Directionally Differentiable Econometric Models

JIN SEO CHO
School of Economics
Yonsei University
50 Yonsei-ro, Seodaemun-gu, Seoul, 120-749, Korea
Email: jinseocho@yonsei.ac.kr

HALBERT WHITE
Department of Economics
University of California, San Diego
9500 Gilman Dr., La Jolla, CA, 92093-0508, U.S.A.
Email: hwhite@weber.ucsd.edu

First version: August 26, 2002. This version: December, 2012

Abstract

We relax the differentiability condition of standard econometric models and instead assume that models are directionally differentiable (Gâteaux differentiable). Using this new condition, we analyze the asymptotic distribution of the extremum estimator and show that its asymptotic distribution can be represented as a function of a Gaussian stochastic process indexed by direction. Our analysis also treats the analysis of differentiable models as a special case. Furthermore, we refine the standard quasi-likelihood ratio, Wald, and Lagrange multiplier test statistics under the directional differentiability condition and examine their asymptotic behaviors. We also interrelate them with the asymptotic behaviors of the standard test statistics obtained under the differentiability condition.

Key Words: Directionally (Gâteaux) differentiable model, Gaussian stochastic process, quasi-likelihood ratio test, Wald test, and Lagrange multiplier test.

JEL Classification: C12, C13, C22, C32.

Acknowledgements: The authors are benefited from discussions with Seung Chan Ahn, In Choi, Horag Choi, Robert Davies, Graham Elliott, Chirok Han, John Hillas, Jung Hur, Hide Ichimura, Yongho Jeon, Estate Khmaladze, Chang-Jin Kim, Chang Sik Kim, Tae-Hwan Kim, Naoto Kunitomo, Bruce Lehmann, Mark Machina, Jaesun Noh, Taeyoung Park, Donggyu Sul, Hung-Jen Wang, Yoshihiro Yajima, Arai Yoichi, Byung Sam Yoo and other participants at the seminars at Sogang University, the University of Auckland, the University of Tokyo, the Econometrics Study Group of the Korean Econometric Society, VUW, and Yonsei University.
1 Introduction

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

Nevertheless, many important econometric models are not differentiable (D). For example, King and Shively (1993) provide important non-differentiable econometric models. More specifically, they attempt to resolve the so-called Davies’ (1977, 1987) identification problem by re-parameterization. When unidentified nuisance parameters are present under the null, they re-parameterize the original models and identify them. These models, however, are only directionally differentiable (D-D), and analysis of them has not yet been available in the literature. As another example, Aigner, Lovell, and Schmidt (1977) specify the stochastic frontier production function models to capture inefficiently produced outputs. Nevertheless, the inference of efficient production using this model has not yet been examined in the literature due to its obscurity under the null of efficient production as we detail below.

The goal of this paper, therefore, is to extend the analysis tool for D models to the level of D-D models. For this, we exploit the fact that D-D models can be analyzed within the framework of the tightness in Billingsley (1999). Specifically, each direction around the parameter of interest can be regarded as an index indicating a particular value of directional derivatives. Tightness here governs the stochastic relationship of the directional derivatives in a way to apply the functional central limit theorem (FCLT) and the uniform law of large numbers (ULLN), so that D-D models can be analyzed in a similar manner to D models with another separate index.

Another goal of this paper is to demonstrate that the D-D model analysis provided here is applicable to numerous examples and deepens our model understanding. For this goal, we revisit some econometric models. As pointed out above, King and Shively’s (1993) re-parameterized models have not yet been fully analyzed. By the invariance principle, the likelihood obtained by the re-parameterization has to be identical to the likelihood obtained without the re-parameterization. This, on the other hand, implies that the analysis of re-parameterized models has to be different from that of standard econometric models. We aim to provide a suitable way to understand only D-D models so that we can achieve the dual goal to reconcile the invariance principle under the framework of the re-parameterization. We also apply our theory to the stochastic frontier production function models in Aigner, Lovell, and Schmidt (1977) and Box-Cox transformation. These models are hard to analyze using the conventional analysis assuming differentiability when testing particular
but important hypotheses as given below. We show that our analysis can clearly resolve this issue, and from this the applicability of our analysis enhances. Finally, we apply our analysis to standard D models and show that we can obtain results theoretically consistent with the conventional analysis when the GMM estimation is applied to D models. Specifically, we show that D-D model analysis reduces to the conventional analysis, so that it can be treated as a special case of D-D model analysis.

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

Before proceeding with our discussion, we introduce mathematical notation used throughout this paper. For any \( x \in \mathbb{R}^r, \|x\| \) stands for the Euclidean norm. \( \{\cdot\} \) and \( \text{cl}(A) \) stand for an indication function and a closure of set \( A \), respectively. Furthermore, we omit the argument of functions, so that, for example, \( f(\cdot) \) is also denoted as \( f \) unless confusion would otherwise arise.

## 2 Differentiable and Directionally Differentiable Models

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

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

Assumption 1 is a standard condition for stationary time-series data. Many economic data satisfy the given condition such as the standard ARMA process, hidden Markov processes, and GARCH processes.

The following models are examined in the current study, and we also provide further conditions for the consistence of the extremum estimator defined below.

**Assumption 2 (Model)** A sum of measurable functions \( \{L_n(\theta) := \sum_{t=1}^n \ell_t(\theta; X^t) : \theta \in \Theta\} \) is given as a model for \( X^n \) such that for each \( t, \ell_t(\cdot; X^t) \) is Lipschitz continuous on \( \Theta \) almost surely-a.e. (a.s.-a.e.), where
for each $t = 1, 2, \ldots, n$, $X^t$ denotes $(X_1, \ldots, X_t)$, and $\Theta$ is a compact and convex set in $\mathbb{R}^r$ with $r \in \mathbb{N}$. □

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

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

We use Assumptions 2 and 3 to have a consistent extreme estimator defined as $\hat{\theta}_n$. That is,

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

This is achieved by mainly applying the ULLN to $n^{-1}L_n$.

There are several comments relevant to these assumptions. Assumption 3(i) requires model identification. Even if models are not identified, we can still maintain model analysis using the framework of Davies (1977, 1987). Nevertheless, accommodating this makes the key aspects of D-D models obscure. Furthermore, unidentified models can be transformed into identified models using the re-parameterization method of King and Shivley (1993) as we detail below. Thus, here we pay attention only to identified models without loss of generality. Next, $\theta_*$ can be on the boundary of $\Theta$ as often ensued in the re-parameterization method of King and Shively (1993). Assumption 3(ii) is provided to permit this. Finally, we suppress $X^t$ in $\ell_t(\cdot; X^t)$ for notational simplicity, so that $\ell_t(\cdot; X^t)$ is abbreviated into $\ell_t$.

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

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

Theorem 1 is straightforward and well known in the literature (e.g., Andrews 1999). We, therefore, do not prove this in the Appendix. Nevertheless, Theorem 1 addresses further questions in terms of its asymptotic distribution.

2.1 Directionally Differentiable Models

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

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

$$Df(\theta; d) := \lim_{h \downarrow 0} \frac{f(\theta + hd) - f(\theta)}{h}.$$
exists in \( \mathbb{R} \), where \( \Delta(\theta) := \{ x \in \mathbb{R}^r : x + \theta \in \text{cl}\{C(\theta)\}, \|x\| = 1 \} \), and \( C(\theta) := \{ x \in \mathbb{R}^r : \exists \theta' \in \Theta, x := \theta + \delta\theta', \delta \in \mathbb{R}^+ \}; \)

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

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

We have several remarks relevant to this definition. We note that the definition of D-D functions is weaker than that of D functions because D functions are D-D although the converse of this statement does not hold. That is, D-D functions can have a different derivative if another direction is chosen. Furthermore, D-D functions can have a continuum number of different partial derivatives depending on \( d \) as there can be a continuum number of different directions if \( r \) is greater than one. One the other hand, if \( f \) is D, the direction \( d \) affects \( Df(\theta; d) \) up to the dimension of \( \Theta \) at most, so that if \( r \) number of different partial derivatives are available, for any direction \( d \), it is possible to represent the tangent hyperplane of \( f \) by their linear combination. As another remark, we note that \( Df(\theta; d) \) is defined on \( \Delta(\theta) \). This requirement is adopted to accommodate Chernoff’s (1954) analysis device using a cone. That is, when obtaining the asymptotic distribution of the extremum estimator, Chernoff (1954) notes that it is essential to approximate the parameter space by a cone \( C(\theta) \). We define \( \Delta(\theta) \) to collect only directions relevant to \( C(\theta) \), so that it can play the role of domain of a Gaussian stochastic process which we use for obtaining the asymptotic behavior of the extremum estimator below. Even if \( \theta \) is on the boundary of \( \Theta \), \( \Delta(\theta) \) is well defined because it does not contain the directions of boundary sides in \( \Delta(\theta) \) by the definition of \( C(\theta) \). Finally, \( \Delta(\theta) \) can be defined using a norm different from the Euclidean norm. As another alternative, \( \tilde{\Delta}(\theta) := \{ x \in \mathbb{R}^r : x + \theta \in \text{cl}\{C(\theta)\}, \|x\|_{\infty} = 1 \} \) can also capture the same directions as given by \( \Delta(\theta) \), where \( \| \cdot \|_{\infty} \) is the uniform norm. Nevertheless, we here focus on \( \Delta(\theta) \) for brevity.

The following theorem clarifies the interrelationship between D-D and D functions.

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

The proof of Theorem 2 is available in Troutman (1996). We note that the linearity condition of \( Df(\theta; d) \) in \( d \) is key to Theorem 2. Without this, any arbitrarily chosen directional derivative cannot be represented as a linear combination of other \( r \) directional derivatives. Before illustrating examples of D-D models, we provide the definition of continuously D-D functions, which also plays another key role in our analysis.

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

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

We note that the first-order directional differentiability (d-diffliity) is a necessary condition for defining twice continuously D-D functions. Furthermore, twice continuously D functions can be obtained from twice continuously D-D functions by imposing further conditions. Mainly, by imposing a quadratic function condition with respect to $d$, we can obtain twice continuously D functions. The following Lemma 1 formally states this.

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

### 2.2 Examples

We illustrate four D-D models: the conditional heteroskedasticity model in King and Shively (1993); the stochastic frontier production function models; the standard D model for GMM estimation; and the Box-Cox transformation.

#### 2.2.1 Example 1: Conditional Heteroskedasticity in Rosenberg (1973) and King and Shively (1993)

King and Shively (1993) examine a model for conditional heteroskedasticity. When a set of economic data $\{ (Y_t, Q'_t) : (Y_t, W_t, R'_t) \}$ is given, where $R_t$ is a $k$-dimensional vector of regressors and others are random variables defined on $\mathbb{R}$, they assume that the data obey a DGP given as

$$Y_n = W^n \alpha_\ast + R^n \beta_\ast + U^n,$$

$$U^n | Q^n \sim N(0, \sigma_\ast^2 \{ I_n + \kappa_\ast \Omega^n(\rho_\ast) \}),$$

where $Y_n := (Y_1, \ldots, Y_n)'$; $U^n := (U_1, \ldots, U_n)'$; $W^n := (W_1, \ldots, W_n)'$; $R^n$ is an $n \times k$ matrix with $R_t'$ at $t$-th row; $Q^n := (W^n, R^n)$; and $\Omega^n(\rho_\ast)$ is an $n \times n$ square matrix with $t$-th row and $t'$-th column element $\Omega^n_{tt'}(\rho_\ast) := W_t W_{t'} \rho_\ast^{t-t'}/(1 - \rho_\ast^2)$. Given this, they specify a model by assuming that $(\gamma_\ast, \sigma_\ast^2, \kappa_\ast, \rho_\ast) := (\alpha_\ast, \beta_\ast', \sigma_\ast^2, \kappa_\ast, \rho_\ast)$ is a unknown parameter in $\Gamma \times [0, \bar{\sigma}^2] \times [0, \bar{\kappa}] \times [0, \bar{\rho}]$, where $\Gamma$ 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 is less than one.
Therefore, for each \((\gamma, \sigma^2, \kappa, \rho)\), its log-likelihood can be written as

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

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

The motivation of this model stems from Rosenberg (1973), who aims to test \(\kappa_s = 0\) to examine whether a systematic risk of an asset is time-varying. If \(\kappa_s \neq 0\), the conditional covariance of \(U^n|Q^n\) depends on \(W^n\), so that testing the hypothesis is another way of testing conditional homoskedasticity.

Nevertheless, if \(\kappa_s = 0\), \(\rho_s\) is not identified, so that Davies’s (1977, 1987) identification problem arises; this renders the asymptotic null distributions of standard tests non-standard as they are often represented as functions of Gaussian stochastic processes. King and Shively (1993) attempt to resolve the unidentified parameter problem by re-parameterizing the original model carefully. That is, they let

\[
\theta_s' := (\theta_1, \theta_2) := (\kappa_s \cos(\rho_s \pi/2), \kappa_s \sin(\rho_s \pi/2))
\]

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

\[
U^n|Q^n \sim N[0, \sigma^2 \{I_n + (\theta_s')^2 \Omega^n(2 \tan^{-1}(\theta_2/\theta_1)/\pi)\}].
\]

Furthermore, the original hypotheses are rephrased into \(H'_0 : \theta_s' \theta_s = 0\) versus \(H'_1 : \theta_s' \theta_s > 0\) by the re-parameterization. As pointed out by King and Shively (1993), the identification problem does no longer arises under \(H'_0\).

Nevertheless, the model obtained by the re-parameterizations is only D-D under \(H'_0\), and the null parameter value is on the boundary. We note that for each \((\gamma, \sigma^2, \theta)\), the relevant log-likelihood is also modified into

\[
L_n(\gamma, \sigma^2, \theta) = -\frac{n}{2} \log (2\pi) - \frac{1}{2} \log (\det [\sigma^2 \{I_n + (\theta^\prime \theta)^{1/2} \Omega^n(2 \tan^{-1}(\theta_2/\theta_1)/\pi)\}])
\]

\[
- \frac{1}{2\sigma^2} U^n(\gamma)' [I_n + (\theta^\prime \theta)^{1/2} \Omega^n(2 \tan^{-1}(\theta_2/\theta_1)/\pi)]^{-1} U^n(\gamma),
\]

and from this, it also follows that

(1) \[
DL_n(\gamma_s, \sigma^2_s, \theta_s; d) = -\frac{nd_s^2}{2\sigma^2_s} - \frac{(d_1^2 + d_2^2)^{1/2}}{2} \text{tr}[\Omega^n (2 \tan^{-1}(d_2/d_1)/\pi)] + \frac{d_s^2}{2\sigma^2_s} U^n'U^n
\]

\[
+ \frac{1}{\sigma^2_s} (Q^n d_s)'U^n + \frac{(d_1^2 + d_2^2)^{1/2}}{2\sigma^2_s} U^n' \Omega^n (2 \tan^{-1}(d_2/d_1)/\pi) U^n,
\]

where \(d := (d_s', d_\sigma, d_1, d_2)'\) such that \(\theta_s = 0\) and \(d'd = 1\). This derivative is not linear with respect to \((d_1, d_2)\), although it is linear with respect to others. Thus, it is not a D model, and we cannot analyze it in...
the conventional way, assuming differentiability. The second-order directional derivative, which we provide here for future reference, also shares this aspect.

\[
D^2 L_n(\gamma_*, \sigma^2, \theta_*; d) = \frac{nd^2}{2\sigma^2} - \frac{d^2}{\sigma^2} \sigma^2 
\]

\[
= 2(\frac{d^2}{\sigma^2} 1 + \frac{d^2}{\sigma^2} 2) \frac{1}{2} \tr [\Omega^n (2 \tan^{-1}(d_2/d_1)/\pi)^2]
\]

\[
- \frac{d^2}{\sigma^2} 1 + \frac{d^2}{\sigma^2} 2 \frac{1}{2} \frac{d^2}{\sigma^2} \sigma^2 
\]

\[
+ \frac{d^2}{\sigma^2} 1 + \frac{d^2}{\sigma^2} 2 \frac{1}{2} \frac{d^2}{\sigma^2} \sigma^2 
\]

\[
- \frac{d^2}{\sigma^2} 1 + \frac{d^2}{\sigma^2} 2 \frac{1}{2} \frac{d^2}{\sigma^2} \sigma^2 
\]

\[
- \frac{d^2}{\sigma^2} 1 + \frac{d^2}{\sigma^2} 2 \frac{1}{2} \frac{d^2}{\sigma^2} \sigma^2 
\]

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

By this modification, \( d_2/d_1 \) is constrained to \([c, \tilde{c}]\). □

### 2.2.2 Example 2: Stochastic Frontier Production Function Models in Aigner, Lovell, and Schmidt (1977) and Stevenson (1980)

Another D-D model is found from the theory of stochastic frontier production function models. Stochastic production function models are often specified for an identically and independently distributed data sequence \( \{Y_t, X_t\} \) as follows:

\[
Y_t = X_t^\prime \beta_* + U_t,
\]

where \( Y_t \in \mathbb{R} \) is the output produced using inputs \( X_t \in \mathbb{R}^k \) such that \( \beta_* \) is an interior element of \( B \subset \mathbb{R}^k \); \( E[U_t^2] < \infty; E[X_{t,j}^2] < \infty \) for \( j = 1, 2, \ldots, k \); and \( E[X_tX_t^\prime] \) is positive-definite. Here, \( U_t \) stands for an
error term, which cannot be explained by a production function and independent of $X_t$. By this component, the production function is called a stochastic production function model. The motivations of this model are well discussed in Aigner, Lovell, and Schmidt (1977).

One of the early areas of interests in this specification is in identifying inefficiently produced outputs. Given output levels subject to the production function and inputs, it should follow that $E[U_t] = 0$ when outputs are efficiently produced. If $E[U_t] < 0$, outputs are inefficiently produced. That is, it must lie below the frontier of production function on average. Aigner, Lovell, and Schmidt (1977) capture this aspect by assuming that $U_t$ is distributed as

$$U_t \equiv V_t - W_t,$$

where $V_t \sim N(0, \tau^2_s)$, $W_t \sim N^T(\mu_s, \sigma^2_s)$, and $V_t$ is independent of $W_t$. Here, we suppose that $\tau_s > 0$, $\sigma_s \geq 0$, and $\mu_s \geq 0$, and $N^T(\mu_s, \sigma^2_s)$ denotes the truncated-normal random variable distributed on $(0, \infty)$ when the original random variable is normally distributed with population mean $\mu_s$ and variance $\sigma^2_s$. The positively valued $W_t$ captures the inefficiency. If $\mu_s = 0$ and $\sigma^2_s = 0$, this collapses to the stochastic production function model in Zellner, Kmenta, and Drèze (1966). Thus, the key to identifying inefficiently produced outputs is in testing $\mu_s = 0$ and $\sigma^2_s = 0$.

The original form introduced by Aigner, Lovell, and Schmidt (1977) originally assumes $\mu_s = 0$, so that the mode of $W_t$ is always achieved at zero, and this aspect has been pointed out as a serious drawback. A natural way to overcome this drawback is to estimate $\mu_s$ as suggested by Stevenson (1980). Having $\mu_s$ be different from zero has often been specified since then (e.g., Dutta, Narasimhan, and Rajiv (1999), Habib and Ljungqvist (2005), and etc.).

Nevertheless, a testing methodology for $\mu_s = 0$ and $\sigma^2_s = 0$ has not been developed because the likelihood value under the null is not defined. More specifically, the unknown parameters are estimated by maximizing the likelihood function. For each $(\beta, \sigma, \mu, \tau)$, the log-likelihood for this specification is given as

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

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

$$\tilde{\mu}_t := \frac{\tau^2 \mu - \sigma^2 (Y_t - X_t' \beta)}{\tau^2 + \sigma^2} \quad \text{and} \quad \tilde{\sigma}^2 := \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2}.$$
Here, we note that the log-likelihood is not well defined when \( \theta_* := (\beta'_*, \mu_*, \sigma_*, \tau_*)' = (\beta'_*, 0, 0, \tau_*) \). This is mainly because \( \mu_* / \sqrt{\sigma_*^2} = 0 / 0 \), so that \( \ln(\Phi(\mu_* / \sqrt{\sigma_*^2})) \) is not defined. Furthermore, if we let

\[
\bar{\mu}_* := \frac{\tau_*^2 \mu_* - \sigma_*^2 U_t}{\tau_*^2 + \sigma_*^2} \quad \text{and} \quad \bar{\sigma}_* := \frac{\tau_*^2 \sigma_*^2}{\tau_*^2 + \sigma_*^2},
\]

\( \bar{\mu}_* / \sqrt{\bar{\sigma}_*^2} = 0 / 0 \), implying that \( \ln(\Phi(\bar{\mu}_* / \sqrt{\bar{\sigma}_*^2})) \) is not defined, either. It follows from this fact that the value of likelihood cannot be computed.

On the other hand, directional limits are well defined. Some tedious algebra shows that for a given \( d := (d'_\beta, d_\mu, d_\sigma, d_r)' \),

\[
\lim_{h \to 0} L_n(\theta_* + hd) = -\frac{n}{2} \ln(\tau_*^2) + \sum_{t=1}^n \ln \left[ \phi \left( \frac{Y_t - X'_t \beta_*}{\sqrt{\tau_*^2}} \right) \right],
\]

which is the log-likelihood value under the null condition. This follows by noting that

\[
\lim_{h \to 0} \Phi \left( \frac{hd\mu_0}{\sqrt{(hd\sigma)^2}} \right) = \Phi \left( \frac{d\mu}{\sqrt{d\sigma}} \right) \quad \text{and} \quad \lim_{h \to 0} \Phi \left( \frac{\bar{\mu}_*(h; d)}{\sqrt{\bar{\sigma}(h; d)^2}} \right) = \Phi \left( \frac{d\mu}{\sqrt{d\sigma}} \right),
\]

where we let

\[
\bar{\sigma}_*(h; d)^2 := \frac{(\tau_* + hd_r)^2 (hd_\sigma)^2}{(\tau_* + dd_r)^2 + (hd_\sigma)^2}
\]

and

\[
\bar{\mu}_*(h; d) := \frac{(\tau_* + hd_r)^2 hd\mu_0 - (hd_\sigma)^2 (Y_t - X'_t (\beta_* + hd_\beta))}{(\tau_* + hd_r)^2 + (hd_\sigma)^2}.\]

Thus, it is still possible to test the parameters of interest using the directional limit.

We can also obtain directional derivatives of \( L_n(\cdot) \) at \( (\beta_*, 0, 0, \tau_*) \), although its partial derivatives are not obtained. Its first and second-order directional derivatives are obtained as follows:

\[
(3) \quad DL_n(\theta_*; d) = \sum_{t=1}^n \frac{1}{\tau_*} \left\{ d_r (U_t^2 - \tau_*^2) + \left[ -d_\mu + X'_t d_\beta - \psi (d_\mu, d_\sigma) \right] \tau_* U_t \right\},
\]

\[
(4) \quad D^2 L_n(\theta_*; d) = \sum_{t=1}^n \frac{1}{\tau_*^2} \left\{ d_\sigma^2 (U_t^2 - \tau_*^2) + d_r^2 \tau_*^2 \right\}
- \sum_{t=1}^n \frac{1}{\tau_*} \left\{ \left[ d_r U_t - (d_\mu - X'_t d_\beta) \tau_* \right] [3d_r U_t - (d_\mu - X'_t d_\beta) \tau_*] \right\}
- \sum_{t=1}^n \frac{1}{\tau_*} \left\{ \psi (d_\mu, d_\sigma) d_r^2 U_t^2 + \psi (d_\mu, d_\sigma) [d_\mu U_t^2 - 4d_r \tau_* U_t + (d_\mu - 2X'_t d_\beta) \tau_*^2] \right\},
\]

where

\[
\psi (d_\mu, d_\sigma) := |d_\sigma| \frac{\phi (d_\mu / |d_\sigma|)}{\Phi (d_\mu / |d_\sigma|)}.
\]
Here, if $\theta^* = (\beta^* \cdot 0, 0, \tau^*)$, $U_t \sim N(0, \tau^2)$. We also note that the first and second-order directional derivatives are neither linear nor quadratic with respect to $d$, respectively, which means that $L_n$ is not differentiable, although their directional derivatives are well defined.

Before moving to the next example, we specify the set of directions. The given condition on $\theta^*$ ($\mu_* \geq 0$ and $\sigma_* \geq 0$) implies that the set of directions has to be specified as

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

By this set, we can represent any $\theta$ in the parameter space of $\Theta$ as a function of direction in $\Delta(\theta^*)$ and the distance between $\theta$ and $\theta^* = (\beta^*, 0, 0, \tau)$.

