\delta})) := \sup_{\{h^{(\delta)}, \phi\}} Q_n(\delta_0 + h^{(\delta)}s_{\delta}, \phi),$$

and its asymptotic null distribution can be obtained by applying Corollary 3. Note that the definition of $\mathcal{QW}_n$ is exactly the same as $\mathcal{W}_n$ except that $\widetilde{h}_n^{(\delta)}(s_{\delta})$ is defined using $Q_n(\cdot)$ instead of $L_n(\cdot)$. If we further let the weight function $\widehat{W}_n(s_{\delta})$ be $s_{\delta}'\widehat{W}_n s_{\delta}$ such that $\widehat{W}_n$ converges to $-\widetilde{\mathbf{A}}_*^{(\delta,\delta)}$ a.s.$-\mathbb{P}$,

$$\mathcal{QW}_n \Rightarrow \sup_{s_{\delta} \in \Delta(\delta_0)} \max[0, s_{\delta}'\widetilde{\mathbf{Z}}^{(\delta)}](-s_{\delta}'\widetilde{\mathbf{A}}_*^{(\delta,\delta)}s_{\delta})^{-1}\max[0, s_{\delta}'\widetilde{\mathbf{Z}}^{(\delta)}].$$

The proof of Corollary 1(*vii*) corroborates that the asymptotic null distribution of $\mathcal{QW}_n$ is equivalent to that of $\mathcal{QLR}_n$ particularly because δ_0 is an interior element. Finally, we define LM test statistic in the GMM context and examine its asymptotic null distribution. For this purpose, we let

$$\mathcal{QLM}_n := \sup_{(s_{\delta}, s_{\phi}) \in \Delta(\delta_0) \times \Delta(\ddot{\phi}_n)} n\widetilde{W}_n(s_{\delta}, s_{\phi}) \max\left[0, \frac{DQ_n(\ddot{\theta}_n; s_{\delta})}{2\widetilde{D}^2 Q_n(\ddot{\theta}_n; s_{\delta}, s_{\phi})}\right]^2,$$

where for each (s_{δ}, s_{ϕ}),

$$\begin{aligned}
\widetilde{D}^2 Q_n(\ddot{\theta}_n; s_{\delta}, s_{\phi}) &:= D\mathbf{g}_n(\ddot{\theta}_n; s_{\delta})'\{-\mathbf{M}_n\}^{-1} D\mathbf{g}_n(\ddot{\theta}_n; s_{\delta}) \\
&\quad - D\mathbf{g}_n(\ddot{\theta}_n; s_{\delta})'\{-\mathbf{M}_n\}^{-1} D\mathbf{g}_n(\ddot{\theta}_n; s_{\phi})\{D\mathbf{g}_n(\ddot{\theta}_n; s_{\phi})'\{-\mathbf{M}_n\}^{-1} D\mathbf{g}_n(\ddot{\theta}_n; s_{\phi})\}^{-1} \\
&\quad \times D\mathbf{g}_n(\ddot{\theta}_n; s_{\phi})'\{-\mathbf{M}_n\}^{-1} D\mathbf{g}_n(\ddot{\theta}_n; s_{\delta}),
\end{aligned}$$

and $\ddot{\theta}_n := (\delta_0, \ddot{\phi}_n)$ such that $\ddot{\phi}_n := \arg\max_{\phi} Q_n(\delta_0, \phi)$. If we let $\widetilde{W}_n(s_{\delta}, s_{\phi}) = s_{\delta}'\widehat{W}_n s_{\delta}$ for each $(s_{\delta}, s_{\phi}) \in \Delta(\delta_0) \times \Delta(\phi_*)$, $\mathcal{QLM}_n \Rightarrow (\widetilde{\mathbf{Z}}^{(\delta)})'(-\widetilde{\mathbf{A}}_*^{(\delta,\delta)})^{-1}(\widetilde{\mathbf{Z}}^{(\delta)})$ by Corollary 4, the interiority condition of δ_0, and the proof of Corollary 1(*vii*), where $\widehat{W}_n$ is the same weight matrix as used for the $\mathcal{QW}_n$ test statistic. $\square$

4 Conclusion

In this paper, we examine data estimation and inference using D-D econometric models and provide conditions under which the extremum estimator behaves regularly. Specifically, we show that the extremum estimator has a distribution different from that of standard D models by showing that it is represented as a function of a Gaussian stochastic process indexed by directions. Furthermore, the analysis assuming D models can be treated as a special case of D-D model analysis.

In addition, we appropriately modify the standard QLR, Wald, and LM test statistics for D-D models. These modifications are provided for general D-D models, and the three test statistics have asymptotic null distributions represented as functionals of the same Gaussian stochastic process. Finally, the three test statistics are asymptotically equivalent under the null if the benchmark model assumption is further imposed, which is popularly used for empirical data examination.

Appendix: Proofs

Proof of Lemma 1: To show the given claim, we show that f is twice continuously D at $\boldsymbol{\theta}_0$. The same proof can also be applied to other parameter values.

If we let $g(h) := f(\boldsymbol{\theta}_0 + h\boldsymbol{d})$, g is twice continuously D from the given condition, so that we can apply the mean-value theorem: for some $\widetilde{h} \geq 0$

$$g(h) = g(0) + g'(0)h + \frac{1}{2}g''(\widetilde{h})h^2,$$

implying that $f(\boldsymbol{\theta}) - f(\boldsymbol{\theta}_0) - Df(\boldsymbol{\theta}_0; \boldsymbol{d})h = \frac{1}{2}D^2 f(\widetilde{\boldsymbol{\theta}}; \boldsymbol{d})h^2$, where $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + h\boldsymbol{d}$, $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + \widetilde{h}\boldsymbol{d}$, and $\widetilde{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$ as $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$. Thus,

$$f(\boldsymbol{\theta}) - f(\boldsymbol{\theta}_0) - Df(\boldsymbol{\theta}_0; \boldsymbol{d})h - \frac{1}{2}D^2 f(\boldsymbol{\theta}_0; \boldsymbol{d})h^2 = \frac{1}{2}D^2 f(\widetilde{\boldsymbol{\theta}}; \boldsymbol{d})h^2 - \frac{1}{2}D^2 f(\boldsymbol{\theta}_0; \boldsymbol{d})h^2.$$

Furthermore, for some $\mathbf{A}(\boldsymbol{\theta}_0) \in \mathbb{R}^r$ and $\mathbf{B}(\boldsymbol{\theta}_0) \in \mathbb{R}^r \times \mathbb{R}^r$, $\mathbf{A}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \mathbf{A}(\boldsymbol{\theta}_0)\boldsymbol{d}h$ and $(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{B}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = D^2 f(\widetilde{\boldsymbol{\theta}}; \boldsymbol{d})h^2$ from the linear and quadratic form conditions, so that

$$f(\boldsymbol{\theta}) - f(\boldsymbol{\theta}_0) - \mathbf{A}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{B}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \frac{1}{2}(D^2 f(\widetilde{\boldsymbol{\theta}}; \boldsymbol{d}) - D^2 f(\boldsymbol{\theta}_0; \boldsymbol{d}))h^2.$$

Therefore,

$$
\begin{aligned}
&\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0}\left|\frac{1}{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^2}\{f(\boldsymbol{\theta})-f(\boldsymbol{\theta}_0)-\mathbf{A}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\mathbf{B}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\}\right|\\
&=\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0}\sup_{\boldsymbol{d}}\left|\frac{1}{2\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^2}(D^2f(\widetilde{\boldsymbol{\theta}};\boldsymbol{d})-D^2f(\boldsymbol{\theta}_0;\boldsymbol{d}))h^2\right|\\
&=\lim_{\widetilde{\boldsymbol{\theta}}\to\boldsymbol{\theta}_0}\sup_{\boldsymbol{d}}\left|\frac{1}{2}(D^2f(\widetilde{\boldsymbol{\theta}};\boldsymbol{d})-D^2f(\boldsymbol{\theta}_0;\boldsymbol{d}))\right|\leq\lim_{\widetilde{\boldsymbol{\theta}}\to\boldsymbol{\theta}_0}\frac{M}{2}\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|=0,
\end{aligned}
$$

where the last inequality follows from the uniformity condition. This completes the proof. ■

Proof of Lemma 2: (*i*) For this purpose, we verify the conditions of Wooldridge and White (1988). First, AC1 of Wooldridge and White (1988) is satisfied by Assumption 6(*iii*) because we can let $n^{-1/2}\sum \ell_t\ (\boldsymbol{\theta}_*;\boldsymbol{d})$ be their $\sum Z_{nt}$. Second, the conditions (i, ii, iii) of AC2 in Wooldridge and White (1988) trivially hold by our assumptions that $\|D\ell_t(\boldsymbol{\theta}_*;\boldsymbol{d})\|_s<\Delta$ uniformly in t, that ν_τ is of size $-1/(1-\gamma)<-1/2$, and that $\{\mathbf{Y}_t\}$ is a strong mixing sequence of size $-sq/(s-q)<-s/(s-2)$ because $s>q\geq 2$, respectively. Third, condition (iv) of AC2 can be easily defined from the fact that $\|\ell_t(\boldsymbol{\theta}_*;\boldsymbol{d})\|_s<\Delta<\infty$ uniformly in t and d. Finally, their condition in Assumption 5 does not need to be proved as our goal is not to obtain the standard normal distribution.

