Directionally Differentiable Econometric Models

JIN SEO CHO

School of Economics

Yonsei University

50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, 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

First version: August, 2002. This version: October, 2016

Abstract

The current paper examines the limit distribution of the quasi-maximum likelihood estimator obtained from a directionally differentiable quasi-likelihood function and represents its limit distribution as a functional of a Gaussian stochastic process indexed by direction. By this, the standard analysis that assumes a differentiable quasi-likelihood function is treated as a special case of our analysis. We also redefine the standard quasi-likelihood ratio, Wald, and Lagrange multiplier test statistics so that their null limit behaviors are regular under our model framework.

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

JEL Classification: C12, C13, C22, C32.

Acknowledgements: The co-editor, Yoon-Jae Whang, and two anonymous referees provided very helpful comments for which we are most grateful. The second author (Halbert White) passed away while the submission 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

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 the log-likelihood function by Taylor's expansion that requires differentiability.

Many important econometric models are not estimated by differentiable quasi-likelihood functions. For example, the likelihood function examined by King and Shively (1993) is not differentiable. They attempt to resolve the so-called Davies's (1977, 1987) identification problem by reparameterizing the likelihood function using the polar coordinates. The consequent likelihood function, however, is not differentiable (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 be conducted in the conventional way due to the non-differentiability of the likelihood function. In addition to these, many quasi-likelihood functions in prior literature are not D, and their model analysis should be different from the standard case.

The goal of this paper is therefore to extend the model scope to include non-differentiable models for quasi-likelihood estimation. Specifically, we suppose that model parameters are estimated by maximizing D-D quasi-likelihood functions. As it turns out, D quasi-likelihood function is a special case of D-D quasi-likelihood functions, so that the analysis assuming D-D quasi-likelihood function generalizes what is obtained by assuming D quasi-likelihood function condition. In more detail, the generalization is achieved by resorting to Billingsley's (1999) asymptotically tight probability measure condition. Each direction around the parameter of interest is regarded as an index indicating a particular value of directional derivatives. The asymptotically tight probability measure condition 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) to the random functions indexed by direction. The large sample distribution of quasi-maximum likelihood (QML) estimator (or M-estimator) obtained by this process turns out to generalize what is obtained by D quasi-likelihood function.

Another goal of this study is to provide test statistics that can be properly used for D-D quasi-likelihood models. We achieve this goal by refining the conventional quasi-likelihood ratio (QLR), Wald, and Lagrange multiplier (LM) test statistics in the D-D function context, and we show that the three test statistics are asymptotically equivalent under the null hypothesis and mild regularity conditions detailed below. By this, we desire to achieve the dual purpose to estimate and infer D-D econometric models.

Our D-D model analysis is applicable to a number of empirically popular econometric models. As an illustration of our analysis, we revisit King and Shively's (1993) reparameterized model and demonstrate that our analysis provides a vehicle that can be efficiently applied to their model analysis. In addition to this, we also collect other model analyses to the Supplement to this study and demonstrate its usefulness (see Cho and White, 2016). They include Aigner, Lovell, and Schmidt's (1977) stochastic frontier production function model and the Box-Cox transformation. The standard generalized method of moments (GMM) estimation is also revisited using the analysis of D-D quasi-likelihood functions.

The approach of the current study is related to the works in prior literature. First, Pollard (1985) examines stochastically differentiable quasi-likelihood functions that are not D although their population analogs are D. The D-D quasi-likelihood function is not stochastically differentiable because the population quasi-likelihood function is D-D, let alone its sample quasi-likelihood function. Second, Andrews (2001) examines inferencing when there is an unidentified parameter under a maintained null hypothesis that is possibly on the boundary of the parameter space. Although D-D quasi-likelihood functions may be derived by reparameterizing the parameter space as in King and Shively (1993), the analysis of here does not assume only D-D quasi-likelihood functions obtained through reparameterization. General D-D quasi-likelihood functions are assumed so that the likelihood function framework as in Aigner, Lovell, and Schmidt (1977) can be properly analyzed using the analysis of here. Finally, Fang and Santos (2014) examine a D-D transform of a consistent estimator and note that D transform is necessary and sufficient for a valid application of the standard bootstrap to the transformation. Instead of assuming the presence of a consistent estimator, we examine a consistent QML estimator obtained from D-D quasi-likelihood functions for our study.

The plan of this paper is as follows. In Section 2, D-D functions are defined and examined, and the D quasi-likelihood function is investigated as a special case of D-D quasi-likelihood functions. We also provide regularity conditions for D-D quasi-likelihood functions and consider the limit distribution of the QML estimator under these conditions. Section 3 considers data inferences using D-D quasi-likelihood functions. For this, we redefine the standard QLR, Wald, and LM test statistics and derive their null limit distributions. Furthermore, Section 4 examines King and Shively's (1993) reparameterized model for demonstration purpose. In the same section, we also conduct Monte Carlo experiments using the same model. Section 5 offers concluding remarks, and formal mathematical proofs are collected to the 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 Directionally Differentiable Quasi-Likelihood Functions

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.* $\square$

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. Next, we suppose the following quasi-likelihood function that is assumed to capture the DGP properties:

Assumption 2 (Qausi-Likelihood Function)**.** *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 quasi-likelihood function 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,* $\mathbf{X}^t$ *denotes* $(\mathbf{X}_1, \cdots, \mathbf{X}_t)$*, and* $\boldsymbol{\Theta}$ *is a compact and convex set in* $\mathbb{R}^r$ *with* $r \in \mathbb{N}$*.* $\square$

This quasi-likelihood function condition is widely used in the literature, and the quasi-maximum likelihood (QML) estimator is a typical estimator obtained from Assumption 2 that captures the key properties of the DGP. We further characterize the DGP by

Assumption 3 (Existence and Identification)**.** *(i) For each* $\boldsymbol{\theta}$*,* $n^{-1}E[L_n(\boldsymbol{\theta})]$ *exists in* $\mathbb{R}$ *and is finite for any n; (ii) For a unique* $\boldsymbol{\theta}_* \in \boldsymbol{\Theta}$*,* $E[n^{-1}L_n(\cdot)]$ *is maximized at* $\boldsymbol{\theta}_* \in \boldsymbol{\Theta}$ *for any n.* $\square$

Using Assumptions 2 and 3, we let the QML estimator $\widehat{\boldsymbol{\theta}}_n$ be such that $L_n(\widehat{\boldsymbol{\theta}}_n) = \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L_n(\boldsymbol{\theta})$.

Several remarks are warranted on these assumptions. 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 renders the key aspects of D-D quasi-likelihood functions obscure. Therefore, we highlight the D-D quasi-likelihood function analysis by using identified models. Second, $\boldsymbol{\theta}_*$ can be on the boundary of $\boldsymbol{\Theta}$ as often ensued by the reparameterization method of King and Shively (1993). Assumption 3(*ii*) permits this. Finally, for notational simplicity, we abbreviate $\ell_t(\,\cdot\,; \mathbf{X}^t)$ into $\ell_t(\cdot)$ from now.

Given Assumptions 1 to 3, the QML estimator is consistent, viz., $\widehat{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}_*$ a.s.–$\mathbb{P}$, and this 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 is that the differentiability condition is not necessary for the consistency of the

QML estimator as Wald (1949) points out. On the other hand, the limit distribution of the QML estimator is dependent upon the differentiability as we discuss in the next subsection.

2.1 Directional Differentiability

The smoothness condition of D-D functions is important for the limit distribution of $\widehat{\boldsymbol{\theta}}_n$. In this subsection, we define the D-D quasi-likelihood function and characterize D quasi-likelihood functions using the D-D quasi-likelihood function.

Definition 1 (D-D Functions)**.** *(i)* $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ *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)* $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ *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)* $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ *is said to be* D-D on $\boldsymbol{\Theta}$*, if for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*,* f *is D-D on* $\Delta(\boldsymbol{\theta})$*.* □

Several remarks are in order on this definition. First, note that the definition of D-D function is weaker than that of D function. D-D functions can have different directional derivatives that is nonlinearly dependent upon the direction, and there can be a continuum number of directions if r is greater than unity. On the other hand, if $f(\cdot)$ is D, $Df(\boldsymbol{\theta}; \boldsymbol{d})$ is represented as a linear combination of r 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 limit distribution of the QML 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 other than 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 $\Delta(\boldsymbol{\theta})$. We proceed our discussions with $\Delta(\boldsymbol{\theta})$.

There exists a regular relationship between D-D and D functions as Troutman (1996, p. 122) describes. That is, if (*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 $\boldsymbol{\theta}$, $Df(\boldsymbol{\theta}; \boldsymbol{d})$ is continuous and linear in $\boldsymbol{d}$, then $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is D on $\boldsymbol{\Theta}$. 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 investing D-D quasi-likelihood functions, we provide the definition of the twice D-D function that also plays another key role in our analysis.

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

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

Note that the first-order directional differentiability is necessary for defining the twice D-D function. Furthermore, for a twice D-D function to be twice D, it is necessary for $D^2 f(\boldsymbol{\theta}; \widetilde{\boldsymbol{d}}; \boldsymbol{d})$ to be bilinear in $\boldsymbol{d}$ and $\widetilde{\boldsymbol{d}}$. We state on this in the Supplement more precisely. From now, we denote $D^2 f(\boldsymbol{\theta}; \widetilde{\boldsymbol{d}}; \boldsymbol{d})$ as $D^2 f(\boldsymbol{\theta}; \boldsymbol{d})$ if $\boldsymbol{d} = \widetilde{\boldsymbol{d}}$.

2.2 Asymptotic Distribution of the QML Estimator

As pointed out in the previous subsection, the most significant difference between D-D and D functions lies in the linearity condition of the directional derivative in $\boldsymbol{d}$. Further regularity conditions are provided for the D-D quasi-likelihood function.

Assumption 4 (D-D Quasi-Likelihood Function)**.** $\ell_t : \boldsymbol{\Theta} \mapsto \mathbb{R}$ *is twice 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.*$-\mathbb{P}$*.* $\square$

We use Assumption 4 to approximate D-D quasi-likelihood functions by a second-order directional Taylor expansion for each direction. For this goal, the following regularity conditions are also imposed:

Assumption 5 (Regular D-Difflity)**.** *(i) For each* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*,* $D\ell_t(\boldsymbol{\theta}; \cdot)$ *and* $D^2 \ell_t(\boldsymbol{\theta}; \cdot)$ *are continuous on* $\Delta(\boldsymbol{\theta})$ *a.s.–*$\mathbb{P}$*; (ii) For each* $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}$*,* $|D\ell_t(\boldsymbol{\theta}; \boldsymbol{d}) - D\ell_t(\boldsymbol{\theta}'; \boldsymbol{d})| \leq M_t \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|$ *and* $|D^2\ell_t(\boldsymbol{\theta}; \boldsymbol{d}) - D^2\ell_t(\boldsymbol{\theta}'; \boldsymbol{d})| \leq M_t \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|$ *uniformly on* $\Delta(\boldsymbol{\theta}) \cap \Delta(\boldsymbol{\theta}')$*, where* $\{M_t\}$ *is a sequence of stationary and ergodic 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* $|D\ell_t(\boldsymbol{\theta}; \boldsymbol{d}_1) - D\ell_t(\boldsymbol{\theta}; \boldsymbol{d}_2)| \leq M_t \|\boldsymbol{d}_1 - \boldsymbol{d}_2\|^{\lambda}$ *and* $|D^2\ell_t(\boldsymbol{\theta}; \boldsymbol{d}_1) - D^2\ell_t(\boldsymbol{\theta}; \boldsymbol{d}_2)| \leq M_t \|\boldsymbol{d}_1 - \boldsymbol{d}_2\|^{\lambda}$*.* $\square$

The example in Section 4 satisfies Assumptions 4 and 5. Here, Assumption 5(*iii*) is assumed to apply the asymptotic tightness and the ULLN to the first and second-order directional derivative, respectively. We detail the asymptotic tightness and the ULLN below, when they are more relevant. If Assumption 5(*iii*) is replaced by the following stronger Assumption 5(*iii*)*, the quasi-likelihood function is twice D a.s.–$\mathbb{P}$ as examined in the Supplement:

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

We let Assumption 5* denote Assumptions 5(*i*, *ii*, and $iii^\star$) going forward when D quasi-likelihood functions are referred. Unless otherwise stated, Assumption 5 stands for Assumptions 5(*i*, *ii*, and *iii*).

We impose further regularity conditions for the limit distribution of the QML estimator.

Assumption 6 (CLT)**.** *(i) For any t, $E[D\ell_t(\boldsymbol{\theta}_*; \boldsymbol{d})] = 0$ uniformly on $\Delta(\boldsymbol{\theta}_*)$; (ii) For any n, $A_*(\boldsymbol{d}) := E[n^{-1}D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{d})]$ is strictly negative and finite uniformly on $\Delta(\boldsymbol{\theta}_*)$; (iii) For any n, $B_*(\boldsymbol{d}, \boldsymbol{d})$ is strictly positive and finite uniformly on $\Delta(\boldsymbol{\theta}_*)$, 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 $\{\mathbf{X}_t \in \mathbb{R}^k : t \in \mathbb{N}\}$ 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 central limit theorem (CLT). Note that Assumption 6(*i*) may not hold uniformly in $\boldsymbol{d}$ if $\boldsymbol{\theta}_*$ is a boundary point of $\boldsymbol{\Theta}$: for some $\boldsymbol{d}$, $E[D\ell_t(\boldsymbol{\theta}_*; \boldsymbol{d})]$ can be strictly negative if $\boldsymbol{\theta}_*$ is a boundary point, although $\boldsymbol{\theta}_*$ maximizes $E[\ell_t(\cdot)]$. If so, the test statistics considered below can be degenerate. We impose Assumption 6(*i*) and prevents this. On the other hand, if $\boldsymbol{\theta}_*$ is an interior element, Assumption 6(*i*) can be derived from the condition that $\boldsymbol{\theta}_*$ maximizes $E[\ell_t(\cdot)]$. 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 the 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 the mixingale process. Hansen (1996a) generalizes this and provides the regularity conditions for the asymptotic tightness of Lipschitz continuous functions. In essence, Assumption 6 is used to apply both CLT and asymptotic tightness to $n^{-1/2}DL_n(\boldsymbol{\theta}_*; \,\cdot\,)$. Finally, the focus of here is different from that of Fang and Santos (2014) in which a directionally differentiable transform of a consistent estimator is examined, whereas our QML estimator maximizes a directionally differentiable quasi-likelihood function.