### 2.2.3 Example 3: Generalized Method of Moments (GMM) in Hansen (1982)

Hansen (1982) examines an estimation method and relevant inference by generalizing the method of moments. One of the regularity conditions for this method is differentiability. We consider the GMM estimator $\hat{\theta}_n$ obtained by maximizing

$$Q_n(\theta) := g_n(X^n; \theta)' \{-M_n\}^{-1} g_n(X^n; \theta),$$

where $\{X_t : t = 1, 2, \cdots\}$ is a sequence of strictly stationary and ergodic random variables; $g_n(X^n; \theta) := n^{-1} \sum_{t=1}^n q(X_t; \theta)$; $q_t := q(X_t; \cdot) : \Theta \mapsto \mathbb{R}^k$ is continuously D a.s.–$\mathbb{P}$ on $\Theta$ given in Assumption 2 ($r \leq k$); for each $\theta \in \Theta$, $q(\cdot; \theta)$ is measurable; $M_n$ is a symmetric and positive definite random matrix a.s.–$\mathbb{P}$ uniformly in $n$, converging to a symmetric and positive definite $M_*$ a.s.–$\mathbb{P}$; for some integrable $m(X_t)$, $\|q_t(\cdot)\|_{\infty} \leq m(X_t)$ and $\|\nabla_\theta q_t(\cdot)\|_{\infty} \leq m(X_t)$; and there is a unique $\theta^*$, which maximizes

$$E[q_t(\theta)'] \{-M_*\}^{-1} E[q_t(\theta)]$$

on the interior part of $\Theta$. We denote the uniform matrix norm by $\| \cdot \|_{\infty}$. Furthermore, without more primitive moment conditions, we simply suppose that $n^{1/2}g_n(X^n; \theta^*) \Rightarrow W \sim N(0, S_*)$ for some positive definite matrix $S_*$. The GMM estimator is widely applied for empirical data.

The given conditions for $Q_n$ do not exactly satisfy the model conditions in Assumption 2. Nevertheless, Bates and White (1985, Theorem 2.6) show that the GMM estimator $\hat{\theta}_n$ is consistent for $\theta^*$ under the given conditions, and it is not hard to apply our theory on D-D models developed below to the GMM estimator. We, therefore, examine how the D-D model analysis can be suited to the analysis of D models by focusing on the GMM estimation.
Directional derivatives play a key role as before. We note that the first-order directional derivative of $g_n(\cdot) := g_n(X^n; \cdot)$ is obtained as

$$D g_n(\theta; d) = \nabla_\theta g_n(X^n; \theta)'d,$$

where $\nabla_\theta g_n(X^n; \theta) := [\nabla_{\theta_1}g_n(X^n; \theta_1), \cdots, \nabla_{\theta_r}g_n(X^n; \theta_r)]'$. As Eq. (5) makes clear, $D g_n(\theta; d)$ is now linear with respect to $d$. Applying the mean-value theorem implies that for each $d$,

$$g_n(\theta; d) = g_n(\theta_*; d) + D g_n(\bar{\theta}; d)(\theta - \theta_*),$$

for some $\bar{\theta}$ between $\theta$ and $\theta_*$. And

$$DQ_n(\theta; d) = -2d'\nabla_\theta g_n(\theta)'M^{-1}_n g_n(\theta),$$

so that for each $d$, $n^{1/2}DQ_n(\theta_*; d) \Rightarrow -2d'\nabla_\theta q_t(\theta_*)'M^{-1}W$ by applying the law of large numbers (LLN) to $\nabla_\theta g_n(\theta_*)$, which converges to $C_* := E[\nabla_\theta q_t(\theta_*)]$ a.s. \(\Rightarrow P\) from that $\|\nabla_\theta q_t(\cdot)\|_{\infty} \leq m(X_t)$. We use these to obtain the asymptotic behavior of the GMM estimator below by following the D-D model analysis.

### 2.2.4 Example 4: Box-Cox's (1964) Transformation

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

$$Y_t = Z_t' \theta_0 + \frac{\theta_1}{\theta_2}(X_t^{\theta_2} - 1) + U_t,$$

where $\{(Y_t, X_t, Z_t) \in \mathbb{R}^{2+k} : t = 1, 2, \cdots\}$ is assumed to be identically and independently distributed; $X_t$ is strictly greater than zero almost surely; and $U_t := Y_t - E[Y_t|Z_t, X_t]$. We further suppose that the unknown parameters are estimated by the nonlinear least squares (NLS) estimation method by maximizing

$$L_n(\theta) := -\frac{1}{2} \sum_{t=1}^{n} \left\{ Y_t - Z_t' \theta_0 - \frac{\theta_1}{\theta_2}(X_t^{\theta_2} - 1) \right\}^2,$$

where $\theta := (\theta_0', \theta_1, \theta_2)' \in \Theta_0 \times \Theta_12$; $\Theta_0$ is a convex and compact set in $\mathbb{R}^k$; and

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

This model specification is selected to avoid an identification problem. We note that $\theta_{1*} = 0$ if and only if $\theta_{2*} = 0$ under the given model assumption. Thus, we can avoid Davies’s (1977, 1987) identification problem which may arise when $\theta_1$ and $\theta_2$ are independent parameters and $\theta_{1*} = 0$. 

11
Nevertheless, if \( \theta_1^* = 0 \) and \( \theta_2^* = 0 \), the model analysis becomes obscure because \( \theta_1^*(X_t^{\theta_2^*} - 1)/\theta_2^* \) has the form of \( 0 \times 0/0 \), and this is not defined, implying that the standard tests cannot be applied.

We can overcome this problem by analyzing the directional limits. For this, we let \( d = (d_0', d_1, d_2)' \) and \( \theta^* = (\theta_0^*, 0, 0)' \) with \( \theta_0^* \) interior to \( \Theta_0 \). Then we can obtain that

\[
L_n(\theta^* + h d) = -\frac{1}{2} \sum_{t=1}^n \left\{ Y_t - Z_t' (\theta_0^* + d_0 h) - \frac{d_1}{d_2} (X_t^{d_2 h} - 1) \right\}^2,
\]

which is now differentiable with respect to \( h \) at 0. Furthermore, this reformulation shows that there lurks an identification problem different from the previous cases. We note that \( d_1/d_2 \) lacks its corresponding distance and disappears when \( h \) is zero, so that \( d_1/d_2 \) is not identified at \( \theta^* = (\theta_0^*, 0, 0)' \). We also derive the following directional derivatives:

\[
(9) \quad DL_n(\theta^*; d) = \sum_{t=1}^n U_t \{ Z_t' d_0 + \frac{d_1}{d_2} \log(X_t) d_2 \}, \quad \text{and}
\]

\[
(10) \quad D^2 L_n(\theta^*; d) = -\sum_{t=1}^n \{ Z_t' d_0 + \frac{d_1}{d_2} \log(X_t) d_2 \}^2 + \frac{d_1}{d_2} \sum_{t=1}^n U_t \{ \log(X_t) \}^2 d_2^2,
\]

so that for given \( d_1/d_2 \), Eqs. (9) and (10) are linear and quadratic in \( (d_0, d_2) \) respectively.

For our future analysis, we further suppose that

\[
\begin{bmatrix}
  n^{-1/2} \sum U_t Z_t' \\
  n^{-1/2} \sum U_t \log(X_t)
\end{bmatrix} \Rightarrow \mathbf{W} := \begin{bmatrix}
  \mathbf{W}_0' \\
  \mathbf{W}_1
\end{bmatrix} \sim N(\mathbf{0}, \mathbf{B}_s),
\]

where \( \mathbf{B}_s \) is a \((k + 1) \times (k + 1)\) positive definite matrix with a finite maximum eigenvalue. In addition, \( E[\log(X_t)^2] < \infty \) and \( E[Z_t'Z_t] < \infty \) are assumed.

Indeed, many nonlinear models share similar features. For example, Table 1 of Cheng, Evans, and Iles (1992) collects numerous nonlinear models with parameter instability problems. Many of them can be analyzed within the framework of directional derivatives. In addition to these, higher-order approximations of the quasi-likelihood function can be further simplified if directional derivatives are used. Cho, Ishida, and White (2011, 2012) and White and Cho (2012) revisit testing neglected linearity using artificial neural networks and show that its analysis requires higher-order approximations than the conventional analysis. In particular, Cho, Ishida, and White (2011, 2012) avoid the difficulties arising from higher-order approximations by using the directional derivatives defined in the current study, and Cho (2011) also examines additional aspects of quasi-maximum likelihood estimation obtained by the same methodology.
2.3 Asymptotic Distribution of the Extremum Estimator

As highlighted by the examples in the previous subsection, the most significant difference between D-D and D models is found in the linearity condition of directional derivatives in \( d \). We provide further regularity conditions for D-D models.

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

We use Assumption 4 to approximate D-D models by a second-order directional Taylor expansion. If models are D-D, Taylor approximations have to be applied for each direction. Furthermore, second-order directional derivatives do not have to be quadratic with respect to \( d \) as the previous examples demonstrate. They can be any non-linear function of \( d \) satisfying the following regularity conditions.

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

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

(iii) For each \( \theta \in \Theta \) and for all \( d_1, d_2 \in \Delta(\theta) \), there is \( \lambda > 0 \) such that \( |D\ell_t(\theta; d_1) - D\ell_t(\theta; d_2)| \leq M_t||d_1 - d_2||^\lambda \) and \( |D^2 \ell_t(\theta; d_1) - D^2 \ell_t(\theta; d_2)| \leq M_t||d_1 - d_2||^\lambda \). □

Assumption 5 will always be assumed together with Assumption 4 so that directional derivatives in Assumption 5 have their own meanings. We note that Assumptions 5(i and ii) are stochastic analogs of the conditions in Theorem 2 and Lemma 1. Assumption 5(iii) is assumed to apply the tightness and the ULLN to the first and second-order directional derivatives, respectively. We detail the tightness and the ULLN below, where they are more relevant. If Assumption 5(iii) is replaced by the following stronger Assumption 5(iii)*, the model is twice continuously D a.s.–\( \mathbb{P} \) by Lemma 1.

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

We let Assumption 5* denote Assumptions 5(i, ii, and iii*) going forward when we need to refer to D models. Unless stated otherwise, Assumption 5 stands for Assumptions 5(i, ii, and iii). Additional regularity conditions are required to obtain the asymptotic behavior of the extremum estimator. For this, we further restrict our prior regularity conditions by assuming...
Assumption 6 (CLT) (i) \( E[D\ell_t(\theta_s; d)] = 0 \) uniformly in \( d \in \Delta(\theta_s) \) and \( t \);
(ii) \( A_s(d) := E[n^{-1}D^2L_n(\theta_s; d)] \) is strictly negative and finite uniformly in \( d \in \Delta(\theta_s) \) and \( n \);
(iii) \( B_s(d, \bar{d}) \) is strictly positive and finite uniformly in \( d \in \Delta(\theta_s) \) and \( n \), where for each \( d, \bar{d} \),
\[
B_s(d, \bar{d}) := \text{acov}\{n^{-1/2}DL_n(\theta_s; d), n^{-1/2}DL_n(\theta_s; \bar{d})\},
\]
and ‘acov’ denotes the asymptotic covariance of given arguments;
(iv) for some \( q > (r - 1)/2(\lambda \gamma) \) and \( s > q \geq 2 \), and for each \( f_t \in \bar{L} \),
\[
\|f_t - E[f_t|\mathcal{F}_{t-\tau}^t]\|_q \leq \nu_\tau,
\]
where \( \bar{L} := \{a_1f_1 + a_2f_2 : f_1, f_2 \in \{D\ell_t(\theta_s; \cdot, d) : d \in \Delta(\theta_s)\}, a_1, a_2 \in \mathbb{R}\}; \nu_\tau \) is of size \(-1/(1 - \gamma)\) with \( 1/2 \leq \gamma < 1 \); \( \mathcal{F}_{t-\tau}^t := \sigma(Y_{t-\tau}, \cdots, Y_{t+\tau}) \); and \( \{Y_t \in \mathbb{R}^k : t = 1, 2, \cdots\} \) is a strong mixing sequence with size \(-sq/(s - q)\). In addition, \( E[M_t^s] < \infty \) and \( \sup_{d \in \Delta(\theta_s)} \sup_{t=1,2,\cdots} \|D\ell_t(\theta_s; d)\|_s < \Delta < \infty \). \( \square \)

Assumption 6(i) is imposed to apply the central limit theorem (CLT). Even if Assumption 6(i) does not hold, the extremum estimator still estimates \( \theta_s \) consistently. Nevertheless, the relevant statistics considered below can be degenerate without the zero first-order condition. By imposing Assumption 6(i), we can prevent this. Assumption 6(iii) is also needed to keep the statistics from being degenerate. For notational simplicity, we denote \( B_s(d, \bar{d}) \) as \( B_s(d) \) below if \( d = \bar{d} \). Assumption 6(iv) is imposed to apply corollary 3.1 of Wooldridge and White (1988) and theorem 4 of Hansen (1996a). It follows mainly from Assumption 6(iv) that \( n^{-1/2}DL_n(\theta_s; \cdot) \) obeys a functional central limit theorem (FCLT). Wooldridge and White (1988) provide regularity conditions for the CLT of near-epoch processes as a special case of mixingale processes. Hansen (1996a) generalizes this notion to the space of Lipschitz continuous functions and further provides the regularity conditions for the tightness of these functions. Assumption 6 is constructed to satisfy the regularity conditions of the CLT and the tightness. We note that the regularity conditions for the CLT are stronger than those for standard \( L^2 \)-near epoch processes. The coefficient \( q \) may have to be greater than 2 depending on the dimension of direction \((r-1)\), the coefficient for the magnitude of dependency (\( \rho \)), and the degree of smoothness of the involved function (\( \lambda \)). Furthermore, it is not hard to verify that \( \{f_t, \sigma(X_t^\ell), \bar{L}\} \) is an \( L^q \)-mixingale class of Hansen (1996a) by proposition 2.9 of Wooldridge and White (1988), where the space of first-order directional derivatives is extended to the space of their linear combinations.

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

\[
L_n(\hat{\theta}_n(d)) = \max_{\theta \in \Theta_s(d)} L_n(\theta),
\]

where \( \Theta_s(d) := \{ \theta' \in \Theta : \theta' = \theta_s + hd, h \in \mathbb{R}^+, d \in \Delta(\theta_s) \} \). We note that the d-extremum estimator is constrained by \( d \). That is, for given \( d \), \( \Theta_s(d) \) is a straight line starting from \( \theta_s \) and ending at the point crossed by the boundary of \( \Theta \) and \( \{ \theta_s + hd, h \in \mathbb{R}^+ \} \). Therefore, \( \Theta_s(d) \subset \Theta \), so that \( L_n(\hat{\theta}_n(d)) \leq L_n(\hat{\theta}_n) \).

The d-extremum estimator \( \hat{\theta}_n(d) \) can also be represented by the distance between \( \theta_s \) and \( \hat{\theta}_n(d) \). By the constraint that \( \theta^* \in \Theta_s(d) \), we can define \( \tilde{h}_n(d) \) such that \( \hat{\theta}_n(d) = \theta_s + \tilde{h}_n(d)d \). By this identity, the asymptotic behavior of \( \hat{\theta}_n(d) \) can be shown to be interrelated with the asymptotic behavior of \( \tilde{h}_n(d) \). We define the space of \( h \) as \( H_s(d) := \{ h \in \mathbb{R}^+ : \theta_s + hd \in \Theta_s(d) \} \), so that

\[
\max_{h \in H_s(d)} L_n(\theta_s + h d) = L_n(\hat{\theta}(d)).
\]

Given this equality, we can apply the directional second-order Taylor approximation to \( L_n(\theta_s + h d) \) using the first and second-order directional derivatives. The following lemma shows their asymptotic behaviors.

**Lemma 2** Given Assumptions 1 to 6, for each \( d \in \Delta(\theta_s) \),

- (i) \( n^{-1/2} DL_n(\theta_s; d) \Rightarrow Z(d) \), where \( Z(d) \sim N(0, B_s(d)) \);
- (ii) \( n^{-1} D^2 L_n(\theta_s; d) \) converges to \( A_s(d) \) a.s.–\( \mathbb{P} \);
- (iii) \( \{ n^{-1/2} DL_n(\theta_s; d), n^{-1} D^2 L_n(\theta_s; d) \} \Rightarrow \{ Z(d), A_s(d) \} \). \( \square \)

Lemma 2 is not hard to prove, and it further forms the basis for obtaining the asymptotic distribution of \( \tilde{h}_n(d) \). For this purpose, we approximate \( L_n \) on \( \Theta_s(d) \) by the mean-value theorem. That is, for some \( \tilde{\theta}_n(d) \in \Theta(d) \),

\[
L_n(\theta_s + h d) = L_n(\theta_s; d) + DL_n(\theta_s; d)h + \frac{1}{2} D^2 L_n(\theta_s(\theta_s + h d); d)h^2,
\]

where \( h \in H_s(d) \) and \( L_n(\theta_s; d) := \lim_{h \downarrow 0} L_n(\theta_s + h d) \). We note that this approximation is carried out on \( H_s(d) \). Therefore, it follows from this and Lemma 2 that for each \( d \in \Delta(\theta_s) \),

\[
2\{ L_n(\tilde{\theta}_n(d)) - L_n(\theta_s; d) \} \Rightarrow \max_{\tilde{h} \in \mathbb{R}^+} [2Z(d)\tilde{h} + A_s(d)\tilde{h}^2],
\]

where \( \tilde{h} \) captures the asymptotic behavior of \( \sqrt{n} h \). The large sample properties of relevant statistics can be stated as in the following theorem.
Theorem 3  Given Assumptions 1 to 6, for each \( d \in \Delta(\theta_*) \),

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

(ii) \( \sqrt{n}(\hat{\theta}_n(d) - \theta_*) \Rightarrow \max[0, G(d)]d \);

(iii) \( 2\{L_n(\hat{\theta}_n(d)) - L_n(\theta_*)\} \Rightarrow \max[0, Y(d)]^2 \), where for each \( d \), \( Y(d) := \{-A_*(d)\}^{1/2}G(d) \).

We can plainly explain the main contents of Theorem 3. The half-normal random variable \( G(d) \) in Theorem 3(i) is obtained as the weak limit of the argument in the right-hand side (RHS) of (12). That is,

\[
\max[0, G(d)] = \arg\max_{\hat{h}(d) \in \mathbb{R}^+} [2Z(d)\hat{h}(d) + A_*(d)\hat{h}(d)^2].
\]

The main reason for involving the ‘max’ operator is that \( \hat{h}(d) \) lies on the positive real line, which is obtained as the limit of the set \( \{\theta_* + \sqrt{n}hd : h \in H_*(d)\} \), so that if \( Z(d) \) is negative, the RHS of Eq. (12) is obtained by letting \( \hat{h}(d) = 0 \). Otherwise, the RHS is obtained by letting \( \hat{h}(d) = G(d) \). In the literature, Chernoff (1954) first approximates a parameter space by a cone, and Self and Liang (1987) and Andrews (1999) develop this to cope with the boundary parameter problem more fundamentally. We apply their approach to D-D models for given \( d \), and the result in Theorem 3(i) is the consequence of this. Given this, Theorems 3(ii and iii) trivially follow from the identity \( \hat{\theta}_n(d) \equiv \theta_* + \hat{h}_n(d)d \) and Eq. (12), respectively.

Nevertheless, the given pointwise results (with respect to \( d \)) in Theorem 3 are not sufficient to yield the asymptotic behavior of \( \hat{\theta}_n \). It is further necessary to consider the stochastic interrelationships between the d-extremum estimators. Certainly, the relationship between the extremum and d-extremum estimators is specified as

\[
L_n(\hat{\theta}_n) = \sup_{d \in \Delta(\theta_*)} L_n(\hat{\theta}_n(d)).
\]

That is, if we let \( \hat{d}_n := \arg\max_{d \in \Delta(\theta_*)} L_n(\hat{\theta}_n(d)) \), we obtain that \( L_n(\hat{\theta}_n) \equiv L_n(\hat{\theta}_n(\hat{d}_n)) \). Given this, the question is how to derive the asymptotic behavior of \( \hat{\theta}_n \) from that of \( \hat{\theta}_n(d) \), and we can answer this by examining the functional relationship between \( \hat{\theta}_n \) and \( \hat{\theta}_n(\cdot) \). The following lemma promotes this examination.

Lemma 3  Given Assumptions 1 to 6,

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

\[
\limsup_{n \to \infty} \mathbb{P}_n \left( \sup_{||d_1 - d_2|| < \delta} n^{-1/2}|DL_n(\theta_*; d_1) - DL_n(\theta_*; d_2)| > \varepsilon \right) < \varepsilon,
\]

where \( \mathbb{P}_n \) is empirical probability measure;
Lemma 3(i) is used to show that the first-order directional derivatives weakly converge to a Gaussian stochastic process indexed by \( d \). Specifically, this shows the tightness in Billingsley (1999). In our time-series data context, theorem 1 of Hansen (1996a) provides sufficient regularity conditions for the tightness, and we exploit them to show Lemma 3(i). If \( \triangle(\theta_\star) \) has a finite number of elements, Lemma 3 is not necessary in deriving the consequences of Theorem 4 given below. For example, if \( \Theta \) is a subset of \( \mathbb{R} \), \( \triangle(\theta_\star) \) has finite elements. On the other hand, Lemma 3(i) is essential to establishing the weak convergence when the dimension of \( \Theta \) is greater than 1. Thus, we suppose that \( \triangle(\theta_\star) \) has an uncountable number of elements in proving Lemma 3(i).

If \( L_n \) is \( D \) on \( \Theta \), it is trivial to show Lemma 3. By Theorem 2, it follows that \( DL_n(\theta_\star; d) = \nabla_\theta L_n(\theta_\star)d \), so that

\[
\sup_{\|d_1 - d_2\| < \delta} n^{-1/2} |D L_n(\theta_\star; d_1) - D L_n(\theta_\star; d_2)| \leq \|n^{-1/2} \nabla_\theta L_n(\theta_\star)\| \delta,
\]

implying that for any \( \varepsilon > 0 \),

\[
\mathbb{P}_n \left( \sup_{\|d_1 - d_2\| < \delta} n^{-1/2} |D L_n(\theta_\star; d_1) - D L_n(\theta_\star; d_2)| > \varepsilon \right) \leq \mathbb{P} \left( \|n^{-1/2} \nabla_\theta L_n(\theta_\star)\| \delta > \varepsilon \right).
\]

Thus, if \( n^{-1/2} \nabla_\theta L_n(\theta_\star) \) obeys the CLT, we can choose \( \delta \) to have the RHS of Eq. (14) be less than \( \varepsilon \), and this is what Lemma 3(i) desires. Likewise, we can also apply the ULLN to the second-order derivatives to show Lemma 3(ii) under the \( D \) model assumption. We note that for each \( d \in \triangle(\theta_\star) \), \( D^2 L_n(\theta_\star; d) = d' \nabla^2_\theta L_n(\theta_\star)d \), so that for a nontrivial norm, \( \| \cdot \|_\infty \) say,

\[
\sup_d |n^{-1} \{d' \nabla^2_\theta L_n(\theta_\star)d - d'E[\nabla^2_\theta L_n(\theta_\star)]d\}| \leq \sup_d d'| n^{-1} \{\nabla^2_\theta L_n(\theta_\star) - E[\nabla^2_\theta L_n(\theta_\star)]\}|_\infty,
\]

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

Lemma 3 extends Theorem 3 to the level of functional space, and from this, the asymptotic distribution of the extremum estimator follows. Before stating this, if \( L_n(\theta_\star) \) is not defined as in Examples 1, 2, and 4, we let \( L_n(\theta_\star) := L_n(\theta_\star; \bar{d}_n) \) unless confusion would otherwise arise.