(*ii*) We can apply the ergodic theorem given Assumptions 1 and 6(*ii*).

(*iii*) Given Lemmas 2(*i* and *ii*), the given weak convergence follows from theorem 3.9 of Billingsley (1999, p. 37). ■

Proof of Theorem 3: (*i*) For given $\boldsymbol{d}$, if we maximize (8) with respect to h subject to $h\geq 0$, it follows that $\widehat{h}_n(\boldsymbol{d})=\max[0,\{-D^2L_n(\bar{\boldsymbol{\theta}}_n(\boldsymbol{d});\boldsymbol{d})\}^{-1}DL_n(\boldsymbol{\theta}_*;\boldsymbol{d})]$ by using Khun-Tucker theorem. We further note that $\bar{\boldsymbol{\theta}}_n(\boldsymbol{d})\to\boldsymbol{\theta}_*$ a.s.$-\mathbb{P}$ as implied by Theorem 1, so that $\sqrt{n}\widehat{h}_n(\boldsymbol{d})\Rightarrow\max[0,\{-A_*(\boldsymbol{d})\}^{-1}\mathcal{Z}(\boldsymbol{d})]$ by Lemma 2(*iii*). The desired result now follows from the definition of $\mathcal{G}(\boldsymbol{d})$.

(*ii*) From the definition of $\widehat{h}_n(\boldsymbol{d})$, $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})\equiv\boldsymbol{\theta}_*+\widehat{h}_n(\boldsymbol{d})\boldsymbol{d}$. Theorem 3(*i*) yields the given result.

(*iii*) Given that $\arg\max_{\widetilde{h}\in\mathbb{R}^+}[2\mathcal{Z}(\boldsymbol{d})+A_*(\boldsymbol{d})\widetilde{h}^2]=\max[0,\mathcal{G}(\boldsymbol{d})]$,

$$\max_{\widetilde{h}\in\mathbb{R}^+}[2\mathcal{Z}(\boldsymbol{d})+A_*(\boldsymbol{d})\widetilde{h}^2]=\max[0,\{-A_*(\boldsymbol{d})\}^{1/2}\mathcal{G}(\boldsymbol{d})]^2.$$

Thus, the desired result follows from (9). ■

Proof of Lemma 3: (*i*) Given the weak convergence of Lemma 2(*i*), if $\{n^{-1/2}\sum D\ell_t(\boldsymbol{\theta}_*;\cdot)\}$ is tight, the desired result follows from the finite dimensional multivariate CLT based on the Cramér-Wold device, which we do not prove from its self-evidence.

The tightness can be proved by verifying the conditions of theorem 4 in Hansen (1996a). First, from the fact that $\{\mathbf{Y}_t\}$ is a strong mixing sequence of $-sq/(s-q)$, for some $\epsilon > 0$, $\alpha_\tau^{-(s-q)/(sq)} = O(\tau^{-1-\epsilon})$, so that $\sum_{\tau=1}^{\infty} \alpha_\tau^{-(s-q)/(sq)} < \infty$. Second, $\|M_t\|_s < \infty$ uniformly in t from the stationarity assumption of $\{M_t\}$ in Assumption 6(*iv*). Third, $\|D\ell_t(\boldsymbol{\theta}_*; \boldsymbol{d})\|_s < \infty$ uniformly in t and $\boldsymbol{d}$ from Assumption 6(*iv*). Fourth, given that ν_τ is of size $-1/(1-\gamma)$, for some $\epsilon > 0$, $\nu_\tau = O(\tau^{-1/(1-\gamma)-\epsilon})$, implying that $\sum_{\tau=1}^{\infty} \nu_\tau^{1-\gamma} < \infty$. Finally, it is already assumed in Assumption 6(*iv*) that $q > (r-1)/(\gamma\lambda)$. The above results verify the conditions in theorem 4 of Hansen (1996a), and the tightness of $\{n^{-1/2} \sum D\ell_t(\boldsymbol{\theta}_*; \cdot\,)$ follows.

(*ii*) By Assumption 5(*iii*), $|n^{-1}D^2L_n(\boldsymbol{\theta}; \boldsymbol{d}_1) - n^{-1}D^2L_n(\boldsymbol{\theta}; \boldsymbol{d}_2)| \leq n^{-1}\sum M_t \|\boldsymbol{d}_1 - \boldsymbol{d}_2\|^\lambda$. Furthermore, we can apply the ergodic theorem to $\{n^{-1}\sum M_t\}$, so that for any $\omega \in F$, $\mathbb{P}(F) = 1$, and $\varepsilon > 0$, there is an $n^*(\omega, \varepsilon)$ such that if $n \geq n^*(\omega, \varepsilon)$, $|n^{-1}\sum M_t - E[M_t]| \leq \varepsilon$, and this implies that $n^{-1}\sum M_t \leq E[M_t] + \varepsilon$. For the same ε, we may let $\delta := \varepsilon/(E[M_t]+\varepsilon)$. Then, $n^{-1}\sum M_t\|\boldsymbol{d}_1 - \boldsymbol{d}_2\|^\lambda \leq \varepsilon$, whenever $\|\boldsymbol{d}_1 - \boldsymbol{d}_2\|^\lambda \leq \delta$, because $n^{-1}\sum M_t \|\boldsymbol{d}_1 - \boldsymbol{d}_2\|^\lambda \leq n^{-1}\sum M_t\delta = n^{-1}\sum M_t\varepsilon/(\varepsilon + E[M_t]) \leq \varepsilon$. That is, for any $\omega \in F$, $\mathbb{P}(F) = 1$ and $\varepsilon > 0$, there is $n^*(\omega, \varepsilon)$ and δ such that if $n \geq n^*(\omega, \varepsilon)$ and $\|\boldsymbol{d}_1 - \boldsymbol{d}_2\|^\lambda \leq \delta$, $|n^{-1}D^2L_n(\boldsymbol{\theta}; \boldsymbol{d}_1) - n^{-1}D^2L_n(\boldsymbol{\theta}; \boldsymbol{d}_2)| < \varepsilon$, which means that $\{n^{-1}D^2L_n(\boldsymbol{\theta}; \cdot\,)\}_{n^*(\omega,\varepsilon)}^{\infty}$ is equicontinuous. Therefore, it follows that $n^{-1}D^2L_n(\boldsymbol{\theta}; \cdot\,)$ converges to A_* uniformly on $\Delta(\boldsymbol{\theta}_*)$ a.s.–$\mathbb{P}$ by Rudin (1976, p. 168). ■

Proof of Theorem 4: (*i*) Given Lemmas 3(*i* and *ii*), the given weak convergence follows from theorem 3.9 of Billingsley (1999, p. 37).

(*ii*) The given result follows from Theorem 3(*i*), Theorem 4(*i*), and the definition of $\mathcal{G}$.

(*iii*) We can apply the CMT to (10).

(*iv*) From the definition of $\widehat{h}_n$, for each $\boldsymbol{d}$, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d}) - \boldsymbol{\theta}_*) = \sqrt{n}\widehat{h}_n(\boldsymbol{d})\boldsymbol{d}$. Theorems 3(*ii*), 4(*i* to *iii*), (10), and the CMT yield the given result. ■

Proof of Corollary 1: For an efficient presentation, we first prove (*vi*) and (*vii*) before proving (*iv*) and (*v*).

(*i*) As the weak convergence is proved for a general function, we verify only the pointwise weak convergence for this case. From the definition of $DL_n(\boldsymbol{\theta}_*; \boldsymbol{d}) = \nabla_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta})'\boldsymbol{d}$, and $n^{-1/2}\nabla_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*) \Rightarrow \mathbf{Z}$ by theorem 1 of Doukhan, Massart, and Rio (1995). Therefore, $n^{-1/2}DL_n(\boldsymbol{\theta}_*; \boldsymbol{d}) \Rightarrow \mathbf{Z}'\boldsymbol{d}$ for every $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$.

(*ii*) We note that $D^2L_n(\boldsymbol{\theta}; \boldsymbol{d}) = \boldsymbol{d}'\nabla^2_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)\boldsymbol{d}$, so that $n^{-1}\nabla^2_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*) \to \mathbf{A}_*$ a.s.–$\mathbb{P}$ by the ergodic theorem. Therefore, by the definition of $\mathcal{G}(\boldsymbol{d})$, the given result follows.

(*iii*) We can use the definition of $\widehat{h}_n(\boldsymbol{d})$. That is, $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d}) = \boldsymbol{\theta}_* + \widehat{h}_n(\boldsymbol{d})\boldsymbol{d}$. The given result follows from the fact that $\sqrt{n}\widehat{h}_n(\boldsymbol{d}) \Rightarrow \max[0, \mathcal{G}(\boldsymbol{d})]$ and Corollary 1(*ii*).