The limit distribution of $\widehat{\boldsymbol{\theta}}_n$ is obtained by using the regularity conditions provided so far. Our plan is to approximate the quasi-likelihood function by a second-order directional Taylor expansion for each direction and relate this to other directional Taylor expansions. Specifically, we first derive the limit distribution of $\widehat{\boldsymbol{\theta}}_n$ for a particular direction $\boldsymbol{d}$ and call it *directional QML estimator* (*DQML estimator*). Next, we examine how this is interrelated with another DQML estimator obtained by using a different direction. For this examination, we first let $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})$ denote the DQML 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 DQML estimator is constrained by $\boldsymbol{d}$: for given $\boldsymbol{d}$, $\boldsymbol{\Theta}_*(\boldsymbol{d})$ is a straight line starting from $\boldsymbol{\theta}_*$ with its endpoint 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 DQML 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 limit 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 $L_n(\boldsymbol{\theta}_* + (\cdot)\boldsymbol{d})$, so that for some $\bar{\boldsymbol{\theta}}_n(\boldsymbol{d}) \in \boldsymbol{\Theta}(\boldsymbol{d})$, the following holds by the mean-value theorem:

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

This approximation can be carried out on $H_*(\boldsymbol{d})$ because $\widehat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_* + o_{\mathbb{P}}(1)$, so that for each $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$,

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

where $n^{-1/2}DL_n(\boldsymbol{\theta}_*; \boldsymbol{d})$ and $n^{-1}D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{d})$ are such that $\{n^{-1/2}DL_n(\boldsymbol{\theta}_*; \boldsymbol{d}), n^{-1}D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{d})\} \Rightarrow \{\mathcal{Z}(\boldsymbol{d}), A_*(\boldsymbol{d})\}$ as shown in the Appendix. Here, $\widetilde{h}$ captures the limit behavior of $\sqrt{n}h$, and the argument of the right side in (2) is simply obtained as $\max[0, \mathcal{G}(\boldsymbol{d})]$, where $\mathcal{G}(\boldsymbol{d}) := \{-A_*(\boldsymbol{d})\}^{-1}\mathcal{Z}(\boldsymbol{d})$ by the Kuhn-Tucker theorem, so that $\sqrt{n}\widehat{h}_n(\boldsymbol{d}) \Rightarrow \max[0, \mathcal{G}(\boldsymbol{d})]$. This also implies that $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d}) - \boldsymbol{\theta}_*) \Rightarrow \max[0, \mathcal{G}(\boldsymbol{d})]\boldsymbol{d}$ by noting that $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d}) = \boldsymbol{\theta}_* + \widehat{h}_n(\boldsymbol{d})\boldsymbol{d}$, and $2\{L_n(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \max[0, \mathcal{Y}(\boldsymbol{d})]^2$, where for each $\boldsymbol{d}$, $\mathcal{Y}(\boldsymbol{d}) := \{-A_*(\boldsymbol{d})\}^{1/2}\mathcal{G}(\boldsymbol{d})$. Note that (1) implies that the quasi-likelihood function may not be stochastically differentiable with respect to $\boldsymbol{d}$ because the given quasi-likelihood function may not be approximated by a second-order expansion with respect to $\boldsymbol{d}$ (see Pollard, 1985).

This pointwise result (with respect to $\boldsymbol{d}$) is not sufficient to derive the limit distribution of the QML estimator. It is necessary to examine the stochastic interrelationship of DQML estimators obtained by using different directions. Note that

$$L_n(\widehat{\boldsymbol{\theta}}_n) = \sup_{\boldsymbol{d}\in\Delta(\boldsymbol{\theta}_*)} L_n(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})). \tag{3}$$

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 limit behavior of $\widehat{\boldsymbol{\theta}}_n$ is derived by examining how $\widehat{\boldsymbol{\theta}}_n$ is asymptotically associated with $\widehat{\boldsymbol{\theta}}_n(\,\cdot\,)$, and for this purpose, we show in the Appendix that $DL_n(\boldsymbol{\theta}_*;\cdot)$ is asymptotically tight: 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 the empirical probability measure. These facts imply that the first-order directional derivative weakly converges to a Gaussian stochastic process indexed by $\boldsymbol{d}$ (e.g., Billingsley, 1999). In addition, $n^{-1}D^2L_n(\boldsymbol{\theta}_*;\cdot)$ satisfies the ULLN under the given conditions provided so far.

If $L_n(\cdot)$ is D, it is trivial to show the asymptotic tightness, because $DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}) = \nabla_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)\boldsymbol{d}$, 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 $\varepsilon > 0$,

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

Thus, if $n^{-1/2}\nabla_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)$ obeys the CLT, we can choose δ to have the right side be less than ε, and this shows the asymptotic tightness. Likewise, we can apply the ULLN to the second-order derivatives: 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 right side can be made as small as possible by applying the law of large numbers.

By the asymptotically tight directional derivatives, we now extend the pointwise limit result for $\sqrt{n}\widehat{h}_n(\boldsymbol{d})$ to the level of functional space, and from this, we obtain the limit distribution of the QML estimator as reported in the following theorem:

Theorem 1. *Given Assumptions 1 to 6, (i)* $\left\{n^{-1/2}DL_n(\boldsymbol{\theta}_*;\,\cdot\,), n^{-1}D^2L_n(\boldsymbol{\theta}_*;\,\cdot\,)\right\} \Rightarrow (\mathcal{Z}(\cdot), A_*(\cdot))$*, 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) if* $\max[0,\mathcal{Y}(\cdot)]^2$ *is uniquely maximized at* $\boldsymbol{d}_*$ *a.s.–*$\mathbb{P}$, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \Rightarrow$

$\max[0, \mathcal{G}(\boldsymbol{d}_*)]\boldsymbol{d}_*$. $\square$

Note that the limit distribution of the QML estimator is now represented as a functional of the Gaussian stochastic process defined on $\Delta(\boldsymbol{\theta}_*)$. Here, Theorem 1(*iv*) follows from the argmax continuous mapping theorem (e.g., Kim and Pollard, 1990; van der Vaart and Wellner, 1996). If $\max[0, \mathcal{Y}(\cdot)]^2$ is flat on $\Delta(\boldsymbol{\theta}_*)$ almost surely, it is hard to think of $\boldsymbol{d}_*$ as the limit of $\widehat{\boldsymbol{d}}_n$. The unique maximization condition on $\boldsymbol{d}_*$ is imposed to prevent this. This result also implies that 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 null limit distribution is not chi-squared.

Many statistics are known to follow limit distributions characterized by a Gaussian stochastic process. For example, Davies (1977,1987), Andrews (2001), Cho and White (2007,2010,2011a), and Baek, Cho, and Phillips (2015) examine these statistics: unidentified parameters yield limit distributions characterized by a Gaussian stochastic process, and Theorem 1 can be thought of as a variational result of prior literature.

Theorem 1 accommodates the standard D quasi-likelihood function as a special case of D-D quasi-likelihood functions. 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* $\boldsymbol{d}, \widetilde{\boldsymbol{d}}$*,* $B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) = \boldsymbol{d}'\mathbf{B}_*\widetilde{\boldsymbol{d}}$*.* $\square$

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 quasi-likelihood function context. Using these assumptions, we can further refine the results in Theorem 1. We let Assumption 6* denote Assumptions 6(*i*, *ii**, *iii**, and *iv*) going forward.

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* $\boldsymbol{\Theta}$*.* $\square$

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. In particular, Corollaries 1(*iv* to *vii*) imply that $\max[0, \mathcal{Y}(\cdot)]^2$ is uniquely maximized at $\boldsymbol{d}_*$ a.s.$-\mathbb{P}$. Cho (2011) exploits the D-D quasi-likelihood function analysis for D quasi-likelihood function estimation and examines other aspects that are not contained in Corollary 1.

3 Testing Using D-D Quasi-Likelihood Functions

In this section, we examine data inference using D-D quasi-likelihood functions. To this end, the standard QLR, Wald, and LM test statistics are redefined to accommodate the feature of directional differentiability.

It is efficient to first specify the role 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 possibly 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 $(\boldsymbol{d}_{\boldsymbol{\lambda}}{}', \boldsymbol{d}_{\boldsymbol{\upsilon}}{}', \boldsymbol{d}_{\boldsymbol{\tau}}{}')'$,

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

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}}) &:= \begin{bmatrix} \mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}}) & \mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \\ \mathbf{B}_*^{(\boldsymbol{\tau},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\tau}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}})' & B_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \end{bmatrix} \\ &:= \begin{bmatrix} \boldsymbol{d}_{\boldsymbol{\lambda}}{}' \mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})} \widetilde{\boldsymbol{d}}_{\boldsymbol{\lambda}} & \boldsymbol{d}_{\boldsymbol{\lambda}}{}' \mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\widetilde{\boldsymbol{d}}_{\boldsymbol{\upsilon}}) & \boldsymbol{d}_{\boldsymbol{\lambda}}{}' \mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \\ \mathbf{B}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{d}_{\boldsymbol{\upsilon}})' \widetilde{\boldsymbol{d}}_{\boldsymbol{\lambda}} & B_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\upsilon}}) & B_*^{(\boldsymbol{\upsilon},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \\ \mathbf{B}_*^{(\boldsymbol{\tau},\boldsymbol{\lambda})}(\boldsymbol{d}_{\boldsymbol{\tau}})' \widetilde{\boldsymbol{d}}_{\boldsymbol{\lambda}} & B_*^{(\boldsymbol{\tau},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\tau}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\upsilon}}) & B_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \end{bmatrix}, \end{aligned} \tag{4}$$

$DL_n^{(\boldsymbol{\lambda})} \in \mathbb{R}^{r_\lambda}$, $DL_n^{(\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) \in \mathbb{R}$, $DL_n^{(\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}}) \in \mathbb{R}$, $\mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})} \in \mathbb{R}^{r_\lambda} \times \mathbb{R}^{r_\lambda}$, $\mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) \in \mathbb{R}^{r_\lambda}$, $\mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}}) \in \mathbb{R}^{r_\lambda}$, $\mathbf{B}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) = \mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}})$, and $\mathbf{B}_*^{(\boldsymbol{\tau},\boldsymbol{\lambda})}(\boldsymbol{d}_{\boldsymbol{\tau}}) = \mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}})$. Thus, it 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. Similarly, we also suppose that $A_*(\boldsymbol{d}) = \boldsymbol{\iota}_3{}'\mathbf{A}_*(\boldsymbol{d})\boldsymbol{\iota}_3$, where

$$\begin{aligned}\mathbf{A}_*(\boldsymbol{d}) &:= \begin{bmatrix} \mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}) & \mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\pi}},\boldsymbol{d}_{\boldsymbol{\tau}}) \\ \mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\tau}},\boldsymbol{d}_{\boldsymbol{\pi}})' & A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}}) \end{bmatrix} \\ &:= \begin{bmatrix} \boldsymbol{d}_{\boldsymbol{\lambda}}{}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\boldsymbol{d}_{\boldsymbol{\lambda}} & \boldsymbol{d}_{\boldsymbol{\lambda}}{}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) & \boldsymbol{d}_{\boldsymbol{\lambda}}{}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}}) \\ \mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{d}_{\boldsymbol{\upsilon}})'\boldsymbol{d}_{\boldsymbol{\lambda}} & A_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) & A_*^{(\boldsymbol{\upsilon},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\upsilon}},\boldsymbol{d}_{\boldsymbol{\tau}}) \\ \mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\lambda})}(\boldsymbol{d}_{\boldsymbol{\tau}})'\boldsymbol{d}_{\boldsymbol{\lambda}} & A_*^{(\boldsymbol{\tau},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\tau}},\boldsymbol{d}_{\boldsymbol{\upsilon}}) & A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}}) \end{bmatrix}, \end{aligned} \tag{5}$$

$\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})} \in \mathbb{R}^{r_\lambda} \times \mathbb{R}^{r_\lambda}$, $\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) \in \mathbb{R}^{r_\lambda}$, $\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}}) \in \mathbb{R}^{r_\lambda}$, $\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{d}_{\boldsymbol{\upsilon}}) = \mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}})$, and $\mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\lambda})}(\boldsymbol{d}_{\boldsymbol{\tau}})$ $= \mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}})$. We also let $\boldsymbol{\pi}$ be orthogonal to $\boldsymbol{\tau}$: for each $\boldsymbol{d}$, $\mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\tau}},\boldsymbol{d}_{\boldsymbol{\pi}}) = \mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\pi}},\boldsymbol{d}_{\boldsymbol{\tau}}) = \mathbf{0}$. This assumption is useful in eliminating the nuisance parameters from our analysis that are asymptotically irrelevant to testing the hypothesis given below. 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 quasi-likelihood function is twice D. These conditions are collected into

Assumption 7 (D-Derivatives)**.** *(i) For each* $\mathbf{d}$, $DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}) = DL_n^{(\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}) + DL_n^{(\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}})$*, 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* $\mathbf{d}$ *and* $\widetilde{\mathbf{d}}$, $B_*(\boldsymbol{d},\widetilde{\boldsymbol{d}}) = \boldsymbol{\iota}_3{}'\mathbf{B}_*(\boldsymbol{d},\widetilde{\boldsymbol{d}})\boldsymbol{\iota}_3$*, where for each* $\boldsymbol{d}$, $\mathbf{B}_*(\boldsymbol{d},\boldsymbol{d})$ *is symmetric and positive definite; (iii) For each* $\boldsymbol{d}$, $A_*(\boldsymbol{d}) = \boldsymbol{\iota}_3{}'\mathbf{A}_*(\boldsymbol{d})\boldsymbol{\iota}_3$*, 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 on* $\Delta(\boldsymbol{\theta}_*)$*; (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* $C(\boldsymbol{\tau}_*) := \{\mathbf{x}\in\mathbb{R}^{r_\tau} : \exists\boldsymbol{\tau}'\in\mathbf{T}, \mathbf{x} := \boldsymbol{\tau}_* + \delta\boldsymbol{\tau}', \delta\in\mathbb{R}^+\}$*; (vi)* $\boldsymbol{\Pi} = \boldsymbol{\Lambda}\times\boldsymbol{\Upsilon}$ *and* $C(\boldsymbol{\pi}_*) = \mathbb{R}^{r_\lambda}\times C(\boldsymbol{\upsilon}_*)$*, where* $C(\boldsymbol{\upsilon}_*) := \{\mathbf{x}\in\mathbb{R}^{r_\upsilon} : \exists\boldsymbol{\upsilon}'\in\boldsymbol{\Upsilon}, \mathbf{x} := \boldsymbol{\upsilon}_* + \delta\boldsymbol{\upsilon}', \delta\in\mathbb{R}^+\}$*; (vii)* $\boldsymbol{\lambda}_*$ *is an interior element of* $\boldsymbol{\Lambda}$*.* □

Assumptions 7(*v* and *vi*) let the parameter space $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$ be the Cartesian products of two separate parameter spaces. We use this property to represent $L_n(\cdot)$ as a sum of two independent functions at the limit as detailed below. Given this structure, we further let $\boldsymbol{\upsilon}$ be the parameter of interest, and the hypotheses of our interest are given as

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

For our future reference, we also let $\boldsymbol{\Theta}_0$ be the parameter space constrained by the null hypotheses. That is, $\boldsymbol{\Theta}_0 := \{(\boldsymbol{\upsilon}', \boldsymbol{\lambda}', \boldsymbol{\tau}')' \in \boldsymbol{\Theta} : \boldsymbol{\upsilon} = \boldsymbol{\upsilon}_0\}$.