**Theorem 4** Given Assumptions 1 to 6,

(i) \( \{n^{-1/2} D L_n(\theta_\star; \cdot), n^{-1} D^2 L_n(\theta_\star; \cdot)\} \Rightarrow (\mathbb{Z}, A_\star) \), where for each \( d \) and \( d' \), \( E[\mathbb{Z}(d) \mathbb{Z}(d')] = B_\star(d, d') \);
(ii) \( \sqrt{n} \hat{\theta}_n \Rightarrow \max[0, G] \); 

(iii) \( 2\{L_n(\hat{\theta}_n) - L_n(\theta_*)\} \Rightarrow \sup_{d \in \Delta(\theta_*)} \max[0, Y(d)]^2 \); 

(iv) \( \sqrt{n}(\hat{\theta}_n - \theta_*) \Rightarrow \max[0, G(d_*)]d_* \), where \( d_* := \arg\max_{d \in \Delta(\theta_*)} \max[0, Z(d)]^2 \{-A_*(d)\}^{-1}. \)

We note that the extremum estimator is now represented as a function of the Gaussian stochastic process defined on \( \Delta(\theta_*) \). Many statistics in the literature are known to follow asymptotic distributions characterized by Gaussian stochastic processes. For example, Davies (1977, 1987) considers unidentified parameter problems and shows that associated statistics and estimators follow asymptotic distributions characterized by Gaussian stochastic processes. Furthermore, Andrews (2001) and Cho and White (2007, 2010, 2011a) provide many other examples in which asymptotic distributions are characterized by Gaussian stochastic processes. We also note that \( \Delta(\theta_*) \) is well defined even when \( \theta_* \) is on the boundary of \( \Theta \) as explained above.

Theorem 4 accommodates the standard D models as a special case of D-D models. For this examination, we impose the following condition, which precludes trivial asymptotic behavior of the extremum estimator.

Assumption 6 (CLT) (ii)* For a symmetric and negative definite matrix \( A_* \) and each \( d \), \( A_*(d) = d'A_*d \); 

(iii)* For a symmetric and positive definite matrix \( B_* \) and each \( d \), \( B_*(d, d) = d'B_*d \).

Assumptions 6(ii) and iii)* correspond to assuming that \( A_* := \lim_{n \to \infty} n^{-1}E[\nabla^2 \theta L_n(\theta_*)] \) and \( B_* := \text{acov}\{n^{-1/2}\nabla \theta L_n(\theta_*)\} \) are negative and positive definite, respectively, under the framework of D models. These assumptions can further refine the results in Theorem 4. We let Assumption 6* denote Assumptions 6(i, ii*, iii*, and iv).

Corollary 1 Given Assumptions 1 to 4, 5*, and 6*,

(i) \( Z \) is linear in \( d \in \Delta(\theta_*) \), so that for each \( d \), \( Z(d) = Z'd \{ -d'A_*d \}^{-1} \) in distribution, where \( Z \sim N(0, B_*) \);

(ii) for each \( d \), \( G(d) = Z'd \{ -d'A_*d \}^{-1} \) in distribution;

(iii) for each \( d \), \( \sqrt{n} (\hat{\theta}_n(d) - \theta_*) \Rightarrow \max[0, Z'd \{ -d'A_*d \}^{-1}]d \);

(iv) \( \sqrt{n}(\hat{\theta}_n - \theta_*) \Rightarrow \max[0, -Z'd_* \{ d_*'A_*d_* \}^{-1}]d_* \), where \( d_* := \arg\max_d \max[0, Z'd]^2 / d'(-A_*)d \);

(v) \( \sqrt{n}(\hat{\theta}_n - \theta_*) \Rightarrow (-A_*)^{-1}Z \), provided that \( \theta_* \) is interior to \( \Theta \);

(vi) \( 2\{L_n(\hat{\theta}_n) - L_n(\theta_*)\} \Rightarrow \sup_{d \in \Delta(\theta_*)} \max[0, Z'd^2 \{ d'(-A_*)d \}]^{-1} \);

(vii) \( 2\{L_n(\hat{\theta}_n) - L_n(\theta_*)\} \Rightarrow Z'(A_*)^{-1}Z \), provided that \( \theta_* \) is interior to \( \Theta \).

We note that the results in Corollary 1 are the same as for the standard case if \( \theta_* \) is interior to \( \Theta \). Nevertheless, our analysis is more primitive because it involves directional derivatives. When stating Corollary 1,
we suppose that \( r > 1 \) because proving it is trivial when \( r = 1 \). In addition to the properties provided in Corollary 1, there are many other aspects of D models as Cho (2011) clarifies using the D-D model analysis provided here.

2.4 Examples

2.4.1 Example 1 (continued)

The first-order directional derivative given in Eq. (1) can be partitioned into three pieces:

\[
DL_n(\gamma_*, \sigma_2^2, \theta_*; d) = Z_{1,n}(d) + Z_{2,n}(d) + Z_{3,n}(d),
\]

where for each \( d \),

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

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

and \( m(d_2/d_1) := 2 \tan^{-1}(d_2/d_1)/\pi \). Given this, the CLT for a martingale difference sequence of McLeish (1974, Theorem 2.3) can be applied to each \( Z_{i,n}(d) \), so that for each \( d \) and \( i = 1, 2, 3 \),

\[
n^{-1/2} Z_i(d) \Rightarrow Z_i(d),
\]

where \( Z_i(d) \sim N(0, B_x(i)(d, d)) \) is independent of \( Z_j(d) \sim N(0, B_x(j)(d, d)) \) \((i \neq j)\), and for each \( (d, \tilde{d}) \),

\[
B_x^{(1)}(d, \tilde{d}) := \frac{1}{\sigma_*^2} d_{\gamma'} E[Q_t Q_t'] \tilde{d}_\gamma,
\]

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

and

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

Furthermore, we note that \( Z_i(d) \) and \( Z_j(d) \) are asymptotically independent, which enables us to apply the weak convergence on their product space by example 1.4.6 of van der Vaart and Wellner (1996, p. 31), implying that

\[
\{Z_{1,n}(d), Z_{2,n}(d), Z_{3,n}(d)\} \Rightarrow \{Z_1(d), Z_2(d), Z_3(d)\},
\]

19
and leading to the conclusion that \( n^{-1/2} DL_n(\gamma_*, \sigma_*^2, \theta_*; d) \Rightarrow \mathcal{Z}(d) := \mathcal{Z}_1(d) + \mathcal{Z}_2(d) + \mathcal{Z}_3(d) \) by the continuous mapping theorem (CMT).

It is not hard to generalize the given pointwise weak convergence to the level of functional space as this can be achieved by showing the tightness of \( n^{-1/2} DL_n(\gamma_*, \sigma_*^2, \theta_*; \cdot) \). As \( \mathcal{Z}_1, n \) and \( \mathcal{Z}_2, n \) are virtually linear with respect to \( \mathbf{Q}_t U_t \) and \( (U_2^2 - \sigma_*^2) \), respectively, their tightness trivially follows. We pay attention only to \( \mathcal{Z}_3, n \) for brevity. For notational simplicity, we let \( \varepsilon_t := W_t U_t, m := m(d_2/d_1) \), and \( \bar{m} := m(d_2/d_1) \). Then for any \( \varepsilon > 0 \), there is \( \delta \) such that

\[
\limsup_{n \to \infty} \mathbb{P}_n \left( \sup_{|m - \bar{m}| < \delta} \left| n^{-1/2} \sum_{t=2}^{n} \varepsilon_t \sum_{t'=1}^{t-1} \varepsilon_{t'} m_{t-t'} - n^{-1/2} \sum_{t=2}^{n} \varepsilon_t \sum_{t'=1}^{t-1} \varepsilon_{t'} \bar{m}_{t-t'} \right| > \varepsilon \right) < \varepsilon.
\]

We note that the sequence \( \{\varepsilon_t \sum_{t'=1}^{t-1} \varepsilon_{t'} m_{t-t'}, \sigma(\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots)\} \) is a martingale difference sequence uniformly in \( m \). We can apply theorem 2 of Hansen (1996a) to show the tightness. Verifying the conditions of his theorem 1 is trivial if \( E[W_t^4] < \Delta^4 < \infty \). That is, Hansen’s (1996a) \( \lambda \) and \( a \) are identical to 1 in our model assumption,

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

uniformly in \( m \), and the Lipschitz constant given as \( M_t := \sum_{\tau=1}^{t-1} (t - \tau) \bar{m}_{t-\tau-1} |\varepsilon_t \varepsilon_\tau| \) satisfies the moment condition. That is,

\[
\limsup_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E[M_t^2] = (\sigma_* \Delta)^4 \left( \frac{1 + 2 \bar{m} - 2 \bar{m}^3 - \bar{m}^4}{(1 - \bar{m})^6(1 + \bar{m})^3} \right) < \infty
\]

by the standard argument that \( |m| \) is uniformly and strictly bounded by 1 and \( E[|\varepsilon_t^2 \varepsilon_\tau \varepsilon_{t'}|] < (\sigma_* \Delta)^4 < \infty \), where \( \bar{m} := \max[|m(\varepsilon)|, |m(\bar{\varepsilon})|] \). From these facts, it follows that

\[
n^{-1/2} DL_n(\gamma_*, \sigma_*^2, \theta_*; \cdot) \Rightarrow \mathcal{Z}
\]

such that for each \( d \) and \( \bar{d} \), \( E[\mathcal{Z}(d) \mathcal{Z}(\bar{d})] = B_*(d, \bar{d}) := B_*^{(1)}(d, \bar{d}) + B_*^{(2)}(d, \bar{d}) + B_*^{(3)}(d, \bar{d}) \).

The asymptotic behavior of the second-order directional derivative is related to \( B_* \). By applying the LLN to the last elements in the first two lines of Eq. (2), we obtain that

\[
D^2 L_n(\gamma_*, \sigma_*^2, \theta_*; d) = -\frac{1}{\sigma_*^2} d_{\gamma} \mathbf{Q}^n d_{\gamma} - \frac{d_{\sigma^2}(d_1^2 + d_2^2)^{1/2}}{\sigma_*^4} \mathbf{U}^n \left[ \Omega^n(m(d_2/d_1)) \right] \mathbf{U}^n
\]

\[
- \frac{\sigma^2 d_{\sigma^2}^2}{2\sigma_*^4} + \frac{(d_1^2 + d_2^2)}{2} \left\{ \text{tr} \left[ \Omega^n(m(d_2/d_1))^2 \right] - \frac{2}{\sigma_*^2} \mathbf{U}^n \mathbf{D}^n(m(d_2/d_1)) \mathbf{U}^n \right\}
\]

\[
- \frac{(d_1^2 + d_2^2)}{\sigma_*^2} \mathbf{U}^n \mathbf{O}^n(m(d_2/d_1)) \mathbf{U}^n + o_p(n),
\]

20
where $D^n(m(d_2/d_1))$ is a diagonal matrix with those of $\Omega^n(m(d_2/d_1))^2$, and $O^n(m(d_2/d_1))$ is a matrix with off-diagonal elements of $\Omega^n(m(d_2/d_1))^2$ and zero diagonal elements, so that

$$D^n(m(d_2/d_1)) + O^n(m(d_2/d_1)) \equiv \Omega^n(m(d_2/d_1))^2.$$  

Applying theorem 3.7.2 of Stout (1974) shows that $n^{-1}D^2L_n(\gamma^*,\sigma_2^*,\theta^*;d) = -B_*(d,d) + o_P(1)$. This aspect is further strengthened by applying the ULLN. That is, $\sup_d |n^{-1}D^2L_n(\gamma^*,\sigma_2^*,\theta^*;d) + B_*(d,d)| = o_P(1)$, leading to the information matrix equality. Therefore,

$$Z_n,1(d) + Z_n,2(d) \Rightarrow Z_1(d),$$

where for each $d$,

$$Z_{n,1}(d) := \frac{d_2}{\tau_2^*} \sum_{t=1}^n (U_t^2 - \tau_2^2) \quad \text{and} \quad Z_{n,2}(d) := \frac{1}{\tau_2^*} \sum_{t=1}^n [X_t^t d_\beta + m(d_\mu, d_\sigma)] U_t.$$  

Given this, for each $d$, McLeish’s (1974, theorem 2.3) CLT can be applied to each component as before, and it follows that for each $d$,

$$n^{-1/2} \begin{bmatrix} Z_{n,1}(d) \\ Z_{n,2}(d) \end{bmatrix} \Rightarrow \begin{bmatrix} Z_1(d) \\ Z_2(d) \end{bmatrix},$$

where for each $d$,

$$\begin{bmatrix} Z_1(d) \\ Z_2(d) \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2d_2^2 & 0 \\ 0 & E[(X_t^t d_\beta + m(d_\mu, d_\sigma))^2] \end{bmatrix} \right).$$
and \( m(d_\mu, d_\sigma) := -[d_\mu + \psi(d_\mu, d_\sigma)] \). It also easily follows that for each \( d \) and \( \bar{d} \),

\[
E[Z_1(d)Z_1(\bar{d})] = 2 \frac{d_\tau \bar{d}_\tau}{\tau_\star^2}, \quad E[Z_1(d)Z_2(\bar{d})] = 0, \quad \text{and}
\]

\[
E[Z_2(d)Z_2(\bar{d})] = \frac{1}{\tau_\star^2} \left[ \begin{array}{c} m(d_\mu, d_\sigma) \\ d_\beta \\ d_\tau \end{array} \right] ' \left[ \begin{array}{ccc} 1 & E[X'_t] & 0 \\ E[X'_t] & 1 & 0 \\ 0 & 0 & 2 \end{array} \right] \left[ \begin{array}{c} m(\bar{d}_\mu, \bar{d}_\sigma) \\ \bar{d}_\beta \\ \bar{d}_\tau \end{array} \right].
\]

We note that the key elements for applying the CLT to \( Z_n, 1(d) \) and \( Z_n, 2(d) \) are \( (U^2_t - \tau_\star^2) \) and \( U_t \), respectively, and \( Z_n, 1(d) \) and \( Z_n, 2(d) \) are linear with respect to other components containing \( d \). From this fact, their tightness trivially follows, so that

\[
n^{-1/2} D L_n(\theta_\star; \cdot) \Rightarrow Z,
\]

where \( Z \) is a zero-mean Gaussian stochastic process such that for each \( d \) and \( \bar{d} \),

\[
E[Z(d)Z(\bar{d})] = B_s(d, \bar{d})
\]

and

\[
B_s(d, \bar{d}) := \frac{1}{\tau_\star^2} \left[ \begin{array}{c} m(d_\mu, d_\sigma) \\ d_\beta \\ d_\tau \end{array} \right] ' \left[ \begin{array}{ccc} 1 & E[X'_t] & 0 \\ E[X'_t] & 1 & 0 \\ 0 & 0 & 2 \end{array} \right] \left[ \begin{array}{c} m(\bar{d}_\mu, \bar{d}_\sigma) \\ \bar{d}_\beta \\ \bar{d}_\tau \end{array} \right].
\]

It is also possible to define \( Z \) as \( Z_1 + Z_2 \).

Here, we can also easily find another Gaussian stochastic process with a distribution equal to that of \( Z \).

If we let \( \bar{Z}(d) := \delta(d)'\Omega_\star^{1/2} \), where for each \( d \),

\[
\delta(d) := \left[ \begin{array}{c} d_\beta \\ m(d_\mu, d_\sigma) \\ d_\tau \end{array} \right]; \quad \Omega_\star := \frac{1}{\tau_\star^2} \begin{bmatrix} E[X, X'] & E[X'_t] & 0 \\ E[X'_t] & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}
\]

it trivially holds that

\[
E[\bar{Z}(d)\bar{Z}(\bar{d})] = \delta(d)'\Omega_\star \delta(\bar{d}),
\]

which is now identical to \( B_s(d, \bar{d}) \). This fact implies that \( Z \) has an identical distribution to \( \bar{Z} \). Furthermore, the associated Gaussian stochastic process is linear with respect to \( W \), and variation of \( d \) does not introduce any variation from the stochastic part.
The probability limit of the second-order directional derivative can also be easily found by applying the LLN. We note that
\[
\frac{1}{n} \sum_{t=1}^{n} U_t^2 = \tau_n + o_p(1), \quad \frac{1}{n} \sum_{t=1}^{n} U_t X_t = o_p(1), \quad \text{and} \quad \frac{1}{n} \sum_{t=1}^{n} X_t X_t' = E[X_t X_t'] + o_p(1),
\]
so that
\[
n^{-1} D^2 L_n(\theta; d) \xrightarrow{d} -\frac{1}{\tau_n} \left\{ 2d^2 + E[(d_\mu - X_t' d_\beta)^2] + \psi(d_\mu, d_\sigma)^2 + 2[d_\beta - E[X_t' d_\beta] \psi(d_\mu, d_\sigma)] \right\},
\]
and this is identical to \(-B(d, d)\). Thus, the information matrix equality holds, and
\[
2\{L_n(\hat{\theta}_n) - L_n(\theta_*)\} \Rightarrow \sup_{d \in \Delta(\theta_*)} [0, \mathcal{Y}(d)]^2
\]
by Theorem 4(iii), where
\[
\mathcal{Y}(d) := \frac{\delta(d)' \Omega_*^{-1/2} W}{\{\delta(d)' \Omega_* \delta(d)\}^{1/2}}.
\]
so that for each \(d\) and \(\bar{d}\),
\[
E[\mathcal{Y}(d) \mathcal{Y}(\bar{d})] = \frac{\delta(d)' \Omega_* \delta(\bar{d})}{\{\delta(d)' \Omega_* \delta(d)\}^{1/2} \{\delta(\bar{d})' \Omega_* \delta(\bar{d})\}^{1/2}}.
\]
This shows that the asymptotic behavior of the likelihood is represented as a function of the Gaussian stochastic process indexed by \(d\).

2.4.3 Example 3 (continued)

Given the first-order directional derivative in Eq. (5) and the regularity conditions for GMM estimation, it is trivial to show that \(\{n^{1/2} D Q_n(\theta_0; \cdot)\}\) is tight using the fact that it is linear with respect to \(d\). Next, by substituting \(g_n\) in Eq. (6) into \(Q_n\), we obtain that for some \(\bar{\theta}\) between \(\theta\) and \(\theta_*\),
\[
n\{Q_n(\theta) - Q_n(\theta_*)\} = -2d' \nabla_{\theta} g_n(\theta) M_n^{-1} \sqrt{n} g_n(\theta_*) \sqrt{n} h - d' \nabla_{\theta} g_n(\theta) M_n^{-1} \nabla_{\theta} g_n(\theta) d(\sqrt{n} h)^2,
\]
and so
\[
\{Q_n(\hat{\theta}_n) - Q_n(\theta_*)\} \Rightarrow \sup_{d} \sup_{h} -2d' C_*' M_*^{-1} Wh - d' C_*' M_*^{-1} C_* d h^2.
\]
This can be transformed into Eq. (12) by letting \(Z(d) := -d' C_*' M_*^{-1} W\) and \(A_s(d) := -d' C_*' M_*^{-1} C_* d\). We note that these derivatives are linear and quadratic in \(d\), respectively. Therefore, we can apply the results in Corollary 1 even if Assumption 2 does hold exactly. That is,
\[
\{Q_n(\hat{\theta}_n) - Q_n(\theta_*)\} \Rightarrow W' M_*^{-1} C_* \{C_*' M_*^{-1} C_*\}^{-1} C_*' M_*^{-1} W
\]
by Corollary 1(vii). Further, Corollary 1(v) implies that
\[
\sqrt{n}(\hat{\theta}_n - \theta^*) \Rightarrow - \{C^*_sM^{-1}_sC_s\}^{-1}C^*_sM^{-1}_sW
\]
\[
\sim N(0, \{C^*_sM^{-1}_sC_s\}^{-1}\{C^*_sM^{-1}_sS_sM^{-1}_sC_s\}\{C^*_sM^{-1}_sC_s\}^{-1}).
\]
These are the same results as given in the standard GMM literature (e.g., Newey and West (1987)).

### 2.4.4 Example 4 (continued)

Given the first and second-order directional derivatives in Eqs. (9) and (10),
\[
n^{-1/2} DL_n(\theta^*; \mathbf{d}) \Rightarrow \mathbf{d}'W(d_1/d_2), \quad \text{and} \quad n^{-1} D^2 L_n(\theta^*; \mathbf{d}) \Rightarrow \mathbf{d}'(A_*(d_1/d_2))\mathbf{d}
\]
a.s.\(\mathbb{P}\), where \(\mathbf{d} \in \Delta(\theta^*) := \{x \in \mathbb{R}^{k+1} : \|x\| = 1\}, W(d_1/d_2) := (W_0', \frac{d_1}{d_2} W_1)', \]
\[
A_*(d_1/d_2) := \begin{bmatrix}
A_{(0,0)}^* & \frac{d_1}{d_2} A_{(0,1)}^* \\
\frac{d_1}{d_2} A_{(1,0)}^* & \frac{d_1}{d_2}^2 A_{(1,1)}^*
\end{bmatrix},
\]
and
\[
A_* := \begin{bmatrix}
A_{(0,0)}^* & A_{(0,1)}^* \\
A_{(1,0)}^* & A_{(1,1)}^*
\end{bmatrix} := \begin{bmatrix}
-E[Z_tZ_t'] & -E[Z_t \log(X_t)] \\
-E[\log(X_t)Z_t'] & -E[\log(X_t)^2]
\end{bmatrix}.
\]
Here, we refined the set of directions to \(\Delta(\theta^*)\) because \(d_1/d_2\) is not identified when \(h = 0\), and we treat \(d_1/d_2\) as a fixed nuisance parameter when maximizing \(L_n\) with respect to \(\mathbf{d}\). More specifically, we decompose the maximization process into a two-step process separately maximized with respect to \(\mathbf{d}\) and \(d_1/d_2\). This leads to
\[
2\{L_n(\hat{\theta}_n) - L_n(\theta^*)\} \Rightarrow \sup_{d_1/d_2 \in [\varepsilon, \bar{\varepsilon}]} \sup_{d \in \Delta(\theta^*)} \max[0, W(d_1/d_2)'d]^2 \{-d'A_*(d_1/d_2)d\}^{-1}
\]
\[
= \sup_{d_1/d_2 \in [\varepsilon, \bar{\varepsilon}]} W(d_1/d_2)'\{-A_*(d_1/d_2)\}^{-1}W(d_1/d_2)
\]
\[
= \sup_{d_1/d_2 \in [\varepsilon, \bar{\varepsilon}]} W'(-A_*)^{-1}W = W'(-A_*)^{-1}W,
\]
where \(\hat{\theta}_n\) is the NLS estimator, and \(n^{-1/2} DL_n(\theta^*; \cdot)\) and \(n^{-1} D^2 L_n(\theta^*; \cdot)\) obey the FCLT and ULLN as before. This yields the weak convergence in the first line. Furthermore, we can apply the proof of Corollary 1(vii) to obtain the weak convergence. We note that \(d_2/d_1\) is treated as a fixed nuisance parameter when maximizing the likelihood function with respect to \(\mathbf{d}\). The other equalities follow from the definitions of \(A_*(d_1/d_2)^{-1}\) and \(W(d_1/d_2)\), in which \(d_2/d_1\) is canceled off by itself."
3 Testing Hypotheses with D-D Models

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

It is efficient first to specify the roles of each parameter when discussing data inferences using D-D models. We partition $\theta$ into $(\pi', \tau')' = (\lambda', \nu', \tau')'$ such that the directional derivatives of $L_n$ with respect to $\lambda \in \mathbb{R}^\lambda$ and $\nu \in \mathbb{R}^\nu$ are linear and non-linear with respect to $d_\lambda$ and $d_\nu$, respectively. The parameter $\tau \in \mathbb{R}^\tau$ consists of other nuisance parameters asymptotically orthogonal to $\pi := (\lambda', \nu')' \in \mathbb{R}^\pi$ in terms of the second-order directional derivative. More specifically, we suppose that for each $d$ the first-order directional derivative can be written as

$$DL_n(\theta_s; d) = d_\lambda' DL_n(\lambda) + DL_n(\nu)(d_\nu) + DL_n(\tau)(d_\tau)$$

for some random variables $(d_\lambda' DL_n(\lambda), DL_n(\nu)(d_\nu), DL_n(\tau)(d_\tau))$ such that for each $(d_\lambda', d_\nu', d_\tau')'$,

$$\frac{1}{\sqrt{n}} \begin{bmatrix} DL_n(\lambda)(d_\pi) \\ DL_n(\nu)(d_\pi) \\ DL_n(\tau)(d_\pi) \end{bmatrix} = \frac{1}{\sqrt{n}} \begin{bmatrix} d_\lambda' DL_n(\lambda) \\ DL_n(\nu)(d_\nu) \\ DL_n(\tau)(d_\tau) \end{bmatrix} \Rightarrow \begin{bmatrix} Z(\pi)(d_\pi) \\ Z(\nu)(d_\nu) \\ Z(\tau)(d_\tau) \end{bmatrix} := \begin{bmatrix} d_\lambda' Z(\lambda) \\ Z(\nu)(d_\nu) \\ Z(\tau)(d_\tau) \end{bmatrix},$$

which follows $N(0, B_\pi(d))$, and $n^{-1/2}(DL_n(\pi), DL_n(\tau)) \Rightarrow (Z(\pi), Z(\tau))$, where for each $d, \tilde{d} \in \Delta(\theta_s)$.

$$B_\pi(d, \tilde{d}) := \begin{bmatrix} B_\pi(\lambda, \nu)(d_\lambda, \tilde{d}_\lambda) & B_\pi(\lambda, \tau)(d_\lambda, \tilde{d}_\lambda) \\ B_\pi(\nu, \lambda)(d_\nu, \tilde{d}_\lambda) & B_\pi(\nu, \tau)(d_\nu, \tilde{d}_\lambda) \\ B_\pi(\tau, \lambda)(d_\nu, \tilde{d}_\nu) & B_\pi(\tau, \tau)(d_\nu, \tilde{d}_\nu) \end{bmatrix}$$

$DL_n(\lambda)$, $DL_n(\nu)(d_\nu)$, and $DL_n(\tau)(d_\tau)$ are defined on $\mathbb{R}^{\pi, \lambda}$, $\mathbb{R}$, and $\mathbb{R}$ respectively; $B_\pi(\lambda, \nu)(d_\nu)$ and $B_\pi(\lambda, \tau)(d_\tau) \in \mathbb{R}^{\pi, \lambda}; B_\pi(\nu, \lambda)(d_\nu) = B_\nu(\lambda, \nu)(d_\nu); B_\pi(\tau, \lambda)(d_\tau) = B_\lambda(\tau, \lambda)(d_\tau).$ Thus,

$$\text{acov} \left\{ n^{-1/2}DL_n(\theta_s; d), n^{-1/2}DL_n(\theta_s; \tilde{d}) \right\} = \iota_\lambda' B_\pi(d, \tilde{d}) \iota_\lambda,$$

where $\iota_\ell$ is the $\ell \times 1$ vector of ones. Accordingly, we also suppose that $A_\pi(d) = \iota_\lambda' A_\pi(d) \iota_\lambda$, where

$$A_\pi(d) := \begin{bmatrix} A_\pi(\lambda, \nu)(d_\lambda, \tilde{d}_\lambda) & A_\pi(\lambda, \tau)(d_\lambda, \tilde{d}_\lambda) \\ A_\pi(\nu, \lambda)(d_\nu, \tilde{d}_\lambda) & A_\pi(\nu, \tau)(d_\nu, \tilde{d}_\lambda) \\ A_\pi(\tau, \lambda)(d_\nu, \tilde{d}_\nu) & A_\pi(\tau, \tau)(d_\nu, \tilde{d}_\nu) \end{bmatrix}.$$
consistent presentation, we also assume that
Similarly, for each $d$, $\pi$ is assumed to be orthogonal to $\tau$. That is, for each $d$, $A_{\pi \tau}(d, \pi) = A_{\pi \tau}(\pi, \tau) = A_{\pi \tau}(\pi, \tau) = 0$. This assumption is useful in reducing the number of nuisance parameters to the set of nuisance parameters relevant to testing hypotheses. In terms of these definitions, we also permit that $r_\nu$, $r_\lambda$, and $r_\tau$ can be zero, so that $\lambda$, $\nu$, or $\tau$ may be absent from the model. If $r_\nu$ and $r_\tau$ are zero, the model is twice continuously differentiable. We formally state these conditions in Assumption 7.

