Econometric Theory, **34**, 2018, 1101–1131.
doi:10.1017/S0266466617000354

DIRECTIONALLY DIFFERENTIABLE ECONOMETRIC MODELS

JIN SEO CHO
Yonsei University

HALBERT WHITE
University of California

The current article 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. In this way, the standard analysis that assumes a differentiable quasi-likelihood function is treated as a special case of our analysis. We also examine and redefine the standard quasi-likelihood ratio, Wald, and Lagrange multiplier test statistics so that their null limit behaviors are regular under our model framework.

1. INTRODUCTION

Differentiability is one of the regularity conditions for analyzing standard econometric models. For example, Wald (1943) proposed it as one of the regularity conditions for his classic test statistic. As another example, Chernoff (1954) examined use of the likelihood ratio (LR) test statistic by approximating the log-likelihood function by Taylor's expansion. Model differentiability is required for the approximation.

Many important econometric models are estimated by nondifferentiable quasi-likelihood functions. For example, the likelihood function examined by King and Shively (1993) is not differentiable. They attempted to resolve the so-called Davies's (1977, 1987) identification problem by reparameterizing the original parameter space through the polar coordinates. The consequent likelihood

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 article and also provided an exemplary guide for writing a quality research paper. The authors 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, Hon Ho Kwok, Taesuk Lee, Bruce Lehmann, Mark Machina, Jaesun Noh, Kosuke Oya, Taeyoung Park, Peter Phillips, Erwann Sbai, Juwon Seo, Donggyu Sul, Denis Tkachenko, Albert K.C. Tsui, Hung-Jen Wang, Yoshihiro Yajima, Byung Sam Yoo, Ping Yu, and other seminar participants at Sogang University, the University of Auckland, the University of Hong Kong, 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). Cho acknowledges support from the Yonsei University Future-leading Research Initiative of 2017 (2017-22-0090). Address correspondence to Jin Seo Cho, School of Economics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea; e-mail: jinseocho@yonsei.ac.kr.

function, however, is not differentiable (D) but only directionally differentiable (D-D). Additionally, Aigner, Lovell, and Schmidt (1977) and Stevenson (1980) specified the stochastic frontier production function model to capture inefficiently produced outputs. Their model, however, is not D under the null of efficient production. In addition to these, there are many other quasi-likelihood functions in prior literature that are not D and require a different model analysis from the standard case.

The goal of this article is, therefore, to extend the model analysis scope to include nondifferentiable models. Specifically, we suppose that model parameters are estimated by maximizing D-D quasi-likelihood functions. As it turns out, the class of D-D functions includes D functions as a special case, so that we may obtain generalized analysis outputs from D-D quasi-likelihood functions. In the current study, we achieve this generalization by associating Billingsley's (1999) asymptotically tight probability measure condition with the score of D-D quasi-likelihood functions. Each direction around the parameter of interest is regarded as an index indicating a particular value of directional derivatives, by which we can apply the functional central limit theorem (FCLT) and the uniform law of large numbers (ULLN) to the first- and second-order directional derivatives, respectively. Through this process, the large sample properties of quasi-maximum likelihood (QML) estimator (or M-estimator) of D quasi-likelihood functions can be generalized to address D-D quasi-likelihood functions.

Another goal of this study is to provide test statistics that can be properly used for data inference via D-D models. The conventional quasi-likelihood ratio (QLR), Wald, and Lagrange multiplier (LM) test statistics are defined by assuming model differentiability, so that they may or may not be proper for D-D models. We examine the test statistics under the D-D model assumption and provide alternatives in case they are not proper for D-D models. As a result, the Wald and LM test statistics are redefined under the D-D model assumption to maintain their testing principles. We also show that the three test statistics are asymptotically equivalent under the null hypothesis and mild regularity conditions detailed below, thereby achieving the dual purpose of estimating D-D models and inferring data through the model estimation.

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 an efficient vehicle for their model analysis. In addition to this, we include other analyses in the Supplement to this study and demonstrate the usefulness of the current analysis (see Cho and White, 2017). They include Aigner et al. (1977) and Stevenson's (1980) stochastic frontier production function model and the Box-Cox transformation. The standard generalized method of moments (GMM) estimation is also revisited using the D-D model analysis.

The approach of the current study is related to the prior literature. First, Pollard (1985) examined stochastically differentiable quasi-likelihood functions that are not D although their population analogs are D. The D-D quasi-likelihood function here is not stochastically differentiable because the population quasi-likelihood function is D-D, let alone its sample analog. Second, Andrews (2001) examined data inference when there is an unidentified parameter under a maintained null hypothesis that is possibly on the boundary of the parameter space. Indeed, one may reparameterize the parameter space to avoid an unidentified model feature and instead employ D-D quasi-likelihood functions. King and Shively's (1993) model analysis is a typical example of this, as detailed below. The analysis here, nevertheless, does not assume D-D quasi-likelihood functions obtained only through reparameterization. General D-D quasi-likelihood functions are assumed throughout this study so that the analysis here can be a vehicle for general model analysis. The models in Aigner et al. (1977) and Stevenson (1980) examined in the Supplement belong to this case. Finally, Fang and Santos (2014) examined a D-D transform of a consistent estimator and noted that a 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.

The plan of this article 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. Section 3 considers data inference using D-D models. For illustration purposes, we exploit King and Shively's (1993) reparameterized model throughout this article, including Monte Carlo experiments using the same model. Section 4 offers concluding remarks, and formal mathematical proofs are collected in the Appendix. In the Supplement to this study, we provide additional examples for D-D model analysis.

Before moving to the next section, we introduce the mathematical notation used throughout this study. 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 article. The following is the data generating process (DGP) condition:

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

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

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

Assumption 2 (Quasi-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 $(\mathbf{X}_1,\ldots,\mathbf{X}_t)$, and $\boldsymbol{\Theta}$ is a compact and convex set in $\mathbb{R}^r$ with $r \in \mathbb{N}$.

This quasi-likelihood function condition is widely used in the literature, and the QML estimator is defined by the quasi-likelihood function: let $\widehat{\boldsymbol{\theta}}_n$ be the QML estimator such that $L_n(\widehat{\boldsymbol{\theta}}_n) = \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_n(\boldsymbol{\theta})$. 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; and (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.

Several remarks are warranted regarding Assumption 3. First, Assumption 3(*i*) requires model identification. Even if models are not identified, D-D model analysis can still be made using the framework of Davies (1977,1987), but it renders the key aspects of D-D quasi-likelihood functions obscure. Therefore, we highlight the D-D model analysis by assuming model identification. Second, $\boldsymbol{\theta}_*$ can be on the boundary of $\boldsymbol{\Theta}$ as often ensured by the reparameterization method of King and Shively (1993). Assumption 3(*ii*) permits this. Finally, we abbreviate $\ell_t(\cdot;\mathbf{X}^t)$ into $\ell_t(\cdot)$ henceforth for notational simplicity.

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. Another implication is that the differentiability condition is not necessary for the consistency of the QML estimator as Wald (1949) noted. On the other hand, the limit distribution of the QML estimator is critically determined by model differentiability. We discuss this in the next subsections.

2.1. Directional Differentiability

In this subsection, we define the D-D function and characterize the D function through the D-D 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

https://doi.org/10.1017/S0266466617000354 Published online by Cambridge University Press

$\Delta(\boldsymbol{\theta})$, *if for all* $\boldsymbol{d} \in \Delta(\boldsymbol{\theta})$, $Df(\boldsymbol{\theta};\boldsymbol{d})$ *exists*; *and* (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. First, the definition of the D-D function is weaker than that of the D function. D-D functions can have different directional derivatives that are nonlinearly dependent upon $\boldsymbol{d}$, 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) noted that it is essential to approximate the parameter space by a cone $C(\boldsymbol{\theta})$ to obtain 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 the domain for a Gaussian stochastic process that is introduced below. Note that even when $\boldsymbol{\theta}$ is on the boundary of $\boldsymbol{\Theta}$, $\Delta(\boldsymbol{\theta})$ can still be defined not to contain the directions of the boundary side. Finally, another norm other than the Euclidean norm can 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 in $\Delta(\boldsymbol{\theta})$. We continue our discussion using $\Delta(\boldsymbol{\theta})$.

A regular relationship exists between D-D and D functions as Troutman (1996, p. 122) described. 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 a key condition for a D-D function to be D. Without this, directional derivatives cannot be represented as linear combinations of other directional derivatives.

We provide the following 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* $\widetilde{\boldsymbol{d}} \in \Delta(\boldsymbol{\theta})$, $D^2 f(\boldsymbol{\theta};\widetilde{\boldsymbol{d}};\boldsymbol{d})$ *exists, where*

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

Note that first-order directional differentiability is necessary to define 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 discuss this in the Supplement more precisely. Henceforth, 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. Example: Conditional Heteroskedasticity

Many econometric models are specified using D-D quasi-likelihood functions. In this subsection, we illustrate King and Shively's (1993) model as a representative example of D-D models and demonstrate that the notion of the D-D function is important in practice. We include other examples in the Supplement.