3.1 Qausi-Likelihood Ratio Test Statistic

The standard QLR test statistic defined by D quasi-likelihood functions can be used for D-D quasi-likelihood functions to test the given hypothesis. 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 $\mathcal{LR}_n^{(2)} := 2\{L_n(\ddot{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\}$. Note that $\mathcal{LR}_n^{(1)}$ tests whether the unknown parameter is $\boldsymbol{\theta}_*$ or not. Although Theorem 1(*iii*) already provides the limit distribution of $\mathcal{LR}_n^{(1)}$, we reexamine this here again by separating $\Delta(\boldsymbol{\theta}_*)$ into $\Delta(\boldsymbol{\pi}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\pi} : \boldsymbol{\pi}_* + \mathbf{x} \in \mathrm{cl}\{C(\boldsymbol{\pi}_*)\}, \|\mathbf{x}\| = 1\ \}$ and $\Delta(\boldsymbol{\tau}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\tau} : \boldsymbol{\tau}_* + \mathbf{x} \in \mathrm{cl}\{C(\boldsymbol{\tau}_*)\}, \|\mathbf{x}\| = 1\ \}$. We denote their representative components as $\boldsymbol{s}_{\boldsymbol{\pi}}(= (\boldsymbol{s}_{\boldsymbol{\lambda}}', \boldsymbol{s}_{\boldsymbol{\upsilon}}')')$ and $\boldsymbol{s}_{\boldsymbol{\tau}}$, respectively. Here, direction $\boldsymbol{s}$ 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 $\Delta(\boldsymbol{\theta}_*)$. This separation is useful in uncovering the limit distribution of $\mathcal{LR}_n^{(2)}$. The following theorem provides the null limit distribution of $\mathcal{LR}_n$:

Theorem 2. *(i) Given Assumptions 1 to 7,* $\mathcal{LR}_n^{(1)} \Rightarrow \mathcal{H}_0 + \mathcal{H}_1 + \mathcal{H}_2$, *where*

$$\mathcal{H}_0 := \sup_{\boldsymbol{s}_{\boldsymbol{\upsilon}}\in\Delta(\boldsymbol{\upsilon}_*)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})]^2;\ \ \mathcal{H}_1 := \mathbf{Z}^{(\boldsymbol{\lambda})\prime}(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})^{-1}\mathbf{Z}^{(\boldsymbol{\lambda})};\ \ \mathcal{H}_2 := \sup_{\boldsymbol{s}_{\boldsymbol{\tau}}\in\Delta(\boldsymbol{\tau}_*)} \max[0, \mathcal{Y}^{(\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}})]^2$$

such that for each $\boldsymbol{s}_{\boldsymbol{\upsilon}} \in \Delta(\boldsymbol{\upsilon}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\upsilon} : \boldsymbol{\upsilon}_* + \mathbf{x} \in \mathrm{cl}\{C(\boldsymbol{\upsilon}_*)\}, \|\mathbf{x}\| = 1\ \}$,

$$\widetilde{\mathcal{Y}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}) := (-\widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}))^{-1/2}\widetilde{\mathcal{Z}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}});\quad \widetilde{\mathcal{Z}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}) := \mathcal{Z}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}) - \mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})'(\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})^{-1}\mathbf{Z}^{(\boldsymbol{\lambda})};$$

$$\widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}) := A_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}) - \mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})'(\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})^{-1}\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}});$$

and for each $\boldsymbol{s}_{\boldsymbol{\tau}}$, $\mathcal{Y}^{(\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}}) := \{-A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}})\}^{-1/2}\mathcal{Z}^{(\boldsymbol{\tau})}(\boldsymbol{s}_{\boldsymbol{\tau}})$*; (ii) Given Assumptions 1 to 3 and* H_0, $\ddot{\boldsymbol{\theta}}_n$ *converges to* $\boldsymbol{\theta}_*$ a.s.–$\mathbb{P}$*; (iii) Given Assumptions 1 to 7 and* H_0, $\mathcal{LR}_n^{(2)} \Rightarrow \mathcal{H}_1 + \mathcal{H}_2$*; and (iv) Given Assumptions 1 to 7, and* H_0, $\mathcal{LR}_n \Rightarrow \mathcal{H}_0$. □

Several remarks are in order on Theorem 2. First, Theorem 2(*iii*) can be understood as a corollary of Theorem 2(*i*). Note that if $r_\upsilon = 0$ as $\boldsymbol{\upsilon}$ is fixed at $\boldsymbol{\upsilon}_*$, Theorem 2(*i*) implies that

$$\mathcal{LR}_n^{(2)} \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\lambda}}\in\Delta(\boldsymbol{\lambda}_*)} \left\{ \frac{\max[0, \boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{Z}^{(\boldsymbol{\lambda})}]^2}{\boldsymbol{s}_{\boldsymbol{\lambda}}{}'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})\boldsymbol{s}_{\boldsymbol{\lambda}}} \right\} + \mathcal{H}_2 = \mathcal{H}_1 + \mathcal{H}_2. \tag{6}$$

Here, the final equality holds because $\Delta(\boldsymbol{\lambda}_*) = \mathbb{R}^{r_\lambda}$ by Assumption 7(*vi*) and the null quasi-likelihood function is differentiable with respect to $\boldsymbol{\lambda}$. Second, the weak limit in Theorem 2(*iii*) is jointly achieved with that of $\mathcal{LR}_n^{(1)}$ because all of these are obtained by applying the continuous mapping theorem (CMT) to Theorem 1(*i*). Furthermore, $\mathcal{H}_1$ and $\mathcal{H}_2$ in $\mathcal{LR}_n^{(2)}$ are identical to those of $\mathcal{LR}_n^{(1)}$. Third, $\widetilde{\mathcal{Z}}^{(\boldsymbol{\upsilon})}(\cdot)$ is obtained by projecting $\mathcal{Z}^{(\boldsymbol{\upsilon})}(\cdot)$ on $\mathbf{Z}^{(\boldsymbol{\lambda})}$ because the QLR test statistic is constructed by minimizing the impact of the parameter estimation error that arises when estimating the unknown nuisance parameter $\boldsymbol{\lambda}_*$. Fourth, 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 $\mathcal{H}_0$, $\mathcal{H}_1$, and $\mathcal{H}_2$ as given in Theorem 2(*i*). This implies that we can ignore the effects of $\boldsymbol{\tau}$ when testing H_0. Fifth, the null limit distribution of the QLR test is more complicated than that in Theorem 2 in a general set-up. For example, if $\boldsymbol{\lambda}_*$ is a boundary parameter or the Cartesian product representation in Assumption 7(*vi*) is not valid, the null limit distribution of the QLR test statistic is obtained as $\mathcal{H}_{01}' - \mathcal{H}_1'$, where

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

and

$$\mathcal{H}_{01}' := \sup_{(\boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\xi}}) \in \Delta(\boldsymbol{\pi}_*)} \left\{ \frac{\max[0, \mathcal{Z}^{(\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) + \boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{Z}^{(\boldsymbol{\lambda})}]^2}{-A_*^{(\boldsymbol{\xi},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}}) + 2\boldsymbol{s}_{\boldsymbol{\lambda}}{}'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\xi})}(\boldsymbol{s}_{\boldsymbol{\xi}})) + \boldsymbol{s}_{\boldsymbol{\lambda}}{}'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})\boldsymbol{s}_{\boldsymbol{\lambda}}} \right\}.$$

Finally, Liu and Shao (2003) provide an alternative characterization of the quasi-likelihood ratio test using the Hellinger distance that obtains the null limit distribution as a functional of a Gaussian process as in Theorem 2. We leave the application of their methodology to the current context as a future research topic to satisfy the space constraint.

3.2 Wald Test Statistic

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

$$\sup_{\boldsymbol{d}} \sup_{h} L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = \sup_{\{\boldsymbol{s}_{\boldsymbol{\upsilon}},\, \boldsymbol{s}_{\boldsymbol{\lambda}},\, \boldsymbol{s}_{\boldsymbol{\tau}}\}} \sup_{\{h^{(\boldsymbol{\upsilon})},\, h^{(\boldsymbol{\lambda})},\, h^{(\boldsymbol{\tau})}\}} L_n(\boldsymbol{\theta}_* + (h^{(\boldsymbol{\upsilon})}\boldsymbol{s}_{\boldsymbol{\upsilon}}{}', h^{(\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}}{}', h^{(\boldsymbol{\tau})}\boldsymbol{s}_{\boldsymbol{\tau}}{}')'), \tag{7}$$

and we apply the Wald testing principle to $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\cdot)$.

For this purpose, we examine the limit distribution of $\widehat{\boldsymbol{h}}_n(\cdot) := (\widehat{h}_n^{(\boldsymbol{\upsilon})}(\cdot), \widehat{h}_n^{(\boldsymbol{\lambda})}(\cdot), \widehat{h}_n^{(\boldsymbol{\tau})}(\cdot))'$. For each $(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) \in \Delta(\boldsymbol{\upsilon}_*) \times \Delta(\boldsymbol{\lambda}_*) \times \Delta(\boldsymbol{\tau}_*)$, we let

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

where for each $(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}})$, $\boldsymbol{\mathcal{Z}}^{(\boldsymbol{\pi})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) := (\mathcal{Z}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}), \mathbf{Z}^{(\boldsymbol{\lambda})\prime}\boldsymbol{s}_{\boldsymbol{\lambda}})'$. Next, for each $(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) \in \Delta(\boldsymbol{\upsilon}_*) \times \Delta(\boldsymbol{\lambda}_*)$, we also let

$$\begin{bmatrix} \dot{\mathcal{G}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}) \\ \dot{\mathcal{G}}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\lambda}}) \end{bmatrix} := \begin{bmatrix} \{-A_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})\}^{-1} \mathcal{Z}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}) \\ \{\boldsymbol{s}_{\boldsymbol{\lambda}}'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})\boldsymbol{s}_{\boldsymbol{\lambda}}\}^{-1} \mathbf{Z}^{(\boldsymbol{\lambda})\prime}\boldsymbol{s}_{\boldsymbol{\lambda}} \end{bmatrix}.$$

These constitute the limit behavior of $\widehat{\boldsymbol{h}}_n(\cdot)$. First, note that both $(\widehat{h}_n^{(\boldsymbol{\upsilon})}(\cdot), \widehat{h}_n^{(\boldsymbol{\lambda})}(\cdot))'$ and $\widehat{h}_n^{(\boldsymbol{\tau})}(\cdot)$ are initially defined on $\Delta(\boldsymbol{\upsilon}_*) \times \Delta(\boldsymbol{\lambda}_*) \times \Delta(\boldsymbol{\tau}_*)$, but supposing that $A_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \boldsymbol{d}_{\boldsymbol{\tau}}) = 0$ renders the maximization process in the right side of (7) be asymptotically separated into two independent maximization procedures. Second, $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\cdot)$ and $\widehat{h}_n^{(\boldsymbol{\lambda})}(\cdot)$ cannot be less than zero. Thus, for each $(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}})$, one of following four different events asymptotically arises: (*i*) $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) > 0$ and $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) > 0$; (*ii*) $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) > 0$, $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$; (*iii*) $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$, $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) > 0$; and (*iv*) $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$, $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$. Note that these four different events are asymptotically determined by the sign of $\boldsymbol{\mathcal{G}}^{(\boldsymbol{\pi})}(\cdot)$ from the fact that it is asymptotically associated with the limit behavior of $(\widehat{h}_n^{(\boldsymbol{\upsilon})}(\cdot), \widehat{h}_n^{(\boldsymbol{\lambda})}(\cdot))$. Furthermore, their signs indicate how the parameter estimation error affects the asymptotic distribution of $(\widehat{h}_n^{(\boldsymbol{\upsilon})}(\cdot), \widehat{h}_n^{(\boldsymbol{\lambda})}(\cdot))$. For example, if $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$ and $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) > 0$, estimating the nuisance parameter $\boldsymbol{\lambda}_*$ does not affect the limit distribution of $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}})$ because $\widehat{\boldsymbol{\lambda}}_n(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = \boldsymbol{\lambda}_*$ from the fact that $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) = 0$, so that it does not have to be associated with the parameter estimation error of $\widehat{\boldsymbol{\lambda}}_n(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}})$, implying that $\sqrt{n}\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}) \Rightarrow \dot{\mathcal{G}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})$. For the cases of (*iii*) and (*iv*), the similar interpretations apply. If both $\boldsymbol{\lambda}_*$ and $\boldsymbol{\upsilon}_*$ are interior elements, both parameter estimation errors for $\boldsymbol{\upsilon}_*$ and $\boldsymbol{\lambda}_*$ cannot be avoided and have to be taken into account in obtaining the limit distribution of $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}})$ and $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}})$.

We now define the Wald test statistic as

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

where $\widetilde{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})$ is such that for each $\boldsymbol{s_\upsilon} \in \Delta(\boldsymbol{\upsilon}_0)$, $L_n(\boldsymbol{\upsilon}_0 + \widetilde{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})\boldsymbol{s_\upsilon}, \widetilde{\boldsymbol{\lambda}}_n(\boldsymbol{s_\upsilon}), \widetilde{\boldsymbol{\tau}}_n(\boldsymbol{s_\upsilon})) = \sup_{\{h^{(\boldsymbol{\upsilon})}, \boldsymbol{\lambda}, \boldsymbol{\tau}\}} L_n(\boldsymbol{\upsilon}_0 + h^{(\boldsymbol{\upsilon})}\boldsymbol{s_\upsilon}, \boldsymbol{\lambda}, \boldsymbol{\tau})$, and $\widehat{W}_n(\cdot)$ is a weight function that estimates a non-random positive function $\widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\cdot)$ say, uniformly on $\Delta(\boldsymbol{\upsilon}_0)$. Note that $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\cdot)$ is equivalent to $\widetilde{h}_n^{(\boldsymbol{\upsilon})}(\cdot)$ under the null from the fact that $\sup_{h^{(\boldsymbol{\upsilon})}, \boldsymbol{\lambda}, \boldsymbol{\tau}} L_n(\boldsymbol{\upsilon}_0 + h^{(\boldsymbol{\upsilon})}\boldsymbol{s_\upsilon}, \boldsymbol{\lambda}, \boldsymbol{\tau})$ is equivalent to $\sup_{\{\boldsymbol{s_\lambda}, \boldsymbol{s_\tau}\}} \sup_{\{h^{(\boldsymbol{\upsilon})}, h^{(\boldsymbol{\lambda})}, h^{(\boldsymbol{\tau})}\}} L_n(\boldsymbol{\upsilon}_0 + h^{(\boldsymbol{\upsilon})}\boldsymbol{s_\upsilon}, \boldsymbol{\lambda}_* + h^{(\boldsymbol{\lambda})}\boldsymbol{s_\lambda}, \boldsymbol{\tau}_* + h^{(\boldsymbol{\tau})}\boldsymbol{s_\tau})$. This equivalency does not hold under the alternative, and from this, the power of the Wald test is acquired. As the weight function is an important component of the Wald test statistic, we formally state its condition as follows:

Assumption 8 (Weight Function I)**.** $\widehat{W}_n(\cdot)$*, which is strictly positive uniformly on* $\Delta(\boldsymbol{\upsilon}_0)$ *and for every* n*, converges to* $\widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\cdot)$ *a.s.–*$\mathbb{P}$ *uniformly on* $\Delta(\boldsymbol{\upsilon}_0)$ *as* n *tends to infinity. That is,* $\sup_{\boldsymbol{s_\upsilon}} |\widehat{W}_n(\boldsymbol{s_\upsilon}) - \widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})| \to 0$ *a.s.–*$\mathbb{P}$. $\square$

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

The null limit distribution of the Wald test statistic is obtained as follows:

Theorem 3. *Given Assumptions 1 to 8, and* H_0*, if* $\max[0, \widetilde{\mathcal{Y}}(\cdot)]^2$ *is uniquely maximized a.s.–*$\mathbb{P}$*,* $\mathcal{W}_n \Rightarrow \mathcal{H}_0$*.* $\square$

Note that the weak limit of the Wald test statistic in Theorem 3 is identical to that given in Theorem 2. We prove Theorem 3 by noting that for each $\boldsymbol{s_\upsilon}$, maximizing $2\{L_n(\boldsymbol{\upsilon}_0 + h^{(\boldsymbol{\upsilon})}\boldsymbol{s_\upsilon}, \boldsymbol{\lambda}, \boldsymbol{\tau}) - L_n(\boldsymbol{\upsilon}_0, \boldsymbol{\lambda}_*, \boldsymbol{\tau}_*)\}$ with respect to $h^{(\boldsymbol{\upsilon})}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\tau}$ is equivalent to

$$\sup_{\{\boldsymbol{s_\lambda}, \boldsymbol{s_\tau}\}} \sup_{\{h^{(\boldsymbol{\upsilon})}, h^{(\boldsymbol{\lambda})}, h^{(\boldsymbol{\tau})}\}} 2\{L_n(\boldsymbol{\upsilon}_0 + h^{(\boldsymbol{\upsilon})}\boldsymbol{s_\upsilon}, \boldsymbol{\lambda}_* + h^{(\boldsymbol{\lambda})}\boldsymbol{s_\lambda}, \boldsymbol{\tau}_* + h^{(\boldsymbol{\tau})}\boldsymbol{s_\tau}) - L_n(\boldsymbol{\upsilon}_0, \boldsymbol{\lambda}_*, \boldsymbol{\tau}_*)\}. \tag{8}$$

For each $(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}, \boldsymbol{s_\tau})$, (8) is approximated by a quadratic function of $(h^{(\boldsymbol{\upsilon})}, h^{(\boldsymbol{\lambda})}, h^{(\boldsymbol{\tau})})$, and the signs of $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}, \boldsymbol{s_\tau})$ and $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}, \boldsymbol{s_\tau})$ result in different approximations as discussed earlier. Using Assumption 7(*vii*), we show that the optimization process in (8) results in the consequence of Theorem 3.

