Journal of Economic Theory and Econometrics, Vol. 23, No. 2, Jun. 2012, 89–112

Quasi-Maximum Likelihood Estimation Revisited Using the Distance and Direction Method*

Jin Seo Cho†

Abstract We examine an asymptotic analysis of differentiable econometric models using the distance and direction (DD) method introduced by Cho and White (2012), in which the conventional analysis for the quasi-maximum likelihood estimation and inference can be treated as a special case. We extend their approach and revisit the conventional quasi-likelihood ratio, Wald, and Lagrange multiplier test statistics through a different perspective. This new perspective is further analyzed in a unified framework, and we exploit this to introduce new classes of test statistics.

Keywords distance and direction method, quasi-maximum likelihood estimation, asymptotic distribution, quasi-likelihood ratio test, Wald test, Lagrange multiplier test

JEL Classification C12, C22, C45, C52

*The author is most grateful to the Editor, Chirok Han, and three anonymous referees for their helpful comments. Also, I gratefully acknowledge helpful discussions with Isao Ishida, Yongho Jeon, Chul Eung Kim, Yong Kim, Kosuke Oya, Taeyoung Park, and seminar participants at the department of Applied Statistics, Yonsei University and the Center for the Study of Finance and Insurance at Osaka University. I also appreciate the research support by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2009-332-B00031).

†School of Economics, Yonsei University, Korea. E-mail: jinseocho@yonsei.ac.kr

Received November 20, 2011, Revised March 19, 2012, Accepted March 24, 2012

1. INTRODUCTION

Many econometric models are differentiable, and they play a central role in analyzing the asymptotic behaviors of many important statistics. On the other hand, there are a number of non-differentiable models. For example, the conditional heteroskedasticity model of King and Shively (1993) is not differentiable under the null of conditional homoskedasticity. In addition, the stochastic frontier production function models of Aigner, Lovell, and Schmidt (1977), further extended by Stevenson (1980), are not differentiable if outputs are efficiently produced. They are only directionally differentiable.

Cho and White (2012) introduce the so-called direction and distance (DD) method to handle directionally differentiable models. When models are only directionally differentiable, the DD method can treat the conventional analysis assuming differentiability as a special case of their approach.

Further aspects of differentiable models can be provided by the DD method. Although Cho and White (2012) examine differentiable models using the DD method, their discussions do not cover genuine aspects of differentiable models, which we can additionally obtain by applying the DD method to differentiability. When the DD method is applied, we can view differentiable models in a different perspective, and this can be further exploited for developing a theory for new statistics.

Therefore, the goal of this paper is in finding new aspects of differentiable models hidden by the conventional analysis. Our specific achievements are providing new classes of statistics testing for regular hypotheses and a unified theory for this provision. In addition, we also provide new properties of differentiable models, which are not present in the previous literature.

Our specific details and plans for this goal are as follows. We focus our discussions on the quasi-maximum likelihood (QML) estimation and related test statistics using the DD method: the quasi-likelihood ratio (QLR), Wald, and Lagrange multiplier (LM) test statistics. By this, we demonstrate that the DD method has the capability of generalizing the analysis of the conventional tests in a unified framework, so that new classes of test statistics can be provided in a way that the three test statistics are elements of each class. Given that the QLR, Wald, and LM tests are not asymptotically most efficient test statistics in the context of model misspecification, this provision can be thought of as natural extensions for having better performing test statistics. In addition, we explore the DD method in a couple of ways to provide new aspects of differentiable models. First, we apply the DD method by assuming that the unknown parameter value is known. Second, we iterate the same process but by replacing the unknown parameter value with the QML estimator constrained by the null hypothesis. Each method yields its own distance and direction estimator, and they are asymptotically interrelated. We detail this interrelationship below.

This paper is organized as follows. We review the conventional assumptions and results of the QML estimation in Section 2. The main goal of this is twofold. This lets our paper be self-contained; and the well-known properties of the QML estimation will be exploited in obtaining further insights of the DD method. Section 3 briefly reviews the DD method and applies this to the QML estimation. This section aims to sharpen the analysis of the QML estimation. Section 4 examines the asymptotic null behaviors of the QLR, Wald, and LM tests by the DD method, and Section 5 further examines the same test statistics using the DD method in a different perspective. Concluding remarks are provided in Section 6, and mathematical proofs are collected in the Appendix.

Before proceeding, some mathematical notations are provided to avoid possible confusions. We denote $f'(\cdot)|_{x_*}$ by $f'(x_*)$ and let $\partial_x f(x)$ and $\partial^2_{x,y} f(x,y)$ denote $(\partial/\partial x)f(x)$ and $(\partial^2/\partial x\partial y)f(x,y)$ for the sake of brevity. Also, for a set B in a Euclidean space, ∂B denotes the bolder of B.

2. THE STANDARD ANALYSIS OF THE QML STATISTICS

We first fix our ideas by stating the conventional assumptions and results of standard econometric models assuming differentiability. The goal of this is twofold: this lets our paper be self-contained, and we use them for deriving the main results of this paper. Because of this, we do not prove the theorems given in this section.

The following assumptions are our benchmark assumptions.

Assumption 1. *(i) 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 $(\Omega, \mathscr{F}, \mathbb{P})$ is a φ-mixing process of size $-r/(2(r-1))$ with $r \geq 2$ or an α-mixing process of size $-r/(r-2)$ with $r > 2$;*

(ii) The sum of functions defined on Θ $L_n(\cdot) := \sum_{t=1}^n \ell_t(\cdot;\mathbf{X}^t)$ is the quasi-likelihood function such that for each t, $\ell_t(\cdot;\mathbf{X}^t) \in \mathscr{C}^{(2)}(\Theta)$ almost surely–$\mathbb{P}$ (a.s.–$\mathbb{P}$), and for each $\theta \in \Theta$, $\ell_t(\theta;\cdot)$ is measurable–$\mathscr{F}_t$, where for each t, we let $\mathbf{X}^t$ be $(\mathbf{X}_1', \cdots, \mathbf{X}_t')'$, $\mathscr{C}^{(2)}(\Theta)$ is a space of twice continuously differentiable functions defined on Θ, Θ is a compact and convex set in $\mathbb{R}^s$ with $s \in \mathbb{N}$, and $\mathscr{F}_t$ is the Borel σ-field generated by the open sets of $\mathbb{R}^{mt}$;

(iii) For each $\theta \in \Theta$ and $n \in \mathbb{N}$, $nE[\overline{L}_n(\theta)]$ exists in $\mathbb{R}$ and is finite, where $\overline{L}_n(\cdot) := n^{-1}L_n(\cdot)$;

(iv) There is a unique $\theta_* \in \Theta$ *such that for some* $L(\cdot) : \Theta \mapsto \mathbb{R}$*,* $L(\cdot) = E[\overline{L}_n(\cdot)]$ *for every* $n \in \mathbb{N}$*, and it has a global maximum at* θ_**;*

(v) For $j = 1, 2, \ldots, s$ *and each* $t \in \mathbb{N}$*, there is* M_t *such that* $|\partial_j \ell_t(\theta_*; \mathbf{X}^t)|^r \leq M_t$ *and* $E[M_t] < \infty$*, where* $\partial_j \ell_t(\theta_*; \mathbf{X}^t) := (\partial/\partial\theta_j)\ell_t(\theta; \mathbf{X}^t)|_{\theta=\theta_*}$*;*

(vi) For $i, j = 1, 2, \ldots, s$*,* $\{\ell_t : \Theta \times \Omega \mapsto \mathbb{R}\}$ *and* $\{\partial_{ij}\ell_t : \Theta \times \Omega \mapsto \mathbb{R}\}$ *are Lipschitz–*L_1 *a.s–*$\mathbb{P}$ *on* Θ*, where* $\partial_{ij}\ell_t := (\partial^2/\partial\theta_i\partial\theta_j)\ell_t(\cdot; \mathbf{X}^t)$*;*

(vii) For $i, j = 1, 2, \ldots, s$*,* $\{\ell_t, \mathscr{F}_t\}$ *and* $\{\partial_{ij}\ell_t, \mathscr{F}_t\}$ *are* L_1*–mixingales on* $(\Theta, \|\cdot\|)$*, where* $\|\cdot\|$ *is the Euclidean norm;*

(viii) There is a positive definite and finite $\mathbf{B}$ *such that for each* $n \in \mathbb{N}$*,* $\mathbf{B} = acov\{n^{-1/2}\nabla_\theta L_n\ (\theta_*)\}$*, where 'acov' denotes the asymptotic covariance of a given argument; and*

(ix) There is $\mathbf{A}(\cdot) : \Theta \mapsto \mathbb{R}^{s\times s}$ *such that for each* θ*,* $\mathbf{A}(\theta)$ *is negative definite, and for each* $n \in \mathbb{N}$*,* $\mathbf{A}(\theta) = E[\nabla^2_\theta \overline{L}_n(\theta)]$.

These assumptions are standard assumptions for the QML estimator, and they are mostly used to estimate conditional mean equations by nesting models in the quasi-likelihood function. In addition, when conditional mean equations are correctly specified with the linear exponential family distributions assumed, the conditional mean equations can be consistently estimated. That is, distributional misspecification does not matter. White (1994), Gourieroux, Monfort, and Trognon (1984), Levine (1983), Engle and Russell (1998) among others provide meaningful implications of this. Further explanations of these conditions can also be found from numerous sources. White (1982), Andrews (1987), Andrews (1988), Gallant and White (1988), Andrews (1992), White (1994), White (2000) among others provide the roles of the relevant assumptions.

We provide the asymptotic behavior of the QML estimator, and for this we specifically let the QML estimator be defined as $\widehat{\theta}_n := \arg\max_{\theta\in\Theta} \overline{L}_n(\theta)$. We state the asymptotic properties of this QML estimator for our future reference.

Theorem 1. *Given Assumption 1,*

(i) $\widehat{\theta}_n \to \theta_*$ *a.s.–*$\mathbb{P}$*; and*

(ii) $\mathbf{B}^{-1/2}\mathbf{A}\sqrt{n}(\widehat{\theta} - \theta_*) \overset{A}{\sim} N(0, \mathbf{I})$*, where* $\mathbf{A} := \mathbf{A}(\theta_*)$.

One of our interests is in exploiting the QML estimator for testing regular hypotheses. We consider the following hypotheses: $\mathscr{H}_0 : \mathbf{R}\theta_* = \mathbf{r}$ versus $\mathscr{H}_1 : \mathbf{R}\theta_* \neq \mathbf{r}$, where $\mathbf{R}$ is a $q \times s$ vector. In handling the hypotheses, three test statistics are most commonly used: quasi-likelihood ratio, Wald, and Lagrange multiplier test statistics, defined as

$$\mathscr{QLR}_n := 2n\{\overline{L}_n(\widehat{\theta}_n) - \overline{L}_n(\widetilde{\theta}_n)\},$$

$$\mathscr{W}_n := n(\mathbf{R}\widehat{\theta}_n - \mathbf{r})'\{\mathbf{R}\widehat{\mathbf{A}}_n^{-1}\widehat{\mathbf{B}}_n\widehat{\mathbf{A}}_n^{-1}\mathbf{R}'\}^{-1}(\mathbf{R}\widehat{\theta}_n - \mathbf{r}), \quad \text{and}$$

$$\begin{aligned}\mathscr{L}\mathscr{M}_n :=& n\nabla'_\theta\bar{L}_n(\widetilde{\theta}_n)\{\nabla^2_\theta\bar{L}_n(\widetilde{\theta}_n)\}^{-1}\mathbf{R}' \\ &\times\{\mathbf{R}\widetilde{\mathbf{A}}_n^{-1}\widetilde{\mathbf{B}}_n\widetilde{\mathbf{A}}_n^{-1}\mathbf{R}'\}^{-1}\mathbf{R}\{\nabla^2_\theta\bar{L}_n(\widetilde{\theta}_n)\}^{-1}\nabla_\theta\bar{L}_n(\widetilde{\theta}_n),\end{aligned}$$

where $\widetilde{\theta}_n$ is the QML estimator constrained by $\mathscr{H}_0$: $\widetilde{\theta}_n := \arg\max_{\theta\in\Theta}\bar{L}_n(\theta)$ such that $\mathbf{R}\theta = \mathbf{r}$. The properties of the test statistics and the constrained QML (CQML) estimator are well known. We provide them for the use of analyzing the DD method given below.

Theorem 1 (Continued)**.** *Given Assumption 1 and* $\mathscr{H}_0$*, if we let* $\mathbf{D} := \mathbf{I}_s - \mathbf{A}^{-1}\mathbf{R}'(\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')^{-1}\mathbf{R}$ *and* $\mathbf{P} := \mathbf{R}'(\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')^{-1}\mathbf{R}\mathbf{A}^{-1}$*,*

(iii) $\widetilde{\theta}_n \to \theta_*$ *a.s.*$-\mathbb{P}$*;*

(iv) $\sqrt{n}(\widetilde{\theta}_n - \theta_*) \overset{A}{\sim} N(0, \mathbf{D}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{D}')$*;*

(v) $\sqrt{n}(\widehat{\theta}_n - \widetilde{\theta}_n) \overset{A}{\sim} N(0, \mathbf{A}^{-1}\mathbf{P}\mathbf{B}\mathbf{P}'\mathbf{A}^{-1})$*;*

(vi) $\sqrt{n}\bar{L}_n(\widetilde{\theta}_n) = \mathbf{P}\sqrt{n}\bar{L}_n(\theta_*) + o_{\mathbb{P}}(1)$.

