Directionally Differentiable Econometric Models

JIN SEO CHO
Department of Economics
Korea University
5-1, Anam, Seongbuk, Seoul, Korea
Email: jinseoch@korea.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: September 26, 2008

Abstract
We relax the differentiability condition for standard econometric models to the level of
directional (Gâteaux) differentiability and analyze asymptotic distribution of extremum esti-
mator. We show that its asymptotic distribution can be represented as a functional of Gaussian
process indexed by direction. Our analysis also treats the differentiable models as a special
case of directionally differentiable models.

For data inference, we refine standard likelihood ratio, Wald, and Lagrange multiplier test
statistics. These refinements also permit the presence of nuisance parameters. Further, from
this, these test statistics are shown to be asymptotically equivalent if the null models are dif-
ferentiable and do not have boundary parameters.

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

JEL Classification: C12, C13, C22, C32.

Acknowledgements: The authors have benefited from discussions with Horag Choi, Robert Davies,
Graham Elliott, Chirok Han, John Hillas, Estate Khmaladze, Chang-Jin Kim, Chang Sik Kim,
Tae-Hwan Kim, Bruce Lehmann, Mark Machina, Jaesun Noh, Donggyu Sul, Byung Sam Yoo
and other participants at the seminars at VUW, the University of Auckland, and the Econometrics
Study Group of the Korean Econometric Society.

Preliminary Version: PLEASE, DON’T CITE OR CIRCULATE WITHOUT AUTHORS’ PER-
MISSION.
1 Introduction

Model differentiability (difflity) is one of the regularity conditions for standard econometric models. For example, Wald (1943) supposes it as one of the regularity conditions for his statistic. As another example, Chernoff (1954) considers use of likelihood ratio (LR) statistic for various statistical models and exploits difflity to approximate log-likelihood functions by the second–order Taylor expansion.

Many important econometric models are not differentiable (diffie) due to various reasons, nevertheless. In particular, models obtained by re-parameterization often turn out not to be diffie. For example, King and Shively (1993) consider numerous examples belonging to this case. King and Shively (1993) run into the so–called Davies’ (1977, 1987) identification problem by re–parameterization. That is, they re–parameterize the models of interests into identified models when nuisance parameters are identified only under the alternative. These models are only directionally differentiable (d–diffie). Therefore, it is now of question how to analyze d–diffie models.

The goal of this paper is to extend the analysis tool for diffie models to the level of d–diffie models. For this, we exploit the fact that d–diffie models can be analyzed in the framework of the tightness of 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 these directional derivatives in a way to apply the functional central limit theorem (FCLT) and uniform law of large numbers (ULLN), so that d-diffie models can be analyzed as diffie models with another separate index.

Another goal of this paper is to revisit the d-diffie model considered by King and Shively (1993), standard diffie models for the generalized method of moments (GMM) estimation of Hansen (1982), and Box and Cox (1964) transformation. Although King and Shively (1993) re–parameterize the models of their interests to avoid identification problems, questions are not yet resolved. By the invariance principle, likelihood after re–parameterization has to be identical to the likelihood before the re–parameterization, which involves a Gaussian process. This in turn implies that model analysis after re–parameterization has to be different from that for standard econometric models. We aim to provide a suitable tool for model analysis when they are only d–diffie. This provision is also intended as a procedure to reconcile the invariance principle under their set–up.
We also demonstrate our analysis using a standard diffle model. The main goal for this is to show how the analysis for d–diffle models can be suited into diffle models, so that diffle model analysis can be regarded as a special case of d-diffle model analysis.

This paper is organized as follows. In Sections 2, d–diffle models are defined and examined, and diffle models are examined as a special case of d–diffle models. We also provide regularity conditions for d–diffle models and consider the asymptotic distribution of extremum estimator under these conditions. Section 3 considers data inference aspect for d–diffle models. For this, we redefine the standard likelihood ratio (LR), Wald, and Lagrange multiplier (LM) statistics appropriately modified for d–diffle models and derive their asymptotic null distributions. Also, a benchmark model is considered which enables these statistics asymptotically equivalent. Section 4 contains conclusion, and all the mathematical proofs are collected in the Appendix.

Before proceeding our discussion, we introduce some mathematical notations used throughout this paper. For any \( x \in \mathbb{R}^r \), \( \|x\| \) stands for the Euclidean norm. \( 1_{\{\cdot\}} \) and \( \text{cl}(A) \) stand for an indication function and a closure of a set \( A \) respectively. Also, unless confusion will otherwise result, we omit the argument of functions given below, so that for example, \( f(\cdot) \) is also denoted as \( f \).

### 2 Differentiable and Directionally Differentiable Models

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

**A1 (DGP):** A sequence of random variables, \( \{X_t \in \mathbb{R}^m\}_{t=1}^n \) \((m \in \mathbb{N})\), defined on a complete probability space \((\Omega, \mathcal{F}, P)\), is a strictly stationary and ergodic process.

Assumption A1 is a standard condition for time series data. Many economic data satisfy the given condition. These include the standard ARMA process, hidden Markov processes, GARCH processes, and so on.

The following model is examined in this paper, and we further provide the conditions for the consistency of an extremum estimator defined below.

**A2: (MODEL)** A sum of measurable functions indexed by \( \theta \), \( \{L_n(\theta) := \sum_{t=1}^n \ell_t(\theta; X_t) : \theta \in \Theta\} \),
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–\(\mathbb{P}\) (a.s.–\(\mathbb{P}\)), where for each \(t\), \(X^t\) denotes \((X_1, \cdots, X_t)\), and \(\Theta\) is a compact and convex set in \(\mathbb{R}^r\) with \(r \in \mathbb{N}\).

A3: (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 an unique \(\theta_* \in \Theta\), \(E[n^{-1}L_n]\) is maximized at \(\theta_* \in \Theta\) uniformly in \(n\).

We use Assumptions A2 and A3 to have a consistent extreme estimator defined as \(\hat{\theta}_n\):

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

This is mainly attained by applying the ULLN to \(n^{-1}L_n\).

Regarding these assumptions, there are several comments. Assumption A3(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–diffle models obscure. Further, unidentified models can be transformed into identified ones using the re-parameterization method of King and Shivley (1993). Thus, we pay attention to only identified models without big loss of generality. Another comment is that \(\theta_*\) can be on the boundary of \(\Theta\) as often ensued by the re–parameterization method in King and Shively (1993). Assumption A3(ii) still permits this. Finally, we compress \(X^t\) in \(\ell_t(\cdot; X^t)\) for notational simplicity, so that \(\ell_t\) denotes \(\ell_t(\cdot; X^t)\).

The consistency of the extremum estimator can be given as follows.

**Theorem 1.** Given Assumptions A1 to A3, \(\hat{\theta}_n\) converges to \(\theta_*\) a.s.–\(\mathbb{P}\).

As Theorem 1 is straightforward and well known in the literature (see Andrews (1999)), we don’t prove it for brevity. Nevertheless, Theorem 1 motivates further questions.

### 2.1 Directionally Differentiable Models

D–diffle functions are mainly related to the asymptotic distribution of \(\hat{\theta}_n\). In this subsection, we define d–diffle models and from this, characterize diffle models.
Definition (D–diffle Function): (i) A function \( f : \Theta \mapsto \mathbb{R} \) is called \textit{directionally (Gâteaux)} differentiable (d–diffle) at \( \theta \) in the direction \( d \in \Delta(\theta) \), 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–diffle 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–diffle on \( \Theta \), if for all \( \theta \in \Theta \), \( f \) is d–diffle on \( \Delta(\theta) \).

Note that the definition of d–diffle functions is weaker than that of diffle functions because a diffle function is d–diffle, though its converse does not hold. That is, d–diffle functions can have a different derivative from another if another direction is chosen. Further, there can be a continuum number of directions if \( r \) is greater than one, so that d–diffle functions can have a continuum number of different derivatives depending on \( d \). For diffle functions, however, the direction \( d \) does affect \( Df(\theta; d) \) up to the dimension of \( \Theta \). For example, if \( \theta \) is an interior element, then for any \( d \in \Delta(\theta) \), \( Df(\theta; d) \) can be written as a linear combination of other \( r \) directional derivatives. More precisely, for any \( d \), there is a vector of \((a_1, \cdots, a_r)\) such that \( Df(\theta; d) = \sum_{k=1}^{r} a_k Df(\theta; d_k) \), where \( d_i \neq d_j \) and \( i, j = 1, \ldots, r \). Thus, considering partial derivatives is sufficient to obtain the tangent hyperplane of \( f \) at \( \theta \). As another remark, we require that \( Df(\theta; d) \) is defined on \( \Delta(\theta) \).