**Assumption 7 (D-Derivatives)** (i) For each $d \in \Delta(\theta_*)$, $DL_n(\theta_*; d) = DL_n(\pi) (d, \pi) + DL_n(\tau) (d, \tau)$, and $n^{-1/2}(DL_n(\pi), DL_n(\tau)) \Rightarrow (\mathbf{Z}(\pi), \mathbf{Z}(\tau))$.

(ii) For each $d, \tilde{d} \in \Delta(\theta_*)$, $B_\nu(\tilde{d}, \tilde{d}) = \nu_{3d} B_\nu(d, \tilde{d}) \nu_{3d}$, where for each $d$, $B_\nu(d, \tilde{d})$ is the symmetric and positive definite matrix given in Eq. (17);

(iii) For each $d \in \Delta(\theta_*)$, $A_\nu(d) = \nu_{3d} A_\nu(d) \nu_{3d}$, where for each $d$, $A_\nu(d)$ is the symmetric and negative definite matrix given in Eq. (18);

(iv) $A_{\nu \pi \tau}(\tau, \pi, \tau) = A_{\nu \pi \tau}(\pi, \tau, \pi) = 0$ uniformly in $d \in \Delta(\theta_*)$;

(v) $\Theta = \Pi \times T$ and $C(\theta_*) = C(\pi) \times C(\tau)$, where $C(\pi) := \{x \in \mathbb{R}^{r_\pi} : \exists \pi' \in \Pi, x := \pi_\pi + \delta \pi', \delta \in \mathbb{R}^+\}$ and $C(\tau) := \{x \in \mathbb{R}^{r_\tau} : \exists \tau' \in T, x := \tau_\tau + \delta \tau', \delta \in \mathbb{R}^+\}$.

Assumption 7(v) restricts the parameter space $\Theta$ into the Cartesian product of separate parameter spaces $\Pi$ and $T$. This will be used to represent $L_n$ as the sum of two independent functions as discussed in Theorem 5 given below.

We further partition $\lambda$ and $\nu$ into $(\delta', \phi')' (\in \mathbb{R}^{r_\delta + r_\phi})$ and $(\xi', \psi')' (\in \mathbb{R}^{r_\xi + r_\psi})$, respectively, so that $Z(\lambda, A_{\nu \pi \lambda})$, and $B(\lambda, A_{\nu \lambda})$ can be accordingly represented as $Z(\lambda) = (Z(\delta)'', Z(\phi)'')'$,

$$A_{\nu \pi \lambda} = \begin{bmatrix} A_{(\delta', \phi')} & A_{(\delta, \phi')} & A_{(\phi, \phi')} \end{bmatrix}$$

and $B_{\nu \pi \lambda} = \begin{bmatrix} B_{(\delta', \phi')} & B_{(\delta, \phi')} & B_{(\phi, \phi')} \end{bmatrix}$.

Similarly, for each $d_\nu$, we also let $Z(\nu)(d_\nu) = Z(\xi)(d_\xi) + Z(\psi)(d_\psi)$, where for each $d_\xi$ and $d_\psi$, $Z(\nu)(d_\nu) := (Z(\xi)(d_\xi), Z(\psi)(d_\psi))' \sim N(0, B_{\nu \xi \psi}(d_\nu, d_\nu))$, and the covariance matrix can be given as

$$B_{\nu \xi \psi}(d_\nu, d_\nu) := \begin{bmatrix} B_{(\xi, \xi)}(d_\xi, d_\xi) & B_{(\xi, \psi)}(d_\xi, d_\psi) \[B_{(\psi, \xi)}(d_\psi, d_\xi) & B_{(\psi, \psi)}(d_\psi, d_\psi) \end{bmatrix}.$$
Assumption 8 (Inference) (i) For each \( d_\nu \) and for the symmetric and positive definite \( B^{(v,v)}_s (d_\nu, d_\nu) \) given in Eq. (19), \( Z^{(v)}(d_\nu) = Z^{(v)'}(d_\nu)' \), such that \( Z^{(v)}(d_\nu) \sim N(\theta, B^{(v,v)}_s (d_\nu, d_\nu)) \);

(ii) For each \( d_\nu \) and for the symmetric and negative definite \( A^{(v,v)}_s (d_\nu) \) given in Eq. (20), \( A^{(v,v)}_s (d_\nu) = \nu_2 A^{(v,v)'}_s (d_\nu) \nu_2 \).

We note that Assumption 8 does not impose any restriction on \( A^{(\lambda, \lambda)}_s \) and \( B^{(\lambda, \lambda)}_s \) in terms of their signs because Assumption 7 will always be assumed together with Assumption 8. It follows from Assumption 7 that both \(-A^{(\lambda, \lambda)}_s\) and \(B^{(\lambda, \lambda)}_s\) are symmetric and positive definite, although Assumptions 8(i) and (ii) are not implied by Assumption 7.

Given the partition of \( \theta \), we let \( \delta \) or \( \xi \) be the parameter of interest, and the hypotheses are given as

\[
\begin{align*}
H'_0 : \delta_* &= \delta_0; \\
H'_1 : \delta_* &\neq \delta_0
\end{align*}
\text{ or }
\begin{align*}
H''_0 : \xi_* &= \xi_0; \\
H''_1 : \xi_* &\neq \xi_0
\end{align*}
\]

We note that the role of the parameters in the first hypotheses (\( H'_0 \) versus \( H'_1 \)) is different from those in the second hypotheses (\( H''_0 \) versus \( H''_1 \)). That is, the directional derivative with respect to \( d_\delta \) is linear, whereas it is nonlinear with respect to \( d_\xi \).

For future reference, we also let \( \Theta_0 \) be the parameter space constrained by the null hypotheses. That is, \( \Theta_0 \) is either \( \Theta_0^{(\delta)} := \{ \theta \in \Theta : \delta = \delta_0 \} \) or \( \Theta_0^{(\xi)} := \{ \theta \in \Theta : \xi = \xi_0 \} \) depending on the hypotheses of interest to the researcher.

We introduce new notation and save space by handling these hypotheses at the same time. The parameter \( \pi \) is reorganized into \( \eta := (\mu', \omega')' \) such that \( \mu \) is either \( \delta \) or \( \xi \), and \( \omega \) is the rest of \( \pi \). Thus, \( \eta \) is either \((\delta', \omega')' = (\delta', \phi', \xi', \psi')' \) or \((\xi', \omega')' = (\xi', \delta', \phi', \psi')' \). We also let \( H, M, \) and \( \Omega \) be the parameter spaces of \( \eta, \mu, \) and \( \omega \), respectively, so that the hypotheses given as \( H'_0 \) versus \( H'_1 \) or \( H''_0 \) versus \( H''_1 \) can be shortened into

\[
H_0 : \mu_* = \mu_0 \text{ versus } H_1 : \mu_* \neq \mu_0,
\]

where \( \mu_0 \) is either \( \delta_0 \) or \( \xi_0 \). We also let \( d_\eta := (d_\mu', d_\omega')' \) and \( Z^{(\eta)}(d_\eta) := (Z^{(\mu)}(d_\mu), Z^{(\omega)}(d_\omega))' \).
respectively, and also reformulate \( A_s^{(\pi, \pi)}(d_\pi) \) and \( B_s^{(\pi, \pi)}(d_\pi, \tilde{d}_\pi) \) into

\[
A_s^{(\eta, \eta)}(d_\mu, d_\omega) := \begin{bmatrix}
A_s^{(\mu, \mu)}(d_\mu) & A_s^{(\mu, \omega)}(d_\mu, d_\omega) \\
A_s^{(\omega, \mu)}(d_\omega, d_\mu) & A_s^{(\omega, \omega)}(d_\omega)
\end{bmatrix}
\]

and

\[
B_s^{(\eta, \eta)}((d_\mu, d_\omega), (\tilde{d}_\mu, \tilde{d}_\omega)) := \begin{bmatrix}
B_s^{(\mu, \mu)}(d_\mu, \tilde{d}_\mu) & B_s^{(\mu, \omega)}(d_\mu, \tilde{d}_\omega) \\
B_s^{(\omega, \mu)}(d_\omega, \tilde{d}_\mu) & B_s^{(\omega, \omega)}(d_\omega, \tilde{d}_\omega)
\end{bmatrix},
\]

respectively. By these reformulations, Assumption 7(\( \nu \)) can also be restructured as given in the following assumption.

**Assumption 8 (Inference)** (iii) \( H = M \times \Omega \) and \( C(\pi_s) = C(\mu_s) \times C(\omega_s) \), where \( C(\mu_s) := \{ x \in \mathbb{R}^r : \exists \mu' \in M, x := \mu_s + \delta \mu', \delta \in \mathbb{R}^+ \} \) and \( C(\omega_s) := \{ x \in \mathbb{R}^{r_w} : \exists \omega' \in \Omega, x := \omega_s + \delta \omega', \delta \in \mathbb{R}^+ \} \). \( \square \)

### 3.1 Likelihood-Ratio Test Statistic

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

\[
\mathcal{L}\mathcal{R}_n := 2\{ L_n(\hat{\theta}_n) - L_n(\hat{\theta}_n) \},
\]

by following the definition in White (1994), where \( \hat{\theta}_n \) is such that \( L_n(\hat{\theta}_n) := \sup_{\theta \in \Theta_n} L_n(\theta) \). Thus, this QLR test statistic can be analyzed similarly to the standard QLR test statistic.

Specifically, we can split \( \mathcal{L}\mathcal{R}_n \) into \( \mathcal{L}\mathcal{R}_n^{(1)} \) and \( \mathcal{L}\mathcal{R}_n^{(2)} \) such that \( \mathcal{L}\mathcal{R}_n^{(1)} := 2\{ L_n(\hat{\theta}_n) - L_n(\theta_s) \} \) and \( \mathcal{L}\mathcal{R}_n^{(2)} := 2\{ L_n(\hat{\theta}_n) - L_n(\theta_s) \} \). Theorem 4(iii) already provides the asymptotic distribution of \( \mathcal{L}\mathcal{R}_n^{(1)} \), and this is further simplified if Assumption 7 is imposed. We examine this by first letting \( \Delta(\pi_s) := \{ x \in \mathbb{R}^{r_\pi} : \pi_s + x \in \text{cl}\{C(\pi_s)\}, \|x\| = 1 \} \), \( \Delta(\tau_s) := \{ x \in \mathbb{R}^{r_\tau} : \tau_s + x \in \text{cl}\{C(\tau_s)\}, \|x\| = 1 \} \) and denoting their representative components by \( s_\pi(= (s_{\lambda}, s_{\nu})) \) and \( s_\tau \), respectively. The different notation \( s \) is used to distinguish it from \( d \) in \( \Delta(\theta_s) \). We note that \( \Delta(\pi_s) \) and \( \Delta(\tau_s) \) are subsets of \( \Delta(\theta_s) \), so that they can be easily defined. The following theorem provides the asymptotic behavior of \( \mathcal{L}\mathcal{R}_n^{(1)} \) using these additional conditions.

**Theorem 5** Given Assumptions 1 to 7,

\[
2\{ L_n(\hat{\theta}_n) - L_n(\theta_s) \} \Rightarrow \mathcal{H}_1 + \mathcal{H}_2 := \sup_{s_\pi \in \Delta(\pi_s)} \max[0, \hat{y}(\pi)(s_\pi)]^2 + \sup_{s_\tau \in \Delta(\tau_s)} \max[0, \hat{y}(\tau)(s_\tau)]^2,
\]
where for each $s_\pi$ and $s_\tau$,
\[
\gamma^{(\pi)}(s_\pi) := \{-t^2_2 A_{\pi}(s_\pi)t_2\}^{-1/2} Z^{(\pi)}(s_\pi), \quad \gamma^{(\tau)}(s_\tau) := \{-A_s^{(\tau,\tau)}(s_\tau)\}^{-1/2} Z^{(\tau)}(s_\tau),
\]
respectively, and $Z^{(\pi)}(s_\pi) := s_\chi^\prime Z^{(\lambda)} + Z^{(\nu)}(s_\nu)$.

Essentially, the orthogonality condition in Assumption 7(iv) and the parameter space condition in Assumption 7(v) separate $L_n$ into the sum of two pieces and yield Theorem 5. This also implies that we can ignore the pieces containing the orthogonal nuisance parameters $\tau$ (resp. $\pi$) when testing the null hypothesis containing only $\pi$ (resp. $\tau$).

Theorem 5 provides motivation to examine the asymptotic distribution of $LR_n^{(2)}$ because its weak limit contains the same component as given in Theorem 5. The following theorem shows the asymptotic behavior of $\hat{\theta}_n$ and $LR_n^{(2)}$.

**Theorem 6** (i) Given Assumptions 1 to 3 and $H_0$, $\hat{\theta}_n$ converges to $\theta_*$ a.s.$-P$;

(ii) Given Assumptions 1 to 8 and $H_0$, $LR_n^{(2)} \Rightarrow H_0 + H_2 := \sup_{s_\omega \in \Delta(\omega_+)} \max[0, \gamma^{(\omega)}(s_\omega)]^2 + H_2$, where for each $s_\omega$, $\gamma^{(\omega)}(s_\omega) := \{-A_s^{(\omega,\omega)}(s_\omega)\}^{-1/2} Z^{(\omega)}(s_\omega)$ and $\Delta(\omega_+):=\{x \in \mathbb{R}^r_+ : \omega_+ + x \in \text{cl}\{C(\omega_+)\}, \|x\| = 1 \}$.

We note that $H_2$ in Theorem 6(ii) is the same random variable as that in Theorem 5(ii).

Several remarks pertaining to Theorem 6 are in order. First, the consequence in Theorem 6(ii) can be understood as a corollary of Theorem 5. For example, if $\mu = \delta$, $\omega = (\phi', \nu')'$, so that it follows that
\[
LR_n^{(2)} \Rightarrow \sup_{(s_\phi, s_\nu) \in \Delta(\omega_+)} \left\{ \frac{\max[0, s_\phi^\prime Z^{(\phi)} + Z^{(\nu)}(s_\nu)]^2}{s_\phi^\prime(-A_s^{(\phi,\phi)})(s_\phi) - A_s^{(\nu,\nu)}(s_\nu) - 2s_\phi^\prime A_s^{(\phi,\nu)}(s_\nu)} \right\} + H_2.
\]
from Theorem 6. In a similar manner, if $\mu = \xi$, $\omega = (\lambda', \psi')'$, so that
\[
LR_n^{(2)} \Rightarrow \sup_{(s_\lambda, s_\psi) \in \Delta(\omega_+)} \left\{ \frac{\max[0, s_\lambda^\prime Z^{(\lambda)} + Z^{(\psi)}(s_\psi)]^2}{s_\lambda^\prime(-A_s^{(\lambda,\lambda)})(s_\lambda) - A_s^{(\psi,\psi)}(s_\psi) - 2s_\lambda^\prime A_s^{(\lambda,\psi)}(s_\psi)} \right\} + H_2.
\]
Second, the weak limit given in Theorem 6(ii) is achieved jointly with that of $LR_n^{(1)}$ because all of these are obtained by applying the CMT to Theorem 4(i). Furthermore, as pointed out above, $H_2$ in $LR_n^{(2)}$ is identical to $H_2$ of $LR_n^{(1)}$. This trivially follows from the fact that they stem from the same sample statistics. From these facts, the asymptotic null distribution of the QLR test statistic immediately follows, and we formally state it in the following theorem.

**Theorem 7** Given Assumptions 1 to 8, and $H_0$, $LR_n \Rightarrow H_1 - H_0$. 

29
The asymptotic null distribution in Theorem 7 can have various forms depending on the model properties under consideration. For example, if \( \mu = \delta \) and \( r_\psi = 0 \), \( \pi = \lambda, \phi = \omega \), and

\[
\mathcal{L}R_n \Rightarrow \sup_{s_\lambda \in \Delta(\lambda_*)} \left\{ \max[0, s_\lambda'Z(\lambda)]^2 \right\} - \sup_{s_\phi \in \Delta(\phi_*)} \left\{ \frac{\max[0, s_\phi'Z(\phi)]^2}{s_\phi'(-A_s(\phi,\phi)s_\phi)} \right\}
\]

under \( H_0 \). Furthermore, if \( \lambda_* \) is an interior element, applying the proof of Corollary 1(vii) shows that

\[
(23) \quad \mathcal{L}R_n \Rightarrow (\tilde{Z}(\delta))'(-\tilde{A}_s(\delta,\delta))^{-1}(\tilde{Z}(\delta)),
\]

where \( \tilde{Z}(\delta) := Z(\delta) - (A_s(\delta,\phi))(A_s(\phi,\phi))^{-1}Z(\phi) \) and \( \tilde{A}_s(\delta,\delta) := A_s(\delta,\delta) - (A_s(\delta,\phi))(A_s(\phi,\phi))^{-1}(A_s(\phi,\phi))' \). Therefore, the same result is obtained as for the standard D model analysis. If \( \lambda_* \) is a boundary parameter, the QLR test statistic has a different asymptotic null distribution depending on the property of the associated parameter space on the boundary. For example, if \( \delta \in [\delta_\ast, \bar{\delta}] \subset \mathbb{R} \) and \( \phi_\ast \) is an interior element of its parameter space, so that \( \tilde{A}_s(\delta,\delta) \) and \( \tilde{Z}(\delta) \) are now scalars, it follows that

\[
\sup_{s_\lambda \in \Delta(\lambda_*)} \left\{ \frac{\max[0, s_\lambda'Z(\lambda)]^2}{s_\lambda'(-A_s(\lambda,\lambda))s_\lambda} \right\} = (\tilde{Z}(\delta))'(-\tilde{A}_s(\delta,\delta))^{-1}(\tilde{Z}(\delta))1_{\tilde{s}_\delta > 0} + (Z(\phi))'(A_s(\phi,\phi))^{-1}(Z(\phi)),
\]

where \( \tilde{s}_\lambda := (\tilde{s}_\delta, \tilde{s}_\phi) \) solves the LHS. Furthermore, it is a standard exercise to show that \( \tilde{Z}(\delta) < 0 \) if and only if \( \tilde{s}_\delta \leq 0 \). Thus,

\[
(24) \quad \mathcal{L}R_n \Rightarrow (-\tilde{A}_s(\delta,\delta))^{-1}\max[0, \tilde{Z}(\delta)]^2
\]

under \( H_0 \) because

\[
\sup_{s_\phi \in \Delta(\phi_*)} \left\{ \frac{\max[0, s_\phi'Z(\phi)]^2}{s_\phi'(-A_s(\phi,\phi))s_\phi} \right\} = (Z(\phi))'(A_s(\phi,\phi))^{-1}(Z(\phi))
\]

by the proof of Corollary 1(vii).

As another example, if \( \mu = \xi \) and \( r_\psi = 0 \), a different asymptotic null distribution is obtained. In this case, \( \lambda = \omega \) and \( \pi = (\lambda', \xi') \), so that

\[
(25) \quad \mathcal{H}_0 = \sup_{s_\lambda \in \Delta(\lambda_*)} \left\{ \frac{\max[0, s_\lambda'Z(\lambda)]^2}{s_\lambda'(-A_s(\lambda,\lambda))s_\lambda} \right\},
\]

and

\[
(26) \quad \mathcal{H}_1 = \sup_{(s_\lambda, s_\xi) \in \Delta(\pi_*)} \left\{ \frac{\max[0, Z(\xi)(s_\xi) + s_\lambda'Z(\lambda)]^2}{-A_s(\xi,\xi)(s_\xi) + 2s_\lambda'(A_s(\lambda,\xi)(s_\xi)) + s_\lambda'(-A_s(\lambda,\lambda))s_\lambda} \right\}
\]

by applying Eq. (22). Therefore, \( \mathcal{H}_1 - \mathcal{H}_0 \) becomes the asymptotic null distribution as given by Theorem 7. Given this, we can obtain a further regularized asymptotic null distribution by imposing an additional parameter space condition.
Assumption 9 (Benchmark) (i) \( r_\psi = 0; \)

(ii) \( \lambda_* \) is an interior element of \( \Lambda; \)

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

(iv) \( C(\pi_*) = \mathbb{R}^{r_\lambda} \times C(\xi_*), \) where \( C(\xi_*) := \{ x \in \mathbb{R}^{r_\lambda} : \exists \xi' \in \Xi, x := \xi_* + \delta \xi', \delta \in \mathbb{R}^+ \} \).

We note that the null model is \( D \) (Assumption 9(ii)) and does not contain a boundary parameter by Assumption 9(ii). Furthermore, Assumption 9(iv) ensures that \( C(\pi_*) \) is a Cartesian product of two cones, and one of them is \( \mathbb{R}^{r_\lambda} \), which follows from the interiority condition in Assumption 9(ii). By adding this condition, we can obtain the following corollary.