On the other hand, under $\mathscr{H}_1$*,*

(vii) $\widetilde{\theta}_n \to \theta_o$*, where* $\theta_o = \arg\max_{\theta\in\Theta} E[n^{-1}L_n(\theta)]$ *such that* $\mathbf{R}\theta = \mathbf{r}$*; and*

(viii) if $\bar{\mathbf{A}}$ *is strictly negative definite and finite, for some* $\overline{\theta}$ *between* θ_* *and* θ_o*,* $\sqrt{n}[(\widetilde{\theta}_n - \theta_o) + \bar{\mathbf{A}}^{-1}\mathbf{R}'(\mathbf{R}\bar{\mathbf{A}}^{-1}\mathbf{R}')^{-1}(\mathbf{R}\theta_* - \mathbf{r})] \overset{A}{\sim} N(0, \bar{\mathbf{D}}\bar{\mathbf{A}}^{-1}\mathbf{B}\,\bar{\mathbf{A}}^{-1}\bar{\mathbf{D}}')$*, where* $\bar{\mathbf{A}} := \mathbf{A}(\overline{\theta})$ *and* $\bar{\mathbf{D}} := \mathbf{I}_s - \bar{\mathbf{A}}^{-1}\mathbf{R}'(\mathbf{R}\bar{\mathbf{A}}^{-1}\mathbf{R}'\,)^{-1}\mathbf{R}$.

We now examine the asymptotic behaviors of the test statistics under the null and alternative hypotheses. For this, the following condition is assumed.

Assumption 2. *(i)* $\operatorname{rank}(\mathbf{R}) = q \le s$*;*

(ii) $\widetilde{\mathbf{A}}_n - \mathbf{A} \overset{a.s.}{\to} 0$*;* $\widetilde{\mathbf{B}}_n - \mathbf{B} \overset{a.s.}{\to} 0$*;* $\widehat{\mathbf{A}}_n - \mathbf{A} \overset{a.s.}{\to} 0$*; and* $\widehat{\mathbf{B}}_n - \mathbf{B} \overset{a.s.}{\to} 0$.

We also let $N_2(\mu, \Sigma; \mathbf{K})$ denote $\overline{\mathscr{Z}}'\mathbf{K}\overline{\mathscr{Z}}$ for notational simplicity, where $\overline{\mathscr{Z}} \sim N(\mu, \Sigma)$. The following theorem contains the asymptotic behaviors of the test statistics.

Theorem 2. *Given Assumptions 1, 2,*

(i) $\mathscr{QLR}_n \overset{A}{\sim} N_2(0, \mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'; -[\mathbf{R}\mathbf{A}^{-1}\mathbf{R}']^{-1})$ *under* $\mathscr{H}_0$*, and for any sequence* $\{c_n\}$ *such that* $c_n = o(n)$*,* $\mathbb{P}[\mathscr{QLR}_n \ge c_n] \to 1$ *under* $\mathscr{H}_1$*;*

(ii) $\mathscr{W}_n \overset{A}{\sim} \mathscr{X}^2_q$ *under* $\mathscr{H}_0$*, and* $\mathbb{P}[\mathscr{W}_n \ge c_n] \to 1$ *under* $\mathscr{H}_1$*;*

(iii) $\mathscr{L}\mathscr{M}_n \overset{A}{\sim} \mathscr{X}^2_q$ *under* $\mathscr{H}_0$*, and* $\mathbb{P}[\mathscr{L}\mathscr{M}_n \ge c_n] \to 1$ *under* $\mathscr{H}_1$*;*

(iv) for any $\varepsilon > 0$*,* $\mathbb{P}[|\mathscr{W}_n - \mathscr{L}\mathscr{M}_n| \ge \varepsilon] \to 0$ *under* $\mathscr{H}_0$*; and*

(v) for any $\varepsilon > 0$*,* $\mathbb{P}[|\mathscr{W}_n - \mathscr{QLR}_n| \ge \varepsilon] \to 0$ *under* $\mathscr{H}_0$*, if further* $\mathbf{A} + \mathbf{B} = 0$.

Theorem 2 states that the most common test statistics are consistent. Also, the QLR, Wald, and LM test statistics are asymptotically equivalent under the null if the information matrix equality holds. This is often called the *trinity* in the literature (*e.g.* Hayashi (2000)).

3. THE QML ESTIMATOR AND THE DD METHOD

We now examine the QML estimation using the DD method in Cho and White (2012). The goal of this re-examination is not in finding out the same results but examining the questions from a different angle and developing other meaningful statistics from this. We do not even hesitate to exploit the consequences in Section 2 to obtain efficient insights of differentiable models. In this section, we focus on the asymptotic behavior of the QML estimator and quasi-likelihood.

The essence of the DD method is in decomposing parameter values using *distance* (d) and *direction* (s). More specifically, for a particular parameter θ and a reference point θ_*, we can find a (d,s) such that

$$\theta \equiv \theta_* + d \cdot s, \tag{1}$$

where $s \in \mathbb{S}^{s-1} := \{s \in \mathbb{R}^s : s's = 1\}$ and $d \in \mathbb{D}(s) := \{d \in \mathbb{R}^+ : \theta_* + ds \in \Theta\}$. Note that $\mathbb{D}(s)$ represents a set of positive real numbers bounded by the distance between θ_* and $\theta \in \partial\Theta$ in the direction of s. As Θ is a convex set with θ_* being an interior element, $\mathbb{D}(s)$ is a non-empty and convex set for each s. Also, there is a one-to-one mapping between θ ($\neq \theta_*$) and (d,s). That is, θ is different from $\widetilde{\theta}$ if and only if (d,s) is different from $(\widetilde{d},\widetilde{s})$, where (d,s) and $(\widetilde{d},\widetilde{s})$ are distances and directions associated with θ and $\widetilde{\theta}$, respectively. Thus, inferring the unknown parameter θ_* can be also delivered through (d,s).

We specifically implement the DD method to the QML estimation. For this, we now rewrite the quasi-likelihood function as a function of (d,s). That is, for each θ, there is a unique (d,s) satisfying eq. (1) and from this, $L_n(\theta) \equiv L_n(\theta_* + d \cdot s)$. Note that the left-hand side (LHS) is now rephrased into another function of (d,s), and $L_n(\,\cdot\,)$ can be maximized instead with respect to (d,s). For this, we first apply the mean-value theorem around 0 with respect to d: for each s and for some $\overline{d}$ between d and 0,

$$\overline{L}_n(\theta_* + d \cdot s) = \overline{L}_n(\theta_*) + \nabla'_\theta \overline{L}_n(\theta_*) s \cdot d + \frac{1}{2} s' \nabla^2_\theta \overline{L}_n(\theta_* + \overline{d} \cdot s) s \cdot d^2. \tag{2}$$

Here, 0 is the reference point corresponding to θ_* because if $d = 0$, $\theta_* + d \cdot s = \theta_*$. We further note that we can apply the central limit theorem (CLT) and uniform law of large numbers (ULLN) to each element in the right-hand side (RHS) of eq. (2): $\{n^{1/2}\nabla_{\theta}\overline{L}_n(\theta_*), \nabla^2_{\theta}\overline{L}_n(\cdot)\} \Rightarrow \{\mathscr{Z}, \mathbf{A}(\cdot)\}$, where $\mathscr{Z} \sim N(0,\mathbf{B})$. Note that $\mathbf{A}(\theta_*) \equiv \mathbf{A}$ by definition. We now apply the continuous mapping theorem to eq. (2) and obtain that

$$n\{\overline{L}_n(\theta_* + d \cdot s) - \overline{L}_n(\theta_*)\} \Rightarrow \mathscr{Z}' s \cdot \delta + \frac{1}{2} s' \mathbf{A}(\theta_* + \bar{d} \cdot s) s \cdot \delta^2, \tag{3}$$

where δ captures the asymptotic behavior of $\sqrt{n}d$.

We can obtain the asymptotic behavior of the QML estimator from this asymptotic behavior. Given that the QML estimator is obtained by maximizing the LHS of eq. (2) with respect to θ, it is now equivalent to maximizing eq. (3) with respect to (d,s). Thus,

$$\sup_{s \in \mathbb{S}^{s-1}} \sup_{d \in \mathbb{D}(s)} n\{\overline{L}_n(\theta_* + d \cdot s) - \overline{L}_n(\theta_*)\}$$
$$\Rightarrow \sup_{s \in \mathbb{S}^{s-1}} \sup_{\delta \in \mathbb{R}^+} \mathscr{Z}' s \cdot \delta + \frac{1}{2} s' \mathbf{A}(\theta_* + \bar{d} \cdot s) s \cdot \delta^2. \tag{4}$$

Here, the space $\mathbb{D}(s)$ is replaced by $\mathbb{R}^+$ because the space for $\sqrt{n}d$ is $\mathbb{R}^+$ at the limit. Also, Theorem 1(i) states that the QML estimator is consistent for θ_*, implying that $\widehat{d}_n(s) := \arg\max_{d \in \mathbb{D}(s)} \overline{L}_n(\theta_* + d \cdot s)$ is consistent for 0 uniformly on $\mathbb{S}^{s-1}$. This is mainly because $0 \in \mathbb{D}(s)$ for every s, so that the asymptotic limit of the QML estimator, which is θ_*, can be also generated from $\theta_* + d \cdot s$ by letting $d = 0$.

This implies several further consequences. First, for every s, when $\bar{d}$ is supposed to be between 0 and $\widehat{d}_n(s)$, it is also dependent upon the value of s but consistent for 0 uniformly on $\mathbb{S}^{s-1}$, because $\widehat{d}_n(s)$ is consistent for 0 uniformly in s. Thus, eq. (4) can be also written as

$$\sup_{s \in \mathbb{S}^{s-1}} \sup_{d \in \mathbb{D}(s)} n\{\overline{L}_n(\theta_* + d \cdot s) - \overline{L}_n(\theta_*)\} \Rightarrow \sup_{s \in \mathbb{S}^{s-1}} \sup_{\delta \in \mathbb{R}^+} \mathscr{Z}' s \cdot \delta + \frac{1}{2} s' \mathbf{A} s \cdot \delta^2. \tag{5}$$

Second, if we let $\widehat{\delta}(s)$ capture the asymptotic behavior of $\sqrt{n}\widehat{d}_n(s)$, its asymptotic version is now obtained as

$$\widehat{\delta}(s) := -\{s'\mathbf{A}s\}^{-1} \max[0, \mathscr{Z}'s], \tag{6}$$

where the 'max' operator is used to accommodate Kuhn-Tucker's theorem. That is, for every s, $\widehat{\delta}(s)$ has to be greater than 0, so that we cannot simply obtain the optimal estimator $\widehat{\delta}(s)$ as if it is an interior element: the boundary parameter problem has to be taken into account. Thus, applying Kuhn-Tucker's theorem yields the optimal solution in eq. (6). Here, we also multiply -1 to the RHS as $\mathbf{A}$ is negative definite uniformly in n, so that $\widehat{\delta}(s)$ cannot be negative.

The pointwise weak convergence in eq. (6) can be further strengthened. More specifically, we can state the result in eq. (6) by functional weak convergence: $\sqrt{n}\widehat{d}_n(\cdot) \Rightarrow \widehat{\delta}(\cdot)$. This trivially follows by the functional CLT (FCLT) that for each s, maximizing the RHS of eq. (2) with respect to $\sqrt{n}\widehat{d}_n(s)$ yields that $\sqrt{n}\widehat{d}_n(\cdot) = \sqrt{n}\dot{d}_n(\cdot) + o_{\mathbb{P}}(1)$, where we let $\dot{d}_n(s) := -\{s'\nabla^2_\theta \overline{L}_n(\theta_*)s\}^{-1} \max[0, \nabla'_\theta \overline{L}_n(\theta_*)s]$. Here, the distance between $\sqrt{n}\widehat{d}_n(\cdot)$ and $\sqrt{n}\dot{d}_n(\cdot)$ is $o_{\mathbb{P}}(1)$ uniformly on $\mathbb{S}^{s-1}$. Their difference is mainly due to $\bar{d}$, which is set to be zero when defining $\dot{d}_n$, and it converges to zero uniformly on $\mathbb{S}^{s-1}$ as discussed above. Their difference therefore becomes negligible in probability uniformly on $\mathbb{S}^{s-1}$. Thus, we can instead focus on $\dot{d}_n(\cdot)$ to derive the weak convergence of $\widehat{d}_n(\cdot)$. Also, the tightness of $\{n^{-1/2}\nabla'_\theta L_n(\theta_*)s\}$ as a function of s trivially holds because s is linearly multiplied to the object obeying the CLT. Furthermore, $n^{-1}\nabla^2_\theta L_n(\cdot)$ obeys the ULLN, so that $n^{-1}\nabla^2_\theta L_n(\cdot) - \mathbf{A}(\cdot) \to 0 - \mathbb{P}$. From these, $\sqrt{n}\widehat{d}_n(\cdot) \Rightarrow \widehat{\delta}(\cdot)$.