This requirement is adopted to accommodate the recommendation of Chernoff (1954) for d–diffle models. When the asymptotic distribution of the extremum estimator is examined, it is essential to approximate the parameter space by a cone \( C(\theta) \). We define \( \Delta(\theta) \) in a way to collect only directions relevant to \( C(\theta) \), so that it can play the role of domain on which a Gaussian process given below is defined. Nevertheless, the unit circle-based \( \Delta(\theta) \) is not the only way we can consider. As an alternative, we can also consider \( \Delta(\theta) := \{ x \in \mathbb{R}^r : x + \theta \in \text{cl}\{C(\theta)\}, \|x\|_\infty = 1 \} \) to capture the same directions, where \( \| \cdot \|_\infty \) is the uniform norm. We focus here to \( \Delta(\theta) \) for a consistent presentation.

The following theorem clarifies the relationship between d–diffle and diffle functions.

Theorem 2 (Troutman (1996, p.122)): If (i) a function \( f : \Theta \mapsto \mathbb{R} \) is d–diffle 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 \( d \in \Delta(\theta) \), then \( f : \Theta \mapsto \mathbb{R} \) is differentiable on \( \Theta \).

The proof of Theorem 2 is Troutman (1996). Note that the linearity condition of \( Df(\theta; d, \cdot) \) in \( d \) is essential. Without this, any arbitrarily chosen directional derivative cannot be represented by a linear combination of other \( r \) directional derivatives. Before illustrating examples, we define continuously \( d \)-differentiable functions, which also plays another key role in our analysis.

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

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

Note that the first-order directional differentiability \( (d \)-differentiability) is necessary to define a twice continuously \( d \)-differentiable function. Also, twice continuously differentiable functions can be obtained from twice continuously \( d \)-differentiable functions by imposing further conditions. Mainly by imposing a quadratic function condition in \( d \), twice continuously differentiable functions can be obtained. Following Lemma 1 establishes this.

**Lemma 1:** If a function \( f : \Theta \mapsto \mathbb{R} \) satisfying the conditions in Theorem 2 is further (i) twice continuously \( d \)-differentiable on \( \Theta \);

(ii) for each \( \theta, \theta' \) and for some \( M < \infty \), \( |D^2f(\theta'; d) - D^2f(\theta; d)| \leq M\|\theta' - \theta\| \) uniformly on \( \Delta(\theta) \cap \Delta(\theta') \); and (iii) for each \( \theta \in \Theta \), \( D^2f(\theta; d) \) is continuous and quadratic in \( d \in \Delta(\theta) \), then \( f : \Theta \mapsto \mathbb{R} \) is twice continuously differentiable on \( \Theta \).

### 2.2 Examples

Three models are considered for illustration. The first is the one in King and Shively (1993); second is a standard differentiable model for GMM estimation; and the final is Box-Cox transformation.

#### 2.2.1 Example 1: King and Shively (1993)

One of the models considered by King and Shively (1993) is specified for a set of economic data set \( \{(Y_t, Q_t') := (Y_t, W_t, R_t') \} \), where \( R_t \) is a \( k \)-dimensional vector of regressors, and others are...
For this, they exploit the polar coordinates: 

\[ \theta \]

random variables defined on \( \mathbb{R} \). These are assumed to obey a DGP given as

\[
Y^n = W^n \alpha_n + R^n \beta_n + U^n, \\
U^n|Q^n \sim N(0, \sigma_n^2 \{I_n + \kappa_n \Omega^n(\rho_n)\}),
\]

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 \( \Omega^n(\rho_n) \) is an \( n \times n \) square matrix with \( t \)-th row and \( t' \)-th column element \( \Omega_{tt'}^n(\rho_n) := W_t W_{t'} \rho_n |t-t'|/(1 - \rho_n^2) \).

Given this, a model is specified by assuming that \( (\gamma_n', \sigma_n^2, \kappa_n, \rho_n) := (\alpha_n, \beta_n', \sigma_n^2, \kappa_n, \rho_n) \) is a unknown parameter in \( \Gamma \times [0, \sigma^2] \times [0, \bar{\kappa}] \times [0, \bar{\rho}] \), where \( \Gamma \) is a compact and convex subset of \( \mathbb{R}^{k+1} \); \( \sigma^2 \) and \( \bar{\kappa} \) are positive real numbers; and \( \bar{\rho} \) is also positive real number but less than one. Therefore, its log–likelihood function can be constructed as follows: for each \( (\gamma, \sigma^2, \kappa, \rho) \),

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

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

The motivation of this model traces from Rosenberg (1973), who aims to test \( \kappa_n = 0 \) to examine whether a systematic risk of an asset is time-varying or not. Nevertheless, if \( \kappa_n = 0 \) then \( \rho_n \) is not identified, so that Davies (1977, 1987) identification problem arises.

King and Shively (1993) attempt to overcome this challenge by re–parameterization method. For this, they exploit the polar coordinates: \( \theta_n' := (\theta_{1n}, \theta_{2n}) := (\kappa_n \cos(\rho_n \pi/2), \kappa_n \sin(\rho_n \pi/2)) \). Then corresponding parameter space for \( \theta \) can be given as \( [0, \bar{\kappa} \cos(\bar{\rho} \pi/2)] \times [0, \bar{\kappa} \sin(\bar{\rho} \pi/2)] \), and

\[
U^n|Q^n \sim N(0, \sigma_n^2 \{I_n + (\theta_n' \theta_n)^{1/2} \Omega^n(2 \tan^{-1}(\theta_{2n}/\theta_{1n})/\pi)\}).
\]

Also, the original hypotheses is rephrased into \( H_0' : \theta_n' \theta_n = 0 \) versus \( H_1' : \theta_n' \theta_n > 0 \). By this re–parameterizations, the identification problem doesn’t arise any longer under \( H_0' \).

Nevertheless, the model after re–parameterizations is only d–diffle under \( H_0' \), and the null parameter value is on the boundary. Note that the relevant log-likelihood function is modified into the following: for each \( (\gamma, \sigma^2, \theta) \),

\[
L_n(\gamma, \sigma^2, \theta) = - \frac{n}{2} \log (2\pi) - \frac{1}{2} \log \left( \det \left[ \sigma^2 \{I_n + (\theta' \theta)^{1/2} \Omega^n(2 \tan^{-1}(\theta_{2}/\theta_{1})/\pi)\} \right] \right) - \frac{1}{2\sigma^2} U^n(\gamma)' \left[ I_n + (\theta' \theta)^{1/2} \Omega^n(2 \tan^{-1}(\theta_{2}/\theta_{1})/\pi) \right]^{-1} U^n(\gamma),
\]
and from this it also follows that

$$DL_n(\gamma_*, \sigma_*^2, \theta_*; d) = -\frac{nd_2^2}{2\sigma_*^2} - \frac{(d_1^2 + d_2^2)^{1/2}}{2} \text{tr}[\Omega^n (2 \tan^{-1}(d_2/d_1)/\pi)] + \frac{d_2^2}{2\sigma_*^2} \Omega^n U^n (2 \tan^{-1}(d_2/d_1)/\pi) U^n,$$

where \( d := (d'_\gamma, d_\sigma, d_1, d_2)' \) such that \( \theta_* = 0 \) and \( d'd = 1 \). This is not linear with respect to \((d_1, d_2)\), although linear with respect to others. Thus, it’s not a diffle model, and we cannot analyze this in a standard way. The second–order directional derivative shares this aspect as well. We provide this here for future reference.

$$D^2L_n(\gamma_*, \sigma_*^2, \theta_*; d) = \frac{nd_2^2}{2\sigma_*^4} - \frac{d_2^2}{\sigma_*^6} \Omega^n U^n - \frac{1}{\sigma_*^2} (Q^n d_\gamma)' (Q^n d_\gamma) - \frac{2d_2^2}{\sigma_*^4} (Q^n d_\gamma)' (Q^n d_\gamma) U^n$$

\( + \frac{(d_1^2 + d_2^2)}{2} \text{tr}[Q^n (2 \tan^{-1}(d_2/d_1)/\pi)^2] 
- \frac{2(d_1^2 + d_2^2)^{1/2}}{\sigma_*^2} (Q^n d_\gamma)' [Q^n (2 \tan^{-1}(d_2/d_1)/\pi)] U^n 
- \frac{d_2^2}{\sigma_*^4} (d_1^2 + d_2^2)^{1/2} \Omega^n [Q^n (2 \tan^{-1}(d_2/d_1)/\pi)] U^n 
- \frac{(d_1^2 + d_2^2)}{\sigma_*^2} \Omega^n [Q^n (2 \tan^{-1}(d_2/d_1)/\pi)]^2 U^n.$$

Note that this is not quadratic in \( d \), either.