Corollary 2  Given Assumptions 1 to 9, (i)

\[
\mathcal{LR}_n^{(1)} \Rightarrow \sup_{s_\xi \in \Delta(\xi_*)} \max[0, \tilde{y}^i(\xi)(s_\xi)]^2 + (Z^o)^/(\lambda_\lambda)^{-1}(Z^o) + \mathcal{H}_2,
\]

where for each \( s_\xi \in \Delta(\xi_*) := \{ x \in \mathbb{R}^{r_\xi} : \xi_* + x \in \text{cl}\{ C(\xi_*) \}, \| x \| = 1 \} \),

\[
\tilde{y}^i(\xi)(s_\xi) := (\tilde{A}^i(\xi)(s_\xi))^{-1/2}(\tilde{Z}(\xi)) \quad \text{and also} \quad \tilde{Z}(\xi)(s_\xi) := Z^o(s_\xi) - A^o(\xi)(s_\xi)(A^o(\lambda_\lambda))^{-1}Z^o;
\]

(ii) if further \( H_0' \) holds, \( \mathcal{LR}_n \Rightarrow \sup_{s_\xi \in \Delta(\xi_*)} \max[0, \tilde{y}^i(\xi)(s_\xi)]^2. \)

We note that \( \tilde{Z}(\xi) \) is obtained by projecting \( Z^o(\xi) \) on \( Z^o(\lambda) \). This projection is needed as the QLR test is constructed to minimize the parameter estimation error arising from estimating the unknown nuisance parameter \( \lambda_* \). Furthermore, Corollary 2(ii) follows by combining Theorem 7, Eqs. (25), (26), and Corollary 2. Here, we note that it also easily follows that \( \mathcal{H}_0 = (Z^o)^/(\lambda_\lambda)^{-1}(Z^o) \) from Eq. (25) and Assumption 9(ii).

3.2 Wald Test Statistic

We can construct another test statistic by combining the D-D model with the testing principle in Wald (1943). The distance between the extremum estimator \( (\hat{\mu}_n) \) and the null parameter value \( (\mu_0) \) provides a clue to distinguish the null from the alternative hypothesis.

Before applying the Wald testing principle, we first examine the asymptotic distribution of the distance between \( \hat{\mu}_n \) and \( \mu_0 \) under the null. We note that the distance between \( \hat{\theta}_n \) and \( \theta_* \) captured by \( \hat{h}_n \) cannot be used to test the null hypothesis in the way given in Theorem 4 because the inference on \( \mu_* \) can be
mixed with that of the other nuisance parameters \( \omega_s \) and \( \tau_s \). The distance \( \hat{h}_n \) needs to be broken into pieces corresponding to \( \mu, \omega, \) and \( \tau \). This can be easily achieved by separating the set of directions into sets of the relevant directions. We note that for any \( \hat{h}_n d \) and \( d \in \Delta(\theta_s) \), there are \( h^{(\mu)}, h^{(\omega)}, h^{(\tau)} \), and \((s_\mu, s_\omega, s_\tau) \) such that \( \hat{h}_n d = (h^{(\mu)} s_\mu', h^{(\omega)} s_\omega', h^{(\tau)} s_\tau') \) if each parameter space of \( \mu, \omega, \) and \( \tau \) can be approximated by a cone and the parameter space of \( \theta \) can also be approximated by a Cartesian product of these cones as given in Assumptions 7 and 8. Thus,

\[
\sup_{d} \sup_{h} L_n(\theta_s + hd) = \sup_{\{s_\mu, s_\omega, s_\tau\}} \sup_{\{h^{(\mu)}, h^{(\omega)}, h^{(\tau)}\}} L_n(\theta_s + (h^{(\mu)} s_\mu', h^{(\omega)} s_\omega', h^{(\tau)} s_\tau')) ,
\]

and

\[
\hat{h}_n (d_n, \hat{h}_n) = \left( \hat{h}_n^{(\mu)}(s_{\mu,n}, \hat{s}_{\omega,n}, \hat{s}_{\tau,n}), \hat{h}_n^{(\omega)}(s_{\mu,n}, \hat{s}_{\omega,n}, \hat{s}_{\tau,n}), \hat{h}_n^{(\tau)}(s_{\mu,n}, \hat{s}_{\omega,n}, \hat{s}_{\tau,n}) \right),
\]

where \( (d_n, \hat{h}_n) \) and \( (\hat{s}_\mu', n, \hat{s}_\omega', n, \hat{s}_\tau', n, \hat{h}_n^{(\mu)}, \hat{h}_n^{(\omega)}, \hat{h}_n^{(\tau)})' \) solve the LHS and RHS of Eq. (27), respectively.

The Wald test statistic can be defined by associating \( \hat{h}_n^{(\mu)} \) with the Wald testing principle. More specifically, we first examine the asymptotic distribution of \( \hat{h}_n := (\hat{h}_n^{(\mu)}, \hat{h}_n^{(\omega)}, \hat{h}_n^{(\tau)})' \) and provide an appropriate definition of a test statistic using Wald’s (1943) principle. Before discussing this distribution, we provide stochastic processes relevant to the limiting behavior of \( \hat{h}_n \). First, for each \((s_\mu, s_\omega, s_\tau) \) in \( \Delta(\mu_s) \times \Delta(\omega_s) \times \Delta(\tau_s) \), we let

\[
\begin{bmatrix}
G^{(\mu)}(s_\mu, s_\omega) \\
G^{(\omega)}(s_\mu, s_\omega) \\
G^{(\tau)}(s_\tau)
\end{bmatrix} = 
\begin{bmatrix}
G^{(\eta)}(s_\mu, s_\omega) \\
G^{(\tau)}(s_\tau)
\end{bmatrix} := 
\begin{bmatrix}
-A^{(\eta,\eta)}(s_\mu, s_\omega) & -A^{(\eta,\tau)}(s_\mu, s_\omega) \\
-A^{(\tau,\eta)}(s_\tau) & -A^{(\tau,\tau)}(s_\tau)
\end{bmatrix},
\]

where for each \((s_\mu, s_\omega) \), \( Z^{(\eta)}(s_\mu, s_\omega) := (Z^{(\mu)}(s_\mu), Z^{(\omega)}(s_\omega))' \). Next, for each \((s_\mu, s_\omega) \) in \( \Delta(\mu_s) \times \Delta(\omega_s) \), we also let

\[
\begin{bmatrix}
\hat{G}^{(\mu)}(s_\mu) \\
\hat{G}^{(\omega)}(s_\omega)
\end{bmatrix} := 
\begin{bmatrix}
\{A^{(\mu,\mu)}(s_\mu)\}^{-1}Z^{(\mu)}(s_\mu) \\
\{A^{(\omega,\omega)}(s_\omega)\}^{-1}Z^{(\omega)}(s_\omega)
\end{bmatrix}.
\]

These constitute the asymptotic behavior of \( \hat{h}_n \) as established by the following Lemma 4.

**Lemma 4** Given Assumptions 1 to 8,

\[
\sqrt{n} \hat{h}_n \Rightarrow 
\begin{bmatrix}
\hat{G}^{(\mu)} \\
\hat{G}^{(\omega)} \\
0
\end{bmatrix} \times 
\begin{bmatrix}
\max[0, \hat{G}^{(\mu)}] \times 1_{\{\hat{G}^{(\mu)} \geq 0 > \hat{G}^{(\omega)}\}} \\
\max[0, \hat{G}^{(\omega)}] \times 1_{\{\hat{G}^{(\omega)} > 0 \geq \hat{G}^{(\mu)}\}} \\
\max[0, \hat{G}^{(\tau)}]
\end{bmatrix}.
\]

We can provide a heuristic explanation of Lemma 4 as follows. First, we note that both \( (\hat{h}_n^{(\mu)}, \hat{h}_n^{(\omega)})' \) and \( \hat{h}_n^{(\tau)} \) are initially defined on \( \Delta(\mu_s) \times \Delta(\omega_s) \times \Delta(\tau_s) \), but their asymptotic counterparts in the RHS of
Eq. (28) are respectively defined on $\Delta(\mu_*) \times \Delta(\omega_*)$ and $\Delta(\tau_*)$. This is mainly due to Assumption 7(iv and v). By supposing that $A^{(n,\tau)}(d_{\eta}, d_{\tau}) = 0$, the maximization in the RHS of Eq. (27) is asymptotically separated into two independent maximization procedures as given in Theorem 5, resulting in the different domains at the limit. Second, we note that $\hat{h}_n^*(\mu)$ and $\hat{h}_n^*(\omega)$ cannot be less than zero although they both can be zero. Thus, for each $(\mu, \omega, \tau)$, we can possibly obtain one of four different events:

(i) $\hat{h}_n^*(\mu)(\mu, \omega, \tau) > 0$ and $\hat{h}_n^*(\omega)(\mu, \omega, \tau) > 0$; (ii) $\hat{h}_n^*(\mu)(\mu, \omega, \tau) > 0$, $\hat{h}_n^*(\omega)(\mu, \omega, \tau) = 0$;

(iii) $\hat{h}_n^*(\mu)(\mu, \omega, \tau) = 0$, $\hat{h}_n^*(\omega)(\mu, \omega, \tau) > 0$; or (iv) $\hat{h}_n^*(\mu)(\mu, \omega, \tau) = 0$, $\hat{h}_n^*(\omega)(\mu, \omega, \tau) = 0$.

These different events are determined by the sign of $G^{(n)}$, and Lemma 4 distinguishes them using the separate indication functions. In particular, for a given direction $(\mu, \omega, \tau)$, if say $\hat{h}_n^*(\omega)(\mu, \omega, \tau) = 0$ but $\hat{h}_n^*(\mu)(\mu, \omega, \tau) > 0$, the nuisance parameter $\omega$ does not affect the asymptotic distribution of $\hat{h}_n^*(\mu)(\mu, \omega, \tau)$, so that $\sqrt{n}\hat{h}_n^*(\mu)(\mu, \omega, \tau) \Rightarrow \hat{G}^*(\mu)$. This follows because $\hat{h}_n^*(\omega)(\mu, \omega, \tau) = 0$ implies that $\hat{\omega}_n(\mu, \omega, \tau) = \omega_*$, so that the nuisance parameter estimation error no longer exists. Finally, if both $\omega_*$ and $\mu_*$ are interior elements, the second term in the RHS of Eq. (28) can be virtually ignored when maximizing $L_n$ over the directions.

Given these facts, we define the Wald test statistic as

$$ W_n := \sup_{\mu \in \Delta(\mu_0)} n\{h_n(\mu)(\mu)\}\{\hat{W}_n(\mu)\}\{\hat{h}_n(\mu)(\mu)\}, $$

where $\hat{h}_n(\mu)(\mu)$ is such that for each $\mu \in \Delta(\mu_0)$,

$$ L_n(\mu + \hat{h}_n(\mu)(\mu), \hat{\omega}_n(\mu), \hat{\tau}_n(\mu)) = \sup_{\{h_n(\mu), \omega, \tau\}} L_n(\mu + h_n(\mu)(\mu), \omega, \tau), $$

and $\hat{W}_n$ is a weight function, which can estimate a non-random positive function $W_*$ uniformly on $\Delta(\mu_0)$. As this weight function is an important component of the Wald test statistic, we formally state its condition in the following assumption.

**Assumption 10 (Weight Function I)** $\hat{W}_n$, which is strictly positive uniformly on $\Delta(\mu_0)$ and $n$, converges to $W_*$ a.s.--$\mathbb{P}$ as $n$ tends to infinity, where $W_*$ is strictly positive and bounded from above uniformly on $\Delta(\mu_0)$. That is, $\sup_{\mu} |\hat{W}_n(\mu) - W_*(\mu)| \to 0$ a.s.--$\mathbb{P}$. \hfill \Box

Under the standard Wald testing principle, the weight function $W_*$ is assumed to be identical to the asymptotic variance of $\sqrt{n}\hat{h}_n(\mu)$. On the other hand, if the parameter of interest is on the boundary, the weight function needs to be carefully chosen because the asymptotic variance of the weak limit of $\sqrt{n}\hat{h}_n(\mu)$ can have different forms.

From these aspects, the asymptotic null distribution of the Wald test statistic can be obtained as follows.
Theorem 8  Given Assumptions 1 to 8, 10, and $H_0$,

$$W_n \Rightarrow \sup_{s_\mu \in \Delta(\mu_0)} G(\mu)(s_\mu, \bar{s}_\omega(s_\mu))W_n(s_\mu)G'(\mu)(s_\mu, \bar{s}_\omega(s_\mu))I\{\min[G(\mu)(s_\mu, \bar{s}_\omega(s_\mu)), \bar{G}(\omega)(s_\mu, \bar{s}_\omega(s_\mu))] > 0\} + \max[0, \hat{G}(\mu)(s_\mu)]W_n(s_\mu)\max[0, \hat{G}(\mu)(s_\mu)]I\{\bar{G}(\mu)(s_\mu, \bar{s}_\omega(s_\mu)) \geq \bar{G}(\omega)(s_\mu, \bar{s}_\omega(s_\mu))\}$$

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

$$(29) \quad \sup_{s_\omega \in \Delta(\omega_0)} G(\mu)(s_\mu, s_\omega)\left(-A_*(\mu, \eta)(s_\mu, s_\omega)\right)G'(\mu)(s_\mu, s_\omega)I\{\min[G(\mu)(s_\mu, s_\omega), \bar{G}(\omega)(s_\mu, s_\omega)] > 0\} = G(\mu)(s_\mu, \bar{s}_\omega(s_\mu))\left(-A_*(\mu, \eta)(s_\mu, \bar{s}_\omega(s_\mu))\right)G'(\mu)(s_\mu, \bar{s}_\omega(s_\mu)) \times I\{\min[G(\mu)(s_\mu, \bar{s}_\omega(s_\mu)), \bar{G}(\omega)(s_\mu, \bar{s}_\omega(s_\mu))] > 0\}$$

and

$$(30) \quad \sup_{s_\omega \in \Delta(\omega_0)} \max[0, \hat{G}(\mu)(s_\mu)]\left(-A_*(\mu, \eta)(s_\mu, s_\omega)\right)\max[0, \hat{G}(\mu)(s_\mu)]I\{\bar{G}(\mu)(s_\mu, \bar{s}_\omega(s_\mu)) \geq \bar{G}(\omega)(s_\mu, \bar{s}_\omega(s_\mu))\} = \max[0, \hat{G}(\mu)(s_\mu)]\left(-A_*(\mu, \eta)(s_\mu, \bar{s}_\omega(s_\mu))\right)\max[0, \hat{G}(\mu)(s_\mu)]I\{\bar{G}(\mu)(s_\mu, \bar{s}_\omega(s_\mu)) \geq \bar{G}(\omega)(s_\mu, \bar{s}_\omega(s_\mu))\}$$

respectively.

Briefly, the proof of Theorem 8 follows from the fact that for each $s_\mu$, the arguments maximizing $L_n(\mu_0 + h(\mu)s_\mu, \omega, \tau)$ are obtained by

$$\sup_{s_\omega, s_\tau = \{h(\mu), \bar{h}(\omega), \bar{h}(\tau)\}} 2\{L_n(\mu_0 + h(\mu)s_\mu, \omega, \tau) + h(\omega)s_\omega, \tau + h(\tau)s_\tau - L_n(\mu_0, \omega, \tau)\}.$$ 

For each $(s_\mu, s_\omega, s_\tau)$, we first approximate Eq. (31) by a quadratic function and note that the signs of $\tilde{h}(\mu)(s_\mu, s_\omega, s_\tau)$ and $\bar{h}(\omega)(s_\mu, s_\omega, s_\tau)$ can result in different approximations as pointed out by Lemma 4. Accordingly, the asymptotic null distribution of the Wald test statistic also depends on how the indication functions of Lemma 4 are combined with the Wald test statistic. Eqs. (29) and (30) explicitly show how these different approximations and maximization processes are combined with the the signs of $\tilde{h}(\mu)(s_\mu, s_\omega, s_\tau)$ and $\bar{h}(\omega)(s_\mu, s_\omega, s_\tau)$ at the limit. More precisely, Eq. (29) considers when both $\tilde{h}(\mu)(s_\mu, s_\omega, s_\tau)$ and $\bar{h}(\omega)(s_\mu, s_\omega, s_\tau)$ are strictly greater than zero at the limit, whereas Eq. (30) considers when $\tilde{h}(\mu)(s_\mu, s_\omega, s_\tau) > 0$ but $\bar{h}(\omega)(s_\mu, s_\omega, s_\tau) = 0$ at the limit. For the other two events, the Wald test converges to zero in probability as we detail in the proof. Theorem 8 captures these different events by using the indication functions and deliver the asymptotic null distribution by applying the CMT.

The given result in Theorem 8 can be further simplified by additionally imposing the benchmark condition in Assumption 9.
Corollary 3 Given Assumptions 1 to 9 and $H_0''$, if $\lim_{n \to \infty} \sup_{s_\xi \in \Delta(\xi_0)} |\hat{W}_n(s_\xi) + \hat{A}_s(\xi_\xi)(s_\xi)| = 0$ a.s.$-\mathbb{P}$, then $W_n \Rightarrow \lim_{n \to \infty} \sup_{s_\xi \in \Delta(\xi_0)} \max\{0, \hat{A}_s(\xi_\xi)(s_\xi)\}^2$.

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

### 3.3 Lagrange Multiplier Test Statistic

The LM test statistic for standard econometric models can be appropriately modified for D-D models where the idea is similar to the original intention of the LM test statistic to test whether the slopes of models are distributed around zero.

We define the LM test statistic as

$$LM_n := \sup_{(s_\mu, s_\omega) \in \Delta(\mu_0) \times \Delta(\omega_n)} n\hat{W}_n(s_\mu, s_\omega) \max\left\{0, \frac{-DL_n(\tilde{\theta}_n; s_\mu)}{D^2L_n(\tilde{\theta}_n; s_\mu, s_\omega)}\right\}^2,$$

where for each $(s_\mu, s_\omega)$, $\Delta(\omega_n) := \{x \in \mathbb{R}^\omega : x + \bar{\omega}_n \in \text{cl}\{C(\bar{\omega}_n)\}, \|x\| = 1\}$,

$$D^2L_n(\tilde{\theta}_n; s_\mu, s_\omega) := D^2L_n(\tilde{\theta}_n; s_\mu) - DL_n(\tilde{\theta}_n; s_\mu, s_\omega)(D^2L_n(\tilde{\theta}_n; s_\omega))^{-1}DL_n(\tilde{\theta}_n; s_\omega, s_\mu),$$

$$DL_n(\tilde{\theta}_n; s_\mu, s_\omega) := \lim_{h \to 0} h^{-1}\{DL_n(\mu_0, \bar{\omega}_n + hs_\omega, \bar{\tau}_n; s_\mu) - DL_n(\tilde{\theta}_n; s_\mu)\},$$

and $\tilde{W}_n$ is another weight function satisfying the following condition.

**Assumption 11 (Weight Function II)** (i) The unknown nuisance parameter $\omega_n$ is interior to $\Omega$;

(ii) A weight function $\tilde{W}_n$, which is strictly positive uniformly on $\Delta(\mu_0) \times \Delta(\omega_n)$ and $n$, converges to $\tilde{W}$ a.s.$-\mathbb{P}$ as $n$ tends to infinity, where $\tilde{W}$ is strictly positive and bounded from above uniformly on $\Delta(\mu_0) \times \Delta(\omega_n)$. That is, $\sup_{(s_\mu, s_\omega) \in \Delta(\mu_0) \times \Delta(\omega_n)} |\tilde{W}_n(s_\mu, s_\omega) - \tilde{W}(s_\mu, s_\omega)| \to 0$ a.s.$-\mathbb{P}$.

Several remarks pertaining to the definition of the LM test statistic are in order. First, the asymptotic structure of the LM test statistic is designed to be close to those of the QLR and Wald test statistics. Each component of the LM test statistic stems from the first and second-order directional derivatives of $L_n(\mu_0 + h(\mu) s_\mu, \bar{\omega}_n + h(\omega) s_\omega, \bar{\tau}_n + h(\tau) s_\tau)$ with respect to $s_\mu$ and $s_\omega$, where $(\mu_0', \omega_n', \bar{\tau}_n') = \tilde{\theta}_n$. Furthermore, the ‘max’ operator in the LM test statistic is related to the behavior of $\sqrt{n}h_n(\mu)(s_\mu, s_\omega, s_\tau)$, which cannot be less than zero. Second, the LM test statistic can be represented as the supremum of a squared random score function if $\tilde{W}_n(s_\mu, s_\omega)$ is asymptotically equivalent to $-n^{-1}D^2L_n(\tilde{\theta}_n; s_\mu, s_\omega)$. Third, we emphasize that
Given Assumptions 1 to 8 and by the CMT. The following theorem establishes this.

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

$$\mathcal{LM}_n \Rightarrow \sup_{(s_\mu, s_\omega) \in \Delta(\mu_n) \times \Delta(\omega_n)} \max[0, \mathcal{G}(\mu)(s_\mu, s_\omega; \Delta)] \mathcal{W}(s_\mu, s_\omega) \max[0, \mathcal{G}(\mu)(s_\mu, s_\omega; \Delta)],$$

where for each $(s_\mu, s_\omega) \in \Delta(\mu_n) \times \Delta(\omega_n)$,

$$\mathcal{G}(\mu)(s_\mu, s_\omega; \Delta) := \{-A^{(\mu, \mu)}_n(s_\mu, s_\omega)\}^{-1} \mathcal{Z}(\mu)(s_\mu; \Delta),$$

$$\mathcal{Z}(\mu)(s_\mu; \Delta) := \mathcal{Z}(\mu)(s_\mu) - (-A^{(\mu, \omega)}_n(s_\mu, s_\omega))(A^{(\omega, \omega)}_n(s_\omega))^{-1} \mathcal{Z}(\omega)(s_\omega),$$

and $\Delta(\mu_n)$ is defined on $\Delta(\mu_n)$ and $\Delta(\omega_n)$ is interior to $\Omega$. The domain $\Delta(\omega_n)$ estimates $\Delta(\omega_n)$, and the interiority condition lets $\Delta(\omega_n)$ be identical to $\Delta(\omega_n)$ for sufficiently large $n$. If $\omega_n$ is on the boundary, $\Delta(\omega_n)$ can be different from $\Delta(\omega_n)$, which may affect the asymptotic null distribution. Assumption 11(i) precludes this possibility.

Given the definition of the LM test statistic, its asymptotic null distribution is straightforwardly obtained by the CMT. The following theorem establishes this.

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

$$\mathcal{LM}_n \Rightarrow \sup_{(s_\mu, s_\omega) \in \Delta(\mu_n) \times \Delta(\omega_n)} \max[0, \mathcal{G}(\mu)(s_\mu, s_\omega)] \mathcal{W}(s_\mu, s_\omega) \max[0, \mathcal{G}(\mu)(s_\mu, s_\omega; \Delta)],$$

where for each $(s_\mu, s_\omega) \in \Delta(\mu_n) \times \Delta(\omega_n)$,

$$\mathcal{G}(\mu)(s_\mu, s_\omega; \Delta) := \{-A^{(\mu, \mu)}_n(s_\mu, s_\omega)\}^{-1} \mathcal{Z}(\mu)(s_\mu; \Delta),$$

$$\mathcal{Z}(\mu)(s_\mu; \Delta) := \mathcal{Z}(\mu)(s_\mu) - (-A^{(\mu, \omega)}_n(s_\mu, s_\omega))(A^{(\omega, \omega)}_n(s_\omega))^{-1} \mathcal{Z}(\omega)(s_\omega),$$

$$\Delta(\mu_n)$$ is defined on $\Delta(\mu_n)$, and $\Delta(\omega_n)$ is such that $\Delta(\omega_n)$ is on the boundary. This is mainly because the weak limit of $n^{-1/2} DL_n(\hat{\theta}_n; \cdot)$ is captured by $\mathcal{Z}(\mu)(\cdot; \Delta)$, which is obtained by the weak limit of the Taylor expansion of $n^{-1/2} DL_n(\hat{\theta}_n; \cdot)$ at $\omega_n$.

Theorem 9 can be further simplified if the benchmark model assumption is additionally imposed. We provide its asymptotic null distribution in the following corollary.

**Corollary 4** Given Assumptions 1 to 9 and $H_0$, if $\sup_{(s_\xi, s_\lambda) \in \Delta(\xi_n) \times \Delta(\lambda_n)} |\mathcal{W}_n(s_\xi, s_\lambda) + A^{(\xi, \xi)}_n(s_\xi)| \to 0$ a.s.–P, $\mathcal{LM}_n \Rightarrow \max_{s_\xi \in \Delta(\xi_n)} \max[0, \mathcal{G}(\xi)(s_\xi)]^2.