Third, the asymptotic behavior of the quasi-likelihood can also be examined by the DD method. Note that the quasi-likelihood is obtained by substituting eq. (6) into eq. (5). That is,

$$\sup_{s\in\mathbb{S}^{s-1}} \sup_{d\in\mathbb{D}(s)} n\{\overline{L}_n(\theta_* + d\cdot s) - \overline{L}_n(\theta_*)\}$$
$$\Rightarrow \sup_{s\in\mathbb{S}^{s-1}} \frac{1}{2}(-s'\mathbf{A}s)\widehat{\delta}(s)^2 = \sup_{s\in\mathbb{S}^{s-1}} -\frac{1}{2}\left(\frac{\max[0,\mathscr{Z}'s]^2}{s'\mathbf{A}s}\right), \quad (7)$$

so that

$$\sup_{s\in\mathbb{S}^{s-1}} \sup_{d\in\mathbb{D}(s)} 2n\{\overline{L}_n(\theta_* + d\cdot s) - \overline{L}_n(\theta_*)\}$$
$$\Rightarrow \sup_{s\in\mathbb{S}^{s-1}} (-s'\mathbf{A}s)\widehat{\delta}(s)^2 = \sup_{s\in\mathbb{S}^{s-1}} -\left(\frac{\max[0,\mathscr{Z}'s]^2}{s'\mathbf{A}s}\right), \quad (8)$$

which has a form similar to the likelihood ratio statistic testing the unknown parameter θ_*. In addition, we used the identity: $L_n(\widehat{\theta}_n) \equiv \sup_{s\in\mathbb{S}^{s-1}} \sup_{d\in\mathbb{D}(s)} L_n(\theta_* + d\cdot s)$ to obtain the above result. Here, we note that the RHS of eq. (8) is a function of a Gaussian process indexed by s. Thus, if we let $\mathscr{G}(\cdot)$ be a Gaussian process such that for each $s, \widetilde{s} \in \mathbb{S}^{s-1}$, $E[\mathscr{G}(s)] = 0$ and

$$E[\mathscr{G}(s)\mathscr{G}(\widetilde{s})] = \frac{s'\mathbf{B}\widetilde{s}}{\sqrt{-s'\mathbf{A}s}\sqrt{-\widetilde{s}'\mathbf{A}\widetilde{s}}},$$

then eq. (8) can alternatively be reformulated into

$$\sup_{s\in\mathbb{S}^{s-1}}\sup_{d\in\mathbb{D}(s)} 2n\{\overline{L}_n(\theta_*+d\cdot s)-\overline{L}_n(\theta_*)\}\Rightarrow \sup_{s\in\mathbb{S}^{s-1}}\max[0,\mathscr{G}(s)]^2. \tag{9}$$

That is, the asymptotic behavior can be written as a function a Gaussian process. This is a different view from that of the conventional method, as it associates the asymptotic behavior in terms of a Gaussian process combined with 'sup' and 'max' operators.

This also lets the distance estimator be represented as a function of a Gaussian process at the limit. More specifically, if we let $\mathscr{Y}(\cdot)$ be a Gaussian process such that for every $s,\widetilde{s}\in\mathbb{S}^{s-1}$, $E[\mathscr{Y}(s)]=0$ and

$$E[\mathscr{Y}(s)\mathscr{Y}(\widetilde{s})]=\frac{s'\mathbf{B}\widetilde{s}}{\{-s'\mathbf{A}s\}\{-\widetilde{s}'\mathbf{A}\widetilde{s}\}}, \tag{10}$$

we can let $\widehat{\delta}(\cdot)$ be $\max[0,\mathscr{Y}(\cdot)]$, so that the Gaussian process $\mathscr{Y}(\cdot)$ is also associated with $\mathscr{G}(\cdot)$. That is, for each $s\in\mathbb{S}^{s-1}$, $\mathscr{Y}(s)\equiv(-s'\mathbf{A}s)^{-1/2}\mathscr{G}(s)$.

The previous literature observes a number of instances in which the asymptotic null behaviors of statistics are represented by functions of Gaussian processes. This includes Davies (1977; 1987), Bierens (1990), Bierens and Ploberger (1997), Andrews (2001), Cho and White (2007; 2009; 2010; 2011a; 2011b; 2012), Cho and Ishida (2010), Cho, Cheong, and White (2011), and Cho, Ishida, and White (2011) among others. They are mostly associated with test statistics not identified under the hypothesis of their interest.

Fourth, the QML estimator can be also associated with the DD estimator. If we let $\widehat{s}_n := \arg\max_{s\in\mathbb{S}^{s-1}}\overline{L}_n(\theta_*+\widehat{d}_n(s)\cdot s)$, $\widehat{\theta}_n\equiv\theta_*+\widehat{d}_n(\widehat{s}_n)\cdot\widehat{s}_n$, so that $\sqrt{n}(\widehat{\theta}_n-\theta_*)\equiv\sqrt{n}\widehat{d}_n(\widehat{s}_n)\cdot\widehat{s}_n$. Also, $\sqrt{n}(\widehat{\theta}_n-\theta_*)\overset{\text{A}}{\sim}N(0,\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$ by Theorem 1(*ii*), implying that $\sqrt{n}\widehat{d}_n(\widehat{s}_n)\cdot\widehat{s}_n\overset{\text{A}}{\sim}N(0,\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$. From this, a couple of implications can be deduced. The first implication is that $\sqrt{n}\widehat{d}_n(\widehat{s}_n)$ has asymptotically zero probability mass at zero. More precisely, if we let $\hat{\mathfrak{s}}_n := \arg\max_{s\in\mathbb{S}^{s-1}}\max[0,\mathscr{G}(s)]^2$, $\hat{\mathfrak{s}}$ becomes the asymptotic weak limit of $\widehat{s}$, and it also follows that $\sqrt{n}(\widehat{\theta}_n-\theta_*)=\sqrt{n}\widehat{d}_n(\widehat{s}_n)\cdot\widehat{s}_n\Rightarrow\widehat{\delta}_n(\hat{\mathfrak{s}})\cdot\hat{\mathfrak{s}}\equiv\max[0,\mathscr{Y}(\hat{\mathfrak{s}})]\cdot\hat{\mathfrak{s}}$, which has to follow $N(0,\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$ in distribution by Theorem 1(*ii*). This implies that $\mathbb{P}(\max[0,\mathscr{Y}(\hat{\mathfrak{s}})]=0)=0$. Therefore, $\sqrt{n}\widehat{d}_n(\widehat{s}_n)\Rightarrow\mathscr{Y}(\hat{\mathfrak{s}})$. The second implication is that $\widehat{d}_n(\widehat{s}_n)\cdot\widehat{s}_n=\{-\nabla^2_\theta L_n(\theta_*)\}^{-1}\nabla_\theta L_n(\theta_*)+o_{\mathbb{P}}(1)$ because $\sqrt{n}\widehat{d}_n(\widehat{s}_n)\cdot\widehat{s}_n=\sqrt{n}(\widehat{\theta}_n-\theta_*)=\{-\nabla^2_\theta L_n(\theta_*)\}^{-1}\nabla_\theta L_n(\theta_*)+o_{\mathbb{P}}(1)$ by the conventional approximation. Thus, we can also obtain that

$$\widehat{\delta}(\hat{\mathfrak{s}})\cdot\hat{\mathfrak{s}}=\mathscr{Y}(\hat{\mathfrak{s}})\cdot\hat{\mathfrak{s}}=\{-\hat{\mathfrak{s}}'\mathbf{A}\hat{\mathfrak{s}}\}^{-1}\mathscr{Z}'\hat{\mathfrak{s}}\cdot\hat{\mathfrak{s}}=-\mathbf{A}^{-1}\mathscr{Z}, \tag{11}$$

and $\hat{\mathfrak{s}}$ satisfying the last equality in eq. (11) is obtained as $-\widehat{\mathfrak{c}}\mathbf{A}^{-1}\mathscr{Z}$, where $\widehat{\mathfrak{c}}$ is a scaler such that $\hat{\mathfrak{s}}'\hat{\mathfrak{s}} = 1$. That is,

$$\widehat{\mathfrak{c}} := \{\mathscr{Z}'(-\mathbf{A})^{-2}\mathscr{Z}\}^{-1/2} \quad \text{and} \quad \widehat{\delta}(\hat{\mathfrak{s}}) = \widehat{\mathfrak{c}}^{-1}. \tag{12}$$

Eq. (11) has an interesting interpretation which cannot be delivered by the conventional approach. When $\mathscr{Y}(\cdot)(\cdot)$ is viewed as a Gaussian random function defined on $\mathbb{S}^{s-1}$, we could have a random variable following a multivariate normal distribution by substituting the randomly chosen $\hat{\mathfrak{s}}$ into the Gaussian random process. That is, the QML estimator selects direction $\hat{\mathfrak{s}}$ so that $\mathscr{Y}(\hat{\mathfrak{s}}) \cdot \hat{\mathfrak{s}}$ becomes a multivariate normal random variable. If direction s is selected in a different way, the consequence can be different.

Finally, the quasi-likelihood ratio statistic testing the unknown parameter θ_* can also be associated with direction s. That is,

$$\begin{aligned} 2n\{\overline{L}_n(\widehat{\theta}_n) - \overline{L}_n(\theta_*)\} &\Rightarrow \sup_{s\in\mathbb{S}^{s-1}} \max[0, \mathscr{G}(s)]^2 \\ &= \mathscr{G}(\hat{\mathfrak{s}})^2 = -\hat{\mathfrak{s}}'\mathbf{A}\hat{\mathfrak{s}}\widehat{\delta}(\hat{\mathfrak{s}})^2 = \mathscr{Z}'(-\mathbf{A})^{-1}\mathscr{Z}, \end{aligned} \tag{13}$$

where the second last equality follows from eq. (12) and the fact that $\hat{\mathfrak{s}} = -\widehat{\mathfrak{c}}\mathbf{A}^{-1}\mathscr{Z}$. This also implies that $\sup_{s\in\mathbb{S}^{s-1}} \max[0, \mathscr{G}(s)]^2 \sim N_2(0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}; -\mathbf{A})$, which is a consistent result to Theorem 2(*iii*). Once again, the important aspect of this derivation is that we now understand the asymptotic behavior as a function of a Gaussian process. We formally state these consequences in the following theorem.

Theorem 3. *Given Assumption 1,*
(i) $\sqrt{n}\widehat{d}_n(\cdot) = \sqrt{n}\dot{d}_n(\cdot) + o_{\mathbb{P}}(1) \Rightarrow \widehat{\delta}(\cdot)$;
(ii) $\sqrt{n}(\widehat{\theta}_n - \theta_*) \equiv \sqrt{n}\widehat{d}_n(\widehat{s}_n) \cdot \widehat{s}_n \Rightarrow \widehat{\delta}(\hat{\mathfrak{s}}) \cdot \hat{\mathfrak{s}} = \mathscr{Y}(\hat{\mathfrak{s}}) \cdot \hat{\mathfrak{s}} \sim N(0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$;
(iii) $\hat{\mathfrak{s}} = \widehat{\mathfrak{c}}\mathbf{A}^{-1}\mathscr{Z}$ *and* $\widehat{\delta}(\hat{\mathfrak{s}}) = \widehat{\mathfrak{c}}^{-1}$; *and*
(iv) $2\{L_n(\widehat{\theta}_n) - L_n(\theta_*)\} \Rightarrow \mathscr{G}(\hat{\mathfrak{s}})^2 \sim N_2(0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}; -\mathbf{A})$.

The main contents of Theorem 3 are already proved while associating the DD method with the conventional approach. We therefore omit the proofs. Also, Theorem 3 provides additional insights of the DD method which Cho and White (2012) do not discuss.

4. THE QML TESTS AND THE DD METHOD

We now move our interests to inferencing and seek other insights of differentiable models using the DD method. Indeed, Theorem 3 enables us to look at the three most popular test statistics as special tests which we can define using the DD method.

We consider two different DD methods by letting two different parameters be our reference parameters. In this section, we let our reference parameter be the unknown parameter θ_* as before and examine the three most popular test statistics. In the next section, we choose the CQML estimator be our reference parameter and explore the DD methods from a different angle.