Given this original set–up, the model needs to be refined further. If \( d_1 \) is zero, then \( d_2/d_1 \) is not defined. We need to have an upper bound for \( d_2/d_1 \) to avoid this. This upper bound is equivalent to having a parameter space for \( \rho \) strictly less than one in terms of Rosenberg’s (1973) framework. Also, we cannot allow that \( d_2 = 0 \). The diagonal elements of \( \Omega^n(0) \) contain the elements mixed by \( 0^0 \), which cannot be defined. We refine this by referring Rosenberg’s (1973) original aim to test for the presence of a time–varying systematic risk, which cannot be attained by having a null model with time–varying variance. We therefore let \( d_2 \) be strictly positive. Having this lower bound is equivalent to letting \( \rho \) be away from zero in terms of the original model. Consequently, our refined parameter space for \( \theta \) is

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

By this modification, \( d_2/d_1 \) is constrained to \([\epsilon, \bar{\epsilon}] \).
2.2.2 Example 2: Generalized Method of Moments (GMM)

Hansen (1982) considers an estimation method generalizing the method of moments and also relevant inference. One of the regularity conditions for this is diffliy. We consider a representative 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; \( g_n(X^n; \theta) := n^{-1} \sum_{t=1}^{n} q(X_t; \theta) \); \( q_t := q(X_t; \cdot) : \Theta \rightarrow \mathbb{R}^k \) is continuously diffle a.s.–\( \mathbb{P} \) on \( \Theta \) satisfying the condition in A2 \( (r \leq k) \); 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

\[
E[q_t(\theta)'] \{-M^*_n\}^{-1} E[q_t(\theta)]
\]

on the interior part of \( \Theta \). We denote the maximum element of a matrix by \( \| \cdot \|_{\infty} \). Further, without providing with more primitive moment conditions, we suppose that \( n^{1/2} g_n(X^n; \theta^*) \Rightarrow W \sim N(0, S_*) \) for some positive definite matrix \( S_* \).

The given conditions for \( Q_n \) do not exactly satisfy the model conditions in A2. Nevertheless, Bates and White (1985, Theorem 2.6) shows that the GMM estimator \( \hat{\theta}_n \) is consistent for \( \theta^* \) under the given conditions, and we can apply our analysis for d-diffle models developed below even to GMM estiamtor, in which the first-order directional derivative plays the key role. Note that the first–order directional derivative of \( g_n(\cdot) := g_n(X^n; \cdot) \) is obtained as

\[
Dg_n(\theta; d) = \nabla_{\theta} g_n(X^n; \theta)' d,
\]

where \( \nabla_{\theta} g_n(X^n; \theta^*) := [\nabla_{\theta_1} g_n(X^n; \theta^*), \cdots, \nabla_{\theta_r} g_n(X^n; \theta^*)]' \). As is clear, \( Dg_n(\theta; d) \) is now linear in \( d \). Applying the mean-value theorem implies that for each \( d \),

\[
g_n(\theta; d) = g_n(\theta^*; d) + Dg_n(\bar{\theta}; d)(\theta - \theta^*)
\]

for some \( \bar{\theta} \) between \( \theta \) and \( \theta^* \), and

\[
DQ_n(\theta; d) = -2d' \nabla_{\theta} g_n(\theta)' M_n^{-1} g_n(\theta),
\]

8
so that for each \( d \), \( n^{1/2}DQ_n(\theta_*; d) \Rightarrow -2d' C_* M_*^{-1} W \) after applying the LLN to \( \nabla \theta g_n(\theta_*) \), which converges to \( C_* := E[\nabla \theta q_t(\theta_*)] \) a.s. from that \( \| \nabla \theta q_t(\cdot) \|_\infty \leq m(X_t) \). We use these to obtain the asymptotic behaviors of the GMM estimator below by the analysis for \( d \)-diffle models.

### 2.2.3 Example 3: Box-Cox Transformation

Applying directional derivatives makes model analysis more sensible for many nonlinear models with irregular properties. For example, Box and Cox (1964) 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

\[
L_n(\theta) := -\frac{1}{2} \sum_{t=1}^{n} \left\{ Y_t - Z_t \theta_0 - \frac{\theta_1}{\theta_2} (X_t^{\theta_2} - 1) \right\}^2 \tag{6}
\]

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

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

This specification is adopted to avoid identification problem when the parameter of interests is \( \theta_{1s} = 0 \). Note that \( \theta_{1s} = 0 \) if and only if \( \theta_{2s} = 0 \). Thus, we can avoid Davies (1977, 1987) identification problem when \( \theta_{1s} = 0 \). Nevertheless, if \( \theta_{1s} = 0 \) and \( \theta_{2s} = 0 \) then the model analysis becomes obscure because \( \theta_{1s}(X_t^{\theta_2} - 1)/\theta_{2s} = 0 \times 0/0 \) is not defined. We overcome this by applying the directional derivatives. 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

\[
L_n(\theta_* + h d) = -\frac{1}{2} \sum_{t=1}^{n} \left\{ Y_t - Z_t'(\theta_{0*} + d_0 h) - \frac{d_1}{d_2} (X_t^{d_2 h} - 1) \right\}^2 ,
\]

which is now differentiable with respect to \( h \) at 0. Further, this reformulation shows that different identification problem lurks behind (6). Note that \( d_1/d_2 \) lacks its corresponding distance and disappears when \( h \) is zero. Thus, \( d_1/d_2 \) is not identified at \( \theta_* = (\theta_{0*}', 0, 0)' \). Also note that

\[
DL_n(\theta_*; d) = \sum_{t=1}^{n} U_t \{ Z_t' d_0 + \frac{d_1}{d_2} \log(X_t) d_2 \} , \tag{7}
\]
and

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

(8)

so that for given \( d_1/d_2 \), (7) and (8) are linear and quadratic in \((d_0, d_2)\) respectively.

For future reference, we further assume that

\[ \left( n^{-1/2} \sum U_t Z_t', n^{-1/2} \sum U_t \log(X_t) \right)' \Rightarrow W := (W_0', W_1') \sim N(0, B_*) \],

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

Many nonlinear models with similar structures are found in the literature. For example, table 1 of Cheng, Evans, and Iles (1992) collects numerous nonlinear models with parameter instability problems together with appropriate re-parameterizations for the problems. Most of them have similar structures to Box-Cox transformation.

### 2.3 Asymptotic Distribution of Extremum Estimator

As contrasted by the examples in the previous subsection, the biggest difference between only d–diffle and diffle models is in the linearity condition of directional derivatives in \( d \). We provide further regularity conditions for d–diffle models.

A4 (D–DIFFLITY): A model \( \ell_t : \Theta \mapsto \mathbb{R} \) is twice continuously d–diffle on \( \Theta \) a.s.\( \Rightarrow \); and for each \( \theta \in \Theta \) and \( d \in \Delta(\theta) \), \( D^2 \ell_t(\cdot; d) \) is continuous on \( \Theta \) a.s.\( \Rightarrow \).

We use assumption A4 to approximate d–diffle models by the second–order directional Taylor expansion. If models are only d–diffle, separate Taylor approximations are needed for different directions. Further, the second–order directional derivatives don’t have to be quadratic with respect to \( d \). It can be any non–linear function of \( d \) satisfying the following regularity conditions.

A5 (REGULAR D–DIFFLITY): (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.\( \Rightarrow \);

(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 $|D \ell_t(\theta; d_1) - D \ell_t(\theta; d_2)| \leq M_1 \|d_1 - d_2\|^\lambda$ and $|D^2 \ell_t(\theta; d_1) - D^2 \ell_t(\theta; d_2)| \leq M_1 \|d_1 - d_2\|^\lambda$.

Assumption A5 will be always assumed with assumption A4 so that directional derivatives in A5 are meaningful. Note that A5(i and ii) are exactly stochastic analogs of the conditions in Theorem 2 and Lemma 1. Assumption A5(iii) is assumed to apply the tightness and ULLN to the first and second–order directional derivatives respectively. We discuss these below in detail when they are more relevant. If assumption A5(iii) is replaced by the following stronger assumption A5(iii)*, then the model is twice continuously diffle a.s.–$\mathbb{P}$ by Lemma 1.

A5: (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 A5* denote A5(i, ii, and iii*) below when diffle models are considered. Otherwise, A5 stands for A5(i, ii, and iii).

Additional regularity conditions are required to provide the asymptotic distribution of the extremum estimator. For this, we further restrict our prior regularity conditions:

A6 (CLT): (i) $E[D \ell_t(\theta_s; d)] = 0$ uniformly in $d \in \Delta(\theta_s)$ and $t$;

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

(iii) $B_*(d, \tilde{d})$ is strictly positive and finite uniformly in $d \in \Delta(\theta_s)$ and $n$, where for each $d, \tilde{d}$,

$$B_*(d, \tilde{d}) := \text{acov}\{n^{-1/2} L_n(\theta_s; d), n^{-1/2} L_n(\theta_s; \tilde{d})\},$$

and ‘acov’ denotes the asymptotic covariance of given arguments;

(iv) for some $q > (r - 1)/(\lambda \gamma)$ and $s > q \geq 2$, and for each $f_t \in \mathbb{L}$,

$$\|f_t - E[f_t | \mathcal{F}_{t-\tau}^{t+\tau}]\|_q \leq \nu_\tau,$$

where $\mathbb{L} := \{a_1 f_1 + a_2 f_2 : f_1, f_2 \in \{D \ell_t(\theta_s; \cdot, d) : d \in \Delta(\theta_s)\}, a_1, a_2 \in \mathbb{R}\}; \nu_\tau$ is of size $-1/(1 - \gamma)$ with $1/2 \leq \gamma < 1$; $\mathcal{F}_{t-\tau}^{t+\tau} := \sigma(Y_{t-\tau}, \cdots, Y_{t+\tau});$ 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 $\sup_{d \in \Delta(\theta_s)} \sup_{t=1,2,\cdots} \|D \ell_t(\theta_s, d)\|_s < \Delta < \infty$. 