(*vi*) By the definition of $\mathcal{Y}$ of Theorem 4, for each $\boldsymbol{d}$, $\mathcal{Y}(\boldsymbol{d}) = \{\boldsymbol{d}'(-A_*)\boldsymbol{d}\}^{-1/2}\mathbf{Z}'\boldsymbol{d}$, so that Theorem 4(*iii*) implies the desired result.

(*vii*) From the fact that $\mathrm{cl}\{C(\boldsymbol{\theta}_*)\} = \bar{\mathbb{R}}^r$, there is $\boldsymbol{d}^\star \in \Delta(\boldsymbol{\theta}_*)$ such that $\max[0, \mathbf{Z}'\boldsymbol{d}^\star] = \mathbf{Z}'\boldsymbol{d}^\star$ and $\boldsymbol{d}^\star = -\boldsymbol{d}$ if $\max[0, \mathbf{Z}'\boldsymbol{d}] = 0$. Thus, the given 'max' operator can be ignored in this case. That is,

$$\boldsymbol{d}_* = \arg \max_{\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)} \boldsymbol{d}'\mathbf{Z}\mathbf{Z}'\boldsymbol{d}\{\boldsymbol{d}'(-\mathbf{A}_*)\boldsymbol{d}\}^{-1}.$$

For notational simplicity, if we let

$$\boldsymbol{v} := \frac{(-\mathbf{A}_*)^{1/2}\boldsymbol{d}}{\{\boldsymbol{d}'(-\mathbf{A}_*)\boldsymbol{d}\}^{1/2}},$$

it follows that $\boldsymbol{v}'\boldsymbol{v} = 1$ and $\boldsymbol{v}'(-\mathbf{A}_*)^{-1/2}\mathbf{Z}\mathbf{Z}'(-\mathbf{A}_*)^{-1/2}\boldsymbol{v} = \boldsymbol{d}'\mathbf{Z}\mathbf{Z}'\boldsymbol{d}\{\boldsymbol{d}'(-\mathbf{A}_*)\boldsymbol{d}\}^{-1}$. Given this, we note that $\max_{\boldsymbol{v}} \boldsymbol{v}'(-\mathbf{A}_*)^{-1/2}\mathbf{Z}\mathbf{Z}'(-\mathbf{A}_*)^{-1/2}\boldsymbol{v} = \max_{\boldsymbol{d}} \boldsymbol{d}'\mathbf{Z}\mathbf{Z}'\boldsymbol{d}\{\boldsymbol{d}'(-\mathbf{A}_*)\boldsymbol{d}\}^{-1}$, and it is equal to the maximum eigenvalue of $(-\mathbf{A}_*)^{-1/2}\mathbf{Z}\mathbf{Z}'(-\mathbf{A}_*)^{-1/2}$, which is equal to $\mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z}$. It is mainly because $\mathrm{rank}((-\mathbf{A}_*)^{-1/2}\mathbf{Z}\mathbf{Z}'(-\mathbf{A}_*)^{-1/2}) = 1$ (so that there is one positive eigenvalue, and other eigenvalues are zero), and the sum of eigenvalues is equal to $\mathrm{tr}[(-\mathbf{A}_*)^{-1/2}\mathbf{Z}\mathbf{Z}'(-\mathbf{A}_*)^{-1/2}] = \mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z}$. These two facts lead to the desired result.

(*iv*) This follows trivially from the definition of $\boldsymbol{d}_*$.

(*v*) By the same reason as in the proof of (*vii*), we can ignore the 'max' operator, so that $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \Rightarrow \mathbf{Z}'\boldsymbol{d}_*\{\boldsymbol{d}_*'(-\mathbf{A}_*)\boldsymbol{d}_*\}^{-1}\boldsymbol{d}_*$.

Given this, also from the proof of (*vii*), if we let

$$\boldsymbol{v}_* := \frac{(-\mathbf{A}_*)^{1/2}\boldsymbol{d}_*}{\{\boldsymbol{d}_*'(-\mathbf{A}_*)\boldsymbol{d}_*\}^{1/2}},$$

$\boldsymbol{v}_*$ is the eigenvector of $(-\mathbf{A}_*)^{-1/2}\mathbf{Z}\mathbf{Z}'(-\mathbf{A}_*)^{-1/2}$ corresponding to the maximum eigenvalue given as $\mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z}$, so that

$$(-\mathbf{A}_*)^{-1/2}\mathbf{Z}\mathbf{Z}'(-\mathbf{A}_*)^{-1/2}\boldsymbol{v}_* = \mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z}\boldsymbol{v}_* \tag{23}$$

by the definition of eigenvector. This implies that

$$\boldsymbol{v}_*'(-\mathbf{A}_*)^{-1/2}\mathbf{Z}\mathbf{Z}'(-\mathbf{A}_*)^{-1/2}\boldsymbol{v}_* = \mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z}\boldsymbol{v}_*'\boldsymbol{v}_* = \mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z} \tag{24}$$

because $\boldsymbol{v}_*'\boldsymbol{v}_* = 1$. Plugging the definition of $\boldsymbol{v}_*$ to the LHS of (24) leads to that

$$\mathbf{Z}'\boldsymbol{d}_*\{\boldsymbol{d}_*'(-\mathbf{A}_*)\boldsymbol{d}_*\}^{-1} = (\boldsymbol{d}_*'\mathbf{Z})^{-1}\mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z}.$$

Thus, $\mathbf{Z}'\boldsymbol{d}_*\{\boldsymbol{d}_*{}'(-\mathbf{A}_*)\boldsymbol{d}_*\}^{-1}\boldsymbol{d}_* = (\boldsymbol{d}_*{}'\mathbf{Z})^{-1}\mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z}\boldsymbol{d}_*$. Also, plugging the definition of $\boldsymbol{v}_*$ to (23) yields that $(\boldsymbol{d}_*{}'\mathbf{Z})^{-1}\mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z}\boldsymbol{d}_* = (-\mathbf{A})^{-1}\mathbf{Z}$. Therefore, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \Rightarrow \mathbf{Z}'\boldsymbol{d}_*\{\boldsymbol{d}_*{}'(-\mathbf{A}_*)\boldsymbol{d}_*\}^{-1}\boldsymbol{d}_*$ $= (-\mathbf{A}_*)^{-1}\mathbf{Z}$. This completes the proof. ■

Proof of Theorem 5: We note that for any $h\boldsymbol{d}$ such that $h \in \mathbb{R}^+$ and $\boldsymbol{d} \in \Delta(\boldsymbol{\pi}_*)$, there are $h^{(\boldsymbol{\pi})} \in \mathbb{R}^+$, $h^{(\boldsymbol{\tau})} \in \mathbb{R}^+$, $\boldsymbol{s}_{\boldsymbol{\pi}} \in \Delta(\boldsymbol{\pi}_*)$, and $\boldsymbol{s}_{\boldsymbol{\tau}} \in \Delta(\boldsymbol{\tau}_*)$ such that $h\boldsymbol{d} = [h^{(\boldsymbol{\pi})}\boldsymbol{s}_{\boldsymbol{\pi}}{}', h^{(\boldsymbol{\tau})}\boldsymbol{s}_{\boldsymbol{\tau}}{}']'$ by Assumption 7. Thus, $L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = L_n(\boldsymbol{\pi}_* + h^{(\boldsymbol{\pi})}\boldsymbol{s}_{\boldsymbol{\pi}}, \boldsymbol{\tau}_* + h^{(\boldsymbol{\tau})}\boldsymbol{s}_{\boldsymbol{\tau}})$, implying that

$$\begin{aligned}
&2\{L_n(\boldsymbol{\pi}_* + h^{(\boldsymbol{\pi})}\boldsymbol{s}_{\boldsymbol{\pi}}, \boldsymbol{\tau}_* + h^{(\boldsymbol{\tau})}\boldsymbol{s}_{\boldsymbol{\tau}}) - L_n(\boldsymbol{\pi}_*, \boldsymbol{\tau}_*)\} \\
&\qquad = 2DL_n(\boldsymbol{\pi}_*, \boldsymbol{\tau}_*; \boldsymbol{s}_{\boldsymbol{\pi}})h^{(\boldsymbol{\pi})} + 2DL_n(\boldsymbol{\pi}_*, \boldsymbol{\tau}_*; \boldsymbol{s}_{\boldsymbol{\tau}})h^{(\boldsymbol{\tau})} \\
&\qquad\quad + D^2L_n(\boldsymbol{\pi}_*, \boldsymbol{\tau}_*; \boldsymbol{s}_{\boldsymbol{\pi}})(h^{(\boldsymbol{\pi})})^2 + D^2L_n(\boldsymbol{\pi}_*, \boldsymbol{\tau}_*; \boldsymbol{s}_{\boldsymbol{\tau}})(h^{(\boldsymbol{\tau})})^2 \\
&\qquad\quad + 2DL_n(\boldsymbol{\pi}_*, \boldsymbol{\tau}_*; \boldsymbol{s}_{\boldsymbol{\pi}}; \boldsymbol{s}_{\boldsymbol{\tau}})h^{(\boldsymbol{\pi})}h^{(\boldsymbol{\tau})} + o_{\mathbb{P}_{\boldsymbol{d}_{\boldsymbol{\pi}}}}(1) + o_{\mathbb{P}_{\boldsymbol{d}_{\boldsymbol{\tau}}}}(1),
\end{aligned}$$