4.1. EXAMINATION OF THE CQML ESTIMATOR

We first examine the asymptotic behavior of the CQML estimator by the DD method. The CQML estimator plays a central role in analyzing the test statistics. First, we represent the hypotheses using distance and direction. That is, we let (d_0, s_0) be such that $\theta_o \equiv \theta_* + d_o \cdot s_o$. Note that θ_o maximizes $E[\overline{L}_n(\cdot)]$ subject to $\mathbf{R}\theta = \mathbf{r}$, so that $\mathbf{R}\theta_o \equiv \mathbf{r}$ by its definition. This implies that $\mathbf{R}\theta_* = \mathbf{r} - d_o \cdot \mathbf{R}s_o$. Also, for any $s \in \mathbb{S}^{s-1}$, $d_o = 0$ if and only if $\theta_o = \theta_*$. Also, $\theta_o = \theta_*$ under $\mathscr{H}_0$. Therefore, we can state that $\mathbf{R}\theta_* = \mathbf{r}$ under $\mathscr{H}_o$. On the other hand, $\theta_o \neq \theta_*$ under $\mathscr{H}_1$, so that for some $d_o > 0$ and $s_o \in \mathbb{S}^{s-1}$, $\mathbf{R}\theta_* = \mathbf{r} - d_o\mathbf{R}s_o$. Thus, the original hypotheses can be also reformulated into the following:

$$\mathscr{H}_0' : \forall\, s \in \mathbb{S}^{s-1},\, d_o = 0 \quad \text{versus}$$
$$\mathscr{H}_1' : \exists (s_o, d_o) \in \mathbb{S}^{s-1} \times \mathbb{R}^+ \text{ such that } \mathbf{R}\theta_* = \mathbf{r} - d_o\mathbf{R}s_o.$$

Second, we reparameterize the null condition $\mathbf{R}\theta = \mathbf{r}$ using distance and direction. For any θ, we can find (d,s) such that $\theta = \theta_* + d \cdot s$ by eq. (1), so that the null condition can also be written as $\mathscr{N} := \{(d,s) \in \mathbb{R}^+ \times \mathbb{S}^{s-1} : \mathbf{R}(\theta_* + d \cdot s) = \mathbf{r}\}$. We note that the elements in $\mathscr{N}$ are different under $\mathscr{H}_0$ and $\mathscr{H}_1$. If $\mathscr{H}_0$ holds, $\mathbf{R}\theta_* = \mathbf{r}$, and this implies that $d\mathbf{R} \cdot s = 0$. As $d \in \mathbb{R}^+$, this can be also rewritten as $\mathbf{R} \cdot s = 0$. That is, the null condition combined with $\mathscr{H}_0$ yields a set of directions orthogonal to $\mathbf{R}$. Here, any positive distance $d > 0$ satisfies the null condition. Thus, a subset of $\mathscr{N}$ satisfying $\mathscr{H}_0$ is now $\mathscr{N}_o := \{(d,s) \in \mathbb{R}^+ \times s : \mathbf{R} \cdot s = 0\}$.

Third, we now consider the asymptotic behavior of the CQML estimator. We focus on its null behavior as it determines the asymptotic null distributions of the tests. The CQML estimator is obtained by maximizing the quasi-likelihood (QL) function subject to the null condition: $\sup_{s \in \mathbb{S}_o^{s-1}} \sup_{d \in \mathbb{D}(s)} \overline{L}_n(\theta_* + d \cdot s)$, where $\mathbb{S}_o^{s-1} := \{s \in \mathbb{S}^{s-1} : \mathbf{R}s = 0\}$. Note that this maximization process satisfies the null condition by letting s be an element in $\mathbb{S}_o^{s-1}$. Given that d is not constrained by $\mathscr{N}_o$, maximizing the QL function with respect to d does not have to be constrained by the null condition. Thus, maximizing the QL function with respect to d is an unconstrained maximization process, implying that for each $s \in \mathbb{S}_o^{s-1}$, $\widehat{d}_n(s)$ is the argument maximizing $\overline{L}_n(\theta_* + d \cdot s)$. On the other hand, s has to be orthogonal to $\mathbf{R}$, and eq. (9) is now converted into

$$\sup_{s \in \mathbb{S}_o^{s-1}} \sup_{d \in \mathbb{D}(s)} 2n\{\overline{L}_n(\theta_* + d \cdot s) - \overline{L}_n(\theta_*)\} \Rightarrow \sup_{s \in \mathbb{S}_o^{s-1}} \max[0, \mathscr{G}(s)]^2. \qquad (14)$$

The only difference of this from eq. (9) is that the space for s is given by $\mathbb{S}_o^{s-1}$. The others are identical to the previous case. Thus, if we let $\widetilde{s}_n := \arg\max_{s \in \mathbb{S}_o^{s-1}} \overline{L}_n(\theta_* + \widehat{d}_n(s) \cdot s)$, $\widetilde{\theta}_n \equiv \theta_* + \widehat{d}_n(\widetilde{s}_n) \cdot \widetilde{s}_n$. On the other hand, Theorem 1(iv) implies that $\sqrt{n}\widehat{d}_n(\widetilde{s}_n) \cdot \widetilde{s}_n \equiv \sqrt{n}(\widetilde{\theta}_n - \theta_*) \overset{\text{A}}{\sim} N(0, \mathbf{DA}^{-1}\mathbf{BA}^{-1}\mathbf{D}')$, which means that $\sqrt{n}\widehat{d}_n(\widetilde{s}_n) \cdot \widetilde{s}_n$ does not have a probability mass at zero, either. In addition, Theorem 3(i) implies that $\sqrt{n}\widehat{d}_n(\,\cdot\,) \Rightarrow \widehat{\delta}(\,\cdot\,) = \max[0, \mathscr{Y}(\,\cdot\,)]$ by the definition of $\widehat{\delta}(\cdot)$, so that if we let $\widetilde{\mathfrak{s}} := \arg\max_{s \in \mathbb{S}_o^{s-1}} \max[0, \mathscr{G}(s)]^2$, $\max[0, \mathscr{Y}(\widetilde{\mathfrak{s}})] = \mathscr{Y}(\widetilde{\mathfrak{s}})$ with probability 1. Therefore, it now follows that

$$\widehat{\delta}(\widetilde{\mathfrak{s}}) \cdot \widetilde{\mathfrak{s}} = \mathscr{Y}(\widetilde{\mathfrak{s}}) \cdot \widetilde{\mathfrak{s}} = \{-\widetilde{\mathfrak{s}}'\mathbf{A}\widetilde{\mathfrak{s}}\}^{-1}\mathscr{Z}'\widetilde{\mathfrak{s}} \cdot \widetilde{\mathfrak{s}} = \mathbf{D}(-\mathbf{A})^{-1}\mathscr{Z}, \qquad (15)$$

and $\widetilde{\mathfrak{s}}$ satisfying the final equality is obtained when $\widetilde{\mathfrak{s}} = -\widetilde{\mathfrak{c}}\mathbf{DA}^{-1}\mathscr{Z}$, where $\widetilde{\mathfrak{c}}$ is a scaler such that $\widetilde{\mathfrak{s}}'\widetilde{\mathfrak{s}} = 1$. That is, $\widetilde{\mathfrak{c}} = \{\mathscr{Z}'\mathbf{A}^{-1}\mathbf{D}'\mathbf{DA}^{-1}\mathscr{Z}\}^{-1/2}$ and $\widehat{\delta}(\widetilde{\mathfrak{s}}) = \widetilde{\mathfrak{c}}^{-1}$. In particular, eqs. (11) and (15) imply that

$$\widehat{\delta}(\widetilde{\mathfrak{s}}) \cdot \widetilde{\mathfrak{s}} = \mathbf{D}\widehat{\delta}(\hat{\mathfrak{s}}) \cdot \hat{\mathfrak{s}}. \qquad (16)$$

This fact and eq. (14) further imply that

$$\begin{aligned} 2n\{\overline{L}_n(\widetilde{\theta}_n) - \overline{L}_n(\theta_*)\} &\Rightarrow \sup_{s \in \mathbb{S}_o^{s-1}} \max[0, \mathscr{G}(s)]^2 = \mathscr{G}(\widetilde{\mathfrak{s}})^2 = -\widetilde{\mathfrak{s}}'\mathbf{A}\widetilde{\mathfrak{s}}\widehat{\delta}(\widetilde{\mathfrak{s}})^2 \\ &= \mathscr{Z}'[\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{R}'(\mathbf{RA}^{-1}\mathbf{R}')^{-1}\mathbf{RA}^{-1}]\mathscr{Z}, \end{aligned}$$

where the last equality is obtained by the definition of $\mathbf{D}$ and eq. (15). Thus, $\sup_{s \in \mathbb{S}_o^{s-1}} \max[0, \mathscr{G}(s)]^2 \sim N_2(0, \mathbf{B}; \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{R}'(\mathbf{RA}^{-1}\mathbf{R}')^{-1}\mathbf{RA}^{-1})$, and the null condition modifies the asymptotic distribution of the parameter estimates. We formally state these results as follows.

Theorem 4. *Given Assumption 1 and* $\mathscr{H}_0$,

(i) $\sqrt{n}(\widetilde{\theta}_n - \theta_*) \equiv \sqrt{n}\widehat{d}_n(\widetilde{s}_n) \cdot \widetilde{s}_n \Rightarrow \widehat{\delta}(\widetilde{\mathfrak{s}}) \cdot \widetilde{\mathfrak{s}} = \mathscr{Y}(\widetilde{\mathfrak{s}}) \cdot \widetilde{\mathfrak{s}} \sim N(0, \mathbf{DA}^{-1}\mathbf{BA}^{-1}\mathbf{D}')$;

(ii) $\widetilde{\mathfrak{s}} = -\widetilde{\mathfrak{c}}\mathbf{DA}^{-1}\mathscr{Z}$ *and* $\widehat{\delta}(\widetilde{\mathfrak{s}}) = \widetilde{\mathfrak{c}}^{-1}$*; and*

(iii) $2\{L_n(\widetilde{\theta}_n) - L_n(\theta_*)\} \Rightarrow \mathscr{G}(\widetilde{\mathfrak{s}})^2$*, which follows* $N_2(0, \mathbf{B}; \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{R}'(\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')^{-1}\mathbf{RA}^{-1})$.

As most proofs are already stated, we do not reiterate the same proof in the Appendix.

4.2. EXAMINATION OF THE TEST STATISTICS

As mentioned above, the asymptotic behavior of the CQML estimator is crucial in analyzing the three most popular test statistics by the DD method. In particular, Theorem 4 contains the essential properties of the CQML estimator. We investigate their asymptotic null behaviors using these properties.

First, we show that the asymptotic null distribution of the QLR statistic can be obtained straightforwardly by combining Theorems 3 and 4. That is, we note that $\mathscr{QLR}_n = 2n\{\overline{L}_n(\widehat{\theta}_n) - \overline{L}_n(\widetilde{\theta}_n)\} = 2n\{\overline{L}_n(\widehat{\theta}_n) - \overline{L}_n(\theta_*)\} - 2\{L_n(\widetilde{\theta}_n) - L_n(\theta_*)\}$, and the asymptotic behaviors of the curly brackets are already given in Theorems 3 and 4. If we combine them,

$$
\begin{aligned}
\mathscr{QLR}_n &= 2n\{\overline{L}_n(\theta_* + \widehat{d}_n(\widehat{s}_n)\cdot\widehat{s}_n) - \overline{L}_n(\theta_* + \widehat{d}_n(\widetilde{s}_n)\cdot\widetilde{s}_n)\} \\
&\Rightarrow \sup_{s\in\mathbb{S}^{s-1}} \max[0,\mathscr{G}(s)]^2 - \sup_{s\in\mathbb{S}_o^{s-1}} \max[0,\mathscr{G}(s)]^2 = \mathscr{G}(\hat{\mathfrak{s}})^2 - \mathscr{G}(\widetilde{\mathfrak{s}})^2 \\
&= -\hat{\mathfrak{s}}'\mathbf{A}\hat{\mathfrak{s}}\widehat{\delta}(\hat{\mathfrak{s}})^2 + \widetilde{\mathfrak{s}}'\mathbf{A}\widetilde{\mathfrak{s}}\widehat{\delta}(\widetilde{\mathfrak{s}})^2 = -\hat{\mathfrak{s}}'\mathbf{A}\hat{\mathfrak{s}}\mathscr{Y}(\hat{\mathfrak{s}})^2 + \widetilde{\mathfrak{s}}'\mathbf{A}\widetilde{\mathfrak{s}}\mathscr{Y}(\widetilde{\mathfrak{s}})^2 \\
&= -\mathscr{Z}'\mathbf{A}^{-1}\mathscr{Z} + \mathscr{Z}'[\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{R}'(\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')\mathbf{R}\mathbf{A}^{-1}]\mathscr{Z} \\
&= -\mathscr{Z}'\mathbf{A}^{-1}\mathbf{R}'(\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')^{-1}\mathbf{R}\mathbf{A}^{-1}\mathscr{Z} \\
&\sim N_2(0, \mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'; -[\mathbf{R}\mathbf{A}\mathbf{R}']^{-1}).
\end{aligned}
$$

That is, the asymptotic null distribution of the QLR test can be represented by the asymptotic behaviors of the distance and direction estimators $(\mathscr{Y}(\cdot), \hat{\mathfrak{s}}, \widetilde{\mathfrak{s}})$. The final term is the result obtained by combining it with Theorem 2(i).