11
Assumption A6(i) is imposed to apply the central limit theorem (CLT). Even if A6(i) doesn’t hold, the extremum estimator can still consistently estimate $\theta_*$ when $\theta_*$ can be on the boundary of $\Theta$. Nevertheless, the relevant statistics considered below can be degenerate without the zero first-order condition. Assuming A6(i) can prevent this. Assumption A6(iii) is also needed to keep the statistics from being degenerate. For notational simplicity, we denote $B_*(d, \tilde{d})$ below by $B_*(d)$ if $d = \tilde{d}$. Assumption A6(iv) is imposed to apply corollary 3.1 of Wooldridge and White (1988) and theorem 4 of Hansen (1996). Mainly from A6(iv) it follows that $n^{-1/2}DL_n(\theta_*; \cdot)$ obeys the functional central limit theorem. Wooldridge and White (1988) provide regularity conditions for the CLT of near-epoch processes as a special case of mixingale processes. Hansen (1996) generalizes this notion to the space of Lipschitz continuous functions and further provides the regularity conditions for the tightness of these functions. Assumption A6 is constructed to satisfy the regularity conditions of the CLT and the tightness. Note that the regularity conditions for CLT is further strengthened than those for the standard $L^2$-near epoch processes. The coefficient $q$ may have to be greater than two 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 ($\lambda$). Also, by proposition 2.9 of Wooldridge and White (1988), it is not hard to verify that $\{f_t, \sigma(X_t^d), \bar{L}\}$ is an $L^q$-mixingale class of Hansen (1996), where the space of the first-order directional derivatives is extended to the space of their linear combinations.

The limiting distribution of $\hat{\theta}_n$ can be derived using the regularity conditions provided so far. Our main plan is to approximate the model by the 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 related to another $d$–extremum estimator in distribution. By 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,

$$L_n(\hat{\theta}_n(d)) = \max_{\theta \in \Theta_*(d)} L_n(\theta),$$

where $\Theta_*(d) := \{\theta' \in \Theta : \theta' = \theta_* + hd, \ h \in \mathbb{R}^+, d \in \Delta(\theta_*)\}$. Note that the $d$–extremum estimator is constrained by $d$. That is, for given $d$, $\Theta_*(d)$ is a straight line starting from $\theta_*$ and ending at the point crossed by the boundary of $\Theta$ and $\{\theta_* + hd, \ h \in \mathbb{R}^+\}$. Certainly, $\Theta_*(d) \subset \Theta$,
so that \(L_n(\hat{\theta}_n(d)) \leq L_n(\hat{\theta}_n)\).

The d–extremum estimator \(\hat{\theta}_n(d)\) can be also represented by the distance between \(\theta_*\) and \(\hat{\theta}_n(d)\). By the constraint that \(\hat{\theta}_n(d) \in \Theta_*(d)\), we can define \(\hat{h}_n(d)\) such that \(\hat{\theta}_n(d) = \theta_* + \hat{h}_n(d) d\). By this identity, the asymptotic behavior of \(\hat{\theta}_n(d)\) is related to the asymptotic behavior of \(\hat{h}_n(d)\).

We define the space of \(h\) as \(H_*(d) := \{h \in \mathbb{R}^+ : \theta_* + h d \in \Theta_*(d)\}\). Then,

\[
\max_{h \in H_*(d)} L_n(\theta_* + h d) = L_n(\theta(d)).
\]

Given this device, we apply the directional second–order Taylor approximation with respect to \(h\). This needs to be related to the asymptotic behaviors of the first and second–order directional derivatives. The following lemma reveals them.

**Lemma 2:** Given Assumptions A1 to A6, for each \(d \in \Delta(\theta_*)\),

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

(ii) \(n^{-1} \mathbb{L}n(\theta_*; d) \Rightarrow A_*(d)\) a.s.–\(\mathbb{P}\);

(iii) \(\{n^{-1/2} DL_n(\theta_*; d), n^{-1} \mathbb{L}^2 n(\theta_*; d)\} \Rightarrow \{\mathcal{Z}(d), A_*(d)\}\).

These are not hard to prove, and they further form the basis for the asymptotic distribution of \(\hat{h}_n(d)\). To show this, we approximate \(L_n\) on \(\Theta_*(d)\) by the mean–value theorem. That is, for some \(\bar{\theta}_n(d) \in \Theta(d)\),

\[
L_n(\theta_* + h d) = L_n(\theta_*) + DL_n(\theta_*; d) h + \frac{1}{2} D^2 L_n(\bar{\theta}_n(d); d) h^2;
\]

where \(h \in H_*(d)\). Note that this approximation is carried out on \(H_*(d)\). Therefore, it follows from this and Lemma 2 that for each \(d \in \Delta(\theta_*)\),

\[
2\{L_n(\hat{\theta}_n(d)) - L_n(\theta_*)\} \Rightarrow \max_{h \in \mathbb{R}^+} \left[2\mathcal{Z}(d)\tilde{h} + A_*(d)\tilde{h}^2\right],
\]

where \(\tilde{h}\) captures the asymptotic behavior of \(\sqrt{n} h\). Further, their large sample properties can be provided as follows.

**Theorem 3:** Given Assumptions A1 to A6, for each \(d \in \Delta(\theta_*)\),

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

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

The main contents of Theorem 3 can be explained plainly as follows. The half–normal random variable $\mathcal{G}(d)$ in Theorem 3(i) is obtained as the weak limit of the argument in the right–hand side (RHS) of (10). That is,

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

The main reason involving the ‘max’ operator is that $\tilde{h}(d)$ lies on the positive real line, which is obtained as the limit of the set $\{\theta_* + \sqrt{n}h d : h \in H_*(d)\}$, so that if $Z(d)$ is negative then the RHS of (10) is obtained by letting $\tilde{h}(d) = 0$. Otherwise, the RHS is obtained by letting $\tilde{h}(d) = \mathcal{G}(d)$.

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 boundary parameter problem more fundamentally. We apply their approach to d–diffie 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 (10) respectively.

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

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

(11)

The question is how to derive the asymptotic behavior of $\hat{\theta}_n$ from that of $\hat{\theta}_n(d)$, and this can be answered by examining the functional relationship between $\hat{\theta}_n$ and $\hat{\theta}_n(\cdot)$. The following lemma promotes this examination.

**Lemma 3:** Given Assumptions A1 to A6,

(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;
(i) for all $\varepsilon > 0$, there is $n(\varepsilon)$ a.s. $-\mathbb{P}$ such that if $n > n(\varepsilon)$, then

$$\sup_{d \in \Delta(\theta_*)} |n^{-1} D^2 L_n(\theta_*; d) - A_*(d)| < \varepsilon.$$  

Lemma 3(i) is used to show that the first–order directional derivatives weakly converge to a Gaussian process indexed by $d$. Specifically, this shows the tightness of Billingsley (1999). Given our time series context, theorem 1 of Hansen (1996) provides sufficient regularity conditions for this, and we exploit this to show Lemma 3(i). Also, if $\Delta(\theta_*)$ is a set of finite number of elements, then Lemma 3 is not necessary in deriving the consequences of Theorem 4 given below. Note that if $\Theta$ is a subset of $\mathbb{R}$, then $\Delta(\theta_*)$ has finite elements. Lemma 3(i) is essential in showing the weak convergence when $\Theta$ has a dimension greater than one. Thus we suppose that $\Delta(\theta_*)$ has uncountable number of elements in proving Lemma 3(ii).

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

$$\sup_{d_1 - d_2 < \delta} n^{-1/2} |DL_n(\theta_*; d_1) - DL_n(\theta_*; d_2)| \leq \left\| n^{-1/2} \nabla_{\theta} L_n(\theta_*) \right\| \delta,$$

implying that for any $\varepsilon > 0$,

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

(12)

Thus, given that $n^{-1/2} \nabla_{\theta} L_n(\theta_*)$ obeys central limit theorem (CLT), we can choose $\delta$ in a way the RHS of (12) to be less than $\varepsilon$ as desired by Lemma 3(i). Likewise, we can also apply the ULLN to the second–order derivatives to show Lemma 3(ii) under the differ model assumption. Note that for each $d \in \Delta(\theta_*)$, $D^2 L_n(\theta_*; d) = d' \nabla^2_{\theta} L_n(\theta_*) d$, so that for a nontrivial metric, say the uniform metric, $\| \cdot \|_{\infty}$,

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

where we can make the RHS as small as we wish by applying the law of large numbers (LLN).