King and Shively (1993) examined 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 assumed

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

where $\mathbf{Y}^n := (Y_1, \ldots, Y_n)'$, $\mathbf{U}^n := (U_1, \ldots, U_n)'$, $\mathbf{W}^n := (W_1, \ldots, W_n)'$, $\mathbf{R}^n$ is an $n \times k$ matrix with $\mathbf{R}_t{}'$ at t-th row, $\mathbf{Q}^n := (\mathbf{W}^n, \mathbf{R}^n)$, and $\mathbf{\Omega}^n(\rho_*)$ is an $n \times n$ square matrix with t-th row and t'-th column element $\mathbf{\Omega}^n_{tt'}(\rho_*) := W_t W_{t'} \rho_*{}^{|t'-t|}/(1-\rho_*{}^2)$. Furthermore, they let $(\gamma_*', \sigma_*^2, \kappa_*, \rho_*) := (\alpha_*, \beta_*', \sigma_*^2, \kappa_*, \rho_*)$ be an unknown parameter in $\mathbf{\Gamma} \times [0, \bar{\sigma}^2] \times [0, \bar{\kappa}] \times [0, \bar{\rho}]$, where $\mathbf{\Gamma}$ 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 $(\gamma, \sigma^2, \kappa, \rho)$, its log-likelihood can be written as

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

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

This model was motivated by Rosenberg (1973), who aimed to test $\kappa_* = 0$ and examined whether a systematic risk of an asset is time-varying. If $\kappa_* \neq 0$, the conditional covariance of $\mathbf{U}^n|\mathbf{Q}^n$ depends on $\mathbf{W}^n$, so that the error exhibits time-varying conditional heteroskedasty. On the other hand, if $\kappa_* = 0$, the error exhibits conditional homoskedasticity, but ρ_* is not identified, leading to Davies's (1977, 1987) identification problem. This renders the null limit distributions of the standard test statistics nonstandard.

King and Shively (1993) attempted 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

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

Furthermore, the original hypotheses are modified into $H_0' : \boldsymbol{\theta}_*{}'\boldsymbol{\theta}_* = 0$ versus $H_1' : \boldsymbol{\theta}_*{}'\boldsymbol{\theta}_* > 0$. Note that the null parameter value is on the boundary of $\mathbf{\Theta}$ and the identification problem no longer arises under H_0'. On the other hand, the reparameterized quasi-likelihood function is not D. It is indeed D-D under H_0': for each $(\gamma, \sigma^2, \boldsymbol{\theta})$, the modified log-likelihood is

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

and from this, θ_{2*}/θ_{1*} has the form of $0/0$ under the null, and this renders $\tan^{-1}(0/0)$ undefined. Nonetheless, this yields the null log-likelihood desired by Rosenberg (1973) and King and Shively (1993): for each $\boldsymbol{d} := (\boldsymbol{d}_\gamma', d_\sigma, d_1, d_2)'$

https://doi.org/10.1017/S0266466617000354 Published online by Cambridge University Press

such that $\boldsymbol{\theta}_* = \mathbf{0}$ and $\boldsymbol{d}'\boldsymbol{d} = 1$,

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

as desired, because $0 \times \tan^{-1}(\cdot) \equiv 0$ on the Euclidean real line. Furthermore,

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

which 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

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

which is not quadratic with respect to $\boldsymbol{d}$, implying the D model analysis cannot be applied for this model. In particular, the limit distribution of the QML estimator must be differently obtained from the standard case.

There are many other D-D models. For example, the stochastic frontier production function model introduced by Aigner et al. (1977) and Stevenson (1980) is also D-D. As another example, if Box-Cox (1964) transformation is used as a regressor, the model is D-D when the regressor does not reduce the prediction error variance. In the Supplement, we analyze them along with the GMM estimation defined by the D model using the method of this study.

2.3. Asymptotic Distribution of the QML Estimator

As noted in Section 2.1, the most significant difference between D-D and D functions lies in the linearity condition of $Df(\boldsymbol{\theta}; \boldsymbol{d})$ in $\boldsymbol{d}$. In this section, we provide further regularity conditions for D-D models. In particular, the smoothness condition of the D-D quasi-likelihood function is important in obtaining the limit distribution of $\widehat{\boldsymbol{\theta}}_n$.

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.$-\mathbb{P}$.

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 conditions are also imposed:

Assumption 5 (Mode of continuity)**.** (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 $\boldsymbol{\theta},\boldsymbol{\theta}' \in \boldsymbol{\Theta}$, $|D\ell_t(\boldsymbol{\theta};\boldsymbol{d}) - D\ell_t(\boldsymbol{\theta}';\boldsymbol{d})| \le M_t\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|$ and $|D^2\ell_t(\boldsymbol{\theta};\boldsymbol{d}) - D^2\ell_t(\boldsymbol{\theta}';\boldsymbol{d})| \le M_t\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|$ uniformly on $\Delta(\boldsymbol{\theta}) \cap \Delta(\boldsymbol{\theta}')$, where $\{M_t\}$ is a sequence of stationary and ergodic variables; and (iii) for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and for all $\boldsymbol{d}_1,\boldsymbol{d}_2 \in \Delta(\boldsymbol{\theta})$, there is $\lambda > 0$ such that $|D\ell_t(\boldsymbol{\theta};\boldsymbol{d}_1) - D\ell_t(\boldsymbol{\theta};\boldsymbol{d}_2)| \le M_t\|\boldsymbol{d}_1-\boldsymbol{d}_2\|^{\lambda}$ and $|D^2\ell_t(\boldsymbol{\theta};\boldsymbol{d}_1) - D^2\ell_t(\boldsymbol{\theta};\boldsymbol{d}_2)| \le M_t\|\boldsymbol{d}_1-\boldsymbol{d}_2\|^{\lambda}$.

The examples mentioned in Section 2.2 satisfy Assumptions 4 and 5. Here, Assumption 5(*iii*) is assumed to apply the asymptotic tightness and ULLN to the first- and second-order directional derivatives, respectively. We detail the asymptotic tightness and 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.–$\mathbb{P}$:

Assumption 5 (Mode of continuity)**.** (iii)* For each $\boldsymbol{\theta}$ and for all $\boldsymbol{d} \in \Delta(\boldsymbol{\theta})$, $D\ell_t(\boldsymbol{\theta};\boldsymbol{d})$ and $D^2\ell_t(\boldsymbol{\theta};\widetilde{\boldsymbol{d}};\boldsymbol{d})$ are 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**) going forward when D quasi-likelihood functions are referenced. 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) $A_*(\boldsymbol{d}) := E[n^{-1}D^2L_n(\boldsymbol{\theta}_*;\ \boldsymbol{d})]$ is strictly negative and finite uniformly on $\Delta(\boldsymbol{\theta}_*)$; (iii) $B_*(\boldsymbol{d},\boldsymbol{d})$ is strictly positive and finite uniformly on $\Delta(\boldsymbol{\theta}_*)$, where for each $\boldsymbol{d},\widetilde{\boldsymbol{d}}$,

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

and "acov" denotes the asymptotic covariance of given arguments; and (iv) For some $q > (r-1)/(\lambda\gamma)$ and $s > q \ge 2$, and for each $f_t \in \bar{\mathbb{L}}$, $\|f_t - E[f_t|\mathcal{F}_{t-\tau}^{t+\tau}]\|_q \le \nu_\tau$, where $\bar{\mathbb{L}} := \{a_1f_1 + a_2f_2 : f_1,f_2 \in \{D\ell_t(\boldsymbol{\theta}_*;\cdot,\boldsymbol{d}) : \boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)\}, a_1,a_2 \in \mathbb{R}\}$; ν_τ is of size $-1/(1-\gamma)$ with $1/2 \le \gamma < 1$; $\mathcal{F}_{t-\tau}^{t+\tau} := \sigma(\mathbf{X}_{t-\tau},\ldots,\mathbf{X}_{t+\tau})$; 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 $\sup_{\boldsymbol{d}\in\Delta(\boldsymbol{\theta}_*)}\sup_{t=1,2,\ldots}\|D\ell_t(\boldsymbol{\theta}_*;\boldsymbol{d})\|_s < \Delta < \infty$.

Some remarks are warranted on Assumption 6. 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*), which 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 $E[\ell_t(\cdot)]$.

https://doi.org/10.1017/S0266466617000354 Published online by Cambridge University Press

Assumption 6(*iii*) is imposed for the same purpose. For notational simplicity, we let $B_*(\boldsymbol{d})$ denote $B_*(\boldsymbol{d},\widetilde{\boldsymbol{d}})$ if $\boldsymbol{d} = \widetilde{\boldsymbol{d}}$ henceforth. 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) provided regularity conditions for the CLT of near-epoch processes as a special case of the mixingale process. Hansen (1996a) generalized this and provided 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, our focus is different from that of Fang and Santos (2014) in which a D-D transform of a consistent estimator is examined. Our interests are in examining the estimator maximizing a D-D quasi-likelihood function.