We can view this in a different angle from the conventional approach. As $\hat{\mathfrak{s}}$ and $\widetilde{\mathfrak{s}}$ are selected to maximize $\max[0,\mathscr{G}(s)]^2$ from $\mathbb{S}^{s-1}$ and $\mathbb{S}_o^{s-1}$ respectively, the asymptotic null distribution of the QLR test is given as $N_2(0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}; -(\mathbf{R}\ \mathbf{A}\mathbf{R}')^{-1})$. If other different rules apply, the same distribution is not necessarily obtained as we can see from the occasions of the Wald and LM tests.

We next examine the Wald test. The asymptotic null behavior of the Wald test is also easily obtained by Theorems 3 and 4. We note that $\sqrt{n}(\widehat{\theta}_n - \widetilde{\theta}_n) = \sqrt{n}(\widehat{\theta}_n - \theta_*) - \sqrt{n}(\widehat{\theta}_n - \theta_*) = \sqrt{n}[\widehat{d}_n(\widehat{s}_n)\cdot\widehat{s}_n - \widehat{d}_n(\widetilde{s}_n)\cdot\widetilde{s}_n]$, so that $\sqrt{n}(\mathbf{R}\widehat{\theta}_n - \mathbf{R}\widetilde{\theta}_n) = \sqrt{n}(\widehat{d}_n(\widehat{s}_n)\mathbf{R}\cdot\widehat{s}_n - \widehat{d}_n(\widetilde{s}_n)\mathbf{R}\cdot\widetilde{s}_n) = \sqrt{n}\widehat{d}_n(\widehat{s}_n)\mathbf{R}\cdot\widehat{s}_n$, where the last equality holds by the fact that $\mathbf{R}\cdot\widetilde{s}_n \equiv 0$ given by the null condition $\mathscr{N}_o$. This also implies that $\sqrt{n}(\mathbf{R}\widehat{\theta}_n - \mathbf{r}) = \sqrt{n}\widehat{d}_n(\widehat{s}_n)\mathbf{R}\cdot\widehat{s}_n \Rightarrow \widehat{\delta}(\hat{\mathfrak{s}})\mathbf{R}\cdot\hat{\mathfrak{s}} = \mathbf{R}\mathscr{Y}(\hat{\mathfrak{s}})\cdot\hat{\mathfrak{s}} \sim N(0, \mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\ \mathbf{R}')$ under $\mathscr{H}_0$ by Theorem 3(i). Thus, if Assumption 2 and $\mathscr{H}_0$ further hold, $\mathscr{W}_n = n\widehat{d}_n(\widehat{s}_n)\widehat{s}_n'\mathbf{R}'[\mathbf{R}\widehat{\mathbf{A}}_n^{-1}\widehat{\mathbf{B}}_n\widehat{\mathbf{A}}_n^{-1}\mathbf{R}']^{-1}\mathbf{R}\widehat{s}_n\widehat{d}_n(\widehat{s}_n) \Rightarrow \mathscr{Y}(\hat{\mathfrak{s}})\hat{\mathfrak{s}}'\mathbf{R}'(\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}')^{-1}\ \mathbf{R}\hat{\mathfrak{s}}\mathscr{Y}(\hat{\mathfrak{s}}) \sim \mathscr{X}_q^2$. That is, we can represent the asymptotic null distribution of the Wald test using the asymptotic null distributions of $(\mathscr{Y}(\cdot), \hat{\mathfrak{s}})$.

As for the QLR test, we can endow the asymptotic null behavior with a different perspective. When $\mathscr{Y}(\cdot)(\cdot)'\mathbf{R}'[\mathbf{R}\mathbf{A}^{-1}\ \mathbf{B}\mathbf{A}^{-1}\mathbf{R}']^{-1}\mathbf{R}(\cdot)\mathscr{Y}(\cdot)$ is viewed as a chi-square process, the conventional Wald test chooses $\hat{\mathfrak{s}}$ so that the asymptotic null distribution of $\mathscr{Y}(\hat{\mathfrak{s}})\hat{\mathfrak{s}}'\mathbf{R}'[\mathbf{R}\ \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}']^{-1}\mathbf{R}\hat{\mathfrak{s}}\mathscr{Y}(\hat{\mathfrak{s}})$ is a chi-square distribution. If a different s is selected, the same chi-square distribution may not be obtained.

The LM test can be also analyzed in a similar manner. In particular, its asymptotic null behavior turns out to have the same structure as the Wald test. For this, we consider another direction estimator $\acute{s}_n$ slightly different from both $\widehat{s}_n$ and $\widetilde{s}_n$. We let

$$\acute{s}_n := \acute{c}_n\{-\nabla^2_\theta \overline{L}_n(\widetilde{\theta})\}^{-1}\sqrt{n}\nabla_\theta \overline{L}_n(\widetilde{\theta}_n), \tag{17}$$

where $\acute{c}_n := \{n\nabla'_\theta \overline{L}_n(\widetilde{\theta}_n)[-\nabla^2_\theta \overline{L}_n(\widetilde{\theta}_n)]^{-1}[-\nabla^2_\theta \overline{L}_n(\widetilde{\theta}_n)]^{-1}\nabla_\theta \overline{L}_n(\widetilde{\theta}_n)\}^{-1/2} > 0$. In particular, note that $\acute{s}_n$ is defined following the structure of $\hat{\mathfrak{s}}$. Here, the last inequality holds because $[-\nabla^2_\theta \overline{L}_n(\widetilde{\theta}_n)]^{-1}[-\nabla^2_\theta \overline{L}_n(\widetilde{\theta}_n)]^{-1}$ is positive definite by Assumption 1(*ix*). Also, $\acute{c}_n$ is defined to have $\acute{s}'_n\acute{s}_n = 1$. For each s, we further let $\acute{d}_n(s) := -\{s'\nabla^2_\theta \overline{L}_n(\widetilde{\theta}_n)s\}^{-1}\max[0, \nabla'_\theta \overline{L}_n(\widetilde{\theta}_n)s]$ by following the structure of $\dot{d}_n(s)$. The only difference between $\dot{d}_n(s)$ and $\acute{d}_n(s)$ is that the unknown parameter θ_* is now replaced by $\widetilde{\theta}_n$. All these are defined to denote the LM test by the distance and direction estimators. That is,

$$\mathscr{L}\mathscr{M}_n = n\acute{d}_n(\acute{s}_n)\acute{s}'_n\mathbf{R}'\{\mathbf{R}\widetilde{\mathbf{A}}_n^{-1}\widetilde{\mathbf{B}}_n\widetilde{\mathbf{A}}_n^{-1}\mathbf{R}'\}^{-1}\mathbf{R}\acute{s}_n\acute{d}_n(\acute{s}_n), \tag{18}$$

and that

$$\acute{d}_n(\acute{s}_n) = -\frac{\max\left[0, \acute{c}_n\nabla'_\theta \overline{L}_n(\widetilde{\theta})[-\nabla^2_\theta \overline{L}_n(\widetilde{\theta}_n)]^{-1}\nabla_\theta \overline{L}_n(\widetilde{\theta})\right]}{\acute{c}_n^2\nabla'_\theta \overline{L}_n(\widetilde{\theta})[-\nabla^2_\theta \overline{L}_n(\widetilde{\theta}_n)]^{-1}\nabla_\theta \overline{L}_n(\widetilde{\theta})} = \acute{c}_n^{-1},$$

where the last equality holds by Assumption 1(*ix*). Note that the distance and direction estimators $(\acute{d}_n(\cdot), \acute{s}_n)$ are not necessarily identical to the previous estimators $(\widehat{d}_n(\cdot), \widehat{s}_n)$ or $(\widehat{d}_n(\cdot), \widetilde{s}_n)$. Nevertheless, their asymptotic behaviors are interrelated under $\mathscr{H}_0$. We verify this by examining the key components constituting $\acute{s}_n$:

$$\nabla_\theta \overline{L}_n(\widetilde{\theta}_n) = \nabla_\theta \overline{L}_n(\theta_* + \widehat{d}_n(\widetilde{s}_n)\cdot\widetilde{s}_n) = \nabla_\theta \overline{L}_n(\theta_*) + \mathbf{A}\widehat{d}_n(\widetilde{s}_n)\cdot\widetilde{s}_n + o_{\mathbb{P}}(\sqrt{n}) \tag{19}$$

under $\mathscr{H}_0$. Here, $\mathbf{A}$ is the asymptotic limit of $\nabla^2_\theta \overline{L}_n(\theta_*)$. Also, $\nabla_\theta \overline{L}_n(\theta_*) = -\mathbf{A}\widehat{d}_n(\widehat{s}_n)\cdot\widehat{s}_n + o_{\mathbb{P}}(\sqrt{n})$ and $\widehat{d}_n(\widetilde{s}_n)\cdot\widetilde{s}_n = \mathbf{D}\widehat{d}_n(\widehat{s}_n)\cdot\widehat{s}_n + o_{\mathbb{P}}(\sqrt{n})$ by constructing the sample analogs of eqs. (11) and (16), respectively. Now, plugging these into eq. (19) yields that $\nabla_\theta \overline{L}_n(\widetilde{\theta}_n) = \mathbf{A}[-\mathbf{I}+\mathbf{D}]\widehat{d}_n(\widehat{s}_n)\cdot\widehat{s}_n + o_{\mathbb{P}}(\sqrt{n}) = -\widehat{d}_n(\widehat{s}_n)\mathbf{R}'(\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')^{-1}\mathbf{R}\cdot\widehat{s}_n + o_{\mathbb{P}}(\sqrt{n})$, where the last equality follows by the definition of $\mathbf{D} := \mathbf{I}_s - \mathbf{A}^{-1}\mathbf{R}'(\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')^{-1}\mathbf{R}$. Therefore, it now follows that $\acute{s}_n = \acute{c}_n\widehat{d}_n(\widehat{s}_n)\mathbf{A}^{-1}\mathbf{R}'(\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')^{-1}\mathbf{R}\cdot\widehat{s}_n + o_{\mathbb{P}}(\sqrt{n})$, and letting n tend to infinity after plugging this into eq. (18) yields that $\mathscr{LM}_n \Rightarrow \mathscr{Y}(\hat{\mathfrak{s}})\hat{\mathfrak{s}}'\mathbf{R}'(\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}')^{-1}\mathbf{R}\hat{\mathfrak{s}}\mathscr{Y}(\hat{\mathfrak{s}}) \sim \mathscr{X}_q^2$ under $\mathscr{H}_0$. This consequence is identical to the Wald test. That is, the asymptotic distribution is now represented by the asymptotic behaviors of distance and direction estimators $(\mathscr{Y}(\cdot),\hat{\mathfrak{s}})$, and they are selected to have a test statistic asymptotically following $\mathscr{X}_q^2$ under $\mathscr{H}_0$. We can therefore say that the LM test is a different test yielding the same asymptotic null distribution as the Wald test. We formally state our derivations in the following theorem.

Theorem 5. *Given Assumptions 1, 2, and* $\mathscr{H}_0$*,*

(i) $\mathscr{QLR}_n \Rightarrow -\hat{\mathfrak{s}}'\mathbf{A}\hat{\mathfrak{s}}\mathscr{Y}(\hat{\mathfrak{s}})^2 + \widetilde{\mathfrak{s}}'\mathbf{A}\widetilde{\mathfrak{s}}\mathscr{Y}(\widetilde{\mathfrak{s}})^2$*, which follows* $N_2(0, \mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'; -(\mathbf{R}\mathbf{A}\mathbf{R}')^{-1})$*;*

(ii) $\mathscr{W}_n \Rightarrow \mathscr{Y}(\hat{\mathfrak{s}})\hat{\mathfrak{s}}'\mathbf{R}'(\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}')^{-1}\mathbf{R}\hat{\mathfrak{s}}\mathscr{Y}(\hat{\mathfrak{s}}) \sim \mathscr{X}_q^2$*; and*

(iii) $\mathscr{LM}_n \Rightarrow \mathscr{Y}(\hat{\mathfrak{s}})\hat{\mathfrak{s}}'\mathbf{R}'(\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}')^{-1}\mathbf{R}\hat{\mathfrak{s}}\mathscr{Y}(\hat{\mathfrak{s}}) \sim \mathscr{X}_q^2$*, where we let* $\widetilde{\mathfrak{s}} := \arg\max_{s\in\mathbb{S}_o^{s-1}} \max[0,\mathscr{G}(s)]^2$.

The biggest message of Theorem 5 is that for a Gaussian process $\mathscr{Y}(\cdot)$ with the covariance eq. (10), QLR, Wald, and LM tests choose $(\hat{\mathfrak{s}},\widetilde{\mathfrak{s}})$ or $\hat{\mathfrak{s}}$ so that they have the standard null distributions, which we can easily tabulate.