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

### 3.4 Examples

**3.4.1 Example 1 (continued)**

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

$$B_s(d, \tilde{d}) = \begin{bmatrix} B_s^{(\pi, \pi)}(d, \tilde{d}) & 0' \\ 0 & \frac{1}{\sigma_2^2} d_\gamma' E[Q_t Q_t'] \tilde{d}_\gamma \end{bmatrix},$$

and

$$B_s^{(\pi, \pi)}(d, \tilde{d}) = \begin{bmatrix} \frac{1}{2\sigma^2} d_\sigma^2 d_\sigma^2 & \frac{1}{2\sigma_2^2} d_\sigma^2 h(d_1, d_2) E[W_t^2] \\ \frac{1}{2\sigma_2^2} d_\sigma^2 h(d_1, d_2) E[W_t^2] & h(d_1, d_2) h(d_1, d_2) \left[ \frac{1}{2} E[W_t^4] + k(d_2/d_1, \tilde{d}_2/\tilde{d}_1) E[W_t^2] \right] \end{bmatrix},$$

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

$$h(d_1, d_2) := \frac{(d_1^2 + d_2^2)^{1/2}}{1 - m(d_2/d_1)^2},$$

and

$$k(d_2/d_1, \tilde{d}_2/\tilde{d}_1) := \frac{m(d_2/d_1)m(\tilde{d}_2/\tilde{d}_1)}{1 - m(d_2/d_1)m(\tilde{d}_2/\tilde{d}_1)}.$$
From these facts, the asymptotic null distributions of the test statistics are straightforwardly derived. First, the asymptotic null distribution of the QLR test statistic is obtained as follows.

\[
\mathcal{L} \mathcal{R}^{(1)}_n \Rightarrow \sup_{s_{\pi} \in \Delta(\pi^*)} \max[0, \mathcal{Y}^{(\pi)}(s_{\pi})]^2 + \mathcal{H}_2
\]

by Theorem 5, where \( \mathcal{Y}^{(\pi)}(s_{\pi}) \) is a standard Gaussian stochastic process with covariance structure

\[
E[\mathcal{Y}^{(\pi)}(s_{\pi}) \mathcal{Y}^{( \pi)}(s_{\pi})] = \frac{\iota_3 \tilde{B}_s^{(\pi, \pi)}(s_{\pi}, \tilde{s}_{\pi}) \iota_3}{\{\iota_3 \tilde{B}_s^{(\pi, \pi)}(s_{\pi}, s_{\pi}) \iota_3\}^{1/2} \{\iota_3 \tilde{B}_s^{(\pi, \pi)}(\tilde{s}_{\pi}, s_{\pi}) \iota_3\}^{1/2}},
\]

and \( \mathcal{H}_2 \) is a chi-squared random variable with degrees of freedom \( k + 1 \). Second,

\[
\mathcal{L} \mathcal{R}^{(2)}_n \Rightarrow \sup_{s_{\sigma^2} \in \{-1, 1\}} \max[0, \mathcal{Y}^{(\sigma^2)}(s_{\sigma^2})]^2 + \mathcal{H}_2
\]

by Theorem 6, where \( \mathcal{Y}^{(\sigma^2)} \) is a standard Gaussian stochastic process having \( E[\mathcal{Y}^{(\sigma^2)}(s_{\sigma^2}) \mathcal{Y}^{(\sigma^2)}(s_{\sigma^2})] = 1 \), implying that it is free of the direction \( s_{\sigma^2} \), so that \( \sup_{s_{\sigma^2} \in \{-1, 1\}} \max[0, \mathcal{Y}^{(\sigma^2)}(s_{\sigma^2})]^2 \) is a chi-squared random variable with one degree of freedom by the proof of Corollary 1(vii). Thus,

\[
(32) \quad \mathcal{L} \mathcal{R}_n \Rightarrow \sup_{s_2/s_1 \in [\underline{c}, \overline{c}]} \max[0, \bar{\mathcal{Y}}^{(\theta)}(s_1, s_2)]^2
\]

by Corollary 2(ii), where \( \bar{\mathcal{Y}}^{(\theta)} \) is a standard Gaussian stochastic process with covariance structure

\[
c(s_2/s_1, \bar{s}_2/\bar{s}_1) = \frac{1}{\{c(s_2/s_1, \bar{s}_2/\bar{s}_1)\}^{1/2} \{c(\bar{s}_2/\bar{s}_1, \bar{s}_2/\bar{s}_1)\}^{1/2}}
\]

and for each \((s_2/s_1, \bar{s}_2/\bar{s}_1)\),

\[
c(s_2/s_1, \bar{s}_2/\bar{s}_1) := \frac{1}{2} \var(W_t^2) + k(s_2/s_1, \bar{s}_2/\bar{s}_1) E[W_t^2]^2.
\]

This structure is homogenous of degree zero with respect to \( s_1 \) and \( s_2 \), so that \( \bar{\mathcal{Y}}^{(\theta)} \) can be equivalently represented as a function of \( s_2/s_1 \). This explains why the maximization of Eq. (32) is taken over \([\underline{c}, \overline{c}]\).

Second, we can also apply the Wald test statistic to this model. By the condition of Corollary 3, an appropriate weight function needs to be chosen, and we define the weight function as

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

where \( \widehat{E}_n[W_t^2] := n^{-1} \sum_{t=1}^n W_t^2 \) and \( \bar{\var}_n[W_t^2] := n^{-1} \sum_{t=1}^n W_t^2 - (n^{-1} \sum_{t=1}^n W_t^2)^2 \). It is trivial to show that this estimator satisfies Assumption 10. Then the Wald test statistic is defined as

\[
\mathcal{W}_n := n \{\hat{h}_n^{(\theta)}(s_2/s_1)\} \widehat{W}_n(s_2/s_1, s_2/s_1) \{\hat{h}_n^{(\theta)}(s_2/s_1)\},
\]

38
and it has the following asymptotic null behavior:

\[(33) \quad \mathcal{W}_n \Rightarrow \sup_{s_2/s_1 \in [c, c]} \max \{0, \tilde{Y}^{(\theta)}(s_1, s_2)\}^2 \]

by Corollary 3, where \(\tilde{h}_n(\theta, s_2)\) is such that

\[L_n(\tilde{\gamma}_n, \tilde{\sigma}_n^2, \tilde{h}_n(\theta, s_2), s_2) = \sup_{(h(\theta), \gamma^2)} L_n(\gamma, \sigma^2, h(\theta), s_1, h(\theta), s_2)\]

and \(s_1^2 + s_2^2 = 1\).

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

\[\mathcal{LM}_n := \sup_{s_2/s_1 \in [c, c]} \left\{ \max \left[ 0, \frac{1}{2} \left( \mathbf{U}^n(\tilde{\gamma}_n) \right)' \mathbf{W}_n^{1/2} \mathbf{U}^n(\tilde{\gamma}_n) \right] \right\}^2 \mathcal{W}_n(s_2/s_1, s_2/s_1), \]

where

\[DL_n(\tilde{\gamma}_n, \tilde{\sigma}_n^2, 0; s_1, s_2) = \{2\tilde{\sigma}_n^2\}^{-1} \left\{ \mathbf{U}^n(\tilde{\gamma}_n)' \mathbf{W}_n(\tilde{\gamma}_n) - \tilde{\sigma}_n^2 \text{tr}(\mathbf{W}_n(\tilde{\gamma}_n)) \right\}; \]

\[\tilde{D}^2 L_n(\tilde{\gamma}_n, \tilde{\sigma}_n^2, 0; s_1, s_2, s_{\sigma^2}) := \frac{1}{2} \left\{ \text{tr}(\mathbf{W}_n(\tilde{\gamma}_n)') \mathbf{U}^n(\tilde{\gamma}_n) \right\} - \frac{2}{\tilde{\sigma}_n^2} \mathbf{U}^n(\tilde{\gamma}_n)' \mathbf{W}_n(\tilde{\gamma}_n) - \mathbf{U}^n(\tilde{\gamma}_n) \sum_{s_{\sigma^2}} \mathbf{U}^n(\tilde{\gamma}_n); \]

\((\tilde{\gamma}_n, \tilde{\sigma}_n^2)\) is such that \(L_n(\tilde{\gamma}_n, \tilde{\sigma}_n^2, 0) = \sup_{(\gamma, \sigma^2)} L_n(\gamma, \sigma^2, 0); \) and the same weight matrix is used as for the Wald test statistic. Here, \(\tilde{D}^2 L_n(\tilde{\gamma}_n, \tilde{\sigma}_n^2, 0; s_1, s_2, s_{\sigma^2})\) is virtually indexed only by \((s_1, s_2)\) because \(s_{\sigma^2}\) disappears from its construction. Corollary 4 now implies that

\[(34) \quad \mathcal{LM}_n \Rightarrow \sup_{s_2/s_1 \in [c, c]} \max \{0, \tilde{Y}^{(\theta)}(s_1, s_2)\}^2, \]

so that all three test statistics are asymptotically equivalent.

The distribution of the Gaussian stochastic process \(\tilde{Y}^{(\theta)}\) can be uncovered by a simulation method. We note that for each \((s_1, s_2)\), it has the same covariance structure as

\[\tilde{Y}^{(\theta)}(s_1, s_2) := \frac{1}{t(s_2/s_1, s_2/s_1)^{1/2}} \left[ \frac{\text{var}(W_i^2)}{2} \right]^{1/2} Z_0 + E[W_i^2] \sum_{j=1}^{\infty} m(s_2/s_1)^2 Z_j, \]

where \(Z_j \sim \text{IID } N(0, 1). \) Due to the IID condition, it is not hard to generate the distribution of \(\tilde{Y}^{(\theta)}\) by a simulation method.
There are several cautions in generating the distribution. First, \( \text{var}(W_t^2) \) and \( E[W_t^2] \) are unknown, and we should instead use their consistent estimators. Second, the index \( j \) must be truncated at a moderately large number so that it does not significantly affect the null distribution. For this examination, we implement a Monte Carlo simulation. The DGP for \( Y_t = U_t \sim \text{IID } N(0,1) \) and \( W_t \sim \text{IID } N(0,1) \), which is independent of \( U_t \). We assume that the parameters other than \( \alpha^*, \sigma^2, \theta_1^*, \) and \( \theta_2^* \) are known and also let \( \xi = 0.5, \bar{c} = 1.5 \). The total number of replications is 2,000. Figure 1 shows the Q-Q plot between \( \sup_{(s_1,s_2)} \max[0, \hat{Y}(\theta)(s_1,s_2)]^2 \) and \( \sup_{(s_1,s_2)} \max[0, \hat{Y}(\theta)(s_1,s_2)]^2 \), where for each \( (s_1, s_2) \),

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

and \( \text{var}_n(W_t^2) \) and \( \hat{E}_n[W_t^2] \) are the method of moments estimators for \( \text{var}(W_t^2) \) and \( E[W_t^2] \), respectively, when \( n = 500 \). We note that the Q-Q line in Figure 1 is also identical to the 45-degree line, which implies that estimating \( \text{var}(W_t^2) \) and \( E[W_t^2] \) does not modify the asymptotic distribution significantly. Figure 2 shows the empirical distributions of the QLR test statistic for various sample sizes and the asymptotic null distribution obtained by the simulation method. As \( n \) becomes large, the empirical distribution of the QLR test statistic approaches the asymptotic distribution.

This aspect becomes more apparent with a reduced number of unknown nuisance parameters. As another simulation, we assume that \( \alpha^* \) is also known. Figure 3 shows that the empirical null distribution of the QLR test statistic is closer to the asymptotic distribution even when the sample size is much smaller than in the previous case.

\[\square\]

### 3.4.2 Example 2 (continued)

We can also test the efficient production hypothesis by the QLR, Wald, and LM test statistics. For this examination, we represent the parameters in the model in terms of the parameters defined in this section. Specifically, we let \( \xi = (\mu, \sigma)' \), \( \lambda = \phi = \beta \), \( \tau = \tau \), and \( \pi = (\beta', \xi')' = (\beta', \mu, \sigma)' \). The hypotheses of interests here are

\[H_0^{\text{eff}} : \xi_\ast = 0 \quad \text{vs.} \quad H_1^{\text{eff}} : \xi_\ast \neq 0.\]

Then, for each \( \mathbf{d} \) and \( \tilde{\mathbf{d}} \),

\[
B_\ast (\mathbf{d}, \tilde{\mathbf{d}}) = \begin{bmatrix}
B_\ast (\pi, \pi) (d_\pi, \tilde{d}_\pi) & 0' \\
0 & \frac{2}{\tau^2} d_\tau' \tilde{d}_\tau
\end{bmatrix},
\]

40
and
\[ B_{\pi}(d_\pi, \tilde{d}_{\pi}) = \frac{1}{\tau_s^2} \begin{bmatrix} d_\beta E[X_iX_i']d_\beta & d_\beta E[X_i']m(d_\mu, d_\sigma) \\ m(d_\mu, d_\sigma)E[X_i]d_\beta & m(d_\mu, d_\sigma)m(d_\mu, d_\sigma) \end{bmatrix}. \]

By the information matrix equality, for each \( d, B_\pi(d) \) is identical to \(-A_\pi(d)\).

We can also identify the asymptotic null distributions of the tests using the theorems in this section. First, we examine the QLR test. Applying Theorem 5 shows that

\[ LR_n^{(1)} := 2\{L_n(\tilde{\theta}_n) - L_n(\theta_*)\} \Rightarrow \sup_{s_\pi \in \Delta(\pi_\ast)} \max[0, \tilde{Y}^{(\pi)}(s_\pi)]^2 + \mathcal{H}_2, \]

where for each \( s_\pi \in \Delta(\pi_\ast) := \{(s'_\beta, s_\sigma) : (s'_\beta s_\beta + s_\mu^2 + s_\sigma^2 = 1, s_\mu > 0, \text{ and } s_\sigma > 0\}, \)

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

\[ Z^{(\pi)}(s_\pi) := s'_\beta Z^{(\beta)} + m(s_\mu, s_\sigma)Z^{(\xi)}, \]

and

\[ \begin{bmatrix} Z^{(\beta)} \\ Z^{(\xi)} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} E[X_tX_t'] & E[X_t'] \\ E[X_t] & 1 \end{bmatrix} \right). \]

We can see that \([Z^{(\beta)'}, Z^{(\xi)'})'\] is the weak limit of \( n^{-1/2}\tau_n^{-1}\sum_{t=1}^n[U_tX_t', U_t]'\). Corollary 2(i) also implies that

\[ LR_n^{(1)} := 2\{L_n(\tilde{\theta}_n) - L_n(\theta_*)\} \Rightarrow \sup_{s_\xi \in \Delta(\xi_\ast)} \max[0, \tilde{Y}^{(\xi)}(s_\xi)]^2 + Z^{(\beta)'E[X_tX_t']^{-1}Z^{(\beta)}} + \mathcal{H}_2, \]

where for each \( s_{\xi_\ast} \in \Delta(\xi_\ast) := \{(s_\mu, s_\sigma) : (s_\mu^2 + s_\sigma^2 = 1, s_\mu > 0, \text{ and } s_\sigma > 0\}, \)

\[ \tilde{Y}^{(\xi)}(s_\xi) := (\tilde{B}^{(\xi)}(s_\xi))^{-1/2}Z^{(\xi)}(s_\xi), \]

\[ \tilde{B}^{(\xi)}(s_\xi) := m(s_\mu, s_\sigma)^2\{1 - E[X_t']E[X_tX_t']^{-1}E[X_t]\}, \]

and also \( Z^{(\xi)}(s_\xi) := m(s_\mu, s_\sigma)\{Z^{(\xi)} - E[X_t']E[X_tX_t']^{-1}Z^{(\beta)}\}. \)

Furthermore, Theorem 6 shows that

\[ LR_n^{(2)} := 2\{L_n(\tilde{\theta}_n) - L_n(\theta_*)\} \Rightarrow \sup_{s_\beta \in \Delta(\beta_\ast)} \max[0, \tilde{Y}^{(\beta)}(s_\beta)]^2 + \mathcal{H}_2, \]

where for each \( s_{\beta_\ast} \in \Delta(\beta_\ast) := \{s_\beta \in \mathbb{R}^k : s'_\beta s_\beta = 1\}, \)

\[ \tilde{Y}^{(\beta)}(s_\beta) := s'_\beta E[X_tX_t']^{-1}Z^{(\beta)} + \mathcal{H}_2. \]

Therefore, Corollary 2(ii) now yields that

\[ LR_n \Rightarrow \sup_{s_\xi \in \Delta(\xi_\ast)} \max \left[ 0, \frac{m(s_\mu, s_\sigma)}{|m(s_\mu, s_\sigma)|} Z^{2} \right]. \]
under $H_0'''$, where $Z \sim N(0, 1)$. Here, we note that $Z$ is derived as \( \{1 - E[X_t' | E[X_t']^i E[X_t']^{-1} E[X_t']^{-1} Z(\beta)\} \), and if we let \( r(x) := \phi(x)/[x \Phi(x)] \),
\[
m(s_\mu, s_\sigma) = \frac{s_\mu}{|m(s_\mu, s_\sigma)|} \left( \frac{1 + r(s_\mu/s_\sigma)}{1 + r(s_\mu/s_\sigma)} \right),
\]
which is $-1$ uniformly on $\Delta(\xi_0)$. From this, the asymptotic null distribution reduces to $\max[0, -Z]^2$, and this implies that $\mathcal{L} \mathcal{R}_n \overset{\lor}{\sim} \max[0, -Z]^2$ under $H_0'''$.

We conduct Monte Carlo simulations to verify this. We have let $(X_t', U_t)' \sim \text{IID } N(\theta_2, I_2)$ and obtained the null distribution of the QLR test statistic. We let the number of replications be $2,000$ as before and examined the null distribution when $n = 50, 100, \text{ and } 200$. Our Monte Carlo simulation results are contained in Figure 4. We can easily see that the null distributions of the QLR test statistics exactly overlap with that of $\max[0, -Z]^2$.

Second, we examine the Wald test statistic using the D-D model analysis. For this, if we let
\[
\hat{W}_n(s_\mu, s_\sigma) := m(s_\mu, s_\sigma)^2 \left\{ 1 - n^{-1} \sum_{t=1}^{n} X_t' (n^{-1} \sum_{t=1}^{n} X_t X_t')^{-1} n^{-1} \sum_{t=1}^{n} X_t \right\},
\]
the LLN implies that
\[
\sup_{s_\mu, s_\sigma} |\hat{W}_n(s_\mu, s_\sigma) - \bar{B}_n(\xi_0)(s_\mu, s_\sigma)| \to 0 \text{ a.s. } - P.
\]
In particular, $m(\cdot, \cdot)^2$ is bounded by $1$ and $2/\pi$ from above and below, respectively. Using $\hat{W}_n(s_\mu, s_\sigma)$, we can let the Wald test statistic be
\[
\mathcal{W}_n := \sup_{s_\mu, s_\sigma} n \{\tilde{h}_n^2(s_\mu, s_\sigma)\} \{\hat{W}_n(s_\mu, s_\sigma)\} \{\tilde{h}_n^2(s_\mu, s_\sigma)\},
\]
where $\tilde{h}_n(s_\mu, s_\sigma)$ is such that for each $(s_\mu, s_\sigma)$,
\[
L_n(\tilde{h}_n(s_\mu, s_\sigma), s_\mu, s_\sigma, \tilde{h}_n(s_\mu, s_\sigma), \tilde{h}_n(s_\mu, s_\sigma)) = \sup_{\{h(\xi), \beta, \tau\}} L_n(h(\xi)(s_\mu, s_\sigma), s_\mu, s_\sigma, \beta, \tau).
\]
Corollary 3 now implies that $\mathcal{W}_n \Rightarrow \sup_{s_\xi \in \Delta(\xi_0)} \max[0, \bar{h}_n(\xi_0)(s_\xi)]^2$, and this is the identical weak limit to the QLR test statistic. Thus, $\mathcal{W}_n \overset{\lor}{\sim} \max[0, -Z]^2$ under $H_0'''$.

Finally, we investigate the LM test statistic. We let
\[
\mathcal{L} \mathcal{M}_n := \sup_{(s_\mu, s_\sigma, s_\beta) \in \Delta(\xi_0) \times \Delta(\beta_n)} \{n \hat{W}_n(s_\mu, s_\sigma, s_\beta) \max \left[ 0, \frac{-DL_n(\tilde{\theta}_n; s_\mu, s_\sigma)}{D^2 L_n(\tilde{\theta}_n; s_\mu, s_\sigma, s_\beta)} \right]^2 \},
\]
\[
\mathcal{L} \mathcal{M}_n := \sup_{(s_\mu, s_\sigma, s_\beta) \in \Delta(\xi_0) \times \Delta(\beta_n)} \{n \hat{W}_n(s_\mu, s_\sigma, s_\beta) \max \left[ 0, \frac{-DL_n(\tilde{\theta}_n; s_\mu, s_\sigma)}{D^2 L_n(\tilde{\theta}_n; s_\mu, s_\sigma, s_\beta)} \right]^2 \},
\]
where \( \tilde{\theta}_n = (\tilde{\beta}_n, 0, \tilde{\tau}_n) \) with \( \tilde{\beta}_n = (\sum_{t=1}^n X_t X_t')^{-1} \sum_{t=1}^n X_t Y_t \), \( \tilde{\tau}_n = (n^{-1} \sum_{t=1}^n \tilde{U}_t^2)^{1/2} \), and \( \tilde{U}_t := Y_t - X_t' \tilde{\beta}_n \); \( \Delta(\tilde{\beta}_n) := \{ \mathbf{x} \in \mathbb{R}^k : \mathbf{x}' \mathbf{x} = 1 \} \);

\[
D \mathcal{L}_n(\tilde{\theta}_n; s_\mu, s_\sigma) = \frac{m(s_\mu, s_\sigma)}{s_n^2} \sum_{t=1}^n \tilde{U}_t;
\]

and

\[
\begin{align*}
-\frac{1}{n} \tilde{D}^2 L_n(\tilde{\theta}_n; s_\mu, s_\sigma, s_\beta) &= \frac{m(s_\mu, s_\sigma)^2}{s_n^2}\{1 - s'_\beta E[X_t]'(s'_\beta E[X_t]s_\beta)^{-1}E[X_t]s_\beta\} + o_p(1).
\end{align*}
\]

It is not hard to show that this LLN also holds uniformly on \( \Delta(\xi_0) \times \Delta(\tilde{\beta}_n) \). Thus, for each \((s_\mu, s_\sigma, s_\beta)\), we may let

\[
\tilde{W}_n(s_\mu, s_\sigma, s_\beta) := \frac{m(s_\mu, s_\sigma)^2}{s_n^2}\left\{1 - n^{-1}\sum_{t=1}^n s'_\beta X_t \left(s'_\beta n^{-1}\sum_{t=1}^n X_tX_t's_\beta\right)^{-1}n^{-1}\sum_{t=1}^n X_t's_\beta\right\}.
\]

Here, applying the proof of Corollary 1(vii) implies that optimizing the objective function with respect to \( s_\beta \) yields that

\[
\sup_{s_\beta \in \Delta(\tilde{\beta}_n)} n\tilde{W}_n(s_\mu, s_\sigma, s_\beta) \max \left[ 0, \frac{-D \mathcal{L}_n(\tilde{\theta}_n; s_\mu, s_\sigma)}{D^2 L_n(\tilde{\theta}_n; s_\mu, s_\sigma, s_\beta)} \right]^2 = \max \left[ 0, \frac{m(s_\mu, s_\sigma)}{m(s_\mu, s_\sigma)} \frac{n^{-1/2}\sum_{t=1}^n \tilde{U}_t}{\{\tau_n^2(1 - E[X_t]'E[X_t]X_t'X_t')^{-1}E[X_t]\}^{1/2}} \right]^2 + o_p(1),
\]

so that

\[
\mathcal{L} \mathcal{M}_n = \sup_{(s_\mu, s_\sigma) \in \Delta(\xi_0)} \max \left[ 0, \frac{m(s_\mu, s_\sigma)}{m(s_\mu, s_\sigma)} \frac{n^{-1/2}\sum_{t=1}^n \tilde{U}_t}{\{\tau_n^2(1 - E[X_t]'E[X_t]X_t'X_t')^{-1}E[X_t]\}^{1/2}} \right]^2 + o_p(1)
\]

under \( H^m_0 \), implying that \( \mathcal{L} \mathcal{M}_n \overset{A}{\sim} \max[0, -Z]^2 \) by noting that \( \frac{m(\cdot \cdot \cdot)}{m(\cdot \cdot \cdot)} = -1 \) on \( \Delta(\xi_0) \) as discussed above and that \( n^{-1/2}\sum_{t=1}^n \tilde{U}_t \sim N(0, \tau_n^2(1 - E[X_t]'E[X_t]X_t'X_t')^{-1}E[X_t]) \). This is exactly what Corollary 4 asserts.