The limit distribution of $\widehat{\boldsymbol{\theta}}_n$ is obtained using the regularity conditions provided thus 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 the *directional QML estimator* (*DQML estimator*). Next, we examine how this is interrelated with another DQML estimator obtained 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 a 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 $\boldsymbol{d}$, $L_n(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})) \le L_n(\widehat{\boldsymbol{\theta}}_n)$.

We can also represent the DQML estimator $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})$ using the distance between $\boldsymbol{\theta}_*$ and $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})$. With 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 $\mathbb{R}^r$, $H_*(\boldsymbol{d})$ must 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:

$$L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = L_n(\boldsymbol{\theta}_*) + DL_n(\boldsymbol{\theta}_*;\boldsymbol{d})h + \frac{1}{2}D^2L_n(\bar{\boldsymbol{\theta}}_n(\boldsymbol{d});\boldsymbol{d})h^2. \tag{3}$$

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 $\boldsymbol{d} \in \Delta(\boldsymbol{\theta}_*)$,

$$2\{L_n(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \max_{\widetilde{h}\in\mathbb{R}^+} [2\mathcal{Z}(\boldsymbol{d})\widetilde{h} + A_*(\boldsymbol{d})\widetilde{h}^2], \tag{4}$$

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 (4) 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 (3) 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 using different directions. Note that

$$L_n(\widehat{\boldsymbol{\theta}}_n) = \sup_{\boldsymbol{d}\in\Delta(\boldsymbol{\theta}_*)} L_n(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d})). \tag{5}$$

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, $n^{-1}D^2L_n(\boldsymbol{\theta}_*;\cdot)$ obeys the ULLN under the given conditions provided thus far.

If $L_n(\cdot)$ is D, it is trivial to show 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 $\varepsilon > 0$,

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

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,

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

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)* $\{n^{-1/2}DL_n(\boldsymbol{\theta}_*;\cdot), n^{-1}D^2L_n(\boldsymbol{\theta}_*;\cdot)\} \Rightarrow (\mathcal{Z}(\cdot), A_*(\cdot))$, *where for each* $\boldsymbol{d}$ *and* $\boldsymbol{d}'$, $E[\mathcal{Z}(\boldsymbol{d})\mathcal{Z}(\boldsymbol{d}')] = B_*(\boldsymbol{d},\boldsymbol{d}')$; *(ii)* $\sqrt{n}\widehat{h}_n(\cdot) \Rightarrow \max[0,\mathcal{G}(\cdot)]$; *(iii)* $2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \sup_{\boldsymbol{d}\in\Delta(\boldsymbol{\theta}_*)}\max[0,\mathcal{Y}(\boldsymbol{d})]^2$; *and (iv)* $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \Rightarrow \max[0,\mathcal{G}(\boldsymbol{d}_*)]\boldsymbol{d}_*$, *provided that* $\max[0,\mathcal{Y}(\cdot)]^2$ *is uniquely maximized at* $\boldsymbol{d}_*$ *a.s.*$-\mathbb{P}$.

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 almost surely flat on $\Delta(\boldsymbol{\theta}_*)$, 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 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, 2011), and Baek, Cho, and Phillips (2015) examined statistics with this feature: unidentified parameters yield a limit distribution characterized by a Gaussian process, and Theorem 1 can be thought of as a variational result of this.

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

Assumptions 6(*ii* and *iii*)* correspond to the assumption that $\mathbf{A}_* := \lim_{n\to\infty}n^{-1}E[\nabla^2_{\boldsymbol{\theta}}L_n(\boldsymbol{\theta}_*)]$ and $\mathbf{B}_* := \text{acov}\{n^{-1/2}\nabla_{\boldsymbol{\theta}}L_n(\boldsymbol{\theta}_*)\}$ 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 $6^\star$ denote Assumptions 6(*i*, *ii**, *iii**, and *iv*) henceforth.

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) $2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \sup_{\boldsymbol{d}\in\Delta(\boldsymbol{\theta}_*)}\max[0,\mathbf{Z}'\boldsymbol{d}]^2\{\boldsymbol{d}'(-\mathbf{A}_*)\boldsymbol{d}\}^{-1}$; *and*
(vii) $2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \mathbf{Z}'(-\mathbf{A}_*)^{-1}\mathbf{Z}$, *provided that* $\boldsymbol{\theta}_*$ *is interior to* $\boldsymbol{\Theta}$.

Corollary 1 is the same consequence as for the standard case if $\boldsymbol{\theta}_*$ is interior to $\boldsymbol{\Theta}$. Our analysis is more primitive because it involves directional derivatives. 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) exploited the D-D quasi-likelihood function analysis for D quasi-likelihood function estimation and examined other aspects that are not contained in Corollary 1.

3. TESTING HYPOTHESES USING D-D QUASI-LIKELIHOOD FUNCTIONS

This section examines data inference using D-D models. To this end, the standard QLR, Wald, and LM test statistics are reviewed and redefined in case there is a need to accommodate 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 nonlinear 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 $\boldsymbol{d},\widetilde{\boldsymbol{d}} \in \Delta(\boldsymbol{\theta}_*)$,

$$\mathbf{B}_*(\boldsymbol{d},\widetilde{\boldsymbol{d}}) := \begin{bmatrix} \mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\pi}},\widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}}) & \mathbf{B}_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\pi}},\widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \\ \mathbf{B}_*^{(\boldsymbol{\tau},\boldsymbol{\pi})}(\boldsymbol{d}_{\boldsymbol{\tau}},\widetilde{\boldsymbol{d}}_{\boldsymbol{\pi}})' & B_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}},\widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \end{bmatrix}$$
$$:= \begin{bmatrix} \boldsymbol{d}_{\boldsymbol{\lambda}}{}'\mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\widetilde{\boldsymbol{d}}_{\boldsymbol{\lambda}} & \boldsymbol{d}_{\boldsymbol{\lambda}}{}'\mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\widetilde{\boldsymbol{d}}_{\boldsymbol{\upsilon}}) & \boldsymbol{d}_{\boldsymbol{\lambda}}{}'\mathbf{B}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \\ \mathbf{B}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{d}_{\boldsymbol{\upsilon}})'\widetilde{\boldsymbol{d}}_{\boldsymbol{\lambda}} & B_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\upsilon}},\widetilde{\boldsymbol{d}}_{\boldsymbol{\upsilon}}) & B_*^{(\boldsymbol{\upsilon},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\upsilon}},\widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \\ \mathbf{B}_*^{(\boldsymbol{\tau},\boldsymbol{\lambda})}(\boldsymbol{d}_{\boldsymbol{\tau}})'\widetilde{\boldsymbol{d}}_{\boldsymbol{\lambda}} & B_*^{(\boldsymbol{\tau},\boldsymbol{\upsilon})}(\boldsymbol{d}_{\boldsymbol{\tau}},\widetilde{\boldsymbol{d}}_{\boldsymbol{\upsilon}}) & B_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{d}_{\boldsymbol{\tau}},\widetilde{\boldsymbol{d}}_{\boldsymbol{\tau}}) \end{bmatrix}, \quad \textbf{(6)}$$

https://doi.org/10.1017/S0266466617000354 Published online by Cambridge University Press

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

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

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

$$\mathbf{A}_*(\boldsymbol{d}) := \begin{bmatrix} \mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\pi})}(\boldsymbol{d_\pi}) & \mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d_\pi},\boldsymbol{d_\tau}) \\ \mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\pi})}(\boldsymbol{d_\tau},\boldsymbol{d_\pi})' & A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \end{bmatrix}$$
$$:= \begin{bmatrix} \boldsymbol{d_\lambda}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\boldsymbol{d_\lambda} & \boldsymbol{d_\lambda}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon}) & \boldsymbol{d_\lambda}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \\ \mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{d_\upsilon})'\boldsymbol{d_\lambda} & A_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon}) & A_*^{(\boldsymbol{\upsilon},\boldsymbol{\tau})}(\boldsymbol{d_\upsilon},\boldsymbol{d_\tau}) \\ \mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\lambda})}(\boldsymbol{d_\tau})'\boldsymbol{d_\lambda} & A_*^{(\boldsymbol{\tau},\boldsymbol{\upsilon})}(\boldsymbol{d_\tau},\boldsymbol{d_\upsilon}) & A_*^{(\boldsymbol{\tau},\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \end{bmatrix}, \quad \textbf{(7)}$$