Lemma 3 extends Theorem 3 to the level of functional space, and from this, the asymptotic distribution of the extremum estimator follows. Specifically, the following Theorem 4 establishes this.
**Theorem 4:** Given Assumptions A1 to A6,

(i) \( \left\{ n^{-1/2}DL_n(\theta_s; \cdot ), n^{-1}D^2L_n(\theta_s; \cdot ) \right\} \Rightarrow (\mathcal{Z}, A_s), \) where for each \( d, d', E[\mathcal{Z}(d)\mathcal{Z}(d')] = B_s(d, d'); \)

(ii) \( \sqrt{n}\tilde{h}_n \Rightarrow \max[0, \mathcal{G}]; \)

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

(iv) \( \sqrt{n}\tilde{\theta}_n \Rightarrow \max[0, \mathcal{G}(d_s)]d_s, \) where \( d_s := \arg \max_{d \in \Delta(\theta_s)} \max[0, \mathcal{Z}(d)]^2\{-A_s(d)\}^{-1}. \)

Note that the extremum estimator is now represented by the Gaussian process given in Lemma 3 in the limit.

Theorem 4 accommodates the standard diffle models as a special case of d–diffle models. For this examination, we impose the following condition, which precludes trivial asymptotic behavior of the extremum estimator when models are diffle.

**A6:** (ii)* For a symmetric and negative definite matrix \( A_s \) and each \( d, A_s(d) = d'A_s d. \)

(iii)* For a symmetric and positive definite matrix \( B_s \) and each \( d, \tilde{d}, B_s(d, \tilde{d}) = d'B_s \tilde{d}. \)

Assumptions A6(ii and iii)* respectively correspond to assuming a positive definite \( B_s := \text{acov} \{n^{-1/2}\nabla_\theta L_n(\theta_s)\} \) and negative definite \( A_s := \lim_{n \to \infty} n^{-1}E[\nabla^2_\theta L_n(\theta_s)] \) for diffle models. These can refine the results of Theorem 4. We let A6* denote A6(i, ii*, iii*, and iv).

**Corollary 1:** Given Assumptions A1 to A4, A5*, and A6*,

(i) \( \mathcal{Z} \) is linear in \( d \in \Delta(\theta_s) \), so that for each \( d, \mathcal{Z}(d) = Z'd \) in distribution, where \( Z \sim N(0, B_s) \);

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

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

(iv) \( \sqrt{n}(\tilde{\theta}_n - \theta_s) \Rightarrow \max[0, Z'd\{d - d'A_s d\}^{-1}]d_s, \) where

\[ d_s := \arg \max_{d} \max[0, Z'd\{d - d'A_s d\}^{-1}]; \]

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

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

(vii) \( 2\{L_n(\hat{\theta}_n) - L_n(\theta_s)\} \Rightarrow Z'(-A_s)^{-1}Z, \) provided that \( \theta_s \) is interior to \( \Theta. \)
Note that the results are the same as in the standard case if \( \theta_* \) is an interior element of \( \Theta \). Nevertheless, our analysis is more primitive as our approach involves directional derivatives. Also, as Corollary 1 is trivial when \( r = 1 \), we suppose only the case \( r > 1 \) in proving Corollary 1.

### 2.4 Examples

#### 2.4.1 Example 1 (continued)

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

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

where for each \( d \),

\[
Z_{1,n}(d) := \frac{d_{\gamma_*}'}{\sigma_*^2} \sum_{t=1}^{n} Q_t U_t,
\]

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

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

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 \( i = 1, 2, 3 \),

\[
Z_i(d) \Rightarrow Z_i(d),
\]

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

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

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

and

\[
B_{\gamma_*}^{(3)}(d, \tilde{d}) := \frac{(d_1^2 + d_2^2)^{1/2} (\tilde{d}_1^2 + \tilde{d}_2^2)^{1/2}}{\{1 - m(d_2/d_1)^2\} \{1 - m(\tilde{d}_2/\tilde{d}_1)^2\}} \frac{m(d_2/d_1)m(\tilde{d}_2/\tilde{d}_1) E[W_t^2]^2}{1 - m(d_2/d_1)m(d_2/d_1)}. \]
Further, note that \( \mathcal{Z}_i(\mathbf{d}) \) and \( \mathcal{Z}_j(\mathbf{d}) \) 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

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

and leading to that \( n^{-1/2}DL_n(\gamma, \sigma^2, \theta; \mathbf{d}) \Rightarrow \mathcal{Z}(\mathbf{d}) := Z_1(\mathbf{d}) + Z_2(\mathbf{d}) + Z_3(\mathbf{d}) \) by the continuous mapping theorem (CMT).

It is not hard to generalize the given pointwise weak convergence to the level of functional space. It can be achieved by showing the tightness of \( n^{-1/2}DL_n(\gamma, \sigma^2, \theta; \mathbf{d}) \). As \( Z_{1,n} \) and \( Z_{2,n} \) are virtually linear with respect to \( Q_tU_t \) and \( (U_t^2 - \sigma^2) \) respectively, their tightness trivially follows. We pay attention only to \( Z_{3,n} \) for brevity. For notational simplicity, we let \( \varepsilon_t := W_tU_t \), \( m := m(d_2/d_1) \), and \( \tilde{m} := m(\tilde{d}_2/\tilde{d}_1) \). Then for any \( \epsilon > 0 \), there is \( \delta \) such that

\[
\limsup_{n \to \infty} \mathbb{P}_n \left( \sup_{|m - \tilde{m}| < \delta} \left| n^{-1/2} \sum_{t=2}^{n} \varepsilon_t \sum_{t' = 1}^{t-1} \varepsilon_{t'} m^{t-t'} - n^{-1/2} \sum_{t=2}^{n} \varepsilon_t \sum_{t' = 1}^{t-1} \varepsilon_{t'} \tilde{m}^{t-t'} \right| > \epsilon \right) < \epsilon. \tag{14}
\]

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 (1996) 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 (1996) \( \lambda \) and \( a \) are identical to 1 in our model assumption,

\[
\limsup_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E[\varepsilon_t^2 \sum_{t' = 1}^{t-1} \varepsilon_{t'} m^{t-t'}] = (\sigma_s \Delta)^4 \left( \frac{m^2}{1 - m^2} \right) < \infty
\]

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

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

by the standard argument using the facts that \( |m| \) is uniformly and strictly bounded by one and that \( E[|\varepsilon_t^2 \varepsilon_{t'} \varepsilon_{t''}|] < (\sigma_s \Delta)^4 < \infty \), where \( \tilde{m} := \max( |m(\varepsilon)|, |m(\bar{\varepsilon})|) \). From these, it follows that \( n^{-1/2}DL_n(\gamma, \sigma^2, \theta; \cdot) \Rightarrow \mathcal{Z} \) such that for each \( \mathbf{d} \) and \( \tilde{\mathbf{d}} \), \( E[\mathcal{Z}(\mathbf{d})\mathcal{Z}(\tilde{\mathbf{d}})] = B_*(\mathbf{d}, \tilde{\mathbf{d}}) := B^{(1)}_*(\mathbf{d}, \tilde{\mathbf{d}}) + B^{(2)}_*(\mathbf{d}, \tilde{\mathbf{d}}) + B^{(3)}_*(\mathbf{d}, \tilde{\mathbf{d}}) \).
The asymptotic behavior of the second–order directional derivative can be related to $B_s$. After applying the LLN to the last elements in the first two lines of (2),

\[
D^2 L_n(\gamma_s, \sigma^2_s; \theta_s; d) = -\frac{1}{\sigma^4_s} d_{\gamma_s} Q^n d_{\gamma_s} - \frac{d_{\sigma_s^2}(d_1^2 + d_2^2)^{1/2}}{\sigma^4_s} U^{n'} \Omega^n(m(d_2/d_1)) \mathbf{U}^n \\
- \frac{nd_{\sigma_s^2}^2}{2\sigma^4_s} + \frac{(d_1^2 + d_2^2)}{2} \left\{ \text{tr} \left[ \Omega^n(m(d_2/d_1))^2 \right] - \frac{2}{\sigma_s^2} U^{n'} D^n(m(d_2/d_1)) \mathbf{U}^n \right\} \\
- \frac{(d_1^2 + d_2^2)}{\sigma^2_s} U^{n'} O^n(m(d_2/d_1)) \mathbf{U}^n + o_P(n),
\]

where $D^n(m(d_2/d_1))$ is a diagonal matrix with the diagonal elements 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 $\Omega^n(m(d_2/d_1))^2$, 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 $n^{-1} D^2 L_n(\gamma_s, \sigma^2_s; \theta_s; d) = -B_s(d, d) + o_P(1)$. This aspect is further strengthened by applying the ULLN. That is, $\sup_d |n^{-1} D^2 L_n(\gamma_s, \sigma^2_s, \theta_s; d) + B_s(d, d)| = o_P(1)$, leading to the information matrix equality. Therefore,

\[
2 \{ L_n(\gamma_n, \sigma^2_n; \hat{\theta}_n) - L_n(\gamma_s, \sigma^2_s, \theta_s) \} \Rightarrow \sup_d \{0, \mathcal{Y}(d)\}^2
\]