where $DL_n(\boldsymbol{\pi}_*, \boldsymbol{\tau}_*; \boldsymbol{s}_{\boldsymbol{\pi}}; \boldsymbol{s}_{\boldsymbol{\tau}})$ is the directional derivative of $DL_n(\,\cdot\,, \cdot\,; \boldsymbol{s}_{\boldsymbol{\pi}})$ with respect to $\boldsymbol{s}_{\boldsymbol{\tau}}$ evaluated at $(\boldsymbol{\pi}_*, \boldsymbol{\tau}_*)$, and $\sup_{\boldsymbol{d}}\ \sup_h L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = \sup_{\{\boldsymbol{s}_{\boldsymbol{\pi}}, \boldsymbol{s}_{\boldsymbol{\tau}}\}} \sup_{\{h^{(\boldsymbol{\pi})}, h^{(\boldsymbol{\tau})}\}} L_n(\boldsymbol{\pi}_* + h^{(\boldsymbol{\pi})}\boldsymbol{s}_{\boldsymbol{\pi}}, \boldsymbol{\tau}_* + h^{(\boldsymbol{\tau})}\boldsymbol{s}_{\boldsymbol{\tau}})$. Therefore,

$$\begin{aligned}
2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} &= \sup_{\boldsymbol{d}} \sup_{h} 2\{L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) - L_n(\boldsymbol{\theta}_*)\} \\
&= \sup_{\boldsymbol{s}_{\boldsymbol{\pi}}}\ \sup_{h^{(\boldsymbol{\pi})}} \{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\pi}})h^{(\boldsymbol{\pi})} + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\pi}})(h^{(\boldsymbol{\pi})})^2 + o_{\mathbb{P}_{\boldsymbol{d}_{\boldsymbol{\pi}}}}(1)\} \\
&\quad + \sup_{\boldsymbol{s}_{\boldsymbol{\tau}}}\ \sup_{h^{(\boldsymbol{\tau})}} \{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\tau}})h^{(\boldsymbol{\tau})} + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\tau}})(h^{(\boldsymbol{\tau})})^2 + o_{\mathbb{P}_{\boldsymbol{d}_{\boldsymbol{\tau}}}}(1)\}, \qquad (25)
\end{aligned}$$

where we exploited the facts that $n^{-1}DL_n(\boldsymbol{\pi}_*, \boldsymbol{\tau}_*; \boldsymbol{s}_{\boldsymbol{\pi}}; \boldsymbol{s}_{\boldsymbol{\tau}})$ has probability limit $\mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\pi})}(\boldsymbol{s}_{\boldsymbol{\tau}}, \boldsymbol{s}_{\boldsymbol{\pi}}) = \mathbf{0}$ by Assumption 7(*iv*) and that $DL_n(\boldsymbol{\theta}_*;\,\cdot\,)$ and $DL_n(\boldsymbol{\theta}_*;\,\cdot\,)$ are $O_{\mathbb{P}}(n^{1/2})$ by Theorem 4(*i*). Given this, the desired result follows by applying the proof of Theorem 4(*iii*) to each piece in the RHS of (25). ■

Proof of Theorem 6: (*i*) We can apply the ULLN.

(*ii*) This follows as a corollary of Theorem 5. ■

Proof of Corollary 2: To show the given claim, we use (25). As it is trivial from the proof of Theorem 5 that $H_{2,n} := \sup_{\boldsymbol{s}_{\boldsymbol{\tau}}}\ \sup_{h^{(\boldsymbol{\tau})}} \{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\tau}})h^{(\boldsymbol{\tau})} + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\tau}})(h^{(\boldsymbol{\tau})})^2\} \Rightarrow \mathcal{H}_2$, we here focus on the weak limit of

$$H_{1,n} := \sup_{\boldsymbol{s}_{\boldsymbol{\pi}}}\ \sup_{h^{(\boldsymbol{\pi})}} \{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\pi}})h^{(\boldsymbol{\pi})} + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\pi}})(h^{(\boldsymbol{\pi})})^2\}.$$

From the fact that for any $h\boldsymbol{d}_{\boldsymbol{\pi}}$ such that $h \in \mathbb{R}^{+}$ and $\boldsymbol{d}_{\boldsymbol{\pi}} \in \Delta(\boldsymbol{\pi}_*)$, there are $h^{(\boldsymbol{\lambda})} \in \mathbb{R}^{+}$, $h^{(\boldsymbol{\xi})} \in \mathbb{R}^{+}$, $\boldsymbol{s}_{\boldsymbol{\lambda}} \in \Delta(\boldsymbol{\lambda}_*)$, and $\boldsymbol{s}_{\boldsymbol{\pi}} \in \Delta(\boldsymbol{\pi}_*)$ such that $h\boldsymbol{d}_{\boldsymbol{\pi}} = [h^{(\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}}{}', h^{(\boldsymbol{\xi})}\boldsymbol{s}_{\boldsymbol{\xi}}{}']'$ and

$$\begin{aligned} H_{1,n} = \sup_{\{\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}\}} \sup_{\{h^{(\boldsymbol{\xi})},h^{(\boldsymbol{\lambda})}\}} & 2DL_n(\boldsymbol{\theta}_*;\boldsymbol{s}_{\boldsymbol{\xi}})h^{(\boldsymbol{\xi})} + 2DL_n(\boldsymbol{\theta}_*;\boldsymbol{s}_{\boldsymbol{\lambda}})h^{(\boldsymbol{\lambda})} + 2DL_n(\boldsymbol{\theta}_*;\boldsymbol{s}_{\boldsymbol{\lambda}};\boldsymbol{s}_{\boldsymbol{\xi}})h^{(\boldsymbol{\lambda})}h^{(\boldsymbol{\xi})} \\ & + D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{s}_{\boldsymbol{\xi}})(h^{(\boldsymbol{\xi})})^2 + D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{s}_{\boldsymbol{\lambda}})(h^{(\boldsymbol{\lambda})})^2, \end{aligned}$$

where $DL_n(\boldsymbol{\theta}_*;\boldsymbol{s}_{\boldsymbol{\lambda}};\boldsymbol{s}_{\boldsymbol{\xi}})$ is the directional derivative of $DL_n(\,\cdot\,;\boldsymbol{s}_{\boldsymbol{\lambda}})$ with respect to $\boldsymbol{s}_{\boldsymbol{\xi}}$ evaluated at $\boldsymbol{\theta}_*$. Given this, it is straightforward to apply the ULLN and FCLT to $H_{1,n}$ by Theorem 4. Therefore,

$$\begin{aligned} H_{1,n} \Rightarrow \mathcal{H}_1 = \sup_{\{\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}\}} \sup_{\{h^{(\boldsymbol{\xi})},\,h^{(\boldsymbol{\lambda})}\}} & 2\mathcal{Z}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})h^{(\boldsymbol{\xi})} + 2\boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{Z}^{(\boldsymbol{\lambda})}h^{(\boldsymbol{\lambda})} + 2\boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})h^{(\boldsymbol{\lambda})}h^{(\boldsymbol{\xi})} \\ & + A_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})(h^{(\boldsymbol{\xi})})^2 + \boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}}(h^{(\boldsymbol{\lambda})})^2, \end{aligned} \tag{26}$$

and there are four different possible cases for the solutions of the RHS of (26) from the fact that $h^{(\boldsymbol{\xi})} \geq 0$ and $h^{(\boldsymbol{\lambda})} \geq 0$. That is, for each $(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}})$ if we let $\widehat{h}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}})$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}})$ maximizes the RHS of (26) either (*i*) $\widehat{h}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$; (*ii*) $\widehat{h}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$; (*iii*) $\widehat{h}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$; or (*iv*) $\widehat{h}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$.