$\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})} \in \mathbb{R}^{r_\lambda} \times \mathbb{R}^{r_\lambda}$, $\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon}) \in \mathbb{R}^{r_\lambda}$, $\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d_\tau}) \in \mathbb{R}^{r_\lambda}$, $\mathbf{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\lambda})}(\boldsymbol{d_\upsilon}) = \mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\upsilon})}(\boldsymbol{d_\upsilon})$, and $\mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\lambda})}(\boldsymbol{d_\tau}) = \mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\tau})}(\boldsymbol{d_\tau})$. We also let $\boldsymbol{\pi}$ be orthogonal to $\boldsymbol{\tau}$: for each $\boldsymbol{d}$, $\mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\pi})}(\boldsymbol{d_\tau},\boldsymbol{d_\pi}) = \mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d_\pi},\boldsymbol{d_\tau}) = \mathbf{0}$. This assumption is useful in eliminating the nuisance parameters from our analysis that are asymptotically irrelevant to testing the hypothesis given below. We also permit r_υ, r_λ, and r_τ to 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 $\boldsymbol{d}$, $DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}) = DL_n^{(\boldsymbol{\pi})}(\boldsymbol{d_\pi}) + DL_n^{(\boldsymbol{\tau})}(\boldsymbol{d_\tau})$, and $n^{-1/2}(DL_n^{(\boldsymbol{\pi})}(\cdot), DL_n^{(\boldsymbol{\tau})}(\cdot)) \Rightarrow (\boldsymbol{\mathcal{Z}}^{(\boldsymbol{\pi})}(\cdot), \mathcal{Z}^{(\boldsymbol{\tau})}(\cdot))$; (ii) for each $\boldsymbol{d}$ and $\widetilde{\boldsymbol{d}}$, $B_*(\boldsymbol{d},\widetilde{\boldsymbol{d}}) = \boldsymbol{\iota}_3{}'\mathbf{B}_*(\boldsymbol{d},\widetilde{\boldsymbol{d}})\boldsymbol{\iota}_3$, 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) $\mathbf{A}_*^{(\boldsymbol{\tau},\boldsymbol{\pi})}(\boldsymbol{d_\tau},\boldsymbol{d_\pi}) = \mathbf{A}_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d_\pi},\boldsymbol{d_\tau}) = \mathbf{0}$ 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 $C(\boldsymbol{\upsilon}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\upsilon} : \exists \boldsymbol{\upsilon}' \in \boldsymbol{\Upsilon}, \mathbf{x} := \boldsymbol{\upsilon}_* + \delta\boldsymbol{\upsilon}', \delta \in \mathbb{R}^+\}$; and (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, we further let $\boldsymbol{\upsilon}$ be the parameter of interest, and the hypotheses of interest are given as

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

For 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. Quasi-Likelihood Ratio Test Statistic

The standard QLR test statistic defined for D quasi-likelihood functions can be used for D-D quasi-likelihood functions without modification. 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 $\boldsymbol{\theta}_*$. Although Theorem 1(*iii*) already provides the limit distribution of $\mathcal{LR}_n^{(1)}$, we reexamine this here 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 $\boldsymbol{s}_{\boldsymbol{\upsilon}} \in \Delta(\boldsymbol{\upsilon}_*) := \{\mathbf{x} \in \mathbb{R}^{r_\upsilon} : \boldsymbol{\upsilon}_* + \mathbf{x} \in \mathrm{cl}\{C(\boldsymbol{\upsilon}_*)\}, \|\mathbf{x}\| = 1\}$,

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

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 regarding 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

$$\mathcal{LR}_n^{(2)} \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\lambda}} \in \Delta(\boldsymbol{\lambda}_*)} \left\{ \frac{\max[0, \boldsymbol{s}_{\boldsymbol{\lambda}}'\mathbf{Z}^{(\boldsymbol{\lambda})}]^2}{\boldsymbol{s}_{\boldsymbol{\lambda}}'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})\boldsymbol{s}_{\boldsymbol{\lambda}}} \right\} + \mathcal{H}_2 = \mathcal{H}_1 + \mathcal{H}_2. \quad \textbf{(8)}$$

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) provided 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.

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 $\boldsymbol{\tau}$, 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:

$$\sup_{\boldsymbol{d}} \sup_{h} L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = \sup_{\{\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}, \boldsymbol{s}_{\boldsymbol{\tau}}\}} \sup_{\{h^{(\boldsymbol{\upsilon})}, h^{(\boldsymbol{\lambda})}, h^{(\boldsymbol{\tau})}\}} L_n(\boldsymbol{\theta}_* + (h^{(\boldsymbol{\upsilon})}\boldsymbol{s}_{\boldsymbol{\upsilon}}{}', h^{(\boldsymbol{\lambda})}\boldsymbol{s}_{\boldsymbol{\lambda}}{}', h^{(\boldsymbol{\tau})}\boldsymbol{s}_{\boldsymbol{\tau}}{}')'), \quad \textbf{(9)}$$

and we can apply Wald's (1943) testing principle to $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\cdot)$.

https://doi.org/10.1017/S0266466617000354 Published online by Cambridge University Press

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 $(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) \in \Delta(\boldsymbol{\upsilon}_*) \times \Delta(\boldsymbol{\lambda}_*) \times \Delta(\boldsymbol{\tau}_*)$, we let

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

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

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

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 $A_*^{(\boldsymbol{\pi},\boldsymbol{\tau})}(\boldsymbol{d_\pi},\boldsymbol{d_\tau}) = 0$ renders the maximization process in the right side of (9) 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 $(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau})$, one of the following four different events asymptotically arises: (*i*) $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) > 0$ and $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) > 0$; (*ii*) $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) > 0$, $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) = 0$; (*iii*) $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) = 0$, $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) > 0$; or (*iv*) $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) = 0$, $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) = 0$. 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 $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) = 0$ and $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) > 0$, estimating the nuisance parameter $\boldsymbol{\lambda}_*$ does not affect the limit distribution of $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau})$ because $\widehat{\boldsymbol{\lambda}}_n(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) = \boldsymbol{\lambda}_*$ from the fact that $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) = 0$, so that it does not have to be associated with the parameter estimation error of $\widehat{\boldsymbol{\lambda}}_n(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau})$, implying that $\sqrt{n}\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}) \Rightarrow \dot{\mathcal{G}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})$. For the cases of (*iii*) and (*iv*), 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 must be taken into account in obtaining the limit distribution of $\widehat{h}_n^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau})$ and $\widehat{h}_n^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau})$.

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 nonrandom 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 $\sup_{\{s_{\lambda},s_{\tau}\}}\sup_{\{h^{(\upsilon)},h^{(\lambda)},h^{(\tau)}\}}L_n(\upsilon_0+h^{(\upsilon)}s_{\upsilon},\lambda_*+h^{(\lambda)}s_{\lambda},\tau_*+h^{(\tau)}s_{\tau})$. 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)**.** For a function $\widehat{W}_n(\cdot)$ that is strictly positive uniformly on $\Delta(\upsilon_0)$ and for every n a.s.–$\mathbb{P}$, $\sup_{s_{\upsilon}\in\Delta(\upsilon_0)}|\widehat{W}_n(s_{\upsilon})-\widetilde{A}_*^{(\upsilon,\upsilon)}(s_{\upsilon})|\to 0$ a.s.–$\mathbb{P}$.

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

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

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

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 s_{υ}, maximizing $2\{L_n(\upsilon_0+h^{(\upsilon)}s_{\upsilon},\lambda,\tau)-L_n(\upsilon_0,\lambda_*,\tau_*)\}$ with respect to $h^{(\upsilon)}$, λ, and τ is equivalent to

$$\sup_{\{s_{\lambda},s_{\tau}\}}\sup_{\{h^{(\upsilon)},h^{(\lambda)},h^{(\tau)}\}}2\{L_n(\upsilon_0+h^{(\upsilon)}s_{\upsilon},\lambda_*+h^{(\lambda)}s_{\lambda},\tau_*+h^{(\tau)}s_{\tau})-L_n(\upsilon_0,\lambda_*,\tau_*)\}. \quad \textbf{(10)}$$

For each $(s_{\upsilon},s_{\lambda},s_{\tau})$, (10) is approximated by a quadratic function of $(h^{(\upsilon)},h^{(\lambda)},h^{(\tau)})$, and the signs of $\widehat{h}_n^{(\upsilon)}(s_{\upsilon},s_{\lambda},s_{\tau})$ and $\widehat{h}_n^{(\lambda)}(s_{\upsilon},s_{\lambda},s_{\tau})$ result in different approximations as discussed earlier. Using Assumption 7(*vii*), we show that the optimization process in (10) results in the consequence of Theorem 3.