5. NEW TESTS AND REINVESTIGATION OF THE TESTS

We showed in the previous section that the asymptotic null distributions of the three most common tests can be represented by the asymptotic behaviors of the distance and direction estimators $(\mathscr{Y}(\cdot),\hat{\mathfrak{s}})$ or $(\mathscr{Y}(\cdot),\hat{\mathfrak{s}},\widetilde{\mathfrak{s}})$. Although this interpretation is useful and insightful, its application is practically infeasible, as the sample version of $(\mathscr{Y}(\cdot),\hat{\mathfrak{s}},\widetilde{\mathfrak{s}})$ is defined by assuming that θ_* is known. That is, $(\widehat{d}_n(\cdot),\widehat{s}_n,\widetilde{s}_n)$ cannot be defined by unknown θ_*, and in this sense we may call it an *infeasible DD estimator*. In this section, we modify our reference parameter into $\widetilde{\theta}_n$ and call the DD estimator defined by this as a *feasible DD estimator*. As the reference parameter is now known, the relevant DD estimator can be obtained without difficulty. We reinforce the relevant theory in this section.

We analyze the QML estimator by the DD method with our reference parameter being the CQML estimator. For this, we consider the following maximization processes: for each $s \in \mathbb{S}^{s-1}$, we let

$$\ddot{d}_n(s) := \arg\max_{d\in\ddot{\mathbb{D}}(s)} \overline{L}_n(\widetilde{\theta}_n + d\cdot s), \quad \text{and} \quad \ddot{s}_n := \arg\max_{s\in\mathbb{S}^{s-1}} \overline{L}_n(\widetilde{\theta}_n + \ddot{d}_n(s)\cdot s), \quad (20)$$

where $\ddot{\mathbb{D}}(s) := \{d \in \mathbb{R}^+ : \widetilde{\theta}_n + ds \in \Theta\}$. Note that $(\ddot{d}_n(\cdot), \ddot{s}_n)$ is defined in a parallel manner to $(\widehat{d}_n(\cdot), \widehat{s}_n)$. The only difference is that the reference parameter is now modified from θ_* to $\widetilde{\theta}_n$, and it trivially follows that

$$\widehat{\theta}_n - \widetilde{\theta}_n \equiv \ddot{d}_n(\ddot{s}_n) \cdot \ddot{s}_n \tag{21}$$

a.s.$-\mathbb{P}$ by their definitions. Thus, $(\ddot{d}_n(\cdot), \ddot{s}_n)$ and $(\widehat{d}_n(\cdot), \widehat{s}_n)$ are not identical estimators unless $\widetilde{\theta}_n = \theta_*$.

Nevertheless, their analyses are almost identical. We show specifically how to obtain the asymptotic behaviors of $(\ddot{d}_n(\cdot), \ddot{s}_n)$. First, for given $s \in \mathbb{S}^{s-1}$, we apply the mean-value theorem as in eq. (2): for some $\bar{d}$ between 0 and d,

$$\overline{L}_n(\widetilde{\theta}_n + d \cdot s) = \overline{L}_n(\widetilde{\theta}_n) + \nabla'_\theta \overline{L}_n(\widetilde{\theta}_n) s \cdot d + \frac{1}{2} s' \nabla^2_\theta \overline{L}_n(\widetilde{\theta}_n + \bar{d} \cdot s) s \cdot d^2. \tag{22}$$

This expansion also implies that for each $s \in \mathbb{S}^{s-1}$ and $\bar{d}_n(s)$ between 0 and $\ddot{d}_n(s)$,

$$\ddot{d}_n(s) = -\frac{\max[0, \nabla'_\theta \overline{L}_n(\widetilde{\theta}_n) s]}{s' \nabla^2_\theta \overline{L}_n(\widetilde{\theta}_n + \bar{d}_n(s) s) s}. \tag{23}$$

Second, we let $\ddot{\delta}(\cdot)$ capture the asymptotic behavior of $\sqrt{n}\ddot{d}_n(\cdot)$. Then,

$$\sup_{d \in \ddot{\mathbf{D}}(s)} n\{\overline{L}_n(\widetilde{\theta}_n + d \cdot s) - \overline{L}_n(\widetilde{\theta}_n)\} \Rightarrow \frac{1}{2}(-s'\mathbf{A}s)\ddot{\delta}(s)^2 = -\frac{1}{2}\left(\frac{\max[0, \ddot{\mathscr{Z}}' s]^2}{s'\mathbf{A}s}\right), \tag{24}$$

as n tends to infinity, where $\ddot{\mathbb{D}}(s) := \{d \in \mathbb{R}^+ : \widetilde{\theta}_n + ds \in \Theta\}$, and $\ddot{\mathscr{Z}}$ is the asymptotic weak limit of $\sqrt{n}\overline{L}_n(\widetilde{\theta}_n)$. That is, $\ddot{\mathscr{Z}} := \mathbf{P}\mathscr{Z}$ by Theorem 1 (*vi*). Here, we obtained the final equality by letting $\ddot{\delta}(s) := \max[0, \ddot{\mathscr{Y}}(s)]$ and $\ddot{\mathscr{Y}}(s) := \{-s'\mathbf{A}s\}^{-1}\ddot{\mathscr{Z}}' s$. As before, $\ddot{\mathscr{Y}}(\cdot)$ can be viewed as a Gaussian process: for each $s, \widetilde{s} \in \mathbb{S}^{s-1}$, $E[\ddot{\mathscr{Y}}(s)] = 0$ and

$$E[\ddot{\mathscr{Y}}(s)\ddot{\mathscr{Y}}(\widetilde{s})] = \frac{s'\mathbf{P}'\mathbf{B}\mathbf{P}\widetilde{s}}{\{-s'\mathbf{A}s\}\{-\widetilde{s}'\mathbf{A}\widetilde{s}\}}.$$

Third, we further apply the continuous mapping theorem to this asymptotic consequence:

$$\sup_{s \in \mathbb{S}^{s-1}} \sup_{d \in \ddot{\mathbf{D}}(s)} n\{\overline{L}_n(\widetilde{\theta}_n + d \cdot s) - \overline{L}_n(\widetilde{\theta}_n)\}$$
$$\Rightarrow \sup_{s \in \mathbb{S}^{s-1}} -\frac{1}{2}\left(\frac{\max[0, \ddot{\mathscr{Z}}' s]^2}{s'\mathbf{A}s}\right) = -\frac{1}{2}\left(\frac{\max[0, \ddot{\mathscr{Z}}' \ddot{\mathfrak{s}}]^2}{\ddot{\mathfrak{s}}'\mathbf{A}\ddot{\mathfrak{s}}}\right), \tag{25}$$

where we let $\ddot{\mathfrak{s}} := \arg\max_{s\in\mathbb{S}^{s-1}} \max[0, \ddot{\mathscr{G}}(s)]^2$ and $\ddot{\mathscr{G}}(s) := \{-s'\mathbf{A}s\}^{-1/2}\ddot{\mathscr{Z}}'s$. That is, $\ddot{\mathscr{G}}(\cdot)$ is a Gaussian process: for each $s, \widetilde{s} \in \mathbb{S}^{s-1}$, $E[\ddot{\mathscr{G}}(s)] = 0$ and

$$E[\ddot{\mathscr{G}}(s)\ddot{\mathscr{G}}(\widetilde{s})] = \frac{s'\mathbf{P}'\mathbf{B}\mathbf{P}\widetilde{s}}{\sqrt{-s'\mathbf{A}s}\sqrt{-\widetilde{s}'\mathbf{A}\widetilde{s}}}.$$

Fourth, we also note that

$$\max[0, \ddot{\mathscr{Z}}'\ddot{\mathfrak{s}}] = \ddot{\mathscr{Z}}'\ddot{\mathfrak{s}}, \tag{26}$$

because $n\{\overline{L}_n(\widehat{\theta}_n) - \overline{L}_n(\widetilde{\theta}_n)\}$ does have zero probability mass at zero by Theorem 2(i) and

$$\sup_{s\in\mathbb{S}^{s-1}} \sup_{d\in\ddot{\mathbf{D}}(s)} n\{\overline{L}_n(\widetilde{\theta}_n + d\cdot s) - \overline{L}_n(\widetilde{\theta}_n)\} = n\{\overline{L}_n(\widehat{\theta}_n) - \overline{L}_n(\widetilde{\theta}_n)\}. \tag{27}$$

Thus, $\ddot{\delta}(\ddot{\mathfrak{s}})\cdot\ddot{\mathfrak{s}} = \ddot{\mathscr{Y}}(\ddot{\mathfrak{s}})\cdot\ddot{\mathfrak{s}} = \{-\ddot{\mathfrak{s}}'\mathbf{A}\ddot{\mathfrak{s}}\}^{-1}\ddot{\mathscr{Z}}'\ddot{\mathfrak{s}}\cdot\ddot{\mathfrak{s}} = (-\mathbf{A})^{-1}\ddot{\mathscr{Z}}$, where the last equality holds by Theorem 1(v). Further, the last equality implies that $\ddot{\mathfrak{s}} = \ddot{\mathfrak{c}}(-\mathbf{A})^{-1}\ddot{\mathscr{Z}}$, where $\ddot{\mathfrak{c}}$ is such that $\ddot{\mathfrak{s}}'\ddot{\mathfrak{s}} = 1$. That is,

$$\ddot{\mathfrak{c}} = \{\ddot{\mathscr{Z}}'(-\mathbf{A})^{-1}(-\mathbf{A})^{-1}\ddot{\mathscr{Z}}\}^{-1/2} \quad \text{and} \quad \ddot{\delta}(\ddot{\mathfrak{s}}) = \ddot{\mathfrak{c}}^{-1}. \tag{28}$$

We summarize the essential results of these steps in the following theorem.

Theorem 6. *Given Assumption 1 and $\mathscr{H}_0$,*

(i) $\sqrt{n}(\widehat{\theta}_n - \widetilde{\theta}_n) \equiv \sqrt{n}\ddot{d}_n(\ddot{s}_n)\cdot\ddot{s}_n \Rightarrow \ddot{\delta}(\ddot{\mathfrak{s}})\cdot\ddot{\mathfrak{s}} = \ddot{\mathscr{Y}}(\ddot{\mathfrak{s}})\cdot\ddot{\mathfrak{s}} \sim N(0, \mathbf{A}^{-1}\mathbf{P}\mathbf{B}\mathbf{P}'\mathbf{A}^{-1})$*; and*

(ii) $\ddot{\mathfrak{s}} = \ddot{\mathfrak{c}}(-\mathbf{A})^{-1}\ddot{\mathscr{Z}}$ *and* $\ddot{\delta}(\ddot{\mathfrak{s}}) = \ddot{\mathfrak{c}}^{-1}$.

The steps above Theorem 6 can be used as intermediate steps to define new test statistics. Note that θ_* is not used here, and instead, $\widetilde{\theta}_n$ is used for Theorem 6, so that its application is practically feasible. We illustrate their applications here, and for this, we provide the following definitions: for each $s \in \mathbb{S}^{s-1}$, we let

$$\ddot{Y}_n(s) := \frac{\sqrt{n}\nabla'_\theta\overline{L}_n(\widetilde{\theta}_n)s}{\{-s'\nabla^2_\theta\overline{L}_n(\widetilde{\theta}_n)s\}} \quad \text{and} \quad \ddot{G}_n(s) := \frac{\sqrt{n}\nabla'_\theta\overline{L}_n(\widetilde{\theta}_n)s}{\{-s'\nabla^2_\theta\overline{L}_n(\widetilde{\theta}_n)s\}^{1/2}}.$$

Also, for each $s \in \mathbb{S}^{s-1}$, we define

$$\ddot{J}_n(s) := \{s'\mathbf{R}'\{\mathbf{R}\widehat{\mathbf{A}}_n^{-1}\widehat{\mathbf{B}}_n\widehat{\mathbf{A}}_n^{-1}\mathbf{R}'\}^{-1}\mathbf{R}s\}^{1/2}\ddot{Y}_n(s) \quad \text{and}$$

$$\widetilde{J}_n(s) := \{s'\mathbf{R}'\{\mathbf{R}\widetilde{\mathbf{A}}_n^{-1}\widetilde{\mathbf{B}}_n\widetilde{\mathbf{A}}_n^{-1}\mathbf{R}'\}^{-1}\mathbf{R}s\}^{1/2}\ddot{Y}_n(s).$$

Furthermore, we let $\mathscr{F}(\mathbb{S}^{s-1}) := \{f : \mathscr{C}(\mathbb{S}^{s-1}) \mapsto \mathbb{R} : f \text{ is continuous.}\}$, where $\mathscr{C}(A)$ is a space of continuous mappings defined on $\mathbb{S}^{s-1}$. This space is considered to provide a family of test statistics whose null behaviors are determined by the Gaussian processes defined above. Specifically, we let these families be $\mathscr{T}_{\ddot{G}} := \{f(\ddot{G}_n) : f : \mathscr{C}(\mathbb{S}^{s-1}) \mapsto \mathbb{R}\}$, $\mathscr{T}_{\ddot{J}} := \{f(\ddot{J}_n) : f : \mathscr{C}(\mathbb{S}^{s-1}) \mapsto \mathbb{R}\}$, and $\mathscr{T}_{\widetilde{J}} := \{f(\widetilde{J}_n) : f : \mathscr{C}(\mathbb{S}^{s-1}) \mapsto \mathbb{R}\}$, respectively. Their asymptotic null behaviors are trivially obtained as follows.