Before completing this example, we provide some remarks relevant to it. Here, we let \( \mu_+ \geq 0 \) so that \( d_\mu \) is always greater than or equal to zero, and this is assumed to avoid numerical failures. It is more general to suppose that \( \mu_+ \) can be negative, so that for some positive \( c > 0 \), \( \mu_+ \in [-c, c] \). For such a case, for example, the asymptotic null distribution of the QLR test is modified into

\[
\mathcal{L} \mathcal{R}_n \Rightarrow \sup_{s_\xi \in \Delta(\xi_0)'} \max \left[ 0, \frac{m(s_\mu, s_\sigma)}{m(s_\mu, s_\sigma)} Z \right]^2,
\]

43
where $\Delta(\xi_0)' := \{(s_\mu, s_\sigma) \in \mathbb{R}^2 : s_\mu^2 + s_\sigma^2 = 1 \text{ and } s_\sigma > 0\}$. Furthermore, it analytically follows that $m(s_\mu, s_\sigma)/|m(s_\mu, s_\sigma)| = -1$ uniformly on $\Delta(\xi_0)'$, so that $\mathcal{L}R_n \Rightarrow \max[0, -Z]^2$, which is the same as for the previous case. Nevertheless, our Monte Carlo experiments assuming the same condition show that the empirical distribution of $\mathcal{L}R_n$ overlaps with that of $Z^2$.

This discrepancy is mainly because the value of $m(s_\mu, s_\sigma)$ sensitively responds to the value of $(s_\mu, s_\sigma)$, so that we obtain that $m(\cdot, \cdot)/|m(\cdot, \cdot)| = \pm1$ numerically on $\Delta(\xi_0)'$. This also implies that

$$\mathcal{L}R_n \Rightarrow \sup_{s_\xi \in \Delta(\xi_0)'} \max \left[0, \frac{m(s_\mu, s_\sigma)}{|m(s_\mu, s_\sigma)|}Z \right]^2 = \sup_{s_\xi \in \Delta(\xi_0)'} \max [0, -Z]^2 = Z^2$$

as we could see from the Monte Carlo experiment. More precisely, if $s_\mu < 0$ and $s_\sigma > 0$, so that we can let $s_\mu = -\sqrt{1 - s_\sigma^2}$, it analytically follows that $\lim_{s_\sigma \to 0} m(-\sqrt{1 - s_\sigma^2}, s_\sigma) = 0$ and for any $s_\sigma > 0$, $m(-\sqrt{1 - s_\sigma^2}, s_\sigma) < 0$. Nevertheless, computing this value requires high precision levels at $s_\sigma \approx 0$, and conventional statistical packages do not provide this precision. Numerically, they compute $m(-\sqrt{1 - s_\sigma^2}, s_\sigma)$ oscillating around zero as $s_\sigma$ converges to 0, implying that $m(\cdot, \cdot)/|m(\cdot, \cdot)| = \pm1$ on $\Delta(\xi_0)'$. Our model is specified to avoid this numerical failure by letting $\mu_o \geq 0$. \hfill \Box

### 3.4.3 Example 3 (continued)

As the model is D, we obtain that $r_o = 0$. Furthermore, supposing that $r_\tau = 0$ does not matter in terms of Theorem 5. Thus, we let $\theta = \pi = \lambda = (\delta', \phi')' = (\mu', \omega')'$.

The objective function $Q_n$ does not satisfy the model condition in Assumption 2. We cannot therefore exactly apply the definition of the QLR test statistic to this model. Nevertheless, we can develop a QLR test-like test statistic using $Q_n$. We let

$$Q\mathcal{L}R_n := \{\sup_{\delta, \phi} Q_n(\delta, \phi) - \sup_{\phi} Q_n(\delta_0, \phi)\}$$

and also let $C_s(\lambda)^{-1}W$ and $C_s(\lambda)^{-1}C_s$ be $Z(\lambda) = (Z(\delta)', Z(\phi')')'$ and $A_s(\delta, \phi)$ of this section, respectively. The asymptotic null distribution of the QLR test statistic is already given in Eq. (23). That is, $Q\mathcal{L}R_n \Rightarrow (\bar{Z}(\delta)' - A_s(\delta, \phi))^{-1}(\bar{Z}(\delta))$.

We can also define a Wald test using the GMM estimator and derive its asymptotic null distribution in a similar manner. For this purpose, we define

$$Q\mathcal{W}_n := \sup_{s_\delta \in \Delta(\delta_0)} n \{\hat{h}_n(\delta)(s_\delta)\} \{\hat{W}_n(\delta)(s_\delta)\} \{\hat{h}_n(\delta)(s_\delta)\},$$

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

$$Q_n(\delta_0 + \hat{h}_n(\delta)(s_\delta), \phi) := \sup_{h(\delta)} Q_n(\delta_0 + h(\delta)s_\delta, \phi)$$
and obtain the asymptotic null distribution by applying Corollary 3. We note that the definition of $QW_n$ is exactly the same as $W_n$ except that $\tilde{h}_n^{i}(s_{\delta})$ is defined using $Q_n$ instead of $L_n$. If we further let the weight function $\tilde{W}_n(s_{\delta})$ be $s_{\delta}'\tilde{W}_n s_{\delta}$ such that $\tilde{W}_n$ converges to $-\tilde{A}_n^{(\delta)}$ a.s. $-\mathbb{P}$, it follows that

$$QW_n \Rightarrow \sup_{s_{\delta} \in \Delta(\delta_0)} \max[0, s_{\delta}'\tilde{Z}^{(\delta)}] (-s_{\delta}'\tilde{A}_n^{(\delta)} s_{\delta})^{-1} \max[0, s_{\delta}'\tilde{Z}^{(\delta)}].$$

The proof of Corollary 1(vii) corroborates that the asymptotic null distribution of $QW_n$ is equivalent to that of $QLR_n$ because of the fact that $\delta_0$ is an interior element.

Finally, we can similarly define an LM test and derive its asymptotic null distribution. For this purpose, we let

$$QLM_n := \sup_{(s_{\delta}, s_{\phi}) \in \Delta(\delta_0) \times \Delta(\phi_0)} n\tilde{W}_n(s_{\delta}, s_{\phi}) \max \left[ 0, \frac{DQ_n(\tilde{\theta}_n; s_{\delta})}{2D^2Q_n(\tilde{\theta}_n; s_{\delta}, s_{\phi})} \right]^2,$$

where for each $(s_{\delta}, s_{\phi})$,

$$\tilde{D}^2Q_n(\tilde{\theta}_n; s_{\delta}, s_{\phi}) := Dg_n(\tilde{\theta}_n; s_{\delta})'\{-M_n\}^{-1}Dg_n(\tilde{\theta}_n; s_{\delta})$$

$$- Dg_n(\tilde{\theta}_n; s_{\phi})'\{-M_n\}^{-1}Dg_n(\tilde{\theta}_n; s_{\phi}) Dg_n(\tilde{\theta}_n; s_{\phi})'\{-M_n\}^{-1}Dg_n(\tilde{\theta}_n; s_{\phi})'\{-M_n\}^{-1}Dg_n(\tilde{\theta}_n; s_{\phi})^{-1}$$

$$\times Dg_n(\tilde{\theta}_n; s_{\phi})'\{-M_n\}^{-1}Dg_n(\tilde{\theta}_n; s_{\delta}),$$

and $\tilde{\theta}_n := (\delta_0, \tilde{\phi}_n)$ such that $\tilde{\phi}_n := \arg \max_{\phi} Q_n(\delta_0, \phi)$. As for the $QW_n$ test statistic, $QLM_n$ is defined by $Q_n$ instead of $L_n$. We also note that $\tilde{D}^2Q_n(\tilde{\theta}_n; s_{\delta}, s_{\phi})$ is defined by the first-order directional derivatives of $g_n$. If we let $\tilde{W}_n(s_{\delta}, s_{\phi}) = s_{\delta}'\tilde{W}_n s_{\delta}$ for each $(s_{\delta}, s_{\phi}) \in \Delta(\delta_0) \times \Delta(\phi_0)$, $QLM_n \Rightarrow (\tilde{Z}^{(\delta)})'(-\tilde{A}_n^{(\delta)})^{-1} (\tilde{Z}^{(\delta)})$ by Corollary 4, the interiority condition of $\delta_0$, and the proof of Corollary 1(vii), where $\tilde{W}_n$ is the same weight function as used for the $QW_n$ test statistic.

3.4.4 Example 4 (continued)

For this model examination, we let $\pi = (\phi', \xi')'$ such that $\omega = \phi = \theta_0$ and $\mu = \xi = \theta_2$, so that $\Omega = \Theta_0$, and $M$ is a closed interval with zero as an interior element. We note that $\theta_{1*} = 0$ if and only if $\theta_{2*} = 0$ from the model assumption. Using these conditions, for given $r := d_1/d_2 \in [\bar{c}, \bar{c}]$, we can apply Corollary 2(ii) to obtain the asymptotic null distribution of the QLR test statistic. That is,

$$LR_n \Rightarrow \sup_{s_{\xi} \in \Delta(\xi_0)} \max[0, \tilde{\gamma}(\xi)(s_{\xi})]^2,$$

where $s_{\xi} := s_2$, $\Delta(\xi_0) := \{-1, 1\}$, and

$$\tilde{\gamma}(\xi)(s_{\xi}) := \frac{s_2^2 \tilde{Z}^{(\xi)}}{(A_0^{(\xi)}\xi)^{1/2}} := \frac{r}{\sqrt{r^2}} \frac{s_2(W_1 - (-A_0^{(0,1)})(-A_0^{(0,0)})^{-1}W_0)}{\left\{(-A_0^{(1,1)}) - (-A_0^{(1,0)})(-A_0^{(0,0)})^{-1}(-A_0^{(1,1)})\right\}^{1/2}}.$$
We note that \( d_1/d_2 \) is canceled off, which explains how the asymptotic null distribution is obtained without being affected by \( r \). From the structure of \( \Delta(\xi_0) \), it also follows that \( \mathcal{LR}_n \Rightarrow \tilde{Z}(\xi)(\tilde{A}_{\xi}^{(\xi,-)})^{-1} \tilde{Z}(\xi) \).

In a similar way, we can apply Corollary 3 for the Wald test statistic. We note that for a given \( r \),
\[
\sqrt{n \tilde{h}_n(\mu)}(s_\xi) \Rightarrow (r \tilde{A}_{\xi}^{(\xi,-)})^{-1} \max[0, s_2 \tilde{Z}(\xi)].
\]
In order to have a Wald test statistic not influenced by the selection of \( r \), we select \( \hat{W}_n \) to be a consistent estimator for \( (r \tilde{A}_{\xi}^{(\xi,-)})^{-1} \). For example, if we let
\[
\hat{W}_n := r^{-1} \left\{ (n^{-1} \sum \log(X_t)^2) - (n^{-1} \sum \log(X_t)Z_t') \left( (n^{-1} \sum Z_tZ_t')^{-1}(n^{-1} \sum Z_t \log(X_t)) \right) \right\},
\]
\[
W_n := n \{ \tilde{h}_n(\mu)(s_\xi) \} \{ \hat{W}_n \} \{ \tilde{h}_n(\mu)(s_\xi) \} \Rightarrow \tilde{Z}(\xi)(\tilde{A}_{\xi}^{(\xi,-)})^{-1} \tilde{Z}(\xi) \text{ by Corollary 3.}
\]

Furthermore, we can apply Corollary 4 to obtain the same asymptotic null distribution of the LM test statistic using the same weight function.

Before moving to the conclusion of this paper, we note that the asymptotic null distributions of the tests are represented as functions of Gaussian processes in general, and this aspect cannot be avoided when the directional differentiability is the key aspect of models. Although we could successfully provide the asymptotic null distributions for the examples we consider here, it can be challenging for other D-D models. Nevertheless, Hansen’s (1996b) weighted bootstrap method can be successfully exploited for such a case as validated by Cho, Cheong, and White (2011) and Cho, Ishida, and White (2011, 2012) in unidentified model contexts.

4 Conclusion

In this paper, we examine data estimation and inference using D-D econometric models and provide conditions under which the extremum estimator behaves regularly. Specifically, we show that the extremum estimator has a distribution different from that of standard D models as it can be represented as a function of a Gaussian stochastic process indexed by directions. Furthermore, this view yields the same results for the extremum estimator as those obtained using D models. This implies that model estimation as those obtained using D models can be understood as a special case of D-D models.

In addition to this, we appropriately modify the standard QLR, Wald, and LM test statistics for D-D models. These modifications are provided for general D-D models in the sense that their null models can be D-D with respect to other nuisance parameters. The three test statistics have null distributions represented as functions of the same Gaussian stochastic processes as derived above, and they turn out to have asymptotically equivalent null distributions under the benchmark model assumptions motivated mainly by Rosenberg (1973) and King and Shively’s (1993) conditional heteroskedasticity model, Aigner, Lovell, and
Schmidt (1977) and Stevenson’s (1980) stochastic frontier production function models, D models for GMM estimation in Hansen (1982), and Box-Cox’s (1964) transformation.

Appendix: Proofs

Proof of Lemma 1: To show the given claim, we show that \( f \) is twice continuously D at \( \theta_0 \). The same proof can also be applied to other parameter values.

If we let \( g(h) := f(\theta_0 + hd) \), \( g \) is twice continuously D from the given condition, so that we can apply the mean-value theorem: for some \( \bar{h} \geq 0 \)

\[
g(h) = g(0) + g'(0)h + \frac{1}{2} g''(\bar{h})h^2,
\]

implying that \( f(\theta) - f(\theta_0) - Df(\theta_0; d)h = \frac{1}{2} D^2 f(\bar{\theta}; d)h^2 \), where \( \theta = \theta_0 + hd, \bar{\theta} = \theta_0 + \bar{h}d, \) and \( \bar{\theta} \to \theta_0 \) as \( \theta \to \theta_0 \).

Thus,

\[
f(\theta) - f(\theta_0) - Df(\theta_0; d)h - \frac{1}{2} D^2 f(\theta_0; d)h^2 = \frac{1}{2} D^2 f(\bar{\theta}; d)h^2 - \frac{1}{2} D^2 f(\theta_0; d)h^2.
\]

Furthermore, for some \( A(\theta_0) \in \mathbb{R}^r \) and \( B(\theta_0) \in \mathbb{R}^r \times \mathbb{R}^r \), \( A(\theta_0)(\theta - \theta_0) = A(\theta_0)dh \) and \( (\theta - \theta_0)'B(\theta_0)(\theta - \theta_0) = D^2 f(\bar{\theta}; d)h^2 \) from the linear and quadratic form conditions, so that

\[
f(\theta) - f(\theta_0) - A(\theta_0)(\theta - \theta_0) - \frac{1}{2}(\theta - \theta_0)'B(\theta_0)(\theta - \theta_0) = \frac{1}{2}(D^2 f(\bar{\theta}; d) - D^2 f(\theta_0; d))h^2.
\]

Therefore,

\[
\lim_{\theta \to \theta_0} \left\{ \frac{1}{\|\theta - \theta_0\|^2} \{ f(\theta) - f(\theta_0) - A(\theta_0)(\theta - \theta_0) - \frac{1}{2}(\theta - \theta_0)'B(\theta_0)(\theta - \theta_0) \} \right\} = 0,
\]

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

\[\blacksquare\]

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

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

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

Proof of Theorem 3: (i) From the fact that \( \sqrt{n} \hat{\theta}_n(d) = \max[0, \{-D^2 L_n(\theta^*_d; d)\}] \) by Eq. (11) and Khun-Tucker theorem, and that \( \hat{\theta}_n(d) \to \theta^* \) a.s. \( \implies \) by Theorem 1, it simply follows that \( \sqrt{n} \hat{\theta}_n(d) \to \max[0, \{-A_*(d)\}] \) by Lemma 2(iii). The desired result now follows from the definition of \( G(d) \).

(ii) From the definition of \( \tilde{h}_n(d) \), \( \tilde{\theta}_n(d) = \theta^* + \tilde{h}_n(d) d \). Theorem 3(i) yields the given result.

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

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

Thus, the desired result follows from Eq. (12).

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

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

(ii) By Assumption 5(iii), \( |n^{-1}D^2 L_n(\theta; d_1) - n^{-1}D^2 L_n(\theta; d_2)| \leq n^{-1} \sum M_t \|d_1 - d_2\|^\lambda \). Furthermore, we can apply the ergodic theorem to \( \{n^{-1} \sum M_t\} \), so that for any \( \omega \in F, P(F) = 1 \), and \( \epsilon > 0 \), there is an \( n^*(\omega, \epsilon) \) such that \( n \geq n^*(\omega, \epsilon), |n^{-1} \sum M_t - E[M_t]| \leq \epsilon \), and this implies that \( n^{-1} \sum M_t \leq E[M_t] + \epsilon \). For the same \( \epsilon \), we may let \( \delta := \epsilon/(E[M_t] + \epsilon) \). Then, \( n^{-1} \sum M_t \|d_1 - d_2\|^\lambda \leq \epsilon \), whenever \( \|d_1 - d_2\|^\lambda \leq \delta \), because \( n^{-1} \sum M_t \|d_1 - d_2\|^\lambda \leq n^{-1} \sum M_t \delta = n^{-1} \sum M_t \delta = n^{-1} \sum M_t \epsilon/(\epsilon + E[M_t] + \epsilon) \leq \delta \).

That is, for any \( \omega \in F, P(F) = 1 \) and \( \epsilon > 0 \), there is \( n^*(\omega, \epsilon) \) and \( \delta \) such that \( n \geq n^*(\omega, \epsilon) \) and \( \|d_1 - d_2\|^\lambda \leq \delta, |n^{-1}D^2 L_n(\theta; d_1) - n^{-1}D^2 L_n(\theta; d_2)| < \epsilon \), which means that \( \{n^{-1}D^2 L_n(\theta; \cdot)\}_{n^*(\omega, \epsilon)}^{\infty} \)
is equicontinuous. Therefore, it follows that \( n^{-1}D^2L_n(\theta; \cdot) \) converges to \( A_* \) uniformly on \( \Delta(\theta_*) \) a.s.\(-\mathbb{P}\) by Rudin (1976, p. 168).

\[ \text{Proof of Theorem 4:} \ \text{(i)} \ \text{Given Lemmas 3(i and ii), the given weak convergence follows from theorem 3.9 of Billingsley (1999, p. 37).} \]

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

\( \text{(iii)} \) We can apply the CMT to Eq. (13).

\( \text{(iv)} \) From the definition of \( \hat{h}_n \), for each \( d \), \( \sqrt{n}(\hat{h}_n(d) - \theta_*) = \sqrt{n}\hat{h}_n(d)d \). Theorems 3(ii), 4(i to iii), Eq. (13), and the CMT yield the given result.

\[ \text{Proof of Corollary 1:} \ \text{For an efficient presentation, we first prove (vi) and (vii) before proving (iv) and (v).} \]

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

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

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

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

\( \text{(vii)} \) From the fact that \( \text{cl}\{C(\theta_*)\} = \mathbb{R}^r \), there is \( d^* \in \Delta(\theta_*) \) such that \( \max[0, \mathcal{Z}'d^*] = \mathcal{Z}'d^* \) and \( d^* = -d \) if \( \max[0, \mathcal{Z}'d] = 0 \). Thus, the given ‘max’ operator can be ignored in this case. That is,

\[ d_* = \arg \max_{d \in \Delta(\theta_*)} d'\mathcal{Z}\mathcal{Z}'d \{d'(-A_*)d\}^{-1}. \]

For notational simplicity, if we let

\[ v := \frac{(-A_*)^{-1/2}d}{\{d'(-A_*)d\}^{1/2}}, \]

it follows that \( v'v = 1 \) and \( v'(-A_*)^{-1/2}\mathcal{Z}\mathcal{Z}'(-A_*)^{-1/2}v = d'\mathcal{Z}\mathcal{Z}'d \{d'(-A_*)d\}^{-1} \). Given this, we note that

\[ \max_v v'(-A_*)^{-1/2}\mathcal{Z}\mathcal{Z}'(-A_*)^{-1/2}v = \max_d d'\mathcal{Z}\mathcal{Z}'d \{d'(-A_*)d\}^{-1}, \]

and it is equal to the maximum eigenvalue of \( (-A_*)^{-1/2}\mathcal{Z}\mathcal{Z}'(-A_*)^{-1/2} \), which is equal to \( \mathcal{Z}'(-A_*)^{-1}\mathcal{Z} \). It is because \( \text{rank}((-A_*)^{-1/2}\mathcal{Z}\mathcal{Z}'(-A_*)^{-1/2}) = 1 \) (so that there is one positive eigenvalue, and other eigenvalues are zero), and the sum of eigenvalues is equal to \( \text{tr}((-A_*)^{-1/2}\mathcal{Z}\mathcal{Z}'(-A_*)^{-1/2}) = \mathcal{Z}'(-A_*)^{-1}\mathcal{Z} \). These two facts lead to the desired result.
(iv) This follows trivially from the definition of $d_s$.

(v) By the same reason as in the proof of (vii), we can ignore the ‘max’ operator, so that $\sqrt{n}(\bar{\theta}_n - \theta_*) \Rightarrow Z' d_s \{d_s'(-A_s) d_s\}^{-1} d_s$.

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

$$v_* := \frac{(-A_s)^{1/2} d_s}{\{d_s'(-A_s) d_s\}^{1/2}},$$

$v_*$ is the eigenvector of $(-A_s)^{-1/2} ZZ'(-A_s)^{-1/2}$ corresponding to the maximum eigenvalue given as $Z'$, so that

$$(-A_s)^{-1/2} ZZ'(-A_s)^{-1/2} v_* = v_* (-A_s)^{-1} Z v_*$$

by the definition of eigenvector. This implies that

$$v_*'(-A_s)^{-1/2} ZZ'(-A_s)^{-1/2} v_* = Z'(-A_s)^{-1} Z v_*' v_* = Z'(-A_s)^{-1} Z$$

because $v_*' v_* = 1$. Plugging the definition of $v_*$ to the LHS of Eq. (36) leads to that

$$Z' d_s \{d_s'(-A_s) d_s\}^{-1} = (d_s'Z)^{-1} Z'(-A_s)^{-1} Z.$$}

Thus, $Z' d_s \{d_s'(-A_s) d_s\}^{-1} d_s = (d_s'Z)^{-1} Z'(-A_s)^{-1} Z d_s$. Also, plugging the definition of $v_*$ to Eq. (35) yields that $(d_s'Z)^{-1} Z'(-A_s)^{-1} Z d_s = (-A_s)^{-1} Z$. Therefore, $\sqrt{n}(\bar{\theta}_n - \theta_*) \Rightarrow Z' d_s \{d_s'(-A_s) d_s\}^{-1} d_s = (-A_s)^{-1} Z$. This completes the proof.