3.3. Lagrange Multiplier Test Statistic

The standard LM test statistic needs to be redefined for D-D quasi-likelihood functions when testing whether the slope of a quasi-likelihood function is asymptotically distributed around zero under the null. We let the LM test statistic be defined as

$$\mathcal{LM}_n:=\sup_{(s_{\upsilon},s_{\lambda})\in\Delta(\upsilon_0)\times\Delta(\ddot{\lambda}_n)} n\max\left[0,\frac{-DL_n(\ddot{\theta}_n;s_{\upsilon})}{\widetilde{D}^2L_n(\ddot{\theta}_n;s_{\upsilon};s_{\lambda})}\right]^2\widetilde{W}_n(s_{\upsilon},s_{\lambda}),$$

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

$$\widetilde{D}^2L_n(\ddot{\theta}_n;s_{\upsilon};s_{\lambda}):=D^2L_n(\ddot{\theta}_n;s_{\upsilon})-DL_n(\ddot{\theta}_n;s_{\upsilon};s_{\lambda})(D^2L_n(\ddot{\theta}_n;s_{\lambda}))^{-1}DL_n(\ddot{\theta}_n;s_{\lambda};s_{\upsilon}),$$
$$DL_n(\ddot{\theta}_n;s_{\upsilon};s_{\lambda}):=\lim_{h\downarrow 0}h^{-1}\{DL_n(\upsilon_0,\ddot{\lambda}_n+hs_{\lambda},\ddot{\tau}_n;s_{\upsilon})-DL_n(\ddot{\theta}_n;s_{\upsilon})\},$$
$$DL_n(\ddot{\theta}_n;s_{\lambda};s_{\upsilon}):=\lim_{h\downarrow 0}h^{-1}\{DL_n(\upsilon_0+hs_{\upsilon},\ddot{\lambda}_n,\ddot{\tau}_n;s_{\lambda})-DL_n(\ddot{\theta}_n;s_{\lambda})\},$$

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

https://doi.org/10.1017/S0266466617000354 Published online by Cambridge University Press

Assumption 9 (Weight function II)**.** For a function $\widetilde{W}_n(\cdot)$ that is strictly positive uniformly on $\Delta(\boldsymbol{\upsilon}_0) \times \Delta(\ddot{\boldsymbol{\lambda}}_n)$ and for every n a.s.–$\mathbb{P}$, $\sup_{(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) \in \Delta(\boldsymbol{\upsilon}_0) \times \Delta(\ddot{\boldsymbol{\lambda}}_n)}$ $|\widetilde{W}_n(\boldsymbol{s}_{\boldsymbol{\upsilon}}, \boldsymbol{s}_{\boldsymbol{\lambda}}) - \widetilde{A}_*^{(\boldsymbol{\upsilon},\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})| \to 0$ 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 the "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 $-n^{-1}\widetilde{D}^2 L_n(\ddot{\boldsymbol{\theta}}_n; \cdot)$. 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, $\mathcal{LM}_n \Rightarrow \sup_{\boldsymbol{s}_{\boldsymbol{\upsilon}} \in \Delta(\boldsymbol{\upsilon}_0)} \max[0, \widetilde{\mathcal{Y}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s}_{\boldsymbol{\upsilon}})]^2$, *provided that* $\max[0, \widetilde{\mathcal{Y}}(\cdot)]^2$ *is uniquely maximized a.s.–*$\mathbb{P}$.

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

3.4. Example: Conditional Heteroskedasticity (Continued)

We continue examining King and Shively's (1993) conditional heteroskedasticity model in this subsection.

For a proper analysis, we further elaborate on the model assumption. 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 separated from zero in terms of the original model. Consequently, our parameter space for $\boldsymbol{\theta}$ is refined into

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

https://doi.org/10.1017/S0266466617000354 Published online by Cambridge University Press

By this modification, d_2/d_1 is constrained to $[\underline{c}, \bar{c}]$, and we can avoid the multifold identification problem of Cho and Ishida (2012), Cho, Ishida, and White (2011, 2014), White and Cho (2012), Baek et al. (2015), and Cho and Phillips (2016).

The first-order directional derivative in (1) can be partitioned into three pieces: $DL_n(\gamma_*, \sigma_*^2, \boldsymbol{\theta}_*; \boldsymbol{d}) = Z_{1,n}(\boldsymbol{d}) + Z_{2,n}(\boldsymbol{d}) + Z_{3,n}(\boldsymbol{d})$, 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 $n^{-1/2} DL_n(\gamma_*, \sigma_*^2, \boldsymbol{\theta}_*; \cdot)$. 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 one 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 $|m(\cdot)|$ is uniformly and strictly bounded by one 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: $n^{-1/2} DL_n(\gamma_*, \sigma_*^2, \boldsymbol{\theta}_*; \cdot) \Rightarrow \mathcal{Z}(\cdot)$, 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 $(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$, $B_*^{(1)}(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) := \frac{1}{\sigma_*^2}\boldsymbol{d}_\gamma{}' E[\mathbf{Q}_t\mathbf{Q}_t{}']\widetilde{\boldsymbol{d}}_\gamma$,

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

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

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 (2), we obtain

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

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 such 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 $n^{-1}D^2L_n(\gamma_*,\sigma_*^2,\ \boldsymbol{\theta}_*;\boldsymbol{d}) = -B_*(\boldsymbol{d},\ \boldsymbol{d}) + o_{\mathbb{P}}(1)$. The ULLN further strengthens this to $\sup_{\boldsymbol{d}}|n^{-1}D^2L_n(\gamma_*,\sigma_*^2,\boldsymbol{\theta}_*;\boldsymbol{d}) + B_*(\boldsymbol{d},\boldsymbol{d})| = o_{\mathbb{P}}(1)$, which also leads to the information matrix equality. This follows mainly because $D^2L_n\ (\gamma_*,\sigma_*^2,\boldsymbol{\theta}_*;\cdot)$ is differentiable on $\Delta(\boldsymbol{\theta}_*)$, so that Assumption 5(*iii*) holds with respect to the second-order directional derivatives. Therefore, $2\{L_n(\widehat{\gamma}_n,\widehat{\sigma}_n^2,\widehat{\boldsymbol{\theta}}_n) - L_n(\gamma_*,\sigma_*^2,\boldsymbol{\theta}_*)\} \Rightarrow \sup_{\boldsymbol{d}}[0,\mathcal{Y}(\boldsymbol{d})]^2$ 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 subsections. Specifically, we let $\boldsymbol{\upsilon} = (\theta_1,\theta_2)'$, $\boldsymbol{\lambda} = \sigma^2$, $\boldsymbol{\tau} = \gamma$, and $\boldsymbol{\pi} = (\sigma^2,\theta_1,\theta_2)'$. Then, for each $\boldsymbol{d}$ and $\widetilde{\boldsymbol{d}}$,

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

and

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

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, $\sigma_*^{-2}\boldsymbol{d}_\gamma{}'E[\mathbf{Q}_t\mathbf{Q}_t{}']\widetilde{\boldsymbol{d}}_\gamma$ is associated only with γ, 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 $\mathbf{B}_*^{(\pi,\pi)}(\boldsymbol{d}_\pi,\widetilde{\boldsymbol{d}}_\pi)$: first, let

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

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

and $q := q(d_1,d_2,\widetilde{d}_1,\widetilde{d}_2) := h\widetilde{h}k$, and note that $\boldsymbol{\iota}_3{}'\widetilde{\mathbf{B}}_*^{(\pi,\pi)}(\boldsymbol{d}_\pi,\widetilde{\boldsymbol{d}}_\pi)\boldsymbol{\iota}_3 = B_*^{(2)}(\boldsymbol{d},\widetilde{\boldsymbol{d}}) + B_*^{(3)}(\boldsymbol{d},\widetilde{\boldsymbol{d}})$ and $\boldsymbol{\iota}_2{}'\ddot{\mathbf{B}}_*^{(\pi,\pi)}(\boldsymbol{d}_\pi,\widetilde{\boldsymbol{d}}_\pi)\boldsymbol{\iota}_2 = B_*^{(2)}(\boldsymbol{d},\widetilde{\boldsymbol{d}})$. Here, the Gaussian stochastic process associated with $\ddot{\mathbf{B}}_*^{(\pi,\pi)}(\boldsymbol{d}_\pi,\widetilde{\boldsymbol{d}}_\pi)$ is independent of that associated with $qE[W_t^2]^2$ because $\widetilde{\mathbf{B}}_*^{(\pi,\pi)}(\boldsymbol{d}_\pi,\widetilde{\boldsymbol{d}}_\pi)$ is block diagonal. Furthermore, $\ddot{\mathbf{B}}_*^{(\pi,\pi)}(\boldsymbol{d}_\pi,\widetilde{\boldsymbol{d}}_\pi)$ 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 is accordingly defined as

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

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

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

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

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

where

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

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

$(\ddot{\gamma}_n, \ddot{\sigma}_n^2)$ is such that $L_n(\ddot{\gamma}_n, \ddot{\sigma}_n^2, \mathbf{0}) = \sup_{(\gamma,\sigma^2)} L_n(\gamma, \sigma^2, \mathbf{0})$, and the same weight matrix is used as for the Wald test statistic. Here, $\widetilde{D}^2 L_n(\ddot{\gamma}_n, \ddot{\sigma}_n^2, \mathbf{0}; s_1, s_2, s_{\sigma^2})$ 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 simulation. 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, simulating $\ddot{\mathcal{Y}}^{(\boldsymbol{\theta})}(\cdot)$ is not hard. When simulating $\ddot{\mathcal{Y}}^{(\boldsymbol{\theta})}(\cdot)$, $\mathrm{var}(W_t^2)$ and $E[W_t^2]$ must 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 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),

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

The empirical distribution of the QLR test statistic approaches the limit distribution as n increases, affirming our theory on the test statistics.

The null limit distribution of the QLR test can be uncovered by several simulation methods. The Monte Carlo method proposed by Dufour (2006) can be used because the model is correctly specified for this model. Hansen's (1996b) weighted bootstrap can also be used to estimate the asymptotic p-values.