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

\[
E[\mathcal{Y}(d) \mathcal{Y}(\tilde{d})] = \frac{B_s(d, \tilde{d})}{\{B_s(d, d)\}^{1/2} \{B_s(\tilde{d}, \tilde{d})\}^{1/2}}.
\]

### 2.4.2 Example 2 (continued)

Given the first–order directional derivative (3) and the regularity conditions for GMM estimation, it is trivial to show that $\{n^{1/2} DQ_n(\theta_s, \cdot)\}$ is tight using the fact that it is linear in $d$. Next, by substituting $g_n$ in (4) into $Q_n$ we obtain that for some $\bar{\theta}$ between $\theta$ and $\theta_s$,

\[
n \{Q_n(\theta) - Q_n(\theta_s)\} = -2d' \nabla_\theta g_n(\bar{\theta})' M_n^{-1} \sqrt{n} g_n(\theta_s) \sqrt{n} h - d' \nabla_\theta g_n(\bar{\theta})' M_n^{-1} \nabla_\theta g_n(\bar{\theta}) d(\sqrt{n} h)^2,
\]

and so

\[
\{Q_n(\hat{\theta}_n) - Q_n(\theta_s)\} \Rightarrow \sup_d \sup_h -2d' C_s' M_s^{-1} W h - d' C_s' M_s^{-1} C_s d h^2.
\]

This structure can be transformed into (10) by letting $Z(d) := -d' C_s' M_s^{-1} W$ and $A_s(d) := -d' C_s' M_s^{-1} C_s d$. Note that these are linear and quadratic in $d$ respectively. Therefore, we can
Note that the results in Corollary 1 even if the model assumption A2 does hold exactly. That is,
\[ \{Q_n(\hat{\theta}_n) - Q_n(\theta^*_s)\} \Rightarrow W' M_s^{-1} C_s \{ -C_s' M_s^{-1} C_s \}^{-1} C_s' M_s^{-1} W \]
by Corollary 1(vii). Further, Corollary 1(v) implies that
\[ \sqrt{n}(\hat{\theta}_n - \theta^*_s) \Rightarrow - \{ C_s' M_s^{-1} C_s \}^{-1} C_s' M_s^{-1} W \]
\[ \sim N(0, \{ C_s' M_s^{-1} C_s \}^{-1} \{ C_s' M_s^{-1} S_s M_s^{-1} C_s \} \{ C_s' M_s^{-1} C_s \}^{-1}). \]
These are the same results as given in the standard GMM literature (e.g. Newey and West (1987)).

2.4.3 Example 3 (continued)

Given the first and second-order derivatives in (7) and (8),
\[ n^{-1/2} DL_n(\theta^*_s; d) \Rightarrow \tilde{d}' W(d_1/d_2), \quad \text{and} \quad n^{-1} D^2 L_n(\theta^*_s; d) \Rightarrow \tilde{d}' \{ A_s(d_1/d_2) \} \tilde{d} \]
a.s. \(-\mathbb{P}\), where \( \tilde{d} \in \tilde{\Delta}(\theta^*_s) := \{ x \in \mathbb{R}^{k+1} : \| x \| = 1 \}, W(d_1/d_2) := (W_{0}', \frac{d_1}{d_2} W_1)' \),
\[ A_s(d_1/d_2) := \begin{bmatrix} A_s^{(0,0)} & \frac{d_1}{d_2} A_s^{(0,1)} \\ \frac{d_1}{d_2} A_s^{(1,0)} & (\frac{d_1}{d_2})^2 A_s^{(1,1)} \end{bmatrix}, \]
and
\[ A_s := \begin{bmatrix} A_s^{(0,0)} & A_s^{(0,1)} \\ A_s^{(1,0)} & A_s^{(1,1)} \end{bmatrix} := \begin{bmatrix} -E[Z_t Z_t'] & -E[Z_t \log(X_t)] \\ -E[\log(X_t) Z_t'] & -E[\log(X_t)^2] \end{bmatrix}. \]

Note that \( \tilde{\Delta}(\theta^*_s) \) is a refined set of directions of \( (d_0', d_2) \) accommodating that \( d_1/d_2 \) is not identified if \( h = 0 \). The weak convergence of \( n^{-1/2} DL_n(\theta^*_s; \cdot) \) and the ULLN of \( n^{-1} D^2 L_n(\theta^*_s; \cdot) \) follow by the same reasons as in Example 2. Thus,
\[ 2\{ L_n(\hat{\theta}_n) - L_n(\theta^*_s) \} \Rightarrow \sup_{d_1/d_2 \in [\epsilon, \tilde{\epsilon}]} \sup_{d \in \tilde{\Delta}(\theta^*_s)} \max[0, W(d_1/d_2)' \tilde{d}]^2 \{ -\tilde{d}' A_s(d_1/d_2) \tilde{d} \}^{-1} \]
\[ = \sup_{d_1/d_2 \in [\epsilon, \tilde{\epsilon}]} W(d_1/d_2)' \{ -A_s(d_1/d_2) \}^{-1} W(d_1/d_2) \]
\[ = \sup_{d_1/d_2 \in [\epsilon, \tilde{\epsilon}]} W'(-A_s)^{-1} W = W'(-A_s)^{-1} W, \]
where \( \hat{\theta}_n \) is the NLS estimator, and we applied the proof of Corollary 1(vii) to show the first equality.
3 Testing Hypotheses with D–diffle Models

In this section, data inference induced by d–diffle models is examined. For this, the standard LR, Wald, and LM test statistics are refined to accommodate d-diffle.

In discussing data inference based on d–diffle models, it is efficient to specify first the roles of each parameter. 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 $\tau$ ($\in \mathbb{R}^r$) consists of other nuisance parameters asymptotically orthogonal to $\pi = (\lambda', \nu')'$ ($\in \mathbb{R}^r$) 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

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

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

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

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

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

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

$DL_n(\lambda), DL_n(\nu)(d_\nu),$ and $DL_n(\tau)(d_\tau)$ are defined on $\mathbb{R}^\lambda \times \mathbb{R}^\lambda \times \mathbb{R}^\lambda$; $B_\ast(\lambda, \lambda)(d_\lambda)$ and $B_\ast(\lambda, \nu)(d_\nu)$ are defined on $\mathbb{R}^\lambda$; $B_\ast(\lambda, \tau)(d_\lambda)$ is defined on $\mathbb{R}^\lambda$; $B_\ast(\nu, \lambda)(d_\nu)$ is defined on $\mathbb{R}^\lambda$; and $B_\ast(\nu, \tau)(d_\nu)$ is defined on $\mathbb{R}^\lambda$. Thus,

$$\text{acov} \left\{ n^{-1/2}DL_n(\theta_\ast; d), n^{-1/2}DL_n(\theta_\ast; \tilde{d}) \right\} = \nu_n' B_\ast(d, \tilde{d}) \nu_n,$$

where $\nu_n$ is the $n \times 1$ vector of ones. Accordingly, we also suppose that $A_\ast(d) = \nu_n' A_\ast(d) \nu_n$, 

21
where

\[
A_s(d)_{(3 \times 3)} := \begin{bmatrix}
A_s^{(\pi, \pi)}(d_\pi) & A_s^{(\pi, \tau)}(d_\pi, d_\tau) \\
A_s^{(\tau, \pi)}(d_\tau, d_\pi) & A_s^{(\tau, \tau)}(d_\tau)
\end{bmatrix}
\]

(16)

\[
:= \begin{bmatrix}
d_\lambda A_s^{(\lambda, \lambda)} d_\lambda & d_\lambda A_s^{(\lambda, \nu)}(d_\nu) & d_\lambda A_s^{(\lambda, \tau)}(d_\tau) \\
A_s^{(\nu, \lambda)}(d_\nu, d_\lambda) & A_s^{(\nu, \nu)}(d_\nu) & A_s^{(\nu, \tau)}(d_\nu, d_\tau) \\
A_s^{(\tau, \lambda)}(d_\tau, d_\lambda) & A_s^{(\tau, \nu)}(d_\tau, d_\nu) & A_s^{(\tau, \tau)}(d_\tau)
\end{bmatrix}
\]

R \in \mathbb{R}^{3 \times 3}; A_s^{(\lambda, \nu)}(d_\nu) \in \mathbb{R}^{r_\nu}; A_s^{(\lambda, \tau)}(d_\tau) = A_s^{(\lambda, \tau)}(d_\tau). Given A_s(d), \pi is assumed to be orthogonal to \tau. That is, for each d, A_s^{(\pi, \pi)}(d_\pi, d_\pi) = A_s^{(\pi, \tau)}(d_\pi, d_\tau) = 0. This assumption is useful in reducing the number of ineffective nuisance parameters for testing purpose. In terms of these definitions, we also permit that \rho, \nu, or \tau may not be present in the model. If \rho, \nu, and \tau are zero, then the model is twice continuously differentiable. We formally contain these conditions in A7.