We examine the asymptotic distribution of each case one by one. First, if $\widehat{h}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$, the RHS of (26) is identical to

$$\sup_{\{\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}\}} [\mathcal{Z}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})\ \boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{Z}^{(\boldsymbol{\lambda})}] \begin{bmatrix} -A_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) & -\boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) \\ -\boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) & -\boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}} \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{Z}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) \\ \boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{Z}^{(\boldsymbol{\lambda})} \end{bmatrix},$$

and maximizing this with respect to $\boldsymbol{s}_{\boldsymbol{\lambda}}$ for a given $\boldsymbol{s}_{\boldsymbol{\xi}}$ yields $\widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})^2 + (\mathbf{Z}^{(\boldsymbol{\lambda})})'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})^{-1}(\mathbf{Z}^{(\boldsymbol{\lambda})})$. For $\widehat{h}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}})$ to be greater than zero, it is necessary that $\widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})$ is greater than zero, too. Second, if $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$ and $\widehat{h}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$, the RHS of (26) is identical to $2\boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{Z}^{(\boldsymbol{\lambda})} + \boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}}$, and maximizing this with respect to $\boldsymbol{s}_{\boldsymbol{\lambda}}$ leads to $(\mathbf{Z}^{(\boldsymbol{\lambda})})'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})(\mathbf{Z}^{(\boldsymbol{\lambda})})$ as its maximum. Also, $\widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})$ cannot be greater than zero. Otherwise, $\widehat{h}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$. Third, if $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$ and $\widehat{h}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$, we can consider $-\boldsymbol{s}_{\boldsymbol{\lambda}}$ as an alternative to $\boldsymbol{s}_{\boldsymbol{\lambda}}$ for the same $\boldsymbol{s}_{\boldsymbol{\xi}}$ while maximizing the RHS of (26) with respect to $\boldsymbol{s}_{\boldsymbol{\lambda}}$ from the fact that $\boldsymbol{\lambda}_*$ is an interior element. Thus, it modifies the given maximization to the first case. Finally, for given $\boldsymbol{s}_{\boldsymbol{\xi}}$ and $\boldsymbol{s}_{\boldsymbol{\lambda}}$, if $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$ and $\widehat{h}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$, $-\boldsymbol{s}_{\boldsymbol{\lambda}}$ can be considered an alternative, as well, and this modifies the maximization to the second case. Therefore, combining all these leads to that $\mathcal{H}_1 = \sup_{\boldsymbol{s}_{\boldsymbol{\xi}} \in \Delta(\boldsymbol{\xi}_*)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})]^2 + (\mathbf{Z}^{(\boldsymbol{\lambda})})'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})(\mathbf{Z}^{(\boldsymbol{\lambda})})$. ■

Proof of Lemma 4: (*i*) We exploit (25) further. First, applying the CMT to Theorem 4(*i*) shows that $(n^{-1/2}DL_n^{(\boldsymbol{\tau})}, n^{-1}D^2L_n^{(\boldsymbol{\tau})}) \Rightarrow (\mathcal{Z}^{(\boldsymbol{\tau})}, A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})})$. Thus,

$$\sup_{\sqrt{n}h^{(\boldsymbol{\tau})}} 2DL_n(\boldsymbol{\theta}_*;\boldsymbol{s_\tau})h^{(\boldsymbol{\tau})} + D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{s_\tau})(h^{(\boldsymbol{\tau})})^2 \Rightarrow \sup_{h^{(\boldsymbol{\tau})}\in\mathbb{R}^+} 2\mathcal{Z}^{(\boldsymbol{\tau})}(\boldsymbol{s_\tau})h^{(\boldsymbol{\tau})} + A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{s_\tau})(h^{(\boldsymbol{\tau})})^2,$$

so that $n^{1/2}\widehat{h}_n^{(\boldsymbol{\tau})}(\boldsymbol{s_\tau}) \Rightarrow \max[0, \{-A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{s_\tau})\}^{-1}\mathcal{Z}^{(\boldsymbol{\tau})}(\boldsymbol{s_\tau})] = \max[0, \mathcal{G}^{(\boldsymbol{\tau})}(\boldsymbol{s_\tau})]$. This holds even as a function of $\boldsymbol{s_\tau}$. That is, $n^{1/2}\widehat{h}_n^{(\boldsymbol{\tau})} \Rightarrow \max[0, \mathcal{G}^{(\boldsymbol{\tau})}]$.

Next, for any $h^{(\boldsymbol{\eta})}\boldsymbol{d_\eta}$ such that $h^{(\boldsymbol{\eta})} \in \mathbb{R}^+$ and $\boldsymbol{d_\eta} \in \Delta(\boldsymbol{\eta}_*)$, there are $h^{(\boldsymbol{\mu})} \in \mathbb{R}^+$, $h^{(\boldsymbol{\omega})} \in \mathbb{R}^+$, and $(\boldsymbol{s_\mu}, \boldsymbol{s_\omega}) \in \Delta(\boldsymbol{\mu}_*) \times \Delta(\boldsymbol{\omega}_*)$ such that $h^{(\boldsymbol{\eta})}\boldsymbol{d_\eta} = [h^{(\boldsymbol{\mu})}\boldsymbol{s_\mu}', h^{(\boldsymbol{\omega})}\boldsymbol{s_\omega}']'$. Therefore,

$$\begin{aligned}
&\sup_{h^{(\boldsymbol{\eta})}}\{2DL_n(\boldsymbol{\theta}_*;\boldsymbol{d_\eta})h^{(\boldsymbol{\eta})} + D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{d_\eta})(h^{(\boldsymbol{\eta})})^2\} \\
=& \sup_{(h^{(\boldsymbol{\mu})},h^{(\boldsymbol{\omega})})} 2DL_n(\boldsymbol{\theta}_*;\boldsymbol{s_\mu})h^{(\boldsymbol{\mu})} + 2DL_n(\boldsymbol{\theta}_*;\boldsymbol{s_\omega})h^{(\boldsymbol{\omega})} + 2DL_n(\boldsymbol{\theta}_*;\boldsymbol{s_\mu};\boldsymbol{s_\omega})h^{(\boldsymbol{\mu})}h^{(\boldsymbol{\omega})} \\
&\qquad + D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{s_\mu})(h^{(\boldsymbol{\mu})})^2 + D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{s_\omega})(h^{(\boldsymbol{\omega})})^2 \\
\Rightarrow& \sup_{(h^{(\boldsymbol{\mu})},h^{(\boldsymbol{\omega})})} 2\mathcal{Z}(\boldsymbol{s_\mu})h^{(\boldsymbol{\mu})} + 2\mathcal{Z}(\boldsymbol{s_\omega})h^{(\boldsymbol{\omega})} + 2A_*^{(\boldsymbol{\omega},\boldsymbol{\mu})}(\boldsymbol{s_\omega},\boldsymbol{s_\mu})h^{(\boldsymbol{\mu})}h^{(\boldsymbol{\omega})} \\
&\qquad + A_*^{(\boldsymbol{\mu},\boldsymbol{\mu})}(\boldsymbol{s_\mu})(h^{(\boldsymbol{\mu})})^2 + A_*^{(\boldsymbol{\omega},\boldsymbol{\omega})}(\boldsymbol{s_\omega})(h^{(\boldsymbol{\omega})})^2.
\end{aligned} \tag{27}$$

Given this, $h^{(\boldsymbol{\mu})}$ and $h^{(\boldsymbol{\omega})}$ on the RHS of (27) are present on the positive Euclidean line, so that there are four different possible inequality constraints. We examine each case one by one. First, if any inequality condition does not bind, $\sqrt{n}(\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega},\boldsymbol{s_\tau}), \widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega}\boldsymbol{s_\tau}))' \Rightarrow \boldsymbol{\mathcal{G}}^{(\boldsymbol{\eta})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega})$ by the standard first-order condition and Lemma 2. This occurs if each component of $\boldsymbol{\mathcal{G}}^{(\boldsymbol{\eta})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega})$ is strictly greater than zero. Second, if $\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega}) < 0$, it simply holds that $\widehat{h}^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega}) = \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega})]$. Thus, $\sqrt{n}(\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega},\boldsymbol{s_\tau}), \widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega},\boldsymbol{s_\tau}))' \Rightarrow (0, \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega})])'$. Likewise, if $\widehat{h}^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega}) = 0$ in the RHS of (27) because $\mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega}) < 0$, $\widehat{h}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega}) = \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega})]$. This implies that $\sqrt{n}(\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega},\boldsymbol{s_\tau}), \widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega},\boldsymbol{s_\tau}))' \Rightarrow (\max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}\ (\boldsymbol{s_\mu},\boldsymbol{s_\omega})], 0)'$. Fourth, it must be the case that $\sqrt{n}(\widehat{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega},\boldsymbol{s_\tau}), \widehat{h}_n^{(\boldsymbol{\omega})}(\boldsymbol{s_\mu},\boldsymbol{s_\omega},\boldsymbol{s_\tau}))' \Rightarrow (0,0)'$ for any other case. Therefore,

$$\sqrt{n}\begin{bmatrix} \widehat{h}_n^{(\boldsymbol{\mu})} \\ \widehat{h}_n^{(\boldsymbol{\omega})} \\ \widehat{h}_n^{(\boldsymbol{\tau})} \end{bmatrix} \Rightarrow \begin{bmatrix} \mathcal{G}^{(\boldsymbol{\mu})} \\ \mathcal{G}^{(\boldsymbol{\omega})} \\ 0 \end{bmatrix} \mathbf{1}_{\{\min[\mathcal{G}^{(\boldsymbol{\mu})},\mathcal{G}^{(\boldsymbol{\omega})}]\geq 0\}} + \begin{bmatrix} \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\omega})}<0\}} \\ \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\omega})}]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\mu})}<0\}} \\ \max[0, \mathcal{G}^{(\boldsymbol{\tau})}] \end{bmatrix}$$

by combining all these and applying Theorem 4(*i*). ■

Proof of Theorem 8: We can approximate (22) by a quadratic function and apply Lemma 4 to obtain that