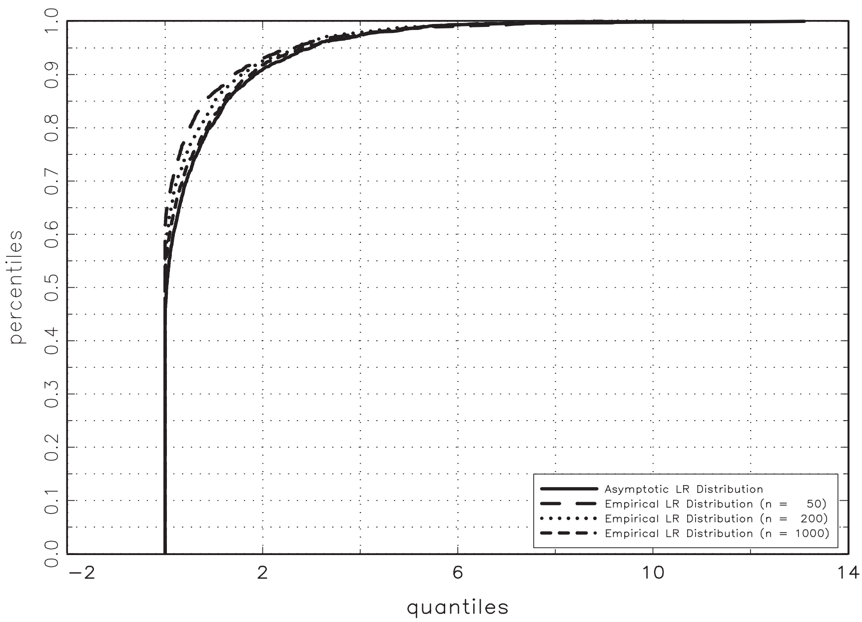

FIGURE 1. Empirical and asymptotic distributions of the QLR test statistic. This figure shows the null limit distribution of the QLR test statistic and the empirical distributions of the QLR test statistic for $n = 50$, 200, and 1,000. The number of iterations is 2,000.

4. 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 a 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 reveal that the three test statistics are asymptotically equivalent under the null and some mild regularity conditions that can be popularly used for empirical examinations.

REFERENCES

Aigner, D., C. Lovell, & P. Schmidt (1977) Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6, 21–37.

Andrews, D. (1999) Estimation when a parameter is on a boundary. *Econometrica* 67, 543–563.

https://doi.org/10.1017/S0266466617000354 Published online by Cambridge University Press

Andrews, D. (2001) Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 69, 683–734.

Baek, Y., J.S. Cho, & P.C.B. Phillips (2015) Testing linearity using power transforms of regressors. *Journal of Econometrics* 187, 376–384.

Billingsley, P. (1999) *Convergence of Probability Measures*. Wiley.

Box, G. & D. Cox (1964) An analysis of transformations. *Journal of the Royal Statistical Society, Series B* 26, 211–252.

Chernoff, H. (1954) On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics* 54, 573–578.

Cho, J.S. (2011) Quasi-maximum likelihood estimation revisited using the distance and direction method. *Journal of Economic Theory and Econometrics* 23, 89–112.

Cho, J.S. & I. Ishida (2012) Testing for the effects of omitted power transformations. *Economics Letters* 117, 287–290.

Cho, J.S., I. Ishida, & H. White (2011) Revisiting tests for neglected nonlinearity using artificial neural networks. *Neural Computation* 23, 1133–1186.

Cho, J.S., I. Ishida, & H. White (2014) Testing for neglected nonlinearity using twofold unidentified models under the null and hexic expansions. In N. Haldrup, M. Meitz, & P. Saikkonen, eds., *Essays in Nonlinear Time Series Econometrics*, pp. 3–27. Oxford University Press.

Cho, J.S. & P.C.B. Phillips (2017) Sequentially Testing Polynomial Model Hypothesis Using the Power Transform of Regressors. *Journal of Applied Econometrics*, forthcoming.

Cho, J.S. & H. White (2007) Testing for regime switching. *Econometrica* 75, 1671–1720.

Cho, J.S. & H. White (2010) Testing for unobserved heterogeneity in exponential and weibull duration models. *Journal of Econometrics* 157, 458–480.

Cho, J.S. & H. White (2011) Generalized runs tests for the IID hypothesis. *Journal of Econometrics* 162, 326–344.

Cho, J.S. & H. White (2017) Supplements to "Directionally Differentiable Econometric Models". School of Economics, Yonsei University. Available at: `http://web.yonsei.ac.kr/jinseocho/pardiff.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, & E. Rio (1995) Invariance principles for absolutely regular empirical processes. *Annales de l'Institut Henri Poincaré, Probabilites et Statistiques* 31, 393–427.

Duofur, J.-M. (2006) Monte carlo tests with nuisance parameters: A general approach to fininte-sample inference and nonstandard asymptotics in ecoonometrics. *Journal of Econometrics* 133, 443–477.

Fang, Z. & A. Santos (2014) Inference on Directionally Differentiable Functions, ArXiv preprint arXiv: 1404.3763.

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. & 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. & D. Pollard (1990) Cube root asymptotics. *Annals of Statistics* 18, 191–219.

Liu, X. & Y. Shao (2003) Asymptotics for likelihood ratio tests under loss of identifiability. *Annals of Statistics* 31, 807–832.

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*. McGraw-Hill.
Stevenson, R. (1980) Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics* 13, 57–66.
Stout, W. (1974) *Almost Sure Convergence*. Academic Press.
Troutman, J. (1996) *Variational Calculus and Optimal Control*. Springer-Verlag.
van der Vaart, A. & J. Weller (1996) *Weak Convergence and Empirical Processes with Applications to Statistics*. 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. & J.S. Cho (2012) Higher-order approximations for testing neglected nonlinearity. *Neural Computation* 24, 273–287.
Wooldridge, J. & H. White (1988) Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* 4, 210–230.

APPENDIX: Proofs

Before proving the main claims of the article, 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)* $n^{-1}D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{d})$ *converges to* $A_*(\boldsymbol{d})$ *a.s.–*$\mathbb{P}$; *and (iii)* $\left\{n^{-1/2}DL_n(\boldsymbol{\theta}_*;\boldsymbol{d}), n^{-1}D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{d})\right\} \Rightarrow \{\mathcal{Z}(\boldsymbol{d}), A_*(\boldsymbol{d})\}$.

LEMMA 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)* $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{d}) - \boldsymbol{\theta}_*) \Rightarrow \max[0, \mathcal{G}(\boldsymbol{d})]\boldsymbol{d}$; *and (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 and (ii) for all* $\varepsilon > 0$, *there is* $n(\varepsilon)$ *a.s.*$-\mathbb{P}$ *such that if* $n > n(\varepsilon)$, $\sup_{\boldsymbol{d}\in\Delta(\boldsymbol{\theta}_*)}|n^{-1}D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{d}) - A_*(\boldsymbol{d})| < \varepsilon$.

Lemma A3(*i*) implies that the first-order directional derivative weakly converges to a Gaussian stochastic process indexed by $\boldsymbol{d}$ (e.g., Billingsley, 1999). In our time-series data context, theorem 1 of Hansen (1996b) provides sufficient regularity conditions for this. Lemma A3(*i*) is used to show the desired weak convergence of the QML estimator with $r > 1$, and 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, Assumption C.1 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 Assumption C.2 in Wooldridge and White (1988) trivially hold

https://doi.org/10.1017/S0266466617000354 Published online by Cambridge University Press

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 Assumption C.2 is easily verified from the fact that $\|\ell_t(\boldsymbol{\theta}_*;\boldsymbol{d})\|_s < \Delta < \infty$ uniformly in d for any t. Finally, their condition in Assumption A.5 is not necessary for Lemma A1 because our goal is not to obtain the standard normal distribution. Their corollary 3.1 implies the desired result.

(*ii*) Given Assumptions 1 and 6(*ii*), we can apply the ergodic theorem.

(*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 (3) with respect to h subject to $h \geq 0$, it follows that $\widehat{h}_n(\boldsymbol{d}) = \max[0, \{-D^2L_n(\bar{\boldsymbol{\theta}}_n(\boldsymbol{d});\boldsymbol{d})\}^{-1}DL_n(\boldsymbol{\theta}_*;\boldsymbol{d})]$ 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 (4). ■

Proof of Lemma A3. (*i*) Given 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 $\varepsilon > 0$, $\alpha_\tau^{-(s-q)/(sq)} = O(\tau^{-1-\varepsilon})$, 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 $\varepsilon > 0$, $\nu_\tau = O(\tau^{-1/(1-\gamma)-\varepsilon})$, 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 desired 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 (5).