3.3 Lagrange Multiplier Test Statistic

The standard LM test statistic can also be appropriately redefined for D-D quasi-likelihood functions to test whether the slopes of quasi-likelihood functions are distributed around zero under the null. We let the LM test statistic be defined as

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

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

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

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

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

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

Assumption 9 (Weight Function II)**.** *A weight function $\widetilde{W}_n(\cdot)$, which is strictly positive uniformly on $\Delta(\boldsymbol{\upsilon}_0)\times\Delta(\ddot{\boldsymbol{\lambda}}_n)$ and for every n, converges to $\widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\cdot)$ a.s.–$\mathbb{P}$ uniformly on $\Delta(\boldsymbol{\upsilon}_0)\times\Delta(\ddot{\boldsymbol{\lambda}}_n)$ as n tends to infinity. That is, $\sup_{(\boldsymbol{s}_{\boldsymbol{\upsilon}},\boldsymbol{s}_{\boldsymbol{\lambda}})\in\Delta(\boldsymbol{\upsilon}_0)\times\Delta(\ddot{\boldsymbol{\lambda}}_n)} |\widetilde{W}_n(\boldsymbol{s}_{\boldsymbol{\upsilon}},\boldsymbol{s}_{\boldsymbol{\lambda}}) - \widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})| \to 0$ 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 null limit 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 $L_n(\boldsymbol{\upsilon}_0+h^{(\boldsymbol{\upsilon})}\boldsymbol{s}_{\boldsymbol{\upsilon}}, \ddot{\boldsymbol{\lambda}}_n + h^{(\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}}, \ddot{\boldsymbol{\tau}}_n + h^{(\boldsymbol{\tau})}\boldsymbol{s}_{\boldsymbol{\tau}})$ with respect to $\boldsymbol{s}_{\boldsymbol{\upsilon}}$ and $\boldsymbol{s}_{\boldsymbol{\lambda}}$, where $(\boldsymbol{\upsilon}_0{}', \ddot{\boldsymbol{\lambda}}_n', \ddot{\boldsymbol{\tau}}_n')' = \ddot{\boldsymbol{\theta}}_n$, and 'max' operator is used to capture the non-negativity property of $\sqrt{n}\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}},\boldsymbol{s}_{\boldsymbol{\lambda}},\boldsymbol{s}_{\boldsymbol{\tau}})$. Second, the LM test statistic is asymptotically the supremum of a squared random score function with respect to $(\boldsymbol{s}_{\boldsymbol{\upsilon}},\boldsymbol{s}_{\boldsymbol{\lambda}})$, provided that $\widetilde{W}_n(\cdot)$ is asymptotically equivalent to $-n^{-1}\widetilde{D}^2L_n(\ddot{\boldsymbol{\theta}}_n;\cdot)$. Third, $\widetilde{W}_n(\cdot)$ is defined on $\Delta(\boldsymbol{\upsilon}_0)\times\Delta(\ddot{\boldsymbol{\lambda}}_n)$, and $\boldsymbol{\lambda}_*$ is an interior element of $\boldsymbol{\Omega}$. Note that the domain $\Delta(\ddot{\boldsymbol{\lambda}}_n)$ estimates $\Delta(\boldsymbol{\lambda}_*)$. The interiority condition lets $\Delta(\ddot{\boldsymbol{\lambda}}_n)$ converge to $\Delta(\boldsymbol{\lambda}_*)$ asymptotically. If $\boldsymbol{\lambda}_*$ is on the boundary, $\Delta(\ddot{\boldsymbol{\lambda}}_n)$ can be different from $\Delta(\boldsymbol{\lambda}_*)$, and the null limit distribution of the LM test statistic is affected by this. Assumption 7(*vii*) precludes this possibility.

The null limit distribution of the LM test statistic is straightforwardly obtained as follows:

Theorem 4. *Given Assumptions 1 to 7, 9, and H_0, if $\max[0,\widetilde{\mathcal{Y}}(\cdot)]^2$ is uniquely maximized a.s.–$\mathbb{P}$, $\mathcal{LM}_n \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\upsilon}}\in\Delta(\boldsymbol{\upsilon}_0)} \max[0,\widetilde{\mathcal{Y}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})]^2$.* □

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

4 Example: 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

$$\mathbf{Y}^n = \mathbf{W}^n \, \alpha_* + \mathbf{R}^n \, \boldsymbol{\beta}_* + \mathbf{U}^n, \qquad \mathbf{U}^n | \mathbf{Q}^n \sim N[\mathbf{0}, \sigma_*^2 \{\mathbf{I}_n + \kappa_* \boldsymbol{\Omega}^n(\rho_*)\}],$$

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 $\boldsymbol{\Omega}^n_{tt'}(\rho_*) := W_t W_{t'} \rho_*{}^{|t'-t|}/(1-\rho_*{}^2)$. 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 $\mathbf{U}^n(\boldsymbol{\gamma}) := \mathbf{Y}^n - \mathbf{Q}^n \boldsymbol{\gamma}$, and $\boldsymbol{\gamma} := (\alpha, \boldsymbol{\beta}')'$.

This model is motivated by Rosenberg (1973), who aims to test $\kappa_* = 0$ and 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 it is conditionally heteroskedastic. On the other hand, if $\kappa_* = 0$, ρ_* is not identified, and Davies's (1977,1987) identification problem arises. This renders the null limit distributions of the standard tests non-standard as they often result in functionals of a Gaussian stochastic process.

King and Shively (1993) attempt to resolve the unidentified parameter problem by reparameterizing the original model using the polar coordinates:

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

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 reparameterization, and the identification problem does no longer arise under H_0'.

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

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

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

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

because $0 \times \tan^{-1}(\cdot) \equiv 0$ on the Euclidean real line. Therefore, the null log-likelihood desired by Rosenberg (1973) and King and Shively (1993) is obtained. We further note that

$$
\begin{aligned}
DL_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*; \boldsymbol{d}) = &- \frac{n d_{\sigma^2}}{2\sigma_*^2} - \frac{(d_1^2 + d_2^2)^{1/2}}{2}\mathrm{tr}[\boldsymbol{\Omega}^n\left(2\tan^{-1}(d_2/d_1)/\pi\right)] + \frac{d_{\sigma^2}}{2\sigma_*^4}\mathbf{U}^{n\prime}\mathbf{U}^n \\
&+ \frac{1}{\sigma_*^2}(\mathbf{Q}^n\boldsymbol{d}_\gamma)'\mathbf{U}^n + \frac{(d_1^2 + d_2^2)^{1/2}}{2\sigma_*^2}\mathbf{U}^{n\prime}\boldsymbol{\Omega}^n\left(2\tan^{-1}(d_2/d_1)/\pi\right)\mathbf{U}^n \qquad (9)
\end{aligned}
$$

that is not linear with respect to (d_1, d_2), implying that the quasi-likelihood function is not D. The second-order directional derivative is also obtained as

$$
\begin{aligned}
D^2L_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*; \boldsymbol{d}) = &\frac{n d_{\sigma^2}^2}{2\sigma_*^4} - \frac{d_{\sigma^2}^2}{\sigma_*^6}\mathbf{U}^{n\prime}\mathbf{U}^n - \frac{1}{\sigma_*^2}(\mathbf{Q}^n\boldsymbol{d}_\gamma)'(\mathbf{Q}^n\boldsymbol{d}_\gamma) \\
&- \frac{2d_{\sigma^2}}{\sigma_*^4}(\mathbf{Q}^n\boldsymbol{d}_\gamma)'\mathbf{U}^n - \sqrt{d_1^2 + d_2^2}\left\{\frac{d_{\sigma^2}}{\sigma_*^4}\mathbf{U}^{n\prime} - \frac{2}{\sigma_*^2}(\mathbf{Q}^n\boldsymbol{d}_\gamma)'\right\}\left[\boldsymbol{\Omega}^n(2\tan^{-1}(d_2/d_1)/\pi)\right]\mathbf{U}^n \\
&+ (d_1^2 + d_2^2)\left\{\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Omega}^n(2\tan^{-1}(d_2/d_1)/\pi)^2\right] - \frac{1}{\sigma_*^2}\mathbf{U}^{n\prime}\left[\boldsymbol{\Omega}^n(2\tan^{-1}(d_2/d_1)/\pi)\right]^2\mathbf{U}^n\right\}. \qquad (10)
\end{aligned}
$$

Note that this is not quadratic with respect to $\boldsymbol{d}$, so that the quasi-likelihood function is not twice D.

We elaborate on the model assumption for an appropriate analysis. If d_1 is zero, d_2/d_1 is not properly 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. We also do not allow that $d_2 = 0$. If it is allowed, the diagonal elements of $\boldsymbol{\Omega}^n(0)$ contain 0^0, so that the model is not again appropriately identified. 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 $[\underline{c}, \bar{c}]$, and we can also avoid the multifold identification problem of Cho and Ishida (2012), Cho, Ishida, and White (2011,2014), Baek, Cho, and Phillips (2015), and Cho and Phillips (2016).

The first-order directional derivative in (9) 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 $\boldsymbol{d}$,

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

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

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 identically and independently distributed (IID) random variables, and $\{U_t W_t \sum_{t'=1}^{t-1} U_{t'}W_{t'}m(d_2/d_1)^{t-t'}\}$ is a martingale difference array (MDA), so that Assumption 6(*iv*) holds for $Z_{1,n}(\boldsymbol{d})$, $Z_{2,n}(\boldsymbol{d})$, and $Z_{3,n}(\boldsymbol{d})$, and the CLT for MDA can be applied to them. Furthermore, the asymptotic tightness also holds for $n^{-1/2}DL_n(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*;\ \cdot\ )$. As $Z_{1,n}(\cdot)$ and $Z_{2,n}(\cdot)$ are linear with respect to $\mathbf{Q}_t U_t$ and $(U_t^2 - \sigma_*^2)$, respectively, it is trivial to show their asymptotic tightness. For the asymptotic tightness of $Z_{3,n}(\cdot)$, we let $\varepsilon_t := W_t U_t$ and $m := m(d_2/d_1)$ and show that $\{n^{-1/2}\sum_{t=2}^{n}\varepsilon_t \sum_{t'=1}^{t-1}\varepsilon_{t'}m^{t-t'}\}$ is asymptotically tight by applying Hansen (1996b). First, his theorem 1 holds if $E[W_t^4] < \Delta^4 < \infty$. Next, his λ and a are identical to 1 in our context, so that

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

for any 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, 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 $\widetilde{\boldsymbol{d}}$, $E[\mathcal{Z}(\boldsymbol{d})\mathcal{Z}(\widetilde{\boldsymbol{d}})] = B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) := B_*^{(1)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) + B_*^{(2)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) + B_*^{(3)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$ with $(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$, $B_*^{(1)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) := \frac{1}{\sigma_*^2}\boldsymbol{d}_{\boldsymbol{\gamma}}{}' E[\mathbf{Q}_t\mathbf{Q}_t{}']\widetilde{\boldsymbol{d}}_{\boldsymbol{\gamma}}$,

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

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

and $\widetilde{m} := m(\widetilde{d}_2/\widetilde{d}_1)$.

The limit behavior of the second-order directional derivative is related to $B_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$. By applying the law of large numbers to (10), we obtain

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

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 $D^2L_n\left(\boldsymbol{\gamma}_*, \sigma_*^2, \boldsymbol{\theta}_*; \cdot\right)$ 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 1(*iii*), where $\mathcal{Y}(\boldsymbol{d}) := \{B_*(\boldsymbol{d}, \boldsymbol{d})\}^{-1/2}\mathcal{Z}(\boldsymbol{d})$, and for each $\boldsymbol{d}$ and $\widetilde{\boldsymbol{d}}$,

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

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 the previous section. Specifically, we let $\boldsymbol{\upsilon} = (\theta_1, \theta_2)'$, $\boldsymbol{\lambda} = \sigma^2$, $\boldsymbol{\tau} = \boldsymbol{\gamma}$, 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

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

where for each (d_1, d_2) and $(\widetilde{d}_1, \widetilde{d}_2)$, $\widetilde{h} := h(\widetilde{d}_1, \widetilde{d}_2)$, $h := h(d_1, d_2) := (d_1^2 + d_2^2)^{1/2}/(1-m^2)$, and $k := k(d_2/d_1, \widetilde{d}_2/\widetilde{d}_1) := m\widetilde{m}/(1-m\widetilde{m})$. Because of the information matrix equality and the fact that $\mathbf{B}_*(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$ is block diagonal, the null limit distribution associated with each block matrix can be separately examined. 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 null limit 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 null limit distributions of the test statistics are more easily obtained by the unique features of $\mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}})$: first, let

$$\widetilde{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}}) := \left[\begin{array}{cc} \ddot{\mathbf{B}}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}}, \widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}}) & \mathbf{0} \\ \mathbf{0}' & qE[W_t^2]^2 \end{array} \right],$$

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

and $q := q(d_1, d_2, \widetilde{d}_1, \widetilde{d}_2) := h\widetilde{h}k$, 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 $qE[W_t^2]^2$ 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}. Using these facts, we can derive the null limit distributions of the three test statistics: first, the null limit distribution of the QLR test statistic is obtained as $\mathcal{LR}_n \Rightarrow \sup_{s_2/s_1 \in [\underline{c}, \bar{c}]} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2)]^2$ by Theorem 2(*iv*), where $\widetilde{\mathcal{Y}}^{(\boldsymbol{\theta})}(\cdot)$ 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 Theorem 3, we let the weight function be

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

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 8, and the Wald test statistic can be 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 Theorem 3, where $\widetilde{h}_n^{(\boldsymbol{\theta})}(s_2/s_1)$ is such that

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

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

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

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

where

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

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

$(\ddot{\boldsymbol{\gamma}}_n, \ddot{\sigma}_n^2)$ is such that $L_n(\ddot{\boldsymbol{\gamma}}_n, \ddot{\sigma}_n^2, \mathbf{0}) = \sup_{(\boldsymbol{\gamma},\sigma^2)} L_n(\boldsymbol{\gamma}, \sigma^2, \mathbf{0})$, and the same weight matrix is used as for the Wald test statistic. Here, $\widetilde{D}^2L_n(\ddot{\boldsymbol{\gamma}}_n, \ddot{\sigma}_n^2, \mathbf{0}; s_1, s_2, s_{\sigma^2})$ is indexed only by (s_1, s_2) because s_{σ^2} disappears by construction. Theorem 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 under the null.

The null limit distributions of the three test statistics can be uncovered by a simulation method. 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 $\ddot{\mathcal{Y}}^{(\boldsymbol{\theta})}(\cdot)$. When simulating $\ddot{\mathcal{Y}}^{(\boldsymbol{\theta})}(\cdot)$, $\mathrm{var}(W_t^2)$ and $E[W_t^2]$ need to be estimated, and the running index j must be truncated at a moderately large level so that it does not significantly affect the null limit distribution.

We conduct Monte Carlo simulations using this method. The DGP for $Y_t = U_t \sim$ IID $N(0,1)$ and $W_t \sim$ IID $N(0,1)$ that 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 $\underline{c} = 0.5$, $\bar{c} = 1.5$. Figure 1 shows the empirical distributions of the QLR test statistic for various sample sizes and the null limit distribution obtained by simulating $\sup_{(s_1,s_2)} \max[0, \widehat{\mathcal{Y}}^{(\boldsymbol{\theta})}(s_1, s_2)]^2$ 2,000 times, 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. The empirical distribution of the QLR test statistic approaches the limit distribution as n increases.