Theorem 7. *Given Assumptions 1, 2, and* $\mathscr{H}_0$,

(i) if $f(\ddot{G}_n) \in \mathscr{T}_{\ddot{G}}$, $f(\ddot{G}_n) \Rightarrow f(\ddot{\mathscr{G}})$;

(ii) if $f(\ddot{J}_n) \in \mathscr{T}_{\ddot{J}}$, $f(\ddot{J}_n) \Rightarrow f(\ddot{\mathscr{J}})$; *and*

(iii) if $f(\widetilde{J}_n) \in \mathscr{T}_{\widetilde{J}}$, $f(\widetilde{J}_n) \Rightarrow f(\ddot{\mathscr{J}})$, *where* $\ddot{\mathscr{J}}(\cdot)$ *is a Gaussian process such that for each* $s, \widetilde{s} \in \mathbb{S}^{s-1}$, $E[\ddot{\mathscr{J}}(s)] = 0$ *and*

$$E[\ddot{\mathscr{J}}(s)\ddot{\mathscr{J}}(\widetilde{s})] = \{s'\mathbf{R}'\{\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'\}^{-1}\mathbf{R}s\widetilde{s}'\mathbf{R}'\{\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'\}^{-1}\mathbf{R}\widetilde{s}\}^{1/2} \times \{-s'\mathbf{A}s\}^{-1}\{-\widetilde{s}'\mathbf{A}\widetilde{s}\}^{-1}s'\mathbf{P}'\mathbf{B}\mathbf{P}\widetilde{s}.$$

There are several remarks relevant to Theorem 7. First, many test statistics can be defined by Theorem 7 by combining functions in $\mathscr{F}(\mathbb{S}^{s-1})$ with the statistics tending to the Gaussian processes under the null. Second, for every $s \in \mathbb{S}^{s-1}$, $\ddot{\mathscr{J}}(s)$ can be also defined as $\{s'\mathbf{R}'\{\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'\}^{-1}\mathbf{R}s\}^{1/2}\ddot{\mathscr{Y}}(s)$. Note that the covariance structure of $\ddot{\mathscr{J}}(\cdot)$ is identical to $\{(\cdot)'\mathbf{R}'\{\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'\}^{-1}\mathbf{R}(\cdot)\}^{1/2}$ $\ddot{\mathscr{Y}}(\cdot)$. Third, the asymptotic null distributions of the new tests may not be so easy to obtain. This difficulty can be easily overcome by applying Hansen's (1996) weighted bootstrap. Finally, we can also understand the asymptotic null behaviors of the three most popular test statistics as a special case of Theorem 7. To understand this better, we first note that for each $s \in \mathbb{S}^{s-1}$, $\acute{d}_n(s) \equiv \max[0, n^{-1/2}\ddot{Y}_n(s)]$ and consider the following lemma.

Lemma 1. *Given Assumptions 1, 2, and* $\mathscr{H}_0$, $\|\ddot{s}_n - \acute{s}_n\| = o_{\mathbb{P}}(1)$.

The proof of Lemma 1 is provided in the Appendix. Using Theorem 7 and Lemma 1 simplifies the asymptotic null behaviors of the three most popular test statistics. We first consider them from the QLR test. We note that

$$
\begin{aligned}
\mathscr{QLR}_n &= 2n\{\bar{L}_n(\widetilde{\theta}_n + \ddot{d}_n(\ddot{s}_n)\cdot \ddot{s}_n) - \bar{L}_n(\widetilde{\theta}_n)\} \\
&= \sup_{s\in\mathbb{S}^{s-1}} 2n\{\bar{L}_n(\widetilde{\theta}_n + \ddot{d}_n(s)\cdot s) - \bar{L}_n(\widetilde{\theta}_n)\} \\
&= \sup_{s\in\mathbb{S}^{s-1}} -\frac{\max[0, \sqrt{n}\nabla_\theta \bar{L}_n(\widetilde{\theta}_n)s]^2}{s'\nabla^2_\theta \bar{L}_n(\widetilde{\theta}_n + \bar{d}_n(s)s)s} \\
&= \sup_{s\in\mathbb{S}^{s-1}} -\frac{\max[0, \sqrt{n}\nabla_\theta \bar{L}_n(\widetilde{\theta}_n)s]^2}{s'\nabla^2_\theta \bar{L}_n(\widetilde{\theta}_n)s} + o_{\mathbb{P}}(1) \\
&= \sup_{s\in\mathbb{S}^{s-1}} \max[0, \ddot{G}_n(s)]^2 + o_{\mathbb{P}}(1),
\end{aligned}
$$

where the second equality holds by the definition of $\ddot{s}_n$, and the third equality follows by combining eqs. (22) and (23). Also, the fourth equality holds by noting that $\sup_{s\in\mathbb{S}^{s-1}} |\bar{d}_n(s)| = o_{\mathbb{P}}(1)$ under $\mathscr{H}_0$, and from this we can conclude that $\mathscr{QLR}_n$ is a test statistic constructed by $f(\cdot)$ in Theorem 7 being $\sup_{s\in\mathbb{S}^{s-1}} \max[0,(\cdot)]^2$ under $\mathscr{H}_0$.

Furthermore, we can also deduce the regular asymptotic interrelationship between distance and direction estimators. Theorems 5 and 7 imply that $\mathscr{QLR}_n = \sup_{s\in\mathbb{S}^{s-1}} \max[0,\ddot{G}_n(s)]^2 + o_{\mathbb{P}}(1) \Rightarrow \sup_{s\in\mathbb{S}^{s-1}} \max[0,\ddot{\mathscr{G}}(s)]^2 = \ddot{\mathscr{G}}(\ddot{\mathfrak{s}})^2 = -\ddot{\mathfrak{s}}'\mathbf{A}\ddot{\mathfrak{s}}$ $\ddot{\delta}(\ddot{\mathfrak{s}})^2 = \ddot{\mathscr{Z}}'(-\mathbf{A})^{-1}\ddot{\mathscr{Z}}$ under $\mathscr{H}_0$, where the third last equality holds by eq. (26), and the last equality holds by eq. (28) and the fact that $\ddot{\mathfrak{s}} = \ddot{\mathfrak{c}}(-\mathbf{A})^{-1}\ddot{\mathscr{Z}}$. From this, Theorem 5($i$) now implies that

$$
\ddot{\mathfrak{s}}'\mathbf{A}\ddot{\mathfrak{s}}\ddot{\mathscr{Y}}(\ddot{\mathfrak{s}})^2 + \widetilde{\mathfrak{s}}'\mathbf{A}\widetilde{\mathfrak{s}}\mathscr{Y}(\widetilde{\mathfrak{s}})^2 \stackrel{d}{=} \hat{\mathfrak{s}}'\mathbf{A}\hat{\mathfrak{s}}\mathscr{Y}(\hat{\mathfrak{s}})^2, \tag{29}
$$

which forms an equality similar to the Pythagorean equation. Note that $\mathscr{Y}(\hat{\mathfrak{s}})$, $\mathscr{Y}(\widetilde{\mathfrak{s}})$, and $\ddot{\mathscr{Y}}(\ddot{\mathfrak{s}})$ capture the distribution of distances between $\widehat{\theta}_n$ and θ_*; between $\widetilde{\theta}_n$ and θ_*; and between $\widehat{\theta}_n$ and $\widetilde{\theta}_n$, respectively. There is no regular interrelationship among these distances unless additional adjustments are made. Eq. (29) shows that a regular relationship is established in the form of Pythagoras equality if their coefficient is adjusted according to those given in eq. (29).

We can also identify f of Theorem 7 constituting the Wald test without difficulties. We note the fact that $(\mathbf{R}\widehat{\theta}_n - \mathbf{r}) = \mathbf{R}(\widehat{\theta}_n - \widetilde{\theta}_n) = \ddot{d}_n(\ddot{s}_n)\mathbf{R}\ddot{s}_n = n^{-1/2}$ $\max[0,\ddot{Y}_n(\ddot{s}_n)]\mathbf{R}\ddot{s}_n$ under $\mathscr{H}_0$, so that we can also rewrite the Wald test statistic into $\mathscr{W}_n := n(\mathbf{R}\widehat{\theta}_n - \mathbf{r})'\{\mathbf{R}\widehat{\mathbf{A}}_n^{-1}\widehat{\mathbf{B}}_n\widehat{\mathbf{A}}_n^{-1}\mathbf{R}'\}^{-1}(\mathbf{R}\widehat{\theta}_n - \mathbf{r}) = \max[0,\ddot{Y}_n(\ddot{s}_n)]^2\ddot{s}_n'\mathbf{R}'\{\mathbf{R}$ $\widehat{\mathbf{A}}_n^{-1}\widehat{\mathbf{B}}_n\widehat{\mathbf{A}}_n^{-1}\mathbf{R}'\}^{-1}\mathbf{R}\ddot{s}_n = \max[0,\ddot{J}_n(\ddot{s}_n)]^2$. Thus, $\mathscr{W}_n$ can be also understood as a test statistic constructed by function f in Theorem 7(ii) being $\max[0,(\cdot)]^2$ such that s is selected by (20). This is asymptotically equivalent to the argument maximizing $\max[0, \ddot{G}_n(\cdot)]^2$ under $\mathscr{H}_0$. Here, we note that the Wald test chooses s in a different way from the QLR test. More specifically, instead of choosing s maximizing the function of interests $\max[0, \ddot{J}_n(\cdot)]^2$, the Wald test chooses s maximizing $\max[0, \ddot{G}_n(\cdot)]^2$ and plugs this back into $\max[0, \ddot{J}_n(\cdot)]^2$. This is designed to have a trivial null distribution. If we apply Theorems 5, 6, and 7(*ii*), $\mathscr{W}_n \Rightarrow \max[0, \ddot{\mathscr{J}}(\ddot{\mathfrak{s}})]^2 = \max[0, \ddot{\mathscr{Y}}(\ddot{\mathfrak{s}})]^2 \ddot{\mathfrak{s}}' \mathbf{R}' \{\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'\}^{-1} \mathbf{R}\ddot{\mathfrak{s}} \sim \mathscr{X}_q^2$ under $\mathscr{H}_0$, and this distribution is easy to tabulate. We further note that

$$\max[0, \mathscr{Y}(\hat{\mathfrak{s}})]^2 \hat{\mathfrak{s}}' \mathbf{R}' \{\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'\}^{-1} \mathbf{R}\hat{\mathfrak{s}} \stackrel{d}{=} \max[0, \ddot{\mathscr{Y}}(\ddot{\mathfrak{s}})]^2 \ddot{\mathfrak{s}}' \mathbf{R}' \{\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'\}^{-1} \mathbf{R}\ddot{\mathfrak{s}}$$

by Theorem 5(*ii*), and this implies that $\max[0, \mathscr{Y}(\hat{\mathfrak{s}})]\mathbf{R}\hat{\mathfrak{s}} \stackrel{d}{=} \max[0, \ddot{\mathscr{Y}}(\ddot{\mathfrak{s}})]\mathbf{R}\ddot{\mathfrak{s}}$, given that they both have the same weight matrix in the middle and that they follow the same chi-square distribution.