(*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 (4) 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 in theorem 3.2.2 of van der Vaart and Wellner (1996) are satisfied by this, leading to the desired result. ■

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

(*i*) As the weak convergence is proved for a general function in Theorem 1, 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 $D^2 L_n(\boldsymbol{\theta};\boldsymbol{d}) = \boldsymbol{d}'\nabla^2_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*)\boldsymbol{d}$, so that $n^{-1}\nabla^2_{\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_*) \to \mathbf{A}_*$ a.s.–$\mathbb{P}$ by the ergodic theorem. Therefore, 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 for 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 and 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

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

by the definition of eigenvector. This implies that

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

because $\boldsymbol{v}_*{}'\boldsymbol{v}_* = 1$. Plugging the definition of $\boldsymbol{v}_*$ to the left side of (A.2) 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 (A.1) 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 $h^{(\boldsymbol{\pi})} \in \mathbb{R}^+$, $h^{(\boldsymbol{\tau})} \in \mathbb{R}^+$, $\boldsymbol{s_\pi} \in \Delta(\boldsymbol{\pi}_*)$, and $\boldsymbol{s_\tau} \in \Delta(\boldsymbol{\tau}_*)$ such that $h\boldsymbol{d} = [h^{(\boldsymbol{\pi})}\boldsymbol{s_\pi}{}', h^{(\boldsymbol{\tau})}\boldsymbol{s_\tau}{}']'$ by Assumption 7. Thus, $L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = L_n(\boldsymbol{\pi}_* + h^{(\boldsymbol{\pi})}\boldsymbol{s_\pi}, \boldsymbol{\tau}_* + h^{(\boldsymbol{\tau})}\boldsymbol{s_\tau})$, implying that

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

where $DL_n(\boldsymbol{\pi}_*, \boldsymbol{\tau}_*; \boldsymbol{s_\pi}; \boldsymbol{s_\tau})$ is the directional derivative of $DL_n(\cdot, \cdot; \boldsymbol{s_\pi})$ with respect to $\boldsymbol{s_\tau}$ evaluated at $(\boldsymbol{\pi}_*, \boldsymbol{\tau}_*)$, and $\sup_{\boldsymbol{d}}\sup_h L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) = \sup_{\{\boldsymbol{s_\pi}, \boldsymbol{s_\tau}\}}\sup_{\{h^{(\boldsymbol{\pi})}, h^{(\boldsymbol{\tau})}\}} L_n(\boldsymbol{\pi}_* + h^{(\boldsymbol{\pi})}\boldsymbol{s_\pi}, \boldsymbol{\tau}_* + h^{(\boldsymbol{\tau})}\boldsymbol{s_\tau})$. Therefore, $\mathcal{LR}_n^{(1)} = \sup_{\boldsymbol{d}}\sup_h 2\{L_n(\boldsymbol{\theta}_* + h\boldsymbol{d}) - L_n(\boldsymbol{\theta}_*)\}$ implies that

$$\begin{aligned}
\mathcal{LR}_n^{(1)} = &\sup_{\boldsymbol{s_\pi}} \sup_{h^{(\boldsymbol{\pi})}} \{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\pi})h^{(\boldsymbol{\pi})} + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s_\pi})(h^{(\boldsymbol{\pi})})^2 + o_{\mathbb{P}_{s_\pi}}(1)\} \\
&+ \sup_{\boldsymbol{s_\tau}} \sup_{h^{(\boldsymbol{\tau})}} \{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\tau})h^{(\boldsymbol{\tau})} + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s_\tau})(h^{(\boldsymbol{\tau})})^2 + o_{\mathbb{P}_{s_\tau}}(1)\},
\end{aligned} \tag{A.3}$$

where we exploited the facts that $n^{-1}DL_n(\boldsymbol{\pi}_*, \boldsymbol{\tau}_*; \boldsymbol{s_\pi}, \boldsymbol{s_\tau})$ has probability limit zero 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 $H_{2,n} := \sup_{\boldsymbol{s_\tau}} \sup_{h^{(\boldsymbol{\tau})}}\{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\tau})h^{(\boldsymbol{\tau})} + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s_\tau})(h^{(\boldsymbol{\tau})})^2\} \Rightarrow \mathcal{H}_2$. Thus, we may focus on the weak limit of $\sup_{\boldsymbol{s_\pi}} \sup_{h^{(\boldsymbol{\pi})}}\{2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\pi})h^{(\boldsymbol{\pi})} + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s_\pi})(h^{(\boldsymbol{\pi})})^2\}$ which is denoted as $H_{01,n}$. From the fact that for any $h\boldsymbol{d_\pi}$ such that $h \in \mathbb{R}^+$ and $\boldsymbol{d_\pi} \in \Delta(\boldsymbol{\pi}_*)$, there are $h^{(\boldsymbol{\lambda})} \in \mathbb{R}^+$, $h^{(\boldsymbol{\upsilon})} \in \mathbb{R}^+$, $\boldsymbol{s_\lambda} \in \Delta(\boldsymbol{\lambda}_*)$, and $\boldsymbol{s_\pi} \in \Delta(\boldsymbol{\pi}_*)$ such that $h\boldsymbol{d_\pi} = [h^{(\boldsymbol{\lambda})}\boldsymbol{s_\lambda}{}', h^{(\boldsymbol{\upsilon})}\boldsymbol{s_\upsilon}{}']'$ and

$$\begin{aligned}
H_{01,n} = &\sup_{\{\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}\}} \sup_{\{h^{(\boldsymbol{\upsilon})}, h^{(\boldsymbol{\lambda})}\}} 2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\upsilon})h^{(\boldsymbol{\upsilon})} + 2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\lambda})h^{(\boldsymbol{\lambda})} + 2DL_n(\boldsymbol{\theta}_*; \boldsymbol{s_\lambda}; \boldsymbol{s_\upsilon})h^{(\boldsymbol{\lambda})}h^{(\boldsymbol{\upsilon})} \\
&+ D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s_\upsilon})(h^{(\boldsymbol{\upsilon})})^2 + D^2L_n(\boldsymbol{\theta}_*; \boldsymbol{s_\lambda})(h^{(\boldsymbol{\lambda})})^2,
\end{aligned}$$

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

$$\begin{aligned}
H_{01,n} \Rightarrow \mathcal{H}_0 + \mathcal{H}_1 = &\sup_{\{\boldsymbol{s_\upsilon}, \boldsymbol{s_\lambda}\}} \sup_{\{h^{(\boldsymbol{\upsilon})}, h^{(\boldsymbol{\lambda})}\}} 2\mathcal{Z}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})h^{(\boldsymbol{\upsilon})} + 2\boldsymbol{s_\lambda}{}'\mathbf{Z}^{(\boldsymbol{\lambda})}h^{(\boldsymbol{\lambda})} + 2\boldsymbol{s_\lambda}{}'\mathbf{A}_*^{(\boldsymbol{\lambda}, \boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})h^{(\boldsymbol{\lambda})}h^{(\boldsymbol{\upsilon})} \\
&+ A_*^{(\boldsymbol{\upsilon}, \boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})(h^{(\boldsymbol{\upsilon})})^2 + \boldsymbol{s_\lambda}{}'\mathbf{A}_*^{(\boldsymbol{\lambda}, \boldsymbol{\lambda})}\boldsymbol{s_\lambda}(h^{(\boldsymbol{\lambda})})^2
\end{aligned} \tag{A.4}$$

by Theorem 1, and there are four different possible cases for the solutions of $(h^{(\boldsymbol{\upsilon})}, h^{(\boldsymbol{\lambda})})$ in the right side of (A.4): for each $(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})$ if we let $\widehat{h}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})$ maximize the right side of (A.4), it follows either (*i*) $\widehat{h}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) > 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) > 0$; (*ii*) $\widehat{h}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) > 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) = 0$; (*iii*) $\widehat{h}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) = 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) > 0$; or (*iv*) $\widehat{h}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) = 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) = 0$.

We examine the limit distribution of each case. First, if $\widehat{h}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) > 0$ and $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) > 0$, the right side of (A.4) is identical to