5 Conclusion

The current study examines the estimation and inference of D-D quasi-likelihood functions and provides conditions under which the QML estimator behaves regularly. Specifically, we show that the QML estimator has a distribution different from that of standard D quasi-likelihood functions by showing that it is represented as a functional of a Gaussian stochastic process indexed by direction. Furthermore, the analysis assuming D quasi-likelihood function can be treated as a special case of D-D quasi-likelihood function analysis. Furthermore, the standard QLR, Wald, and LM test statistics are redefined to fit the structure of D-D quasi-likelihood functions. These modifications are provided for general D-D quasi-likelihood functions, and we show that the three test statistics possess null limit distributions represented as functionals of the same Gaussian stochastic process. We further obtain that the three test statistics are asymptotically equivalent under the null and some mild regularity conditions that are popularly used for empirical examinations.

Appendix: Proofs

Before proving the main claims of the paper, we provide the following preliminary lemmas:

Lemma A1. *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) $n^{-1} D^2 L_n(\boldsymbol{\theta}_*; \boldsymbol{d})$ converges to $A_*(\boldsymbol{d})$ a.s.–$\mathbb{P}$; (iii) $\{n^{-1/2} DL_n(\boldsymbol{\theta}_*; \boldsymbol{d}), n^{-1} D^2 L_n(\boldsymbol{\theta}_*; \boldsymbol{d})\} \Rightarrow \{\mathcal{Z}(\boldsymbol{d}), A_*(\boldsymbol{d})\}$.* □

Lemma A2. *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 $\boldsymbol{d}$, $\mathcal{Y}(\boldsymbol{d}) := \{-A_*(\boldsymbol{d})\}^{1/2} \mathcal{G}(\boldsymbol{d})$.* □

Lemma A3. *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 the empirical probability measure; (ii) for all $\varepsilon > 0$, there is $n(\varepsilon)$ a.s.$-\mathbb{P}$ such that if $n > n(\varepsilon)$, $\sup_{\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)} |n^{-1} D^2 L_n(\boldsymbol{\theta}_*; \boldsymbol{d}) - A_*(\boldsymbol{d})| < \varepsilon$. □

Lemma A3(*i*) implies that the first-order directional derivative weakly forms a Gaussian stochastic process indexed by ***d*** (e.g., Billingsley, 1999). In our time-series data context, theorem 1 of Hansen (1996b) provides sufficient regularity conditions for this. As Lemma A3(*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 A3(*i*).

Proof of Lemma A1: (*i*) To show the given claim, 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 $\|D\ell_t(\boldsymbol{\theta}_*; \boldsymbol{d})\|_s < \Delta$ for any t, ν_τ is of size $-1/(1-\gamma) < -1/2$, and $\{\mathbf{Y}_t\}$ 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 d for any t. 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 A1(*i* and *ii*), the given weak convergence follows from theorem 3.9 of Billingsley (1999, p. 37). ■

Proof of Lemma A2: (*i*) For given $\boldsymbol{d}$, if we maximize (1) with respect to h subject to $h \geq 0$, it follows that $\widehat{h}_n(\boldsymbol{d}) = \max[0, \{-D^2 L_n(\bar{\boldsymbol{\theta}}_n(\boldsymbol{d}); \boldsymbol{d})\}^{-1} DL_n(\boldsymbol{\theta}_*; \boldsymbol{d})]$ by using Khun-Tucker theorem. We further note that $\bar{\boldsymbol{\theta}}_n(\boldsymbol{d}) \to \boldsymbol{\theta}_*$ a.s.$-\mathbb{P}$ because $\widehat{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}_*$ a.s.–$\mathbb{P}$, so that $\sqrt{n}\widehat{h}_n(\boldsymbol{d}) \Rightarrow \max[0, \{-A_*(\boldsymbol{d})\}^{-1}\mathcal{Z}(\boldsymbol{d})]$ by Lemma A1(*iii*). The desired result now follows from the definition of $\mathcal{G}(\boldsymbol{d})$.

(*ii*) From the definition of $\widehat{h}_n(\boldsymbol{d})$, $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d}) \equiv \boldsymbol{\theta}_* + \widehat{h}_n(\boldsymbol{d})\boldsymbol{d}$. Lemma A2(*i*) yields the given result.

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

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

Thus, the desired result follows from (2). ■

Proof of Lemma A3: (*i*) Given the weak convergence of Lemma A1(*i*), if $\{n^{-1/2}\sum D\ell_t(\boldsymbol{\theta}_*; \cdot)\}$ is asymptotically tight, the desired result follows from the finite dimensional multivariate CLT based on the Cramér-Wold device, which we do not prove by its self-evidence.

The asymptotic 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, $\|M_t\|_s < \infty$ for any t from the stationarity assumption of $\{M_t\}$ in Assumption 6(*iv*). Third, $\|D\ell_t(\boldsymbol{\theta}_*; \boldsymbol{d})\|_s < \infty$ uniformly in $\boldsymbol{d}$ for any t 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 $q > (r-1)/(\gamma\lambda)$. These results verify the conditions in theorem 4 of Hansen (1996a), and the asymptotic 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 $A_*(\cdot)$ uniformly on $\Delta(\boldsymbol{\theta}_*)$ a.s.$-\mathbb{P}$ by Rudin (1976, p. 168). ■

Proof of Theorem 1: (*i*) Given Lemmas A3(*i* and *ii*), the given weak convergence follows from theorem 3.9 of Billingsley (1999, p. 37).

(*ii*) The given result follows from Lemma A2(*i*), Theorem 1(*i*), and the definition of $\mathcal{G}(\cdot)$.

(*iii*) We can apply the CMT to (3).

(*iv*) To prove the given claim, we apply the argmax continuous mapping theorem in van der Vaart and Wellner (1996). Note that $\Delta(\boldsymbol{\theta}_*)$ is bounded, and $L_n(\widehat{\boldsymbol{\theta}}_n(\cdot)) - L_n(\boldsymbol{\theta}_*) \Rightarrow \max[0, \mathcal{Y}(\cdot)]^2$ by (2) and Lemma A3. Given that $D\ell_t(\boldsymbol{\theta}; \cdot)$ and $D^2\ell_t(\boldsymbol{\theta}; \cdot)$ are continuous by Assumption 5, $\max[0, \mathcal{Y}(\cdot)]^2$ must be continuous on $\Delta(\boldsymbol{\theta}_*)$ almost surely. Furthermore, $\Delta(\boldsymbol{\theta}_*)$ is a subset of a compact space, so that it trivially follows that $\boldsymbol{d}_*$ is tight, and $\widehat{\boldsymbol{d}}_n$ is uniformly tight. Given that $\boldsymbol{d}_*$ is unique almost surely, the regularity conditions for the argmax continuous mapping theorem are satisfied. Theorem 3.2.2 of van der Vaart and Wellner (1996) holds by this. ■

Proof of Corollary 1: For an efficient proof, we 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 $n^{-1}\nabla^2_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*) \rightarrow \mathbf{A}_*$ a.s.–$\mathbb{P}$ by the ergodic theorem. Therefore, the given result follows from the definition of $\mathcal{G}(\boldsymbol{d})$.

(*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 $\mathcal{Y}(\cdot)$ of Theorem 1, for each $\boldsymbol{d}$, $\mathcal{Y}(\boldsymbol{d}) = \{\boldsymbol{d}'(-A_*)\boldsymbol{d}\}^{-1/2}\mathbf{Z}'\boldsymbol{d}$, so that Theorem 1(*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 a single positive eigenvalue, and the 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

$$(-\mathbf{A}_*)^{-1/2}\mathbf{Z}\mathbf{Z}'(-\mathbf{A}_*)^{-1/2}\boldsymbol{v}_* = \mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z}\boldsymbol{v}_* \tag{11}$$

by the definition of eigenvector. This implies that

$$\boldsymbol{v}_*'(-\mathbf{A}_*)^{-1/2}\mathbf{Z}\mathbf{Z}'(-\mathbf{A}_*)^{-1/2}\boldsymbol{v}_* = \mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z}\boldsymbol{v}_*'\boldsymbol{v}_* = \mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z} \tag{12}$$

because $\boldsymbol{v}_*'\boldsymbol{v}_* = 1$. Plugging the definition of $\boldsymbol{v}_*$ to the left side of (12) 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 (11) 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 2: (*i*) 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

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

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, $\mathcal{LR}_n^{(1)} = \sup_{\boldsymbol{d}}\sup_h 2\{L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) - L_n(\boldsymbol{\theta}_*)\}$ implies that

$$\begin{aligned}
\mathcal{LR}_n^{(1)} = &\sup_{\boldsymbol{s}_{\boldsymbol{\pi}}}\sup_{h^{(\boldsymbol{\pi})}}\{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\pi}})h^{(\boldsymbol{\pi})} + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\pi}})(h^{(\boldsymbol{\pi})})^2 + o_{\mathbb{P}_{\boldsymbol{d}_{\boldsymbol{\pi}}}}(1)\} \\
&+ \sup_{\boldsymbol{s}_{\boldsymbol{\tau}}}\sup_{h^{(\boldsymbol{\tau})}}\{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\tau}})h^{(\boldsymbol{\tau})} + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\tau}})(h^{(\boldsymbol{\tau})})^2 + o_{\mathbb{P}_{\boldsymbol{d}_{\boldsymbol{\tau}}}}(1)\},
\end{aligned} \tag{13}$$

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 $\mathbf{0}$ 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 1(*i*).

Given this, note that $H_{2,n} := \sup_{\boldsymbol{s}_{\boldsymbol{\tau}}} \sup_{h^{(\boldsymbol{\tau})}}\{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\tau}})h^{(\boldsymbol{\tau})} + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\tau}})(h^{(\boldsymbol{\tau})})^2\} \Rightarrow \mathcal{H}_2$. Thus, we may focus on the weak limit of $\sup_{\boldsymbol{s}_{\boldsymbol{\pi}}} \sup_{h^{(\boldsymbol{\pi})}}\{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\pi}})h^{(\boldsymbol{\pi})} + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\pi}})(h^{(\boldsymbol{\pi})})^2\}$ that is denoted as $H_{01,n}$. 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 $h^{(\boldsymbol{\lambda})} \in \mathbb{R}^+$, $h^{(\boldsymbol{v})} \in \mathbb{R}^+$, $\boldsymbol{s}_{\boldsymbol{\lambda}} \in \Delta(\boldsymbol{\lambda}_*)$, and $\boldsymbol{s}_{\boldsymbol{\pi}} \in \Delta(\boldsymbol{\pi}_*)$ such that $h\boldsymbol{d}_{\boldsymbol{\pi}} = [h^{(\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}}{}', h^{(\boldsymbol{v})}\boldsymbol{s}_{\boldsymbol{v}}{}']'$ and

$$H_{01,n} = \sup_{\{\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}\}} \sup_{\{h^{(\boldsymbol{v})}, h^{(\boldsymbol{\lambda})}\}} 2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{v}})h^{(\boldsymbol{v})} + 2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\lambda}})h^{(\boldsymbol{\lambda})} + 2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\lambda}}; \boldsymbol{s}_{\boldsymbol{v}})h^{(\boldsymbol{\lambda})}h^{(\boldsymbol{v})}$$
$$+ D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{v}})(h^{(\boldsymbol{v})})^2 + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\lambda}})(h^{(\boldsymbol{\lambda})})^2,$$

where $DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\lambda}}; \boldsymbol{s}_{\boldsymbol{v}})$ is the directional derivative of $DL_n(\,\cdot\,; \boldsymbol{s}_{\boldsymbol{\lambda}})$ with respect to $\boldsymbol{s}_{\boldsymbol{v}}$ evaluated at $\boldsymbol{\theta}_*$. Given this, if we apply the ULLN and FCLT to $H_{01,n}$ by Theorem 1,

$$H_{01,n} \Rightarrow \mathcal{H}_0 + \mathcal{H}_1 = \sup_{\{\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}\}} \sup_{\{h^{(\boldsymbol{v})},\, h^{(\boldsymbol{\lambda})}\}} 2\mathcal{Z}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}})h^{(\boldsymbol{v})} + 2\boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{Z}^{(\boldsymbol{\lambda})}h^{(\boldsymbol{\lambda})} + 2\boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}})h^{(\boldsymbol{\lambda})}h^{(\boldsymbol{v})}$$
$$+ A_*^{(\boldsymbol{v},\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}})(h^{(\boldsymbol{v})})^2 + \boldsymbol{s}_{\boldsymbol{\lambda}}{}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}}(h^{(\boldsymbol{\lambda})})^2, \tag{14}$$

and there are four different possible cases for the solutions of $(h^{(\boldsymbol{v})},\, h^{(\boldsymbol{\lambda})})$ in the right side of (14): for each $(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}})$ if we let $\widehat{h}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}})$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}})$ maximize the right side of (14), it follows either (*i*) $\widehat{h}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$; (*ii*) $\widehat{h}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$; (*iii*) $\widehat{h}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$; or (*iv*) $\widehat{h}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$.

We examine the limit distribution of each case one by one. First, if $\widehat{h}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$, the right side of (14) is identical to

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

and maximizing this with respect to $\boldsymbol{s}_{\boldsymbol{\lambda}}$ for a given $\boldsymbol{s}_{\boldsymbol{v}}$ yields $\widetilde{\mathcal{Y}}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}})^2 + (\mathbf{Z}^{(\boldsymbol{\lambda})})'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})^{-1}(\mathbf{Z}^{(\boldsymbol{\lambda})})$. Second, if $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$ and $\widehat{h}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$, the right side of (14) 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, $\widetilde{\mathcal{Y}}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}})$ cannot be greater than zero. Otherwise, it must follow that $\widehat{h}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) > 0$ that is contradictory. Third, if $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$ and $\widehat{h}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) = 0$, for the same $\boldsymbol{s}_{\boldsymbol{v}}$, $\boldsymbol{s}_{\boldsymbol{\lambda}}$ cannot be optimal to maximizing the right side of (14) from the fact that $\boldsymbol{\lambda}_*$ is an interior element of $\boldsymbol{\Lambda}$. We can ignore the case in which $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{v}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) =$

0. Therefore, combining the first two cases, we obtain that $\mathcal{H}_0 = \sup_{\boldsymbol{s_\upsilon} \in \Delta(\boldsymbol{\upsilon}_*)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})]^2$ and $\mathcal{H}_1 = (\mathbf{Z}^{(\boldsymbol{\lambda})})'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})(\mathbf{Z}^{(\boldsymbol{\lambda})})$. This also implies that it is necessary for $\widetilde{\mathcal{Y}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})$ to be greater than zero, if $\widehat{h}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda})$ is greater than zero.

(*ii*) We can apply the ULLN.

(*iii*) This follows as a corollary of (*i*).

(*iv*) The desired result is implied by (*i*), (*iii*), and continuous mapping. ■

Proof of Theorem 3: (*i*) We exploit (13) further. First, applying the CMT to Theorem 1(*i*) shows that $(n^{-1/2}DL_n^{(\boldsymbol{\tau})}(\cdot), n^{-1}D^2L_n^{(\boldsymbol{\tau})}(\cdot)) \Rightarrow (\mathcal{Z}^{(\boldsymbol{\tau})}(\cdot), A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\cdot))$. 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 $\sqrt{n}\widehat{h}_n^{(\boldsymbol{\tau})}(\boldsymbol{s_\tau}) \Rightarrow \max[0, \{-A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{s_\tau})\}^{-1}\mathcal{Z}^{(\boldsymbol{\tau})}(\boldsymbol{s_\tau})] = \max[0, \mathcal{G}^{(\boldsymbol{\tau})}(\boldsymbol{s_\tau})]$. This holds even as a function of $\boldsymbol{s_\tau}$. That is, $\sqrt{n}\widehat{h}_n^{(\boldsymbol{\tau})}(\cdot) \Rightarrow \max[0, \mathcal{G}^{(\boldsymbol{\tau})}(\cdot)]$.