Finally, the LM test can be understood in a similar way. For examining this, note that $\mathscr{L}\mathscr{M}_n = n\acute{d}_n(\acute{s}_n)\acute{s}_n' \mathbf{R}' \{\mathbf{R}\widetilde{\mathbf{A}}_n^{-1}\widetilde{\mathbf{B}}_n\widetilde{\mathbf{A}}_n^{-1}\mathbf{R}'\}^{-1} \mathbf{R}\acute{s}_n \acute{d}_n(\acute{s}_n) = \max[0, \ddot{Y}_n(\acute{s}_n)]^2$ $\acute{s}_n' \mathbf{R}' \{\mathbf{R}\widetilde{\mathbf{A}}_n^{-1}\widetilde{\mathbf{B}}_n \ \widetilde{\mathbf{A}}_n^{-1}\mathbf{R}'\}^{-1} \mathbf{R}\acute{s}_n = \max[0, \widetilde{J}_n(\acute{s}_n)]^2$ under $\mathscr{H}_0$. Thus, we can also view $\mathscr{L}\mathscr{M}_n$ as a test statistic constructed by function $f(\cdot)$ in Theorem 7(*iii*) being $\max[0, (\cdot)]^2$ such that s is selected according to eq. (17). This is a different function from the Wald test, because $\ddot{J}_n(\cdot)$ is different from $\widetilde{J}_n(\cdot)$ and s is selected by eq. (17). Nevertheless, their difference is asymptotically negligible. That is, $\ddot{J}_n(\cdot)$ and $\widetilde{J}_n(\cdot)$ are asymptotically identical functions, because both $(\widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n)$ and $(\widetilde{\mathbf{A}}_n, \widetilde{\mathbf{B}}_n)$ estimate $(\mathbf{A}, \mathbf{B})$ consistently by Assumption 2, so that $\sup_{s \in \mathbb{S}^{s-1}} |\ddot{J}_n(s) - \widetilde{J}_n(s)| = o_{\mathbb{P}}(1)$. Thus, it follows that $\mathscr{L}\mathscr{M}_n = \max[0, \ddot{J}_n(\acute{s}_n)]^2 + o_{\mathbb{P}}(1)$. Further, Lemma 1 implies that $\|\acute{s}_n - \ddot{s}_n\| = o_{\mathbb{P}}(1)$, so that $\mathscr{L}\mathscr{M}_n = \max[0, \ddot{J}_n(\ddot{s}_n)]^2 + o_{\mathbb{P}}(1)$. That is, $\mathscr{L}\mathscr{M}_n = \mathscr{W}_n + o_{\mathbb{P}}(1)$. This implies that both Wald and LM tests asymptotically use the same function for Theorem 7, although their finite versions are different. Therefore, it follows that $\mathscr{L}\mathscr{M}_n \Rightarrow \max[0, \ddot{\mathscr{Y}}(\ddot{\mathfrak{s}})]^2 \ddot{\mathfrak{s}}' \mathbf{R}' \{\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'\}^{-1} \mathbf{R}\ddot{\mathfrak{s}} \sim \mathscr{X}_q^2$ under $\mathscr{H}_0$. We summarize these results formally as follows.

Theorem 8. *Given Assumptions 1, 2, and* $\mathscr{H}_0$,

(i) $\mathscr{Q}\mathscr{L}\mathscr{R}_n \Rightarrow \sup_{s \in \mathbb{S}^{s-1}} \max[0, \ddot{\mathscr{G}}(s)]^2 = \ddot{\mathscr{G}}(\ddot{s})^2$, *which follows* $N_2(0, \mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'; -(\mathbf{R}\mathbf{A}^{-1}\mathbf{R}')^{-1})$;

(ii) $\mathscr{W}_n \Rightarrow \max[0, \ddot{\mathscr{J}}(\ddot{\mathfrak{s}})]^2 = \max[0, \ddot{\mathscr{Y}}(\ddot{\mathfrak{s}})]^2 \ddot{\mathfrak{s}}' \mathbf{R}' \{\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'\}^{-1} \mathbf{R}\ddot{\mathfrak{s}} \sim \mathscr{X}_q^2$; *and*

(iii) $\mathscr{L}\mathscr{M}_n \Rightarrow \max[0, \ddot{\mathscr{J}}(\ddot{\mathfrak{s}})]^2 = \max[0, \ddot{\mathscr{Y}}(\ddot{\mathfrak{s}})]^2 \ddot{\mathfrak{s}}' \mathbf{R}' \{\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'\}^{-1} \mathbf{R}\ddot{\mathfrak{s}} \sim \mathscr{X}_q^2$.

Furthermore,

(iv) $\ddot{\mathfrak{s}}'\mathbf{A}\ddot{\mathfrak{s}}\ddot{\mathscr{Y}}(\ddot{\mathfrak{s}})^2 + \widetilde{\mathfrak{s}}'\mathbf{A}\widetilde{\mathfrak{s}}\mathscr{Y}(\widetilde{\mathfrak{s}})^2 \overset{d}{=} \hat{\mathfrak{s}}'\mathbf{A}\hat{\mathfrak{s}}\mathscr{Y}(\hat{\mathfrak{s}})^2$*; and*

(v) $\max[0, \mathscr{Y}(\hat{\mathfrak{s}})]\mathbf{R}\hat{\mathfrak{s}} \overset{d}{=} \max[0, \ddot{\mathscr{Y}}(\ddot{\mathfrak{s}})]\mathbf{R}\ddot{\mathfrak{s}}$.

Theorem 8 now implies that the asymptotic null distributions of the three most popular tests can be obtained by Theorem 7, as they are one of the tests in $\mathscr{T}_{\ddot{G}}$, $\mathscr{T}_{\ddot{J}}$, and $\mathscr{T}_{\widetilde{J}}$.

Given that our model is possibly misspecified, the QLR, Wald, and LM tests are not the most efficient tests. Thus, having classes of tests containing these test statistics as elements can be thought of as a natural extension for having better performing tests. The DD methods serves this purpose.

6. CONCLUSION

The current paper revisits the QML estimation and inference using the DD method introduced by Cho and White (2012). The asymptotic behaviors of the QML estimator is examined in a different way from the conventional approach, and also the QLR, Wald, and LM test statistics are reexamined in the framework of the DD method. Their null behaviors are specifically derived by the DD method, and we also explore the implications of the new framework posited by the DD method. Furthermore, this approach treats the most common three test statistics as special cases of new classes of tests, which we define using the DD method.

APPENDIX

Proof of Theorem 7: (*i*) We note that as a function of s, $\sqrt{n}\nabla'_\theta \overline{L}_n(\widetilde{\theta}_n)s \Rightarrow \ddot{\mathscr{Z}}'s$, which is linear with respect to s, and $\sup_{s\in\mathbb{S}^{s-1}} |s'\nabla^2_\theta \overline{L}_n(\widetilde{\theta})s - s'\mathbf{A}s| = o_{\mathbb{P}}(1)$ under $\mathscr{H}_0$. Thus, it follows that $\ddot{G}_n(\cdot) = \{-s'\nabla^2_\theta \overline{L}_n(\widetilde{\theta})s\}^{-1}\sqrt{n}\nabla'_\theta \overline{L}_n(\widetilde{\theta}_n)s \Rightarrow \ddot{\mathscr{G}}(\cdot)$ under $\mathscr{H}_0$. In particular, the tightness of $\{\sqrt{n}\nabla'_\theta \overline{L}_n(\widetilde{\theta}_n)s\}$ as a sequence of functions of s trivially holds because it is a sequence of linear functions. We now apply the continuous mapping theorem to obtain that $f(\ddot{G}_n) \Rightarrow f(\ddot{\mathscr{G}})$.

(*ii*) By the same reason as for $\ddot{G}_n(\cdot)$, $\ddot{Y}_n(\cdot) \Rightarrow \ddot{\mathscr{Y}}(\cdot)$. Further, $\widehat{\mathbf{A}}_n \overset{\text{a.s.}}{\to} \mathbf{A}$ and $\widehat{\mathbf{B}}_n \overset{\text{a.s.}}{\to} \mathbf{B}$ by Assumption 2. This implies that $\ddot{J}_n(\cdot) \Rightarrow \{(\cdot)'\mathbf{R}'\{\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{R}'\}^{-1}\mathbf{R}(\cdot)\}^{1/2}\ddot{\mathscr{Y}}(\cdot)$ by applying the continuous mapping theorem. For each $s \in \mathbb{S}^{s-1}$, we let $\mathscr{J}(s) := \{s'\mathbf{R}'\{\mathbf{R}\mathbf{A}^{-1}\mathbf{B}\ \mathbf{A}^{-1}\mathbf{R}'\}^{-1}\mathbf{R}s\}^{1/2}\ddot{\mathscr{Y}}(s)$ and apply the continuous mapping theorem to $\ddot{Y}_n(\cdot)$ again to obtain that $f(\ddot{Y}_n) \Rightarrow f(\ddot{\mathscr{Y}})$.

(iii) The only difference between $\ddot{J}_n(\cdot)$ and $\widetilde{J}_n(\cdot)$ is that $\widetilde{\mathbf{A}}_n$ and $\widetilde{\mathbf{B}}_n$ replace $\widehat{\mathbf{A}}_n$ and $\widehat{\mathbf{B}}_n$, respectively. Nevertheless, $\widetilde{\mathbf{A}}_n \stackrel{\text{a.s.}}{\to} \mathbf{A}$ and $\widetilde{\mathbf{B}}_n \stackrel{\text{a.s.}}{\to} \mathbf{B}$ by Assumption 2, so that we can claim the same property for $f(\widetilde{J}_n)$ as in (ii). This completes the proof. ■

Proof of Lemma 1: By eq. (17), $\acute{s}_n := \acute{c}_n\{-\nabla^2_\theta \overline{L}_n(\widetilde{\theta})\}^{-1}\sqrt{n}\nabla_\theta \overline{L}_n(\widetilde{\theta}_n)$, which is obtained as

$$\arg\max_{s\in\mathbb{S}^{s-1}} -\frac{\max[0, \sqrt{n}\nabla'_\theta \overline{L}_n(\widetilde{\theta}_n)s]^2}{2s'\nabla^2_\theta \overline{L}_n(\widetilde{\theta}_n)s}.$$

Also, using eqs. (22) and (23) implies that

$$\ddot{s}_n = \arg\max_{s\in\mathbb{S}^{s-1}} -\frac{\max[0, \sqrt{n}\nabla'_\theta \overline{L}_n(\widetilde{\theta}_n)s]^2}{2s'\nabla^2_\theta \overline{L}_n(\widetilde{\theta}_n + \bar{d}_n(s)s)s}.$$

Here, we note that for every $s \in \mathbb{S}^{s-1}$, $\bar{d}_n(s)$ is between 0 and $\ddot{d}_n(s)$, and $\sup_{s\in\mathbb{S}^{s-1}} \ddot{d}_n(s) = o_\mathbb{P}(1)$ from eq. (23). Thus, $\sup_{s\in\mathbb{S}^{s-1}} \bar{d}_n(s) = o_\mathbb{P}(1)$, so that

$$\ddot{s}_n = \arg\max_{s\in\mathbb{S}^{s-1}} -\frac{\max[0, \sqrt{n}\nabla'_\theta \overline{L}_n(\widetilde{\theta}_n)s]^2}{2s'\nabla^2_\theta \overline{L}_n(\widetilde{\theta}_n)s} + o_\mathbb{P}(1).$$

In other words, $\ddot{s}_n - \acute{s}_n = o_\mathbb{P}(1)$. The desired result follows from this. ■

REFERENCES

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

Andrews, D., 1987, Consistency in nonlinear econometric models: a generic uniform law of large numbers, Econometrica 55, 1465–1472.

Andrews, D., 1988, Laws of large numbers for dependent non-identically distributed random variables, Econometric Theory 5, 458–467.

Andrews, D., 1992, Generic uniform convergency, Econometric Theory 8, 1241–257.

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

Bierens, H., 1990, A consistent conditional moment test of functional form, Econometrica 58, 1443–1458.

Bierens, H. and Ploberger, W., 1997, Asymptotic theory of integrated conditional moment tests, Econometrica 65, 1129–1152.

Cho, J.S., Cheong, T., and White, H., 2011, Experiences with the weighted bootstrap in testing for unobserved heterogeneity in exponential and Weibull duration models, Journal of Economic Theory and Econometrics 22:2, 60–91.

Cho, J.S. and Han, C., 2009, Testing for the mixture hypothesis of geometric distributions, Journal of Economic Theory and Econometrics 20:3, 31–55.

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

Cho, J.S. and Ishida, I., 2010, Testing for the effects of omitted power transformations, Economics Letters, forthcoming.

Cho, J.S. and White, H., 2007, Testing for regime switching, Econometrica 75, 1671–1720.

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

Cho, J.S. and White, H., 2011a, Testing correct model specification using extreme leaning machines, Neurocomputing 74, 2552–2565.

Cho, J.S. and White, H., 2011b, Generalized runs tests for the IID hypothesis, Journal of Econometrics 162, 326–344.

Cho, J.S. and White, H., 2012, Directionally differentiable econometric models, Discussion Paper, School of Economcis, Yonsei University.

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.

Engle, R. and Russell, J., 1998, Autoregressive conditional duration: a new model for irregularly spaced transaction data, Ecoometrica 66, 1127–1162.

Gallant, R. and White, H., 1988, A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models, Basil Blackwell.

Gourieroux, C., Monfort, A, and Trognon, A., 1984, Pseudo-maximum likelihood method: theory, Econometrica 52, 681–700.

Hansen, B., 1996, Inference when a nuisance parameter is not identified under the null hypothesis, Econometrica 64, 413-430.

Hayashi, F., 2000, Econometrics, Princeton University Press.

King, M. and Shively, T., 1993, Locally optimal testing when a nuisance parameter is present only under the alternative, Review of Economics and Statistics 75, 1–7.

Levine, D., 1983, A remark on serial correlation in maximum likelihood, Journal of Econometrics 23, 337–342.

Stevenson, R., 1980, Likelihood functions for generalized stochastic frontier estimation, Journal of Econometrics 13, 57–66.

White, H., 1982, Maximum likelihood estimation of misspecified models, Econometrica 50, 1–25.

White, H., 1994, Estimation, Inference and Specification Analysis, Cambridge University Press.

White, H., 2000, Asymptotic Theory for Econometricians, Academic Press.