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

and maximizing this with respect to $\boldsymbol{s_\lambda}$ for a given $\boldsymbol{s_\upsilon}$ yields $\widetilde{\mathcal{Y}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})^2 + (\mathbf{Z}^{(\boldsymbol{\lambda})})'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})^{-1}(\mathbf{Z}^{(\boldsymbol{\lambda})})$. Second, if $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) > 0$ and $\widehat{h}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) = 0$, the right side of (A.4) is identical to $2\boldsymbol{s_\lambda}'\mathbf{Z}^{(\boldsymbol{\lambda})} + \boldsymbol{s_\lambda}'\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})}\boldsymbol{s_\lambda}$, and maximizing this with respect to $\boldsymbol{s_\lambda}$ leads to $(\mathbf{Z}^{(\boldsymbol{\lambda})})'(-\mathbf{A}_*^{(\boldsymbol{\lambda},\boldsymbol{\lambda})})(\mathbf{Z}^{(\boldsymbol{\lambda})})$ as its maximum. Also, $\widetilde{\mathcal{Y}}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon})$ cannot be greater than zero. Otherwise, it must follow that $\widehat{h}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) > 0$ which is contradictory. Third, if $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) = 0$ and $\widehat{h}^{(\boldsymbol{\upsilon})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) = 0$, for the same $\boldsymbol{s_\upsilon}$, $\boldsymbol{s_\lambda}$ cannot be optimal to maximizing the right side of (A.4) from the fact that $\boldsymbol{\lambda}_*$ is an interior element of $\boldsymbol{\Lambda}$. We can ignore the case in which $\widehat{h}^{(\boldsymbol{\lambda})}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}) = 0$. 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 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*) We let $H_{00,n} := \sup_{\boldsymbol{s_\lambda}} \sup_{h^{(\boldsymbol{\lambda})}} \{2DL_n(\boldsymbol{\theta}_*;\boldsymbol{s_\lambda})h^{(\boldsymbol{\lambda})} + D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{s_\lambda})(h^{(\boldsymbol{\lambda})})^2\}$ and note that $\mathcal{LR}_n^{(2)} = H_{00,n} + H_{2,n} + o_{\mathbb{P},\boldsymbol{s_\lambda}}(1) + o_{\mathbb{P},\boldsymbol{s_\tau}}(1)$. Furthermore, $H_{00,n} = \sup_{\{\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}\}} \sup_{\{h^{(\boldsymbol{\upsilon})},h^{(\boldsymbol{\lambda})}\}} 2DL_n(\boldsymbol{\theta}_*;\boldsymbol{s_\upsilon})h^{(\boldsymbol{\upsilon})} + 2DL_n(\boldsymbol{\theta}_*;\boldsymbol{s_\lambda})h^{(\boldsymbol{\lambda})} + 2DL_n(\boldsymbol{\theta}_*;\boldsymbol{s_\lambda};\boldsymbol{s_\upsilon})h^{(\boldsymbol{\lambda})}h^{(\boldsymbol{\upsilon})} + D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{s_\upsilon})(h^{(\boldsymbol{\upsilon})})^2 + D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{s_\lambda})(h^{(\boldsymbol{\lambda})})^2$ subject to $h^{(\boldsymbol{\upsilon})} = 0$. Note that $H_{00,n} = H_{01,n}$ without the constraint $h^{(\boldsymbol{\upsilon})} = 0$. Therefore, $H_{00,n} \Rightarrow \mathcal{H}_1$ given the fact that $H_{01,n} \Rightarrow \mathcal{H}_0 + \mathcal{H}_1$ as given in (A.4). Thus, $\mathcal{LR}_n^{(2)} \Rightarrow \mathcal{H}_1 + \mathcal{H}_2$ from the fact that $(H_{01,n}, H_{2,n}) \Rightarrow (\mathcal{H}_0 + \mathcal{H}_1, \mathcal{H}_2)$.

(*iv*) The desired result holds by continuous mapping because it follows that $(\mathcal{LR}_n^{(1)}, \mathcal{LR}_n^{(2)}) \Rightarrow (\mathcal{H}_0 + \mathcal{H}_1 + \mathcal{H}_2, \mathcal{H}_1 + \mathcal{H}_2)$ as shown in (*i*) and (*iii*). ■

Proof of Theorem 3. We exploit (A.3) 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,

https://doi.org/10.1017/S0266466617000354 Published online by Cambridge University Press

$$\sup_{h^{(\pi)}}\{2DL_n(\boldsymbol{\theta}_*;\boldsymbol{d_\pi})h^{(\pi)}+D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{d_\pi})(h^{(\pi)})^2\}$$
$$=\sup_{(h^{(\upsilon)},h^{(\lambda)})} 2\{DL_n(\boldsymbol{\theta}_*;\boldsymbol{s_\upsilon})h^{(\upsilon)}+DL_n(\boldsymbol{\theta}_*;\boldsymbol{s_\lambda})h^{(\lambda)}+DL_n(\boldsymbol{\theta}_*;\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})h^{(\upsilon)}h^{(\lambda)}\}$$
$$+D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{s_\upsilon})(h^{(\upsilon)})^2+D^2L_n(\boldsymbol{\theta}_*;\boldsymbol{s_\lambda})(h^{(\lambda)})^2$$
$$\Rightarrow \sup_{(h^{(\upsilon)},h^{(\lambda)})} 2\{\mathcal{Z}(\boldsymbol{s_\upsilon})h^{(\upsilon)}+\boldsymbol{s}'_{\boldsymbol{\lambda}}\mathbf{Z}^{(\lambda)}h^{(\lambda)}+\boldsymbol{s}'_{\boldsymbol{\lambda}}\mathbf{A}_*^{(\lambda,\upsilon)}(\boldsymbol{s_\upsilon})h^{(\upsilon)}h^{(\lambda)}\}$$
$$+A_*^{(\upsilon,\upsilon)}(\boldsymbol{s_\upsilon})(h^{(\upsilon)})^2+\boldsymbol{s}'_{\boldsymbol{\lambda}}\mathbf{A}_*^{(\lambda,\lambda)}\boldsymbol{s_\lambda}(h^{(\lambda)})^2. \tag{A.5}$$

Given this, $h^{(\upsilon)}$ and $h^{(\lambda)}$ have to be at least greater than or equal to zero. This implies that the following four different inequality constraints can be possibly associated with this maximization process: first, if any equality condition does not bind, $\sqrt{n}(\widehat{h}_n^{(\upsilon)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}),\widehat{h}_n^{(\lambda)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda}\boldsymbol{s_\tau}))' \Rightarrow \boldsymbol{\mathcal{G}}^{(\pi)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})$ by the standard first-order condition and Lemma A1. This occurs if every component of $\boldsymbol{\mathcal{G}}^{(\pi)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})$ is strictly greater than zero. Second, if $\mathcal{G}^{(\upsilon)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})<0$, it simply holds that $\widehat{h}^{(\lambda)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})=\max[0,\dot{\mathcal{G}}^{(\lambda)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})]$, and $\sqrt{n}(\widehat{h}_n^{(\upsilon)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}),\widehat{h}_n^{(\lambda)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}))' \Rightarrow (0,\max[0,\dot{\mathcal{G}}^{(\lambda)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})])'$. Third, if $\widehat{h}^{(\lambda)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})=0$ in the right side of (A.5) because $\mathcal{G}^{(\lambda)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})<0$, $\widehat{h}^{(\upsilon)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})=\max[0,\dot{\mathcal{G}}^{(\upsilon)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})]$. This implies that $\sqrt{n}(\widehat{h}_n^{(\upsilon)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}),\widehat{h}_n^{(\lambda)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}))' \Rightarrow (\max[0,\dot{\mathcal{G}}^{(\upsilon)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda})],0)'$. Fourth, it must follow that $\sqrt{n}(\widehat{h}_n^{(\upsilon)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}),\widehat{h}_n^{(\lambda)}(\boldsymbol{s_\upsilon},\boldsymbol{s_\lambda},\boldsymbol{s_\tau}))' \Rightarrow (0,0)'$ for any other case. Therefore, if we combine all these and apply Theorem 1(*i*),

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

and this implies that

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

under H_0. Therefore, it now follows that

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

where for each $\boldsymbol{s_\upsilon}$,

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

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

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

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

$$\begin{aligned} DL_n(\ddot{\boldsymbol{\theta}}_n; \mathbf{s}_{\boldsymbol{\upsilon}}) - DL_n(\boldsymbol{\theta}_*; \mathbf{s}_{\boldsymbol{\upsilon}}) &= \{DL_n(\boldsymbol{\upsilon}_0, \dot{\boldsymbol{\lambda}}_n(\mathbf{s}_{\boldsymbol{\upsilon}}), \boldsymbol{\tau}_*; \mathbf{s}_{\boldsymbol{\upsilon}}; \ddot{\mathbf{s}}_{\boldsymbol{\lambda},n})\}\{\ddot{h}_n^{(\boldsymbol{\lambda})}(\ddot{\mathbf{s}}_{\boldsymbol{\lambda},n})\} \\ &= DL_n(\boldsymbol{\upsilon}_0, \dot{\boldsymbol{\lambda}}_n(\mathbf{s}_{\boldsymbol{\upsilon}}), \boldsymbol{\tau}_*; \mathbf{s}_{\boldsymbol{\upsilon}}; \ddot{\mathbf{s}}_{\boldsymbol{\lambda},n}) \\ &\quad \times \{-D^2 L_n(\boldsymbol{\upsilon}_0, \bar{\boldsymbol{\lambda}}_n(\mathbf{s}_{\boldsymbol{\upsilon}}), \boldsymbol{\tau}_*; \ddot{\mathbf{s}}_{\boldsymbol{\lambda},n})\}^{-1} DL_n(\boldsymbol{\theta}_*; \ddot{\mathbf{s}}_{\boldsymbol{\lambda},n}), \end{aligned} \quad \textbf{(A.6)}$$

where $(\ddot{h}_n^{(\boldsymbol{\lambda})}(\ddot{\mathbf{s}}_{\boldsymbol{\lambda},n}), \ddot{\mathbf{s}}_{\boldsymbol{\lambda},n}) := \arg\sup_{h^{(\boldsymbol{\lambda})}, \mathbf{s}_{\boldsymbol{\lambda}}} L_n(\boldsymbol{\upsilon}_0, \boldsymbol{\lambda}_* + h^{(\boldsymbol{\lambda})}\mathbf{s}_{\boldsymbol{\lambda}}, \boldsymbol{\tau}_*)$, and the last equality follows from the mean-value theorem: there is $\bar{\boldsymbol{\lambda}}_n(\mathbf{s}_{\boldsymbol{\upsilon}})$ such that (A.6) holds. Given this and Theorem 2(ii), we can apply the ULLN:

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

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

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

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

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

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

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

https://doi.org/10.1017/S0266466617000354 Published online by Cambridge University Press