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

$$\begin{aligned}
&\sup_{h^{(\boldsymbol{\pi})}}\{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{d_\pi})h^{(\boldsymbol{\pi})} + D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{d_\pi})(h^{(\boldsymbol{\pi})})^2\} \\
&\quad = \sup_{(h^{(\boldsymbol{\upsilon})}, h^{(\boldsymbol{\lambda})})} 2\{DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\upsilon})h^{(\boldsymbol{\upsilon})} + DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\lambda})h^{(\boldsymbol{\lambda})} + DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda})h^{(\boldsymbol{\upsilon})}h^{(\boldsymbol{\lambda})}\} \\
&\qquad\qquad + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s_\upsilon})(h^{(\boldsymbol{\upsilon})})^2 + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s_\lambda})(h^{(\boldsymbol{\lambda})})^2 \\
&\quad \Rightarrow \sup_{(h^{(\boldsymbol{\upsilon})}, h^{(\boldsymbol{\lambda})})} 2\{\mathcal{Z}(\boldsymbol{s_\upsilon})h^{(\boldsymbol{\upsilon})} + \boldsymbol{s}_{\boldsymbol{\lambda}}'\mathbf{Z}^{(\boldsymbol{\lambda})}h^{(\boldsymbol{\lambda})} + \boldsymbol{s}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})h^{(\boldsymbol{\upsilon})}h^{(\boldsymbol{\lambda})}\} \\
&\qquad\qquad + A_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})(h^{(\boldsymbol{\upsilon})})^2 + \boldsymbol{s}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\boldsymbol{s_\lambda}(h^{(\boldsymbol{\lambda})})^2. \qquad (15)
\end{aligned}$$

Given this, $h^{(\boldsymbol{\upsilon})}$ and $h^{(\boldsymbol{\lambda})}$ have to be at least greater than or equal to zero, so that the following four different inequality constraints can be possibly associated with this maximization process: first, if any equality condition does not bind, $\sqrt{n}(\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}, \boldsymbol{s_\tau}), \widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}\boldsymbol{s_\tau}))' \Rightarrow \boldsymbol{\mathcal{G}}^{(\boldsymbol{\pi})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda})$ by the standard first-order condition and Lemma A1. This occurs if every component of $\boldsymbol{\mathcal{G}}^{(\boldsymbol{\pi})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda})$ is strictly greater than zero. Second, if $\mathcal{G}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}) < 0$, it simply holds that $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}) = \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda})]$, and $\sqrt{n}(\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}, \boldsymbol{s_\tau}), \widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}, \boldsymbol{s_\tau}))' \Rightarrow (0, \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda})])'$. Third, if $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}) = 0$ in the right side of (15) because $\mathcal{G}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}) < 0$, $\widehat{h}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}) = \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda})]$. This implies that $\sqrt{n}(\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}, \boldsymbol{s_\tau}), \widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}, \boldsymbol{s_\tau}))' \Rightarrow (\max\,[0, \dot{\mathcal{G}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda})], 0)'$. Fourth, it must follow that

$\sqrt{n}(\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}), \widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}))' \Rightarrow (0,0)'$ for any other case. Therefore, if we combine all these and apply Theorem 1(i),

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

and this implies that

$$\sqrt{n}\widetilde{h}_n^{(\boldsymbol{\upsilon})}(\cdot) \Rightarrow \max[0, \mathcal{G}^{(\boldsymbol{\upsilon})}(\cdot)]\mathbf{1}_{\{\min[\mathcal{G}^{(\boldsymbol{\upsilon})}(\cdot), \mathcal{G}^{(\boldsymbol{\lambda})}(\cdot)] \geq 0\}} + \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\upsilon})}(\cdot)]\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\upsilon})}(\cdot) \geq 0 > \mathcal{G}^{(\boldsymbol{\lambda})}(\cdot)\}}$$

under H_0. Therefore, it now follows that

$$\begin{aligned} \mathcal{W}_n \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\upsilon}} \in \Delta(\boldsymbol{\upsilon}_0)} & \max[0, \mathcal{G}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \bar{\boldsymbol{s}}_{\boldsymbol{\lambda}}(\boldsymbol{s}_{\boldsymbol{\upsilon}}))]^2 \widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})\mathbf{1}_{\{\min[\mathcal{G}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \bar{\boldsymbol{s}}_{\boldsymbol{\lambda}}(\boldsymbol{s}_{\boldsymbol{\upsilon}})), \mathcal{G}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \bar{\boldsymbol{s}}_{\boldsymbol{\lambda}}(\boldsymbol{s}_{\boldsymbol{\upsilon}}))] > 0\}} \\ & + \max[0, \dot{\mathcal{G}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})]^2 \widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}\mathbf{1}_{\{\mathcal{G}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \bar{\boldsymbol{s}}_{\boldsymbol{\lambda}}(\boldsymbol{s}_{\boldsymbol{\upsilon}})) \geq 0 > \mathcal{G}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \bar{\boldsymbol{s}}_{\boldsymbol{\lambda}}(\boldsymbol{s}_{\boldsymbol{\upsilon}}))\}}, \end{aligned}$$

where for each $\boldsymbol{s}_{\boldsymbol{\upsilon}}$,

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

Here, for given $\boldsymbol{s}_{\boldsymbol{\upsilon}}$, optimizing process with respect to $\bar{\boldsymbol{s}}_{\boldsymbol{\lambda}}(\boldsymbol{s}_{\boldsymbol{\upsilon}})$ lets $\mathcal{G}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \bar{\boldsymbol{s}}_{\boldsymbol{\lambda}}(\boldsymbol{s}_{\boldsymbol{\upsilon}})) > 0$ because $\boldsymbol{\lambda}_*$ is an interior element of $\boldsymbol{\Lambda}$. Therefore,

$$\mathcal{W}_n \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\upsilon}} \in \Delta(\boldsymbol{\upsilon}_0)} \max[0, \mathcal{G}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \bar{\boldsymbol{s}}_{\boldsymbol{\lambda}}(\boldsymbol{s}_{\boldsymbol{\upsilon}}))]^2 \widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})} = \sup_{\boldsymbol{s}_{\boldsymbol{\upsilon}} \in \Delta(\boldsymbol{\upsilon}_*)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})]^2 = \mathcal{H}_0,$$

and this completes the proof. ■

Proof of Theorem 4: For notational simplicity, we suppose that $\boldsymbol{\tau}_*$ is known. By Theorem 2(i), this supposition simplifies our proof without losing generality.

To show the given claim, we derive the convergence limit of each component that constitutes the LM test statistic. First, there is n^* a.s.$-\mathbb{P}$ such that if $n > n^*$, $\Delta(\ddot{\boldsymbol{\lambda}}_n) = \Delta(\boldsymbol{\lambda}_*)$. We note that $\boldsymbol{\lambda}_*$ is an interior element by Assumption 7(vii), so that $\Delta(\boldsymbol{\lambda}_*) = \{\mathbf{x} \in \mathbb{R}^{r_\omega} : \|\mathbf{x}\| = 1\}$, and further for an open ball with radius $\varepsilon > 0$ denoted as $B(\boldsymbol{\lambda}_*, \varepsilon)$ such that $B(\boldsymbol{\lambda}_*, \varepsilon) \subset \boldsymbol{\Lambda}$, there is $n(\varepsilon)$ a.s.–$\mathbb{P}$, so that if $n > n(\varepsilon)$, $\ddot{\boldsymbol{\lambda}}_n \in B(\boldsymbol{\lambda}_*, \varepsilon)$ by Theorem 2(ii). This implies that $\ddot{\boldsymbol{\lambda}}_n$ is an interior element, too. Thus, if we let $n^* > n(\varepsilon)$,

$\Delta(\ddot{\boldsymbol{\lambda}}_n) = \{\mathbf{x} \in \mathbb{R}^{r_\omega} : \|\mathbf{x}\| = 1\}$, which is also $\Delta(\boldsymbol{\lambda}_*)$. Second, $n^{-1/2}DL_n(\ddot{\boldsymbol{\theta}}_n; \cdot) \Rightarrow \ddot{\mathcal{Z}}^{(\boldsymbol{\upsilon})}(\cdot; \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}})$. Applying the mean-value theorem shows that for each $\boldsymbol{s}_{\boldsymbol{\upsilon}}$ there is $\dot{\boldsymbol{\lambda}}(\boldsymbol{s}_{\boldsymbol{\upsilon}})$ such that

$$\begin{aligned} DL_n(\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\upsilon}}) - DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\upsilon}}) &= \{DL_n(\boldsymbol{\upsilon}_0, \dot{\boldsymbol{\lambda}}_n(\boldsymbol{s}_{\boldsymbol{\upsilon}}), \boldsymbol{\tau}_*; \boldsymbol{s}_{\boldsymbol{\upsilon}}; \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda},n})\}\{\ddot{h}_n^{(\boldsymbol{\lambda})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\lambda},n})\} \\ &= DL_n(\boldsymbol{\upsilon}_0, \dot{\boldsymbol{\lambda}}_n(\boldsymbol{s}_{\boldsymbol{\upsilon}}), \boldsymbol{\tau}_*; \boldsymbol{s}_{\boldsymbol{\upsilon}}; \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda},n})\{-D^2L_n(\boldsymbol{\upsilon}_0, \bar{\boldsymbol{\lambda}}_n(\boldsymbol{s}_{\boldsymbol{\upsilon}}), \boldsymbol{\tau}_*; \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda},n})\}^{-1}DL_n(\boldsymbol{\theta}_*; \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda},n}), \end{aligned} \tag{16}$$

where $(\ddot{h}_n^{(\boldsymbol{\lambda})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\lambda},n}), \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda},n}) := \arg\sup_{h^{(\boldsymbol{\lambda})}, \boldsymbol{s}_{\boldsymbol{\lambda}}} L_n(\boldsymbol{\upsilon}_0, \boldsymbol{\lambda}_* + h^{(\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{\tau}_*)$, and the last equality follows from the mean-value theorem: there is $\bar{\boldsymbol{\lambda}}_n(\boldsymbol{s}_{\boldsymbol{\upsilon}})$ such that (16) holds. Given this and Theorem 2(*ii*), we can also apply the ULLN:

$$\sup_{\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}} |n^{-1}DL_n(\boldsymbol{\upsilon}_0, \dot{\boldsymbol{\lambda}}_n(\boldsymbol{s}_{\boldsymbol{\upsilon}}), \boldsymbol{\tau}_*; \boldsymbol{s}_{\boldsymbol{\upsilon}}; \boldsymbol{s}_{\boldsymbol{\lambda}}) - \boldsymbol{s}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})| \xrightarrow{\mathbb{P}} 0, \quad \text{and}$$

$$\sup_{\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}} |n^{-1}D^2L_n(\boldsymbol{\upsilon}_0, \bar{\boldsymbol{\lambda}}_n(\boldsymbol{s}_{\boldsymbol{\upsilon}}), \boldsymbol{\tau}_*; \boldsymbol{s}_{\boldsymbol{\lambda}}) - \boldsymbol{s}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}}| \xrightarrow{\mathbb{P}} 0.$$

Furthermore, it trivially holds that $n^{-1/2}(DL_n(\boldsymbol{\theta}_*; \cdot), DL_n(\boldsymbol{\theta}_*; \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda},n})) \Rightarrow (\mathcal{Z}^{(\boldsymbol{\upsilon})}(\cdot), \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}}'\mathbf{Z}^{(\boldsymbol{\lambda})})$ by the facts that $n^{-1/2}(DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\upsilon}}), DL_n(\boldsymbol{\theta}_*; \boldsymbol{s}_{\boldsymbol{\lambda}}))$ (as functions of $\boldsymbol{s}_{\boldsymbol{\upsilon}}$ and $\boldsymbol{s}_{\boldsymbol{\lambda}}$, respectively) weakly converges to $(\mathcal{Z}^{(\boldsymbol{\upsilon})}(\cdot), (\cdot)'\mathbf{Z}^{(\boldsymbol{\lambda})})$ and that $\max[0, DL_n\ (\boldsymbol{\theta}_*; \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda},n})]^2\{-D^2L_n(\boldsymbol{\theta}_*; \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda},n})\}^{-1} \Rightarrow \max[0, \mathcal{Y}^{(\boldsymbol{\lambda})}(\ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}})]^2$, where $\ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}} := \arg\sup_{\boldsymbol{s}_{\boldsymbol{\lambda}} \in \Delta(\boldsymbol{\lambda}_*)} \max[0, \mathcal{Y}^{(\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\lambda}})]^2$. Thus, it follows that $n^{-1/2}\ DL_n(\ddot{\boldsymbol{\theta}}_n; \cdot) \Rightarrow \ddot{\mathcal{Z}}^{(\boldsymbol{\upsilon})}(\cdot; \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}})$ by applying the CMT, where

$$\ddot{\mathcal{Z}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}; \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}}) := \mathcal{Z}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}) - (\ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}))(\ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}})^{-1}\mathbf{Z}^{(\boldsymbol{\lambda})\prime}\ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}}.$$

We also note that $\ddot{\mathcal{Z}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}; \ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}}) = \mathcal{Z}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}) - \mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})'\{\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\}^{-1}\mathbf{Z}^{(\boldsymbol{\lambda})}$ by the definition of $\ddot{\boldsymbol{s}}_{\boldsymbol{\lambda}}$ and Corollaries 1(*iv* and *v*), and the final entry was defined as $\widetilde{\mathcal{Z}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})$ earlier. Third, we apply the ULLN and obtain that $\sup_{(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}})} |n^{-1}\widetilde{D}^2\ (\ddot{\boldsymbol{\theta}}_n; \boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) - \widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}})| \to 0$ a.s.–$\mathbb{P}$, where $\widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) := A_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}) - \boldsymbol{s}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})(\boldsymbol{s}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}})^{-1}\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})'\boldsymbol{s}_{\boldsymbol{\lambda}}$. Therefore, it follows that

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

Here, note that $\inf_{\boldsymbol{s}_{\boldsymbol{\lambda}} \in \Delta(\boldsymbol{\lambda}_*)}\{-\widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}})\} = -A_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}) + \sup_{\boldsymbol{s}_{\boldsymbol{\lambda}} \in \Delta(\boldsymbol{\lambda}_*)}\{\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})'\boldsymbol{s}_{\boldsymbol{\lambda}}\}\{\boldsymbol{s}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}}\}^{-1}\{\boldsymbol{s}_{\boldsymbol{\lambda}}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})\} = -\{A_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}}) - \mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})'\{\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\}^{-1}\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})\} = -\widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})$. Therefore, $\mathcal{LM}_n \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\upsilon}} \in \Delta(\boldsymbol{\upsilon}_0)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})]^2$ as desired. ■

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.

BAEK, Y., CHO, J., AND PHILLIPS, P.C.B. (2015): "Testing Linearity Using Power Transforms of Regressors," *Journal of Econometrics*, 187, 376–384.

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

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

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 ISHIDA, I. (2012): "Testing for the Effects of Omitted Power Transformations," *Economics Letters*, 117, 287–290.

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. (2014): "Testing for Neglected Nonlinearity Using Twofold Unidentified Models under the Null and Hexic Expansions," *Essays in Nonolinear Time Series Econometrics*. 3–27. Oxford: Oxford University Press.

CHO, J. AND PHILLIPS, P.C.B. (2016): "Sequentially Testing Polynomial Model Hypothesis Using the Power Transform of Regressors," Discussion Paper, School of Economics, Yonsei University.

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. AND WHITE, H. (2016): "Supplements to "Directionally Differentiable Econometric Models"," School of Economics, Yonsei University. <http://web.yonsei.ac.kr/jinseocho/research.htm>.