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

$$2 \{L_n(\pi_* + h^{(\pi)} s_\pi, \tau_* + h^{(\tau)} s_\tau) - L_n(\pi_*, \tau_*)\}$$

$$= 2 DL_n(\pi_*, \tau_*; s_\pi) h^{(\pi)} + 2 DL_n(\pi_*, \tau_*; s_\tau) h^{(\tau)}$$

$$+ D^2 L_n(\pi_*, \tau_*; s_\pi)(h^{(\pi)})^2 + D^2 L_n(\pi_*, \tau_*; s_\tau)(h^{(\tau)})^2$$

$$+ 2 DL_n(\pi_*, \tau_*; s_\pi; s_\tau) h^{(\pi)} h^{(\tau)} + o_{p_{d\pi}}(1) + o_{p_{d\tau}}(1),$$

where $DL_n(\pi_*, \tau_*; s_\pi; s_\tau)$ is the directional derivative of $DL_n(\cdot, \cdot; s_\pi)$ with respect to $s_\tau$ evaluated at $(\pi_*, \tau_*)$, and

$$\sup_{d} \sup_{h} L_n(\theta_* + hd) = \sup_{\{s_\pi, s_\tau\}} \sup_{\{\theta^{(\pi)}, \theta^{(\tau)}\}} L_n(\pi_* + h^{(\pi)} s_\pi, \tau_* + h^{(\tau)} s_\tau).$$
Therefore,

\[ 2\{L_n(\hat{\theta}_n) - L_n(\theta_*)\} = \sup_{d} \sup_{h} 2\{L_n(\theta_* + h d) - L_n(\theta_*)\} \]

\[ = \sup_{\theta_*} \sup_{h} \{2DL_n(\theta_*; s_\pi)h(\pi) + D^2L_n(\theta_*; s_\pi)(h(\pi))^2 + o_{d\pi}(1)\} \]

\[ + \sup_{\theta_*} \sup_{h(\pi)} \{2DL_n(\theta_*; s_\pi)h(\pi) + D^2L_n(\theta_*; s_\pi)(h(\pi))^2 + o_{d\pi}(1)\}, \]

where we exploited the facts that \( n^{-1}DL_n(\pi_*, \tau_*, s_*; s_\pi; s_\pi) \) has probability limit \( A_1(\tau, \pi)(s_\pi, s_\pi) = 0 \) by Assumption 7(iv) and that \( DL_n(\theta_*; \cdot) \) and \( DL_n(\theta_*; \cdot) \) are \( O(n^{1/2}) \) by Theorem 4(i). Given this separation, the desired result follows by applying the proof of Theorem 4(iii) to each piece in the RHS of Eq. (39). \( \square \)

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

(ii) This follows as a corollary of Theorem 5. \( \square \)

**Proof of Corollary 2:** To show the given claim, we use Eq. (39). As it is trivial from the proof of Theorem 5 that \( H_{2,n} := \sup_{\theta_*} \sup_{h(\pi)} \{2DL_n(\theta_*; s_\pi)h(\pi) + D^2L_n(\theta_*; s_\pi)(h(\pi))^2\} \Rightarrow H_2 \), we here focus on the weak limit of

\[ H_{1,n} := \sup_{\theta_*} \sup_{h(\pi)} \{2DL_n(\theta_*; s_\pi)h(\pi) + D^2L_n(\theta_*; s_\pi)(h(\pi))^2\}. \]

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

\[ H_{1,n} = \sup_{\theta_*} \sup_{h(\lambda)h(\xi)} 2DL_n(\theta_*; s_\lambda')h(\lambda) + 2DL_n(\theta_*; s_\lambda')h(\lambda) + 2DL_n(\theta_*; s_\lambda')h(\lambda)h(\xi) \]

\[ + D^2L_n(\theta_*; s_\lambda')h(\lambda)^2 + D^2L_n(\theta_*; s_\lambda')h(\lambda)^2, \]

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

\[ H_{1,n} \Rightarrow H_1 = \sup_{\theta_*} \sup_{h(\lambda)} \{2Z(\xi)(s_\xi)h(\xi) + 2s_\lambda Z(\lambda)(h(\lambda) + 2s_\lambda A_0(\lambda, s_\xi)(s_\xi)h(\lambda)h(\xi) \]

\[ + A_0(\xi, s_\xi)(h(\xi))^2 + s_\lambda A_0(\lambda, s_\xi)(s_\lambda)^2 \}

and there are four different possible cases for the solutions of the RHS of Eq. (40) from the fact that \( h(\xi) \geq 0 \) and \( h(\lambda) \geq 0 \). That is, for each \( (s_\xi, s_\lambda) \) if we let \( \hat{h}(\xi)(s_\xi, s_\lambda) \) and \( \hat{h}(\lambda)(s_\xi, s_\lambda) \) maximizes the RHS of Eq. (40) either (i) \( \hat{h}(\xi)(s_\xi, s_\lambda) > 0 \) and \( \hat{h}(\lambda)(s_\xi, s_\lambda) > 0 \); (ii) \( \hat{h}(\xi)(s_\xi, s_\lambda) > 0 \) and \( \hat{h}(\lambda)(s_\xi, s_\lambda) = 0 \); (iii) \( \hat{h}(\xi)(s_\xi, s_\lambda) = 0 \) and \( \hat{h}(\lambda)(s_\xi, s_\lambda) > 0 \); or (iv) \( \hat{h}(\xi)(s_\xi, s_\lambda) = 0 \) and \( \hat{h}(\lambda)(s_\xi, s_\lambda) = 0 \).
We examine the asymptotic distribution of each case one by one. First, if \( \hat{h}(\xi)(s_\xi, s_\lambda) > 0 \) and \( \hat{h}(\lambda)(s_\xi, s_\lambda) > 0 \), the RHS of Eq. (40) is identical to

\[
\sup_{s_\xi, s_\lambda} \left[ Z(\xi)(s_\xi) s_\lambda' Z(\lambda)^{-1} \right] \left[ Z(\xi)(s_\xi) \right]^{-1}\left[ Z(\xi)(s_\xi) s_\lambda' Z(\lambda)^{-1} \right],
\]

and maximizing this with respect to \( s_\lambda \) for a given \( s_\xi \) yields \( \hat{Y}(\xi)(s_\xi)^2 + (Z(\lambda))'(-A_\lambda^\prime A_\lambda)^{-1}(Z(\lambda)) \). For \( \hat{h}(\xi)(s_\xi, s_\lambda) \) to be greater than zero, it is necessary that \( \hat{Y}(\xi)(s_\xi) \) is greater than zero, too. Second, if \( \hat{h}(\lambda)(s_\xi, s_\lambda) > 0 \) and \( \hat{h}(\xi)(s_\xi, s_\lambda) = 0 \), the RHS of Eq. (40) is identical to \( 2s_\lambda' Z(\lambda) + s_\lambda' A_\lambda^\prime A_\lambda s_\lambda \), and maximizing this with respect to \( s_\lambda \) leads to \( (Z(\lambda))'(-A_\lambda^\prime A_\lambda)(Z(\lambda)) \) as its maximum. Also, \( \hat{Y}(\xi)(s_\xi) \) cannot be greater than zero. Otherwise, \( \hat{h}(\xi)(s_\xi, s_\lambda) > 0 \). Third, if \( \hat{h}(\lambda)(s_\xi, s_\lambda) = 0 \) and \( \hat{h}(\xi)(s_\xi, s_\lambda) > 0 \), we can consider \( -s_\lambda \) as an alternative to \( s_\lambda \) for the same \( s_\xi \) while maximizing the RHS of Eq. (40) with respect to \( s_\lambda \) from the fact that \( A_\lambda \) is an interior element. Thus, it modifies the given maximization to the first case. Finally, for given \( s_\xi \) and \( s_\lambda \), if \( \hat{h}(\lambda)(s_\xi, s_\lambda) = 0 \) and \( \hat{h}(\xi)(s_\xi, s_\lambda) = 0 \), \( -s_\lambda \) can be considered an alternative, as well, and this modifies the maximization to the second case. Therefore, combining all these leads to that \( H_1 = \sup_{s_\xi \Delta(\xi, s)} \max[0, \hat{Y}(\xi)(s_\xi)]^2 + (Z(\lambda))'(-A_\lambda^\prime A_\lambda)(Z(\lambda)). \)

Proof of Lemma 4: (i) We exploit Eq. (39) further. First, applying the CMT to Theorem 4(i) shows that

\[
(n^{-1/2} DL_n(\tau), n^{-1} D^2 L_n(\tau)) \Rightarrow (Z(\tau), A_\lambda^\prime(\tau, \tau)).
\]

Thus,

\[
\sup_{\sqrt{n}h(\tau)^*} 2DL_n(\theta_\tau; s_\tau)h(\tau) + D^2 L_n(\theta_\tau; s_\tau)(h(\tau))^2 \Rightarrow \sup_{h(\tau) \in \mathbb{R}^+} 2Z(\tau)(s_\tau)h(\tau) + A_\lambda^\prime(\tau, \tau)(s_\tau)(h(\tau))^2,
\]

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

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

\[
(41) \sup_{h(\eta)} \left\{ 2DL_n(\theta_\eta; d_{\eta})h(\eta) + D^2 L_n(\theta_\eta; d_{\eta})(h(\eta))^2 \right\}
\]

\[
= \sup_{(h(\mu), h(\omega))} 2DL_n(\theta_\tau; s_\mu)h(\mu) + D^2 L_n(\theta_\tau; s_\mu)(h(\mu))^2 + 2DL_n(\theta_\tau; s_\omega)h(\omega) + D^2 L_n(\theta_\tau; s_\omega)(h(\omega))^2 \]
\[
\Rightarrow \sup_{(h(\mu), h(\omega))} 2Z(s_\mu)h(\mu) + 2Z(s_\omega)h(\omega) + 2A_\lambda^\prime(\mu, \mu)(s_\mu)h(\mu) + A_\lambda^\prime(\omega, \omega)(s_\omega)h(\omega) \]
\[
+ A_\lambda^\prime(\mu, \mu)(s_\mu)h(\mu)^2 + A_\lambda^\prime(\omega, \omega)(s_\omega)h(\omega)^2.
\]
Given this, \( h^{(\mu)} \) and \( h^{(\omega)} \) on the RHS of Eq. (41) are present on the positive Euclidean line, so that there are four different possible inequality constraints. We examine each case one by one. First, if any inequality condition does not bind, \( \sqrt{n} \left( \hat{h}^{(\mu)}(s_\mu, s_\omega, s_\tau), \hat{h}^{(\omega)}(s_\mu, s_\omega s_\tau) \right) \Rightarrow \mathcal{G}^{(\eta)}(s_\mu, s_\omega) \) by the standard first-order condition and Lemma 2. This occurs if each component of \( \mathcal{G}^{(\eta)}(s_\mu, s_\omega) \) is strictly greater than zero. Second, if \( \mathcal{G}^{(\mu)}(s_\mu, s_\omega) < 0 \), it simply holds that \( \hat{h}^{(\omega)}(s_\mu, s_\omega) = \max[0, \hat{\mathcal{G}}^{(\omega)}(s_\mu, s_\omega)] \). Thus, \( \sqrt{n} \left( \hat{h}^{(\mu)}(s_\mu, s_\omega, s_\tau), \hat{h}^{(\omega)}(s_\mu, s_\omega, s_\tau) \right) \Rightarrow (0, \max[0, \hat{\mathcal{G}}^{(\omega)}(s_\mu, s_\omega)])' \). Likewise, if \( \hat{h}^{(\omega)}(s_\mu, s_\omega) = 0 \) in the RHS of Eq. (41) because \( \mathcal{G}^{(\omega)}(s_\mu, s_\omega) < 0 \), \( \hat{h}^{(\mu)}(s_\mu, s_\omega) = \max[0, \hat{\mathcal{G}}^{(\mu)}(s_\mu, s_\omega)] \). This implies that \( \sqrt{n} \left( \hat{h}^{(\mu)}(s_\mu, s_\omega, s_\tau), \hat{h}^{(\omega)}(s_\mu, s_\omega, s_\tau) \right) \Rightarrow (\max[0, \hat{\mathcal{G}}^{(\mu)}(s_\mu, s_\omega)], 0)' \). Fourth, it must be the case that \( \sqrt{n} \left( \hat{h}^{(\mu)}(s_\mu, s_\omega, s_\tau), \hat{h}^{(\omega)}(s_\mu, s_\omega, s_\tau) \right) \Rightarrow (0, 0)' \) for any other case. Therefore,

\[
\sqrt{n} \begin{bmatrix}
\hat{h}^{(\mu)} \\
\hat{h}^{(\omega)} \\
\hat{h}^{(\tau)}
\end{bmatrix} \Rightarrow \begin{bmatrix}
\mathcal{G}^{(\mu)} \\
\mathcal{G}^{(\omega)} \\
0
\end{bmatrix} I_{\{\min[\mathcal{G}^{(\mu)}, \mathcal{G}^{(\omega)}] \geq 0\}} + \begin{bmatrix}
\max[0, \hat{\mathcal{G}}^{(\mu)}]I_{\{\mathcal{G}^{(\omega)} < 0\}} \\
\max[0, \hat{\mathcal{G}}^{(\omega)}]I_{\{\mathcal{G}^{(\mu)} < 0\}} \\
\max[0, \mathcal{G}^{(\tau)}]
\end{bmatrix}
\]

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

\[\text{Proof of Theorem 8:} \text{ We can approximate Eq. (31) by a quadratic function and apply Lemma 4 to obtain that}\]

\[\sup_{\{h^{(\mu)}, \omega, \tau\}} 2 \{ L_n(\mu_0 + h^{(\mu)} s_\mu, \omega, \tau) - L_n(\mu_0, \omega, \tau) \} \]

\[= \sup_{\{s_\mu, s_\omega, s_\tau\}} \sup_{\{h^{(\mu)}, \omega, \tau\}} 2 \{ L_n(\mu_0 + h^{(\mu)} s_\mu, \omega, \tau, s_\omega, s_\tau) - L_n(\mu_0, \omega, \tau) \} \]

\[\Rightarrow \sup_{s_\omega} \mathcal{G}^{(\eta)}(s_\mu, s_\omega)'( -A^{(\eta)}(s_\mu, s_\omega)) \mathcal{G}^{(\eta)}(s_\mu, s_\omega) I_{\{\min[\mathcal{G}^{(\mu)}, \mathcal{G}^{(\omega)}] \geq 0\}} \]

\[+ \sup_{s_\omega} \max[0, \hat{\mathcal{G}}^{(\mu)}(s_\mu)]( -A^{(\mu, \mu)}(s_\mu)) \max[0, \hat{\mathcal{G}}^{(\mu)}(s_\mu)] I_{\{\mathcal{G}^{(\mu)} \geq 0\}} \]

\[+ \sup_{s_\omega} \max[0, \hat{\mathcal{G}}^{(\omega)}(s_\omega)]( -A^{(\mu, \omega)}(s_\omega)) \max[0, \hat{\mathcal{G}}^{(\omega)}(s_\omega)] I_{\{\mathcal{G}^{(\omega)} \geq 0\}} \]

\[+ \sup_{s_\tau} \max[0, \mathcal{G}^{(\tau)}(s_\tau)]( -A^{(\tau, \tau)}(s_\tau)) \max[0, \mathcal{G}^{(\tau)}(s_\tau)],\]

which is identical to

\[
\mathcal{G}^{(\eta)}(s_\mu, \bar{s}_\omega)(s_\mu)'</span>\left(-A^{(\eta, \eta)}(s_\mu, \bar{s}_\omega(s_\mu))\right)\mathcal{G}^{(\eta)}(s_\mu, \bar{s}_\omega(s_\mu)) I_{\{\min[\mathcal{G}^{(\mu)}, \mathcal{G}^{(\omega)}] \geq 0\}} \\
+ \max[0, \hat{\mathcal{G}}^{(\mu)}(s_\mu)]( -A^{(\mu, \mu)}(s_\mu)) \max[0, \hat{\mathcal{G}}^{(\mu)}(s_\mu)] I_{\{\mathcal{G}^{(\mu)} \geq 0\}} \\
+ \max[0, \hat{\mathcal{G}}^{(\omega)}(\bar{s}_\omega)]( -A^{(\omega, \omega)}(\bar{s}_\omega)) \max[0, \hat{\mathcal{G}}^{(\omega)}(\bar{s}_\omega)] I_{\{\mathcal{G}^{(\omega)} \geq 0\}} \\
+ \sup_{s_\tau} \max[0, \mathcal{G}^{(\tau)}(s_\tau)]( -A^{(\tau, \tau)}(s_\tau)) \max[0, \mathcal{G}^{(\tau)}(s_\tau)]
\]
by the definition of $\bar{s}_\omega(s_\mu)$ and $\bar{s}_\omega(s_\mu)$, where for each $s_\mu$, $\bar{s}_\omega(s_\mu)$ is such that

$$\max[0, \dot{G}^{(\omega)}(\bar{s}_\omega)](\dot{A}^{(\omega)}(\bar{s}_\omega)) = \sup_{s_\omega \in \Delta(\omega)} \max[0, \dot{G}^{(\omega)}(s_\omega)](\dot{A}^{(\omega)}(s_\omega)) \cdot$$

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

$$\sqrt{n}h_n^{(\mu)}(s_\mu) \Rightarrow G^{(\mu)}(s_\mu, \bar{s}_\omega(s_\mu)) \cdot$$

This also holds as a function of $s_\mu$, so that Theorem 8 follows by applying the CMT to the Wald test statistic again.

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

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

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

$$DL_n(\hat{\theta}_n; s_\mu) - DL_n(\theta_s; s_\mu) = \{DL_n(\mu_0, \bar{\omega}_n(s_\mu), \tau_s; s_\mu; \bar{\omega}_n(s_\mu))\} \{\bar{h}_n^{(\omega)}(\bar{\omega}_n, s_\mu)\}$$

$$= DL_n(\mu_0, \bar{\omega}_n(s_\mu), \tau_s; s_\mu; \bar{\omega}_n(s_\mu)) \{D^2L_n(\mu_0, \bar{\omega}_n(s_\mu), \tau_s; s_\mu; \bar{\omega}_n(s_\mu))\}^{-1} DL_n(\theta_s; \bar{\omega}_n(s_\mu)),$$

where we define $(\bar{h}_n^{(\omega)}(\bar{\omega}_n, s_\mu), \bar{\omega}_n)$ to satisfy that $L_n(\mu_0, \omega_\ast + \bar{h}_n^{(\omega)}(\bar{\omega}_n)\bar{\omega}_n, \tau_s) = \sup_{\bar{\omega}_n} \sup_{\omega_\ast} L_n(\mu_0, \omega_\ast + \bar{h}_n^{(\omega)}(\bar{\omega}_n)\bar{\omega}_n, \tau_s)$, and the last equality follows from the mean-value theorem: there is $\bar{\omega}_n(s_\mu)$ such that Eq. (11) holds. Given this and Theorem 6(i), we can also apply the ULLN to $n^{-1}DL_n(\mu_0,$
Moreover, it trivially holds that 
\[ n^{-1/2} (DL_n (\theta_s; \cdot), DL_n (\theta_s; \tilde{s}_w, n)) \Rightarrow (Z(\omega), \tilde{Z}(\omega)) \] 
by the facts that 
\[ n^{-1/2} (DL_n (\theta_s; \cdot), DL_n (\theta_s; \tilde{s}_w, n)) \Rightarrow (Z(\omega), \tilde{Z}(\omega)) \] 
as functions of \((s_\mu, s_\omega)\) and that 
\[ \max[0, DL_n (\theta_s; \tilde{s}_w, n)]^2 (\tilde{D}^2 (\theta_s; \tilde{s}_w, n))^{-1} \Rightarrow \max[0, \tilde{Y}(\tilde{\omega})]^2. \] 
Thus, we can obtain that 
\[ n^{-1/2} DL_n (\tilde{\bar{\theta}}; \cdot) \Rightarrow \tilde{Z}(\mu) (\cdot; \tilde{s}_w) \] 
by applying the CMT. Third, we can also apply the ULLN and obtain that 
\[ \sup_{(s_\mu, s_\omega)} |n^{-1} \tilde{D}^2 (\tilde{\bar{\theta}}; s_\mu, s_\omega) - \tilde{A}_n^{(s_\mu, s_\omega)}| \rightarrow 0 \ a.s.-\mathbb{P}. \] 
Given these three facts and Assumption 11(ii), the desired result now holds by the CMT.

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

\[ LM_n \Rightarrow \sup_{s_\xi \in \Delta(\xi)} \left( \frac{\max[0, \tilde{Z}(\xi)(s_\xi; \tilde{s}_\xi)]^2}{\inf_{s_\lambda \in \Delta(\lambda)} \{-\tilde{A}_n^{(\xi)}(s_\xi, s_\lambda)\} \{\tilde{A}_n^{(\xi)}(s_\xi, s_\lambda)\}} \right), \]

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

\[
\inf_{s_\lambda \in \Delta(\lambda)} \{-\tilde{A}_n^{(\xi)}(s_\xi, s_\lambda)\} = -\tilde{A}_n^{(\xi)}(s_\xi) - \sup_{s_\lambda \in \Delta(\lambda)} \{-\tilde{A}_n^{(\xi)(\lambda)}(s_\xi) s_\lambda\} (-s_\lambda \tilde{A}_n^{(\xi)(\lambda)} s_\lambda)^{-1} (-s_\lambda \tilde{A}_n^{(\xi)(\lambda)}(s_\xi)) = -\tilde{A}_n^{(\xi)(\lambda)}(s_\xi).
\]

Next, \[ \tilde{Z}(\xi)(s_\xi; \tilde{s}_\lambda) = Z(\xi)(s_\xi) - \{-\tilde{A}_n^{(\xi)(\lambda)}(s_\xi) s_\lambda\} (-s_\lambda \tilde{A}_n^{(\xi)(\lambda)} s_\lambda)^{-1} (-s_\lambda \tilde{A}_n^{(\xi)(\lambda)}(s_\xi)) = -\tilde{A}_n^{(\xi)(\lambda)} s_\lambda, \] and we already proved in the proof of Corollary 1(v) that \[ \tilde{s}_\lambda \tilde{A}_n^{(\xi)(\lambda)} s_\lambda^{-1} (-s_\lambda \tilde{A}_n^{(\xi)(\lambda)}(s_\xi)) = -\tilde{A}_n^{(\xi)(\lambda)}^{-1} s_\lambda^{-1} \tilde{Z}(\lambda), \] implying that \[ \tilde{Z}(\xi)(s_\xi; \tilde{s}_\lambda) = Z(\xi)(s_\xi) - \{-\tilde{A}_n^{(\xi)(\lambda)}(s_\xi) s_\lambda\} (-s_\lambda \tilde{A}_n^{(\xi)(\lambda)}(s_\xi)) = -\tilde{A}_n^{(\xi)(\lambda)}^{-1} \tilde{Z}(\lambda). \] These two facts imply the desired result.

**References**


Figure 1: Q-Q PLOT BETWEEN $\sup \max[0, \hat{\gamma}(\theta)(s_1, s_2)]^2$ AND $\sup \max[0, \hat{\gamma}(\theta)(s_1, s_2)]^2$. The horizontal line stands for the empirical quantiles of $\sup \max[0, \hat{\gamma}(\theta)(s_1, s_2)]^2$, and the vertical line stands for the empirical quantiles of $\sup \max[0, \hat{\gamma}(\theta)(s_1, s_2)]^2$. Here, $\hat{\gamma}(\theta)(s_1, s_2)$ is computed by assuming that $E[W_t^2]$ and $\text{var}(W_t^2)$ are known, whereas $\hat{\gamma}(\theta)(s_1, s_2)$ is computed by estimating $E[W_t^2]$ and $\text{var}(W_t^2)$ by the method of moments. The sample size is 500, and the number of iterations is 2,000. The 45-degree line shows that estimating $E[W_t^2]$ and $\text{var}(W_t^2)$ by the method of moments does not modify the asymptotic distribution.
Figure 2: **EMPIRICAL AND ASYMPTOTIC DISTRIBUTIONS OF THE QLR TEST STATISTIC.** This figure shows the asymptotic null distribution of the QLR test statistic, which is obtained by the simulation method, and the empirical distributions of the QLR test statistic for various sample sizes: \( n = 50, 200, \) and \( 1,000. \) The number of iterations for obtaining the empirical distributions is 2,000. We can see that the empirical distribution approaches the asymptotic null distribution as the sample size increases. The value of the QLR test statistic is computed by estimating \( \alpha^*, \sigma^2, \theta^*_1, \) and \( \theta^*_2. \)
Figure 3: Empirical and Asymptotic Distributions of the QLR Test Statistic. This figure shows the asymptotic null distribution of the QLR test statistic, which is obtained by the simulation method, and the empirical distributions of the QLR test statistic for various sample sizes: $n = 50, 100, \text{ and } 500$. The number of iterations for obtaining the empirical distributions is 2,000. We can see that the empirical distribution approaches the asymptotic null distribution as the sample size increases. Further, the convergence speed is faster than the previous case. Here, the value of the QLR test statistic is computed by estimating $\sigma^2_*, \theta_{1*}$, and $\theta_{2*}$. 
Figure 4: **Empirical and Asymptotic Distributions of the QLR Test Statistic.** This figure shows the asymptotic null distribution of the QLR test statistic, which is obtained as $\max[0, -Z]^2$, and the empirical distributions of the QLR test statistic for various sample sizes: $n = 50, 100, \text{and} 500$. The number of iterations for obtaining the empirical distributions is 2,000. We can see that the empirical distributions almost overlap with the asymptotic null distribution even when the sample size is as small as 50.