A7 (D-DERIVATIVES): (i) For each d, \text{DL}_n(\theta_\pi; d) = \text{DL}_n^{(\pi)}(d_\pi) + \text{DL}_n^{(\tau)}(d_\tau), and n^{-1/2}(\text{DL}_n^{(\pi)}, \text{DL}_n^{(\tau)}) \Rightarrow (\mathcal{Z}^{(\pi)}, \mathcal{Z}^{(\tau)});

(ii) For each d, \tilde{d} \in \Delta(\theta_\pi), B_s(d, \tilde{d}) = \nu_3^t B_s(d, \tilde{d}) \nu_3, where for each d, B_s(d, d) is the symmetric and positive definite matrix given in (15);

(iii) For each d \in \Delta(\theta_\pi), A_s(d) = \nu_3^t A_s(d) \nu_3, where for each d, A_s(d) is the symmetric and negative definite matrix given in (16);

(iv) A_s^{(\tau, \pi)}(d_\tau, d_\pi) = A_s^{(\pi, \tau)}(d_\pi, d_\tau) = 0 uniformly in d \in \Delta(\theta_\pi);

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

Condition A7(v) restricts the parameter space \Theta into the Cartesian product of separate parameter spaces \Pi and T. This will be used to represent \text{L}_n as the sum of two independent functions as discussed in Theorem 5 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 \text{Z}^{(\lambda)}, A_s^{(\lambda, \lambda)}, and B_s^{(\lambda, \lambda)} can be accordingly represented as \text{Z}^{(\lambda)} = (\text{Z}^{(\delta')}, \text{Z}^{(\phi')})',

\[
A_s^{(\lambda, \lambda)} = \begin{bmatrix}
A_s^{(\delta, \delta)} & A_s^{(\delta, \phi)} \\
A_s^{(\phi, \delta)} & A_s^{(\phi, \phi)}
\end{bmatrix}
\]

and

\[
B_s^{(\lambda, \lambda)} = \begin{bmatrix}
B_s^{(\delta, \delta)} & B_s^{(\delta, \phi)} \\
B_s^{(\phi, \delta)} & B_s^{(\phi, \phi)}
\end{bmatrix}.
\]
Similarly, for each \( d_v \) we also let \( Z^{(v)}(d_v) = Z^{(\xi)}(d_\xi) + Z^{(\psi)}(d_\psi) \), where for each \( d_\xi \) and \( d_\psi \), \( Z^{(v)}(d_v) := (Z^{(\xi)}(d_\xi), Z^{(\psi)}(d_\psi))' \sim N(0, B_{s(v,v)}^{(\xi,\psi)}(d_v, d_v)) \), whose covariance can be written as

\[
B_{s(v,v)}^{(\xi,\psi)}(d_v, \tilde{d}_v) := \begin{bmatrix}
B_{s}^{(\xi,\xi)}(d_\xi, \tilde{d}_\xi) & B_{s}^{(\xi,\psi)}(d_\xi, \tilde{d}_\psi) \\
B_{s}^{(\psi,\xi)}(d_\psi, \tilde{d}_\xi) & B_{s}^{(\psi,\psi)}(d_\psi, \tilde{d}_\psi)
\end{bmatrix}.
\]  

(17)

This also implies that \( B_{s(v,v)}^{(\xi,\psi)}(d_v, d_v) = B_{s}^{(\xi,\xi)}(d_\xi, d_\xi) + 2B_{s}^{(\xi,\psi)}(d_\xi, d_\psi) + B_{s}^{(\psi,\psi)}(d_\psi, d_\psi) \). For a consistent presentation we also assume that \( A_{s(v,v)}^{(\xi,\psi)}(d_v) = \nu_2 A_{s}^{(\xi,\psi)}(d_v) \nu_2 \) for a negative definite

\[
A_{s(v,v)}^{(\xi,\psi)}(d_v) := \begin{bmatrix}
A_{s}^{(\xi,\xi)}(d_\xi) & A_{s}^{(\xi,\psi)}(d_\xi, d_\psi) \\
A_{s}^{(\psi,\xi)}(d_\psi, d_\xi) & A_{s}^{(\psi,\psi)}(d_\psi)
\end{bmatrix}.
\]  

(18)

These assumptions are formally contained in A8.

A8 (INERENCE): (i) For each \( d_v \) and for the symmetric and positive definite \( B_{s(v,v)}^{(\xi,\psi)}(d_v, d_v) \) given in (17), \( Z^{(v)}(d_v) = Z^{(v)}(d_v)' \nu_2 \) such that \( Z^{(v)}(d_v) \sim N(0, B_{s(v,v)}^{(\xi,\psi)}(d_v, d_v)) \);

(ii) For each \( d_v \) and for the symmetric and negative definite \( A_{s(v,v)}^{(\xi,\psi)}(d_v) \) given in (18), \( A_{s(v,v)}^{(\xi,\psi)}(d_v) = \nu_2 A_{s}^{(\xi,\psi)}(d_v) \nu_2 \).

Note that assumption A8 does not impose any restriction to \( A_{s}^{(\lambda,\lambda)} \) and \( B_{s}^{(\lambda,\lambda)} \) in terms of their signs. This is because assumption A7 will be always assumed together with assumption A8. It follows from A7 that both \( -A_{s}^{(\lambda,\lambda)} \) and \( B_{s}^{(\lambda,\lambda)} \) are symmetric and positive definite, although assumptions A8(i and ii) are not implied by A7.

Given the partition of \( \theta \), we let \( \delta \) or \( \xi \) be the parameter of interest, so that null hypotheses are given as follows:

\[
\begin{cases}
H_0' : \delta_s = \delta_0; \\
H_1' : \delta_s \neq \delta_0
\end{cases} \quad \text{or} \quad \begin{cases}
H_0'' : \xi_s = \xi_0; \\
H_1'' : \xi_s \neq \xi_0
\end{cases}
\]

Note that the role of the parameters in the first hypotheses (\( H_0' \) versus \( H_1' \)) is different from that in the second hypotheses (\( H_0'' \) versus \( H_1'' \)). That is, the directional derivative with respect to \( d_\delta \) is linear, but that with respect to \( d_\xi \) is nonlinear.

For future reference, we 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 interests.
For brevity, we introduce a new notation and handle these null 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 for $\eta, \mu,$ and $\omega$ respectively and suppose that the null is

$$
\begin{cases}
H_0 : \mu_* = \mu_0; \\
H_1 : \mu_* \neq \mu_0,
\end{cases}
$$

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 $A_*(\pi, \pi)(d_\pi)$ and $B_*(\pi, \pi)(d_\pi, d_\pi)$ into

$$
A_*(\eta, \eta)(d_\mu, d_\omega) := \begin{bmatrix} A_*(\mu, \mu)(d_\mu) & A_*(\mu, \omega)(d_\mu, d_\omega) \\ A_*(\omega, \mu)(d_\omega, d_\mu) & A_*(\omega, \omega)(d_\omega) \end{bmatrix}
$$

and

$$
B_*(\eta, \eta)((d_\mu, d_\omega), (d_\mu, d_\omega)) := \begin{bmatrix} B_*(\mu, \mu)(d_\mu, d_\mu) & B_*(\mu, \omega)(d_\mu, d_\omega) \\ B_*(\omega, \mu)(d_\omega, d_\mu) & B_*(\omega, \omega)(d_\omega, d_\omega) \end{bmatrix}
$$

respectively. Given this reformulation, we restructure assumption $A7(v)$ into the following.

A8 (INERENCE): (iii) $H = M \times \Omega$ and $C(\pi_*) = C(\mu_*) \times C(\omega_*)$, where $C(\mu_*) := \{x \in \mathbb{R}^{r_\mu} : \exists \mu' \in M, x := \mu_* + \delta \mu', \delta \in \mathbb{R}^+\}$ and $C(\omega_*) := \{x \in \mathbb{R}^{r_\omega} : \exists \omega' \in \Omega, x := \omega_* + \delta \omega', \delta \in \mathbb{R}^+\}$.

3.1 Likelihood Ratio Statistic

The standard LR test statistic for different models can serve for $d$–different models. We define LR statistic as

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

where $\hat{\theta}_n$ is such that $L_n(\hat{\theta}_n) := \sup_{\theta \in \Theta_n} L_n(\theta)$, which is the same as the standard LR statistic. Thus, this LR statistic can be analyzed in a similar manner to the standard LR statistic as well. Specifically, $\mathcal{LR}_n$ can be split into $\mathcal{LR}_n^{(1)}$ and $\mathcal{LR}_n^{(2)}$ such that $\mathcal{LR}_n^{(1)} := 2\{L_n(\hat{\theta}_n) - L_n(\theta_*)\}$ and $\mathcal{LR}_n^{(2)} := 2\{L_n(\hat{\theta}_n) - L_n(\theta_*)\}$, and it turns out that the given results in Theorem 4(ii) is significantly simplified by assumption A7. We can separate LR statistic into a sum of two components, one of which has the same probability limit as $\mathcal{LR}_n^{(2)}$. We examine this by first letting
\[ \Delta(\pi_s) := \{ x \in \mathbb{R}^r : \pi_s + x \in \text{cl}\{C(\pi_s)\}, \|x\| = 1 \} \] and \[ \Delta(\tau_s) := \{ x \in \mathbb{R}^r : \tau_s + x \in \text{cl}\{C(\tau_s)\}, \|x\| = 1 \} \] and denoting their representative components by \( s_\pi = (s_\lambda, s_v) \) and \( s_\tau \) respectively. The different notation \( s \) is used to distinguish them from \( d \) defined on \( \Delta(\theta_s) \). Note that \( \Delta(\pi_s) \) and \( \Delta(\tau_s) \) are subsets of \( \Delta(\theta_s) \), so that they can be defined without difficulty.