$$\begin{aligned}
&\sup_{\{h^{(\mu)},\boldsymbol{\omega},\boldsymbol{\tau}\}} 2\{L_n(\boldsymbol{\mu}_0 + h^{(\boldsymbol{\mu})}\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{\omega}, \boldsymbol{\tau}) - L_n(\boldsymbol{\mu}_0, \boldsymbol{\omega}_*, \boldsymbol{\tau}_*)\} \\
&= \sup_{\{\boldsymbol{s}_{\boldsymbol{\omega}},\boldsymbol{s}_{\boldsymbol{\tau}}\}} \sup_{\{h^{(\mu)},h^{(\omega)},h^{(\tau)}\}} 2\{L_n(\boldsymbol{\mu}_0 + h^{(\boldsymbol{\mu})}\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{\omega}_* + h^{(\boldsymbol{\omega})}\boldsymbol{s}_{\boldsymbol{\omega}}, \boldsymbol{\tau}_* + h^{(\boldsymbol{\tau})}\boldsymbol{s}_{\boldsymbol{\tau}}) - L_n(\boldsymbol{\mu}_0, \boldsymbol{\omega}_*, \boldsymbol{\tau}_*)\} \\
&\Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\omega}}} \boldsymbol{\mathcal{G}}^{(\boldsymbol{\eta})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}})'(-\mathbf{A}_*^{(\boldsymbol{\eta},\boldsymbol{\eta})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}}))\boldsymbol{\mathcal{G}}^{(\boldsymbol{\eta})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \boldsymbol{s}_{\boldsymbol{\omega}})\mathbf{1}_{\{\min[\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}},\boldsymbol{s}_{\boldsymbol{\omega}}),\mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}},\boldsymbol{s}_{\boldsymbol{\omega}})]>0\}} \\
&\quad + \sup_{\boldsymbol{s}_{\boldsymbol{\omega}}} \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}})](-A_*^{(\boldsymbol{\mu},\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}))\max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}})]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}},\boldsymbol{s}_{\boldsymbol{\omega}})\geq 0>\mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}},\boldsymbol{s}_{\boldsymbol{\omega}})\}} \\
&\quad + \sup_{\boldsymbol{s}_{\boldsymbol{\omega}}} \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\omega}})](-A_*^{(\boldsymbol{\omega},\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\omega}}))\max[0, \dot{\mathcal{G}}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\omega}})]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}},\boldsymbol{s}_{\boldsymbol{\omega}})\geq 0>\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}},\boldsymbol{s}_{\boldsymbol{\omega}})\}} \\
&\quad + \sup_{\boldsymbol{s}_{\boldsymbol{\tau}}} \max[0, \mathcal{G}^{(\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}})](-A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}}))\max[0, \mathcal{G}^{(\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}})],
\end{aligned}$$

which is identical to

$$\begin{aligned}
&\boldsymbol{\mathcal{G}}^{(\boldsymbol{\eta})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))'(-\mathbf{A}_*^{(\boldsymbol{\eta},\boldsymbol{\eta})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}})))\boldsymbol{\mathcal{G}}^{(\boldsymbol{\eta})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))\mathbf{1}_{\{\min[\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}},\bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}})),\mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}},\bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))]>0\}} \\
&+ \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}})](-A_*^{(\boldsymbol{\mu},\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}))\max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}})]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}},\widetilde{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))\geq 0>\mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}},\widetilde{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))\}} \\
&+ \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})](-A_*^{(\boldsymbol{\omega},\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}))\max[0, \dot{\mathcal{G}}^{(\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}},\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\omega}}))\geq 0>\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}},\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{\omega}))\}} \\
&+ \sup_{\boldsymbol{s}_{\boldsymbol{\tau}}} \max[0, \mathcal{G}^{(\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}})](-A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}}))\max[0, \mathcal{G}^{(\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}})]
\end{aligned}$$

by the definition of $\bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}})$ and $\widetilde{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}})$, where for each $\boldsymbol{s}_{\boldsymbol{\mu}}$, $\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}})$ is such that

$$\begin{aligned}
&\max[0, \dot{\mathcal{G}}^{(\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})](-A_*^{(\boldsymbol{\omega},\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}))\max[0, \dot{\mathcal{G}}^{(\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}},\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))\geq 0>\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}},\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))\}} \\
&= \sup_{\boldsymbol{s}_{\boldsymbol{\omega}}\in\Delta(\boldsymbol{\omega}_*)} \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\omega}})](-A_*^{(\boldsymbol{\omega},\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\omega}}))\max[0, \dot{\mathcal{G}}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\omega}})]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}},\boldsymbol{s}_{\boldsymbol{\omega}})\geq 0>\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}},\boldsymbol{s}_{\boldsymbol{\omega}})\}}.
\end{aligned}$$

Furthermore, $\sqrt{n}\widetilde{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}})$ weakly converges to different limits $\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))$ and $\max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}})]$ depending on the events $\{\min[\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}})), \mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))] > 0\}$ and $\{\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \widetilde{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}})) \geq 0 > \mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \widetilde{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))\}$, respectively. Otherwise, $\sqrt{n}\widetilde{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}) \Rightarrow 0$. Thus,

$$\begin{aligned}
\sqrt{n}\widetilde{h}_n^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}) \Rightarrow &\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}}, \bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))\mathbf{1}_{\{\min[\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}},\bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}})),\mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}},\bar{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))]>0\}} \\
&+ \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}})]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\mu})}(\boldsymbol{s}_{\boldsymbol{\mu}},\widetilde{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))\geq 0>\mathcal{G}^{(\boldsymbol{\omega})}(\boldsymbol{s}_{\boldsymbol{\mu}},\widetilde{\boldsymbol{s}}_{\boldsymbol{\omega}}(\boldsymbol{s}_{\boldsymbol{\mu}}))\}}.
\end{aligned}$$

This also holds as a function of $\boldsymbol{s}_{\boldsymbol{\mu}}$, so that Theorem 8 follows by applying the CMT to the Wald test statistic again. ■

Proof of Corollary 3: We already proved that $\sqrt{n}\widetilde{h}_n^{(\boldsymbol{\xi})}(\boldsymbol{s_\mu}) \Rightarrow \max[0, (-\widetilde{A}^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s_\xi}))^{-1/2}\widetilde{\mathcal{Y}}^{(\boldsymbol{\xi})}(\boldsymbol{s_\xi})]$ under the same environment in the proof of Corollary 2. Applying the CMT to the Wald statistic completes the proof. ■

Proof of Theorem 9: Before proving the claim, we suppose that $\boldsymbol{\tau}_*$ is known for brevity. By Theorem 5, this supposition simplifies our proof without losing generality.

To show the given claim, we derive the convergence limit of each component constituting the LM test statistic. First, there is n^* a.s.$-\mathbb{P}$ such that if $n > n^*$, $\Delta(\ddot{\boldsymbol{\omega}}_n) = \Delta(\boldsymbol{\omega}_*)$. We note that $\boldsymbol{\omega}_*$ is an interior element by Assumption 11(*i*), so that $\Delta(\boldsymbol{\omega}_*) = \{\mathbf{x} \in \mathbb{R}^{r_\omega} : \|\mathbf{x}\| = 1\}$, and further for an open ball with radius $\varepsilon > 0$ such that $B(\boldsymbol{\omega}_*, \varepsilon) \subset \boldsymbol{\Omega}$, there is $n(\varepsilon)$ a.s.–$\mathbb{P}$, so that if $n > n(\varepsilon)$, $\ddot{\omega}_n \in B(\boldsymbol{\omega}_*, \varepsilon)$ by Theorem 6(*i*). This implies that $\ddot{\omega}_n$ is an interior element, too. Thus, if we let $n^* > n(\varepsilon)$, $\Delta(\ddot{\boldsymbol{\omega}}_n) = \{\mathbf{x} \in \mathbb{R}^{r_\omega} : \|\mathbf{x}\| = 1\}$, which is also $\Delta(\boldsymbol{\omega}_*)$. Second, $n^{-1/2}DL_n(\ddot{\boldsymbol{\theta}}_n; \cdot) \Rightarrow \ddot{\mathcal{Z}}^{(\boldsymbol{\mu})}(\cdot\,; \ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})$. Applying the mean-value theorem shows that for each $\boldsymbol{s_\mu}$ there is $\dot{\boldsymbol{\omega}}(\boldsymbol{s_\mu})$ such that

$$DL_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s_\mu}) - DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\mu}) = \{DL_n(\boldsymbol{\mu}_0, \dot{\boldsymbol{\omega}}_n(\boldsymbol{s_\mu}), \boldsymbol{\tau}_*; \boldsymbol{s_\mu}; \ddot{\boldsymbol{s}}_{\boldsymbol{\omega},n})\}\{\ddot{h}_n^{(\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega},n})\}$$
$$= DL_n(\boldsymbol{\mu}_0, \dot{\boldsymbol{\omega}}_n(\boldsymbol{s_\mu}), \boldsymbol{\tau}_*; \boldsymbol{s_\mu}; \ddot{\boldsymbol{s}}_{\boldsymbol{\omega},n})\{-D^2L_n(\boldsymbol{\mu}_0, \bar{\boldsymbol{\omega}}_n(\boldsymbol{s_\mu}), \boldsymbol{\tau}_*; \ddot{\boldsymbol{s}}_{\boldsymbol{\omega},n})\}^{-1}DL_n(\boldsymbol{\theta}_*; \ddot{\boldsymbol{s}}_{\boldsymbol{\omega},n}),$$