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.

FANG, Z. AND SANTOS, A. (2014): "Inference on Directionally Differentiable Functions," ArXiv preprint arXiv: 1404.3764.

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

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.

KIM, J. AND POLLARD, D. (1990). "Cube Root Asymptotics," *Annals of Statistics*, 18, 191–219.

LIU, X. AND SHAO, Y. (2003): "Asymptotics for Likelihood Ratio Tests under Loss of Identifiability," *Annals of Statistics*, 31 807–832.

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

POLLARD, D. (1985): "New Ways to Prove Central Limit Theorem," *Econometric Theory*, 1, 295–131.

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.

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.

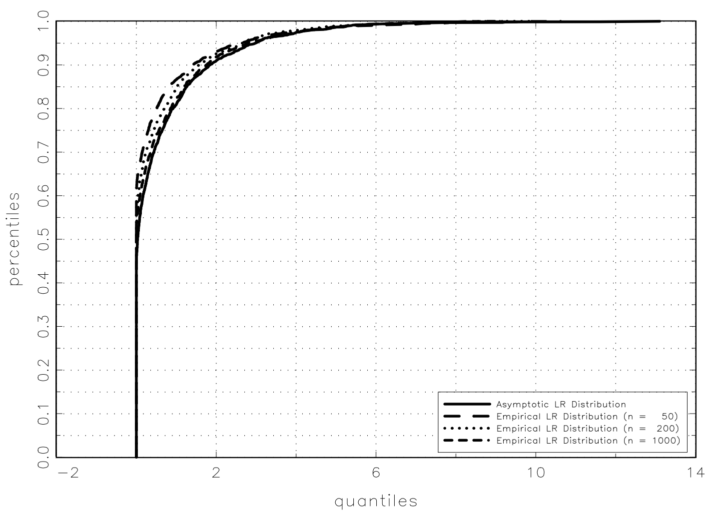

Figure 1: EMPIRICAL AND ASYMPTOTIC DISTRIBUTIONS OF THE QLR TEST STATISTIC. This figure shows the null limit distribution of the QLR test statistic and the empirical distributions of the QLR test statistic for $n = 50$, 200, and 1,000. The number of iterations is 2,000.

Supplements to "Directionally Differentiable Econometric Models"

JIN SEO CHO
School of Economics
Yonsei University
50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, 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

This version: October, 2016

Abstract

We illustrate analyzing directionally differentiable econometric models and provide technical details that are not contained in Cho and White (2016).

Key Words: directionally differentiable quasi-likelihood function, Gaussian stochastic process, quasi-likelihood ratio test, Wald test, and Lagrange multiplier test statistics, stochastic frontier production function, GMM estimation, Box-Cox transform.

JEL Classification: C12, C13, C22, C32.

Acknowledgements: The co-editor, Yoon-Jae Whang, and two anonymous referees provided very helpful comments for which we are most grateful. 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

This note illustrates analysis of econometric models formed by directionally differentiable (D-D) quasi-likelihood functions and provides technical details that are not contained in Cho and White (2016). All theorems, assumptions, and corollaries are those in Cho and White (2016) unless otherwise stated.

2 Examples

In this section, we illustrate the analysis of directionally differentiable econometric models using the stochastic frontier production function in Aigner, Lovell, and Schmidt (1977) and Stevenson (1980), Box and Cox's (1964) transformation, and the standard generalized methods of moments (GMM) estimation in Hansen (1982).

2.1 Example 1: Stochastic Frontier Production Function Models

A D-D quasi-likelihood function is found from the theory of stochastic frontier production function models. Stochastic production function models are often specified for identically and independently distributed (IID) observations $\{Y_t, \mathbf{X}_t\}$ as

$$Y_t = \mathbf{X}_t'\boldsymbol{\beta}_* + U_t,$$

where $Y_t \in \mathbb{R}$ is the output produced by 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, if $E[U_t] < 0$, 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 whether $\mu_* = 0$ 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 letting μ_* be different from zero, and the model with unknown μ_* has been popularly specified for empirical data analysis since then (e.g., Dutta, Narasimhan, and Rajiv (1999), Habib and Ljungqvist (2005), and etc.).

Nevertheless, we do not find a methodology testing whether $\mu_* = 0$ and $\sigma_*^2 = 0$ in prior literature to the best of our knowledge. This is mainly because the likelihood value is not identified 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 identified 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 identified. 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 identified by the model. Even further, this model is not differentiable (D). This can be verified by examining the first-order directional derivative. Some tedious algebra shows that for 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

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

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

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

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

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 D. Therefore, this model cannot be analyzed as for the standard D likelihood function. We examine this model by letting

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

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

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) central limit theorem (CLT) can be applied to $Z_{1,n}(\boldsymbol{d})$ and $Z_{2,n}(\boldsymbol{d})$: for each $\boldsymbol{d}$,

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

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 $\boldsymbol{d}$,

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

and $\mathbf{W} \sim N(\mathbf{0}_{k+2}, \mathbf{I}_{k+2})$, 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 $\widetilde{\mathcal{Z}}(\cdot) \stackrel{\mathrm{d}}{=} \mathcal{Z}(\cdot)$. Furthermore, $\widetilde{\mathcal{Z}}(\cdot)$ is linear with respect to $\mathbf{W}$. This feature makes 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 obtained. 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 law of large numbers (LLN):

$$\frac{1}{n}\sum_{t=1}^{n} U_t^2 = \tau_* + o_{\mathbb{P}}(1), \quad \frac{1}{n}\sum_{t=1}^{n} U_t\mathbf{X}_t = o_{\mathbb{P}}(1), \quad \text{and}$$

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{X}_t\mathbf{X}_t' = E[\mathbf{X}_t\mathbf{X}_t'] + o_{\mathbb{P}}(1).$$

This implies that

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

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 1(*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 null limit distribution can be obtained using this, although the log-likelihood is not properly identified under the null.

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

$$H_0:\boldsymbol{\upsilon}_*=\mathbf{0}\quad\text{versus}\quad H_1:\boldsymbol{\upsilon}_*\neq\mathbf{0}.$$

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 null limit distributions of the test statistics are identified by the theorems in Cho and White (2016). First, we apply the QLR test. Applying Theorem 2 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 $\boldsymbol{s}_{\boldsymbol{\pi}}\in\Delta(\boldsymbol{\pi}_*):=\{(\boldsymbol{s}_{\boldsymbol{\beta}}',s_\mu,s_\sigma)'\in\mathbb{R}^{k+2}:\boldsymbol{s}_{\boldsymbol{\beta}}'\boldsymbol{s}_{\boldsymbol{\beta}}+s_\mu^2+s_\sigma^2=1, s_\mu>0,\text{ and } s_\sigma>0\}$,

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

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

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

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

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

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

and also $\widetilde{\mathcal{Z}}^{(\boldsymbol{v})}(\boldsymbol{s_v}) := m(s_\mu, s_\sigma)\{Z^{(\boldsymbol{v})} - E[\mathbf{X}_t]'E[\mathbf{X}_t\mathbf{X}_t']^{-1}\mathbf{Z}^{(\boldsymbol{\beta})}\}$. Furthermore, Theorem 2 shows that

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

where for each $\boldsymbol{s_\beta} \in \Delta(\boldsymbol{\beta}_*) := \{\boldsymbol{s_\beta} \in \mathbb{R}^k : \boldsymbol{s}_{\boldsymbol{\beta}}'\boldsymbol{s_\beta} = 1\}$, $\mathcal{Y}^{(\boldsymbol{\beta})}(\boldsymbol{s_\beta}) := \{\boldsymbol{s}_{\boldsymbol{\beta}}'E[\mathbf{X}_t\mathbf{X}_t']\boldsymbol{s_\beta}\}^{-1/2}\mathbf{Z}^{(\boldsymbol{\beta})\prime}\boldsymbol{s_\beta}$, and applying Theorem 2(*iii*) 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, Theorem 2(*iv*) now yields that

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

under H_0, where $Z := \{1 - E[\mathbf{X}_t]'E[\mathbf{X}_t\mathbf{X}_t']^{-1}E[\mathbf{X}_t]\}^{-1/2}\{Z^{(\boldsymbol{v})} - E[\mathbf{X}_t]'E[\mathbf{X}_t\mathbf{X}_t']^{-1}\mathbf{Z}^{(\boldsymbol{\beta})}\} \sim N(0,1)$. 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 $\Delta(\boldsymbol{v}_0)$. Thus, the null limit 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 limit 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 1. Note that the null limit distributions of the QLR test statistics exactly overlap with that of $\max[0, -Z]^2$.

The null limit distribution of the QLR test can be uncovered by several simulation methods. The Monte Carlo method proposed by Dufour (2006) can also be used as the model is correctly specified. Although the null likelihood is not identified, the directional limits obtained under the null can be used to form the QLR test statistic. Hansen's (1996) weighted bootstrap can also be used to estimate the asymptotic p-values.

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 $\sup_{s_\mu, s_\sigma} |\widehat{W}_n(s_\mu, s_\sigma) - \widetilde{B}_*^{(\boldsymbol{v},\boldsymbol{v})}(s_\mu, s_\sigma)| \to 0$ 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 defined as

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

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

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

Theorem 3 now implies that $\mathcal{W}_n \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{v}} \in \Delta(\boldsymbol{v}_0)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{v})}(\boldsymbol{s}_{\boldsymbol{v}})]^2$, 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 the LM test statistic be defined as

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

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

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

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

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

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

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

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

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

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

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

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 $\Delta(\boldsymbol{v}_0)$ 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 Theorem 4 asserts.

Before moving to the next example, some remarks are in order. Here, we assume $\mu_* \geq 0$ so that d_μ is always greater than or equal to zero, and this is assumed to avoid the failure in numerical simulation. It is more general to suppose that μ_* can be negative, so that for some positive $c > 0$, $\mu_* \in [-c, c]$. For such a case, for example, the null limit distribution of the QLR test is modified into

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

where $\Delta(\boldsymbol{v}_0)' := \{(s_\mu, s_\sigma) \in \mathbb{R}^2 : s_\mu^2 + s_\sigma^2 = 1 \text{ and } s_\sigma > 0\}$. Furthermore, it analytically follows that $m(s_\mu, s_\sigma)/|m(s_\mu, s_\sigma)| = -1$ uniformly on $\Delta(\boldsymbol{v}_0)'$, so that $\mathcal{LR}_n \Rightarrow \max[0, -Z]^2$, which is the same as for the previous case. Nevertheless, our Monte Carlo experiments assuming the same condition showed that the empirical distribution of $\mathcal{LR}_n$ exactly overlaps with that of Z^2 under the null.

This discrepancy is mainly because the value of $m(s_\mu, s_\sigma)$ sensitively responds to the value of (s_μ, s_σ), so that we obtain that $m(\,\cdot\,,\,\cdot\,)/|m(\,\cdot\,,\,\cdot\,)| = \pm 1$ numerically on $\Delta(\boldsymbol{v}_0)'$. This also implies that

$$\mathcal{LR}_n \Rightarrow \sup_{\boldsymbol{s_v} \in \Delta(\boldsymbol{v}_0)'} \max\left[0, \frac{m(s_\mu, s_\sigma)}{|m(s_\mu, s_\sigma)|} Z\right]^2 = \sup_{\boldsymbol{s_v} \in \Delta(\boldsymbol{v}_0)'} \max\,[0, -Z, Z]^2 = Z^2$$

as can be easily verified by Monte Carlo experiments. More precisely, if $s_\mu < 0$ and $s_\sigma > 0$, so that we can let $s_\mu = -\sqrt{1 - s_\sigma^2}$, it analytically follows that $\lim_{s_\sigma \downarrow 0} m(-\sqrt{1 - s_\sigma^2}, s_\sigma) = 0$ and for any $s_\sigma > 0$, $m(-\sqrt{1 - s_\sigma^2}, s_\sigma) < 0$. Nevertheless, computing this value requires pretty high level of precision around $s_\sigma = 0$, and standard statistical packages do not often provide this level of precision. Numerically, they compute $m(-\sqrt{1 - s_\sigma^2}, s_\sigma)$ oscillating around zero as s_σ converges to 0, so that $m(\,\cdot\,,\,\cdot\,)/|m(\,\cdot\,,\,\cdot\,)|$ is obtained as ± 1 on $\Delta(\boldsymbol{v}_0)'$. Our parameter space restriction is imposed to avoid this numerical failure by letting $\mu_* \geq 0$.

2.2 Example 2: 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:

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

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, $\boldsymbol{\theta} := (\boldsymbol{\theta}_0', \theta_1, \theta_2)' \in \boldsymbol{\Theta}_0 \times \boldsymbol{\Theta}_{12}$, $\boldsymbol{\Theta}_0$ is a convex and compact set in $\mathbb{R}^k$, and

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

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 (1) 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 (1996), 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 avoid the identification problem by re-parameterizing the model using θ_1 and θ_2 as given in (1). If $\theta_{2*} = 0$, θ_{1*} has to be zero by the model condition on $\mathbf{\Theta}_{12}$, and 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 null model is not properly obtained from the model. Note that $\theta_{1*}(X_t^{\theta_{2*}} - 1)/\theta_{2*} = 0 \times 0/0$, implying that the standard tests cannot be applied.

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

$$DL_n(\boldsymbol{\theta}_*; \boldsymbol{d}) = \sum_{t=1}^{n} U_t\{\mathbf{Z}_t{}'\boldsymbol{d}_0 + \log(X_t)d_1\}, \quad \text{and} \tag{2}$$

$$D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{d}) = -\sum_{t=1}^{n}\{\mathbf{Z}_t{}'\boldsymbol{d}_0 + \log(X_t)d_1\}^2 + \sum_{t=1}^{n} U_t\{\log(X_t)\}^2 d_1 d_2, \tag{3}$$

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 properly obtained from the model.

As a remark on this, this reformulation implies that there is a hidden identification problem associated with d_1/d_2. Note that d_1/d_2 lacks its corresponding distance and disappears if h is zero, so that d_1/d_2 is not identified at $\boldsymbol{\theta}_* = (\boldsymbol{\theta}_{0*}', 0, 0)'$.

Using the first and second-order directional derivatives in (2) and (3),

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

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 maximization 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 1(*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 null limit distribution because d_2 vanishes at the limit. We also note that the limit result is the same as what is obtained when an identified model is D.

For the model inference, we let $\boldsymbol{\pi} = (\boldsymbol{\lambda}', \boldsymbol{\upsilon}')'$ such that $\boldsymbol{\lambda} = \boldsymbol{\theta}_0$ and $\boldsymbol{\upsilon} = \theta_2$, 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, Theorem 2(*iv*) can be applied. That is,

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

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

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

Note that $s_1/|s_1| = \pm 1$, and from this,

$$\mathcal{LR}_n \Rightarrow \widetilde{\mathcal{Z}}^{(\boldsymbol{v})}(\widetilde{A}_*^{(\boldsymbol{v},\boldsymbol{v})})^{-1}\widetilde{\mathcal{Z}}^{(\boldsymbol{v})}.$$

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

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

then

$$\mathcal{W}_n := n\{\widetilde{h}_n^{(\boldsymbol{v})}(\boldsymbol{s_v})\}\{\widehat{W}_n\}\{\widetilde{h}_n^{(\boldsymbol{v})}(\boldsymbol{s_v})\} \Rightarrow \widetilde{\mathcal{Z}}^{(\boldsymbol{v})}(\widetilde{A}_*^{(\boldsymbol{v},\boldsymbol{v})})^{-1}\widetilde{\mathcal{Z}}^{(\boldsymbol{v})}$$

by Theorem 3. Finally, Theorem 4 obtains the same null limit distribution for the LM test statistic using the same weight function.