**Theorem 5:** Given Assumptions A1 to A7,

\[ 2\{I_n(\hat{\theta}_n) - I_n(\theta_s)\} \Rightarrow \mathcal{H}_1 + \mathcal{H}_2 := \sup_{s_\pi \in \Delta(\pi_s)} \max[0, \mathcal{J}^{(\pi)}(s_\pi)]^2 + \sup_{s_\tau \in \Delta(\tau_s)} \max[0, \mathcal{J}^{(\tau)}(s_\tau)]^2, \]

where for each \( s_\pi \) and \( s_\tau \),

\[ \mathcal{J}^{(\pi)}(s_\pi) := \{-\mathbf{t}_2^\prime \mathbf{A}_s^{(\pi,\pi)}(s_\pi) \mathbf{t}_2\}^{-1/2} \mathbf{Z}(\pi)(s_\pi), \hspace{1cm} \mathcal{J}^{(\tau)}(s_\tau) := \{-\mathbf{A}_s^{(\tau,\tau)}(s_\tau)\}^{-1/2} \mathbf{Z}(\tau)(s_\tau), \]

respectively, and \( \mathbf{Z}(\pi)(s_\pi) := s_\lambda^\prime \mathbf{Z}(\lambda) + \mathbf{Z}(v)(s_v) \).

The orthogonality condition A7(iv) and parameter space assumption A7(v) separate \( I_n \) into two pieces and yield Theorem 5. Thus, we can ignore the pieces containing the orthogonal nuisance parameters \( \tau \) (resp. \( \pi \)) in testing the null involving only \( \pi \) (resp. \( \tau \)).

Theorem 5 enables us to identify the asymptotic null distribution of \( \mathcal{L} \mathcal{R}_n \) as the difference between the weak limits of \( \mathcal{L} \mathcal{R}_n^{(1)} \) and \( \mathcal{L} \mathcal{R}_n^{(2)} \). That is,

\[ \mathcal{L} \mathcal{R}_n^{(2)} \Rightarrow \mathcal{H}_0 + \mathcal{H}_2 := \sup_{s_\omega \in \Delta(\omega_s)} \max[0, \mathcal{J}^{(\omega)}(s_\omega)]^2 + \mathcal{H}_2, \]

where for each \( s_\omega \), \( \mathcal{J}^{(\omega)}(s_\omega) := \{-\mathbf{A}_s^{(\omega,\omega)}(s_\omega)\}^{-1/2} \mathbf{Z}(\omega)(s_\omega) \) and \( \Delta(\omega_s) := \{ x \in \mathbb{R}^r : \omega_s + x \in \text{cl}\{C(\omega_s)\}, \|x\| = 1 \} \).

There are several comments relevant to Theorem 6. First, the consequence in Theorem 6(ii)

\[ \mathcal{L} \mathcal{R}_n^{(2)} \Rightarrow \mathcal{H}_0 + \mathcal{H}_2 := \sup_{s_\phi, s_v \in \Delta(\omega_s)} \max[0, \mathcal{J}^{(\phi)}(s_\phi)^2 + \mathbf{Z}(\phi)(s_v)]^2 \]

In a similar manner, if \( \mu = \xi \), then \( \omega = (\lambda', \psi')' \), so that

\[ \mathcal{L} \mathcal{R}_n^{(2)} \Rightarrow \mathcal{H}_0 + \mathcal{H}_2 := \sup_{s_\lambda, s_\psi \in \Delta(\omega_s)} \max[0, \mathbf{s}_\lambda' \mathbf{Z}(\lambda) + \mathbf{Z}(\psi)(s_\psi)]^2 \]

In a similar manner, if \( \mu = \xi \), then \( \omega = (\lambda', \psi')' \), so that

\[ \mathcal{L} \mathcal{R}_n^{(2)} \Rightarrow \mathcal{H}_0 + \mathcal{H}_2 := \sup_{s_\lambda, s_\lambda \in \Delta(\omega_s)} \max[0, \mathbf{s}_\lambda' \mathbf{Z}(\lambda) + \mathbf{Z}(\psi)(s_\psi)]^2 \]

In a similar manner, if \( \mu = \xi \), then \( \omega = (\lambda', \psi')' \), so that

\[ \mathcal{L} \mathcal{R}_n^{(2)} \Rightarrow \mathcal{H}_0 + \mathcal{H}_2 := \sup_{s_\lambda, s_\lambda \in \Delta(\omega_s)} \max[0, \mathbf{s}_\lambda' \mathbf{Z}(\lambda) + \mathbf{Z}(\psi)(s_\psi)]^2 \]
Second, the given weak convergence in Theorem 6(ii) is jointly attained with that of $\mathcal{L}\mathcal{R}_n^{(1)}$ because all these can be obtained by applying the CMT to Theorem 4(i). Also, $\mathcal{H}_2$ in $\mathcal{L}\mathcal{R}_n^{(2)}$ is identical to $\mathcal{H}_2$ of $\mathcal{L}\mathcal{R}_n^{(1)}$, as well as they both having the same distribution. This trivially follows from the fact that they trace from the same sample statistic. From these therefore the asymptotic null distribution of the LR statistic immediately follows.

**Theorem 7:** Given Assumptions A1 to A8, and $H_0$, $\mathcal{L}\mathcal{R}_n \Rightarrow \mathcal{H}_1 - H_0$.

The asymptotic null distribution in Theorem 7 can be of various forms depending on the model of consideration. For example, if $\mu = \delta$ and $r_v = 0$, then $\pi = \lambda$, $\phi = \omega$, and under $H'_0$,

$$\mathcal{L}\mathcal{R}_n \Rightarrow \sup_{s_\lambda \in \Delta(\lambda_0)} \left\{ \frac{\max[0, s_\lambda'Z^{(\lambda)}]^2}{s_\lambda'(-A^{(\lambda, \lambda)}_s)s_\lambda} \right\} - \sup_{s_\phi \in \Delta(\phi)} \left\{ \frac{\max[0, s_\phi'Z^{(\phi)}]^2}{s_\phi'(-A^{(\phi, \phi)}_s)s_\phi} \right\}. $$

Further, if $\lambda_s$ is an interior element, then applying the proof of Corollary 1(vii) shows that

$$\mathcal{L}\mathcal{R}_n \Rightarrow (\bar{Z}^{(\delta)})'(-\bar{A}_{s}^{(\delta, \delta)})^{-1}(\bar{Z}^{(\delta)}), \quad (21)$$

where $\bar{Z}^{(\delta)} := Z^{(\delta)} - (A_{s}^{(\delta, \phi)})(A_{s}^{(\phi, \phi)})^{-1}Z^{(\phi)}$ and $\bar{A}_{s}^{(\delta, \delta)} := A_{s}^{(\delta, \delta)} - (A_{s}^{(\delta, \phi)})(A_{s}^{(\phi, \phi)})^{-1}(A_{s}^{(\phi, \delta)})'$. Thus, the same result is obtained as for standard diffle models without boundary parameters. If $\lambda_s$ is a boundary parameter, then the LR statistic has a different asymptotic null distribution depending on the property of the associated parameter space at the boundary. For example, if $\delta \in [\delta_s, \delta]$ $\in \mathbb{R}$ (thus $\bar{A}_{s}^{(\delta, \delta)}$ and $\bar{Z}^{(\delta)}$ are now scalars), and $\phi_s$ is an interior element of its parameter space, then

$$\sup_{s_\lambda \in \Delta(\lambda_s)} \left\{ \frac{\max[0, s_\lambda'Z^{(\lambda)}]^2}{s_\lambda'(-A_{s}^{(\lambda, \lambda)}_s)s_\lambda} \right\} = (\bar{Z}^{(\delta)})'(-\bar{A}_{s}^{(\delta, \delta)})^{-1}(\bar{Z}^{(\delta)})1_{\{z_{\delta} > 0\}} + (Z^{(\phi)})'(-A_{s}^{(\phi, \phi)})^{-1}(Z^{(\phi)}),$$

where $\bar{Z}^{(\delta)} := (\bar{s}^{(\delta)}, \bar{s}^{(\phi)})'$ solves the LHS. Further, it’s a standard exercise to show that $\bar{Z}^{(\delta)} < 0$ if and only if $\bar{s}_\delta \leq 0$. Thus,

$$\mathcal{L}