where we define $(\ddot{h}_n^{(\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega},n}), \ddot{\boldsymbol{s}}_{\boldsymbol{\omega},n})$ to satisfy that $L_n(\boldsymbol{\mu}_0, \boldsymbol{\omega}_* + \ddot{h}_n^{(\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega},n})\ddot{\boldsymbol{s}}_{\boldsymbol{\omega},n}, \boldsymbol{\tau}_*) = \sup_{\boldsymbol{s_\omega}} \sup_{h^{(\boldsymbol{\omega})}} L_n(\boldsymbol{\mu}_0, \boldsymbol{\omega}_* + h^{(\boldsymbol{\omega})}\boldsymbol{s_\omega}, \boldsymbol{\tau}_*)$, and the last equality follows from the mean-value theorem: there is $\bar{\boldsymbol{\omega}}_n(\boldsymbol{s_\mu})$ such that (8) holds. Given this and Theorem 6(*i*), we can also apply the ULLN to $n^{-1}DL_n(\boldsymbol{\mu}_0, \dot{\boldsymbol{\omega}}_n(\boldsymbol{s_\mu}), \boldsymbol{\tau}_*; \cdot\,; \cdot)$ and $n^{-1}D^2L_n(\boldsymbol{\mu}_0, \bar{\boldsymbol{\omega}}_n(\boldsymbol{s_\mu}), \boldsymbol{\tau}_*; \cdot)$ to obtain $A_*^{(\boldsymbol{\mu},\boldsymbol{\omega})}$ and $A_*^{(\boldsymbol{\omega},\boldsymbol{\omega})}$ as their probability limits, respectively. Furthermore, it trivially holds that $n^{-1/2}(DL_n\,(\boldsymbol{\theta}_*; \cdot), DL_n(\boldsymbol{\theta}_*; \ddot{\boldsymbol{s}}_{\boldsymbol{\omega},n})) \Rightarrow (\mathcal{Z}^{(\boldsymbol{\mu})}, \mathcal{Z}^{(\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}}))$ by the facts that $n^{-1/2}\ (DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\mu}), DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\omega})) \Rightarrow (\mathcal{Z}^{(\boldsymbol{\mu})}, \mathcal{Z}^{(\boldsymbol{\omega})})$ as functions of $(\boldsymbol{s_\mu}, \boldsymbol{s_\omega})$ and that $\max[0, DL_n(\boldsymbol{\theta}_*; \ddot{\boldsymbol{s}}_{\boldsymbol{\omega},n})]^2\{-D^2L_n(\boldsymbol{\theta}_*; \ddot{\boldsymbol{s}}_{\boldsymbol{\omega},n})\}^{-1} \Rightarrow \max[0, \mathcal{Y}^{(\boldsymbol{\omega})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})]^2$. Thus, we can obtain that $n^{-1/2}\ DL_n(\ddot{\boldsymbol{\theta}}_n; \cdot) \Rightarrow \ddot{\mathcal{Z}}^{(\boldsymbol{\mu})}(\cdot\,; \ddot{\boldsymbol{s}}_{\boldsymbol{\omega}})$ by applying the CMT. Third, we can also apply the ULLN abd obtain that $\sup_{(\boldsymbol{s_\mu}, \boldsymbol{s_\omega})} |n^{-1}\widetilde{D}^2(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s_\mu}, \boldsymbol{s_\omega}) - \widetilde{A}_*^{(\boldsymbol{\mu},\boldsymbol{\mu})}(\boldsymbol{s_\mu}, \boldsymbol{s_\omega})| \to 0$ a.s.–$\mathbb{P}$. Given these three facts and Assumption 11(*ii*), the desired result now holds by the CMT. ■

Proof of Corollary 4: From Theorem 9, it follows that

$$\mathcal{LM}_n \Rightarrow \sup_{\boldsymbol{s_\xi} \in \Delta(\boldsymbol{\xi}_0)} \left(\frac{\max[0, \ddot{\mathcal{Z}}^{(\boldsymbol{\xi})}(\boldsymbol{s_\xi}; \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}})]^2}{\inf_{\boldsymbol{s_\lambda} \in \Delta(\boldsymbol{\lambda}_*)}\{-\widetilde{A}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s_\xi}, \boldsymbol{s_\lambda})\}} \right),$$

and we separately examine the asymptotic behaviors of the numerator and the denominator in the parenthesis. First,

$$\begin{aligned}
&\inf_{\boldsymbol{s}_{\boldsymbol{\lambda}}\in\Delta(\boldsymbol{\lambda}_*)}\{-\widetilde{A}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}},\boldsymbol{s}_{\boldsymbol{\lambda}})\}\\
&\quad= - A_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) - \sup_{\boldsymbol{s}_{\boldsymbol{\lambda}}\in\Delta(\boldsymbol{\lambda}_*)}\{-\mathbf{A}_*^{(\boldsymbol{\xi},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}})'\boldsymbol{s}_{\boldsymbol{\lambda}}\}\{-\boldsymbol{s}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}}\}^{-1}\{-\boldsymbol{s}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})\}\\
&\quad= - \{A_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) - \mathbf{A}_*^{(\boldsymbol{\xi},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}})'\{\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\}^{-1}\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})\} = -\widetilde{A}_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}).
\end{aligned}$$

Next, $\ddot{\mathcal{Z}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}};\ddot{\mathbf{s}}_{\boldsymbol{\lambda}}) = \mathcal{Z}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) - \{-\mathbf{A}_*^{(\boldsymbol{\xi},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}})'\ddot{\mathbf{s}}_{\boldsymbol{\lambda}}\}\{-\ddot{\mathbf{s}}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\ddot{\mathbf{s}}_{\boldsymbol{\lambda}}\}^{-1}\{\ddot{\mathbf{s}}_{\boldsymbol{\lambda}}'\mathbf{Z}^{(\boldsymbol{\lambda})}\}$, and we already proved in the proof of Corollary 1(v) that $\ddot{\mathbf{s}}_{\boldsymbol{\lambda}}\{-\ddot{\mathbf{s}}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\ddot{\mathbf{s}}_{\boldsymbol{\lambda}}\}^{-1}\{\ddot{\mathbf{s}}_{\boldsymbol{\lambda}}'\mathbf{Z}^{(\boldsymbol{\lambda})}\} = \{-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\}^{-1}\mathbf{Z}^{(\boldsymbol{\lambda})}$, implying that $\ddot{\mathcal{Z}}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}};\ddot{\mathbf{s}}_{\boldsymbol{\lambda}}) = \mathcal{Z}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) - \{-\mathbf{A}_*^{(\boldsymbol{\xi},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\xi}})'\}\{-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\}^{-1}\mathbf{Z}^{(\boldsymbol{\lambda})}$. These facts imply the desired result. ■

References

AIGNER, D., LOVELL, C., AND SCHMIDT, P. (1977): "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, 6, 21–37.

ANDREWS, D. (1994): "Empirical Process Methods in Econometrics," in R. Engle and D. McFadden, eds. *Handbook of Econometrics*, 4, 2247–2294.

ANDREWS, D. (1999): "Estimation When a Parameter is on a Boundary," *Econometrica*, 67, 543–563.

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

BILLINGSLEY, P. (1999): *Convergence of Probability Measures*. New York: Wiley.

BOX, G. AND COX, D. (1964): "An Analysis of Transformations," *Journal of the Royal Statistical Society*, *Series B*, 26, 211–252.

CHERNOFF, H. (1954): "On the Distribution of the Likelihood Ratio," *The Annals of Mathematical Statistics*, 54, 573–578.

CHENG, R., B. EVANS, AND T. ILES. (1992): "Embedded Models in Non-Linear Regression," *Journal of the Royal Statistical Society*, *Series B*, 54, 877–888.

CHO, J. (2011): "Quasi-Maximum Likelihood Estimation Revisited Using the Distance and Direction Method," *Journal of Economic Theory and Econometrics*, 23, 89–112.

CHO, J. AND WHITE, H. (2007): "Testing for Regime Switching," *Econometrica*, 75, 1671–1720.

CHO, J. AND WHITE, H. (2010): "Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Econometrics*, 157, 458–480.

CHO, J. AND WHITE, H. (2011): "Generalized Runs Tests for the IID Hypothesis," *Journal of Econometrics*, 162, 326–344.

CHO, J., CHEONG, T., AND WHITE, H. (2011): "Experience with the Weighted Bootstrap in Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Economic Theory and Econometrics*, 22:2, 60–91.

CHO, J., ISHIDA, I., AND WHITE, H. (2011): "Revisiting Tests for Neglected Nonlinearity Using Artificial Neural Networks," *Neural Computation*, 23, 1133–1186.

CHO, J., ISHIDA, I., AND WHITE, H. (2013): "Testing for Neglected Nonlinearity Using Twofold Unidentified Models under the Null and Hexic Expansions," *Essays in Nonolinear Time Series Econometrics*. Forthcoming. Oxford: Oxford University Press.