2.3 Example 3: Generalized Method of Moments (GMM)

Hansen (1982) examines an estimation method by generalizing the method of moments estimation that requires 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 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 $E[\mathbf{q}_t(\cdot)]'\{-\mathbf{M}_*\}^{-1}E[\mathbf{q}_t(\cdot)]$ 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 conditions in Assumption 2. Even so, our D-D analysis can be easily adapted to the GMM estimation 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

$$D\mathbf{g}_n(\boldsymbol{\theta};\boldsymbol{d}) = \nabla_{\boldsymbol{\theta}}\mathbf{g}_n(\mathbf{X}^n;\boldsymbol{\theta})'\boldsymbol{d}, \tag{4}$$

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 (4) 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 $\boldsymbol{d}$,

$$\mathbf{g}_n(\boldsymbol{\theta};\boldsymbol{d}) = \mathbf{g}_n(\boldsymbol{\theta}_*;\boldsymbol{d}) + D\mathbf{g}_n(\bar{\boldsymbol{\theta}};\boldsymbol{d})(\boldsymbol{\theta}-\boldsymbol{\theta}_*). \tag{5}$$

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 CLT. Here, we applied the LLN to obtain that $\nabla_{\boldsymbol{\theta}}\mathbf{g}_n(\boldsymbol{\theta}_*)$ converges to $\mathbf{C}_* := E[\nabla_{\boldsymbol{\theta}}\mathbf{q}_t(\boldsymbol{\theta}_*)]$ a.s.$-\mathbb{P}$ by the fact that $\|\nabla_{\boldsymbol{\theta}}\mathbf{q}_t(\,\cdot\,)\|_\infty \leq m(\mathbf{X}_t)$. We below use these facts and the vehicles for D-D analysis to obtain the asymptotic behavior of the GMM estimator.

Given (4), it is trivial to show that $\{n^{1/2}DQ_n(\boldsymbol{\theta}_*;\,\cdot\,)\}$ is asymptotically 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 (5) 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 let $\mathcal{Z}(\boldsymbol{d}) := -\boldsymbol{d}'\mathbf{C}_*{}'\mathbf{M}_*^{-1}\mathbf{W}$ 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, by applying Corollary 1(*v*), we obtain that

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

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

As the objective function is D, we simply let $\boldsymbol{\theta} = \boldsymbol{\pi} = (\boldsymbol{\upsilon}',\boldsymbol{\lambda}')'$ for testing the hypothesis. Furthermore, the objective function $Q_n(\cdot)$ does not satisfy the condition in Assumption 2. Therefore, the definition of the

QLR test statistic cannot be exactly applied to this case. Nevertheless, a QLR test-like test statistic can be defined. We let

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

and also let $\mathbf{C}_*'\{-\mathbf{M}_*\}^{-1}\mathbf{W}$ and $\mathbf{C}_*'\{-\mathbf{M}_*\}^{-1}\mathbf{C}_*$ be $\mathbf{Z}^{(\boldsymbol{\pi})} = (\mathbf{Z}^{(\boldsymbol{\upsilon})\prime}, \mathbf{Z}^{(\boldsymbol{\lambda})\prime})'$ and $\mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}$ in Cho and White (2016), respectively. The null limit distribution of the QLR test statistic is obtained as

$$\mathcal{QLR}_n \Rightarrow (\widetilde{\mathbf{Z}}^{(\boldsymbol{\upsilon})})'(-\widetilde{\mathbf{A}}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})})^{-1}(\widetilde{\mathbf{Z}}^{(\boldsymbol{\upsilon})}),$$

where $\widetilde{\mathbf{Z}}^{(\boldsymbol{\upsilon})} := \mathbf{Z}^{(\boldsymbol{\upsilon})} - (\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})})(\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})^{-1}\mathbf{Z}^{(\boldsymbol{\lambda})}$ and $\widetilde{\mathbf{A}}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})} := \mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})} - (\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})})(\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})^{-1}(\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})})'$ by applying Corollary 1. We can also define the Wald test statistic using the GMM estimator and derive its null limit distribution as before. That is,

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

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

$$Q_n(\boldsymbol{\upsilon}_0 + \widetilde{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})\boldsymbol{s}_{\boldsymbol{\upsilon}}, \widetilde{\boldsymbol{\lambda}}_n(\boldsymbol{s}_{\boldsymbol{\upsilon}})) := \sup_{\{h^{(\boldsymbol{\upsilon})},\boldsymbol{\lambda}\}} Q_n(\boldsymbol{\upsilon}_0 + h^{(\boldsymbol{\upsilon})}\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{\lambda}),$$

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

$$\mathcal{QW}_n \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\upsilon}}\in\Delta(\boldsymbol{\upsilon}_0)} \max[0, \boldsymbol{s}_{\boldsymbol{\upsilon}}'\widetilde{\mathbf{Z}}^{(\boldsymbol{\upsilon})}](-\boldsymbol{s}_{\boldsymbol{\upsilon}}'\widetilde{\mathbf{A}}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}\boldsymbol{s}_{\boldsymbol{\upsilon}})^{-1}\max[0, \boldsymbol{s}_{\boldsymbol{\upsilon}}'\widetilde{\mathbf{Z}}^{(\boldsymbol{\upsilon})}].$$

The proof of Corollary 1(*vii*) corroborates that the null limit distribution of $\mathcal{QW}_n$ is equivalent to that of $\mathcal{QLR}_n$ particularly because $\boldsymbol{\upsilon}_0$ is an interior element. Finally, we define the LM test statistic in the GMM context and examine its null limit distribution. For this purpose, we let

$$\mathcal{QLM}_n := \sup_{(\boldsymbol{s}_{\boldsymbol{\upsilon}},\boldsymbol{s}_{\boldsymbol{\lambda}})\in\Delta(\boldsymbol{\upsilon}_0)\times\Delta(\ddot{\boldsymbol{\lambda}}_n)} n\widetilde{W}_n(\boldsymbol{s}_{\boldsymbol{\upsilon}},\boldsymbol{s}_{\boldsymbol{\lambda}})\max\left[0, \frac{DQ_n(\ddot{\boldsymbol{\theta}}_n;\boldsymbol{s}_{\boldsymbol{\upsilon}})}{2\widetilde{D}^2Q_n(\ddot{\boldsymbol{\theta}}_n;\boldsymbol{s}_{\boldsymbol{\upsilon}},\boldsymbol{s}_{\boldsymbol{\lambda}})}\right]^2,$$

where for each $(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda})$,

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

and $\ddot{\boldsymbol{\theta}}_n := (\boldsymbol{\upsilon}_0, \ddot{\boldsymbol{\lambda}}_n)$ such that $\ddot{\boldsymbol{\lambda}}_n := \arg\max_{\boldsymbol{\lambda}} Q_n(\boldsymbol{\upsilon}_0, \boldsymbol{\lambda})$. If we let $\widetilde{W}_n(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}) = \boldsymbol{s_\upsilon}'\widehat{W}_n \boldsymbol{s_\upsilon}$ for each $(\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}) \in \Delta(\boldsymbol{\upsilon}_0) \times \Delta(\boldsymbol{\lambda}_*)$,

$$
\mathcal{QLM}_n \Rightarrow (\widetilde{\mathbf{Z}}^{(\boldsymbol{\upsilon})})'(-\widetilde{\mathbf{A}}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})})^{-1}(\widetilde{\mathbf{Z}}^{(\boldsymbol{\upsilon})})
$$

by Theorem 4, the interiority condition of $\boldsymbol{\upsilon}_0$, 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.

Indeed, many nonlinear models share the similar features. For example, table 1 of Cheng, Evans, and Iles (1992) collects a number of nonlinear models with parameter instability problems. Many of them can be analyzed using the approach of the current study. Furthermore, D-D analysis tools simplify dimensional complexities that arise when higher-order approximations are needed for model analysis. Cho, Ishida, and White (2011,2014) 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.

3 Differentiable Model and Directionally Differentiable Model

In this section, we provide sufficient conditions for a twice D-D function to be twice differentiable.

Theorem C1. *If a function $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is (i) 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 linear in $\boldsymbol{d} \in \Delta(\boldsymbol{\theta})$, then $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is D on $\boldsymbol{\Theta}$.* □

Proof of Theorem C1: Refer to Troutman (1996, p. 122). ■

Theorem C2. *In addition to the conditions in Theorem C1, if a function $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is (i) twice D-D on $\boldsymbol{\Theta}$; (ii) for each $\boldsymbol{\theta}, \boldsymbol{\theta}'$ and for some $M < \infty$, $|D^2 f(\boldsymbol{\theta}'; \widetilde{\boldsymbol{d}}; \boldsymbol{d}) - D^2 f(\boldsymbol{\theta}; \widetilde{\boldsymbol{d}}; \boldsymbol{d})| \leq M\|\boldsymbol{\theta}' - \boldsymbol{\theta}\|$ uniformly on $\Delta(\boldsymbol{\theta}) \cap \Delta(\boldsymbol{\theta}') \times \Delta(\boldsymbol{\theta}) \cap \Delta(\boldsymbol{\theta}')$; and (iii) for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the directional derivative of $Df(\boldsymbol{\theta}; \boldsymbol{d})$ with respect to $\widetilde{\boldsymbol{d}}$ is linear in $\widetilde{\boldsymbol{d}} \in \Delta(\boldsymbol{\theta})$, then $f : \boldsymbol{\Theta} \mapsto \mathbb{R}$ is twice D on $\boldsymbol{\Theta}$.* □

Proof of Theorem C2: To show the given claim, we note that $f(\cdot)$ is differentiable on $\mathbf{\Theta}$ by Theorem C1 and denote the gradient of $f(\cdot)$ as $\mathbf{A}(\cdot)$. We next show that for some $\mathbf{B}(\cdot)$,

$$\lim_{\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|\to 0} \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|=1} \frac{1}{\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|} \left| \mathbf{A}(\widetilde{\boldsymbol{\theta}})'(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - \mathbf{A}(\boldsymbol{\theta}_0)'(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - (\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)'\mathbf{B}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) \right| = 0.$$

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

$$g'(h) = g'(0) + g''(\bar{h})h,$$

implying that $Df(\boldsymbol{\theta}_0 + h\widetilde{\boldsymbol{d}}; \boldsymbol{d}) = Df(\boldsymbol{\theta}_0; \boldsymbol{d}) + D^2 f(\boldsymbol{\theta}_0; \boldsymbol{d}; \widetilde{\boldsymbol{d}})h\bar{h}$, where

$$D^2 f(\boldsymbol{\theta}_0; \widetilde{\boldsymbol{d}}; \boldsymbol{d}) := \lim_{h\downarrow 0} \frac{Df(\boldsymbol{\theta}_0 + h\widetilde{\boldsymbol{d}}; \boldsymbol{d}) - Df(\boldsymbol{\theta}_0; \boldsymbol{d})}{h}.$$

Given this, note that the given conditions imply that $Df(\boldsymbol{\theta}_0; \boldsymbol{d}) = \mathbf{A}(\boldsymbol{\theta}_0)'\boldsymbol{d}$ and $D^2 f(\boldsymbol{\theta}_0; \widetilde{\boldsymbol{d}}; \boldsymbol{d}) = \widetilde{\boldsymbol{d}}'\mathbf{B}(\boldsymbol{\theta}_0)\boldsymbol{d}$. Therefore, if we let $\widetilde{\boldsymbol{\theta}} := \boldsymbol{\theta}_0 + h\widetilde{\boldsymbol{d}}$, then $\mathbf{A}(\widetilde{\boldsymbol{\theta}})'\boldsymbol{d} = \mathbf{A}(\boldsymbol{\theta}_0)'\boldsymbol{d} + h\widetilde{\boldsymbol{d}}'\mathbf{B}(\boldsymbol{\theta}_0 + \bar{h}\widetilde{\boldsymbol{d}})\boldsymbol{d}$, so that

$$\mathbf{A}(\widetilde{\boldsymbol{\theta}})'\boldsymbol{d} - \mathbf{A}(\boldsymbol{\theta}_0)'\boldsymbol{d} - h\widetilde{\boldsymbol{d}}'\mathbf{B}(\boldsymbol{\theta}_0)\boldsymbol{d} \leq h\widetilde{\boldsymbol{d}}'\mathbf{B}(\boldsymbol{\theta}_0 + \bar{h}\widetilde{\boldsymbol{d}})\boldsymbol{d} - h\widetilde{\boldsymbol{d}}'\mathbf{B}(\boldsymbol{\theta}_0)\boldsymbol{d},$$

implying that

$$\frac{1}{h}|\mathbf{A}(\widetilde{\boldsymbol{\theta}})'\boldsymbol{d} - \mathbf{A}(\boldsymbol{\theta}_0)'\boldsymbol{d} - h\widetilde{\boldsymbol{d}}'\mathbf{B}(\boldsymbol{\theta}_0)\boldsymbol{d}| \leq \frac{1}{h}|\widetilde{\boldsymbol{d}}'[\mathbf{B}(\boldsymbol{\theta}_0 + \bar{h}\widetilde{\boldsymbol{d}}) - \mathbf{B}(\boldsymbol{\theta}_0)]\boldsymbol{d}| \leq M \cdot \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|,$$

where the last inequality follows from the uniform bound condition. We further note that $h = \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|$. This implies that

$$\lim_{\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|\to 0} \frac{1}{\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|}|\mathbf{A}(\widetilde{\boldsymbol{\theta}})'\boldsymbol{d} - \mathbf{A}(\boldsymbol{\theta}_0)'\boldsymbol{d} - h\widetilde{\boldsymbol{d}}'\mathbf{B}(\boldsymbol{\theta}_0)\boldsymbol{d}| \leq \lim_{\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|\to 0} M \cdot \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = 0.$$

This completes the proof. ■

References

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

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

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. AND WHITE, H. (2016): "Directionally Differentiable Econometric Models," Discussion Paper, School of Economics, Yonsei University.

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. (2014): "Testing for Neglected Nonlinearity Using Twofold Unidentified Models under the Null and Hexic Expansions," *Essays in Nonolinear Time Series Econometrics*. 3–27. 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.

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.

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

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.

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

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

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

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

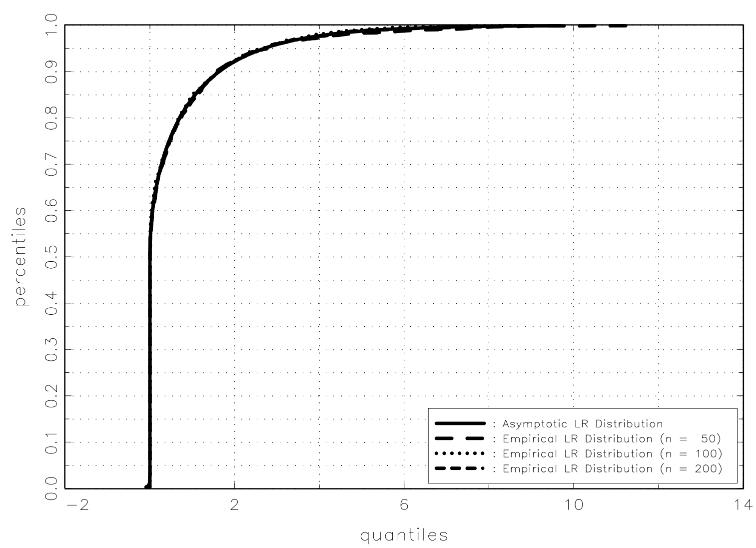

Figure 1: EMPIRICAL AND ASYMPTOTIC DISTRIBUTIONS OF THE QLR TEST STATISTIC. This figure shows the null limit distribution of the QLR test statistic, which is obtained as $\max[0, -Z]^2$, and the empirical distributions of the QLR test statistic for various sample sizes: $n = 50$, 100, and 500. The number of iterations for obtaining the empirical distributions is 2,000. We can see that the empirical distributions almost overlap with the null limit distribution even when the sample size is as small as 50.