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

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

DOUKHAN, P., P. MASSART, AND E. RIO. (1995): "Invariance Principles for Absolutely Regular Empirical Processes," *Annales de l'Institut Henri Poincaré, Probabilites et Statistiques*, 31, 393–427.

DUOFUR, J.-M. (2006): "Monte Carlo Tests with Nuisance Parameters: A General Approach to Fininte-Sample Inference and Nonstandard Asymptotics in Ecoonometrics," *Journal of Econometrics*, 133, 443–477.

DUTTA, S., NARASIMHAN, O., AND RAJIV, S. (1999): "Success in High-Technology Markets: Is Marketing Capability Critical?" *Marketing Science*, 18, 547–598.

GALLANT, R. AND WHITE, H. (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. New York: Basil Blackwell.

HABIB, M. AND LJUNGQVIST, A. (2005): "Firm Values and Managerial Incentives: A Stochastic Frontier Approach," *Journal of Business*, 78, 2053–2094.

HANSEN, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HANSEN, B. (1996a): "Stochastic Equicontinuity for Unbounded Dependent Heterigeneous Arrays," *Econometric Theory*, 12, 347–359.

HANSEN, B. (1996b): "Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–440.

KING, M. AND T. SHIVELY. (1993): "Locally Optimal Testing When a Nuisance Parameter is Present Only Under the Alternative," *The Review of Economics and Statistics*, 75, 1–7.

MCLEISH, D. (1974): "Dependent Central Limit Theorem and Invariance Principles," *The Annals of Probability*, 2, 620–628.

NEWEY, W. AND WEST, K. (1987): "Hypothesis Testing with Efficient Menthod of Moment Estimation,“ *International Economic Review*, 28, 777–787.

ROSENBERG, B. (1973): "The Analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regression," *Annals of Economic and Social Measurement*, 2, 399–428.

RUDIN, W. (1976): *Principles of Mathematical Analysis*. New York: McGraw-Hill.

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

SELF, S. AND LIANG, K. (1987): "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions," *Journal of the American Statistical Association*, 82, 605–610.

STEVENSON, R. (1980): "Likelihood Functions for Generalized Stochastic Frontier Estimation," *Journal of Econometrics*, 13, 57–66.

STOUT, W. (1974): *Almost Sure Convergence*. New York: Academic Press.

TROUTMAN, J. (1996): *Variational Calculus and Optimal Control*. New York: Springer-Verlag.

VAN DER VAART, A. AND WELLER, J. (1996): *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer-Verlag.

WALD, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54, 426–482.

WALD, A. (1949): "Note on the Consistency of the Maximum Likelihood Estimate," *The Annals of Mathematical Statistics*, 20, 596–601.

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

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

WHITE, H. AND J.S. CHO (2012): "Higher-Order Approximations for Testing Neglected Nonlinearity," *Neural Computation*, 24, 273–287.

WOOLDRIDGE, J. AND WHITE, H. (1988): "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory*, 4, 210–230.

ZELLNER, A., KMENTA, J., AND DRÈZE, J. (1966): "Specification and Estimation of Cobb-Douglas Production Functions," *Econometrica*, 34, 784–795.

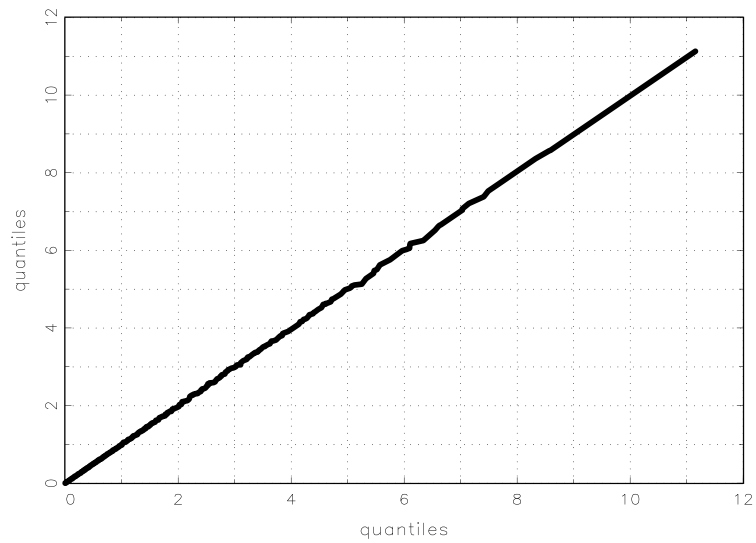

Figure 1: Q-Q PLOT BETWEEN $\sup \max[0, \ddot{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2)]^2$ AND $\sup \max[0, \widehat{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2)]^2$. The horizontal line stands for the empirical quantiles of $\sup \max[0, \widehat{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2)]^2$, and the vertical line stands for the empirical quantiles of $\sup \max[0, \widehat{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2)]^2$. Here, $\ddot{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2)$ is computed by assuming that $E[W_t^2]$ and $\mathrm{var}(W_t^2)$ are known, whereas $\widehat{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2)$ is computed by estimating $E[W_t^2]$ and $\mathrm{var}(W_t^2)$ by the method of moments. The sample size is 500, and the number of iterations is 2,000. The 45-degree line shows that estimating $E[W_t^2]$ and $\mathrm{var}(W_t^2)$ by the method of moments does not modify the asymptotic distribution.

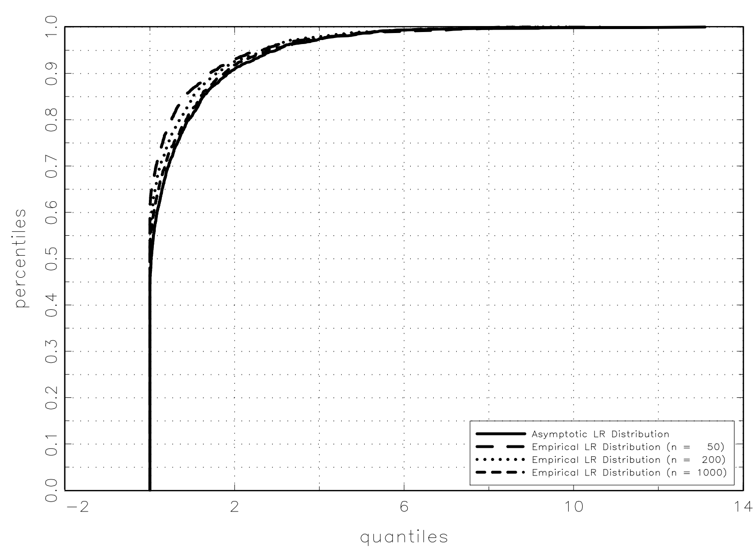

Figure 2: EMPIRICAL AND ASYMPTOTIC DISTRIBUTIONS OF THE QLR TEST STATISTIC. This figure shows the asymptotic null distribution of the QLR test statistic, which is obtained by the simulation method, and the empirical distributions of the QLR test statistic for various sample sizes: n = 50, 200, and 1,000. The number of iterations for obtaining the empirical distributions is 2,000. We can see that the empirical distribution approaches the asymptotic null distribution as the sample size increases. The value of the QLR test statistic is computed by estimating α_*, σ_*^2, θ_{1*}, and θ_{2*}.

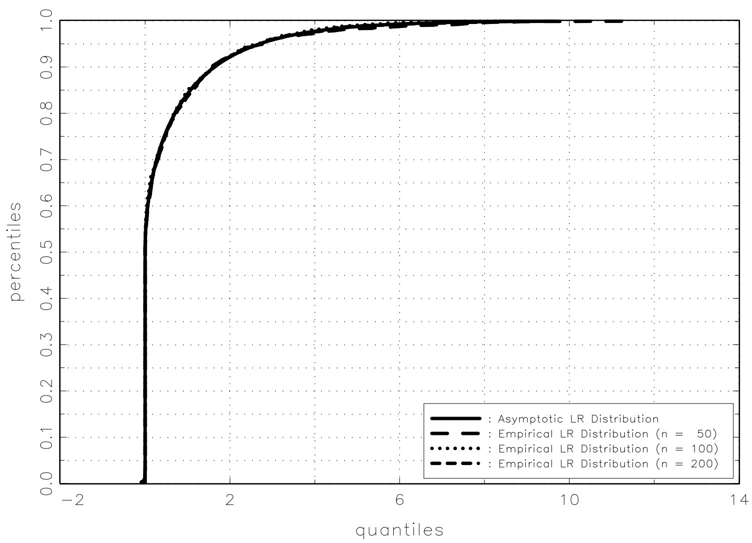

Figure 3: EMPIRICAL AND ASYMPTOTIC DISTRIBUTIONS OF THE QLR TEST STATISTIC. This figure shows the asymptotic null distribution of the QLR test statistic, which is obtained as $\max[0, -Z]^2$, and the empirical distributions of the QLR test statistic for various sample sizes: $n = 50$, 100, and 500. The number of iterations for obtaining the empirical distributions is 2,000. We can see that the empirical distributions almost overlap with the asymptotic null distribution even when the sample size is as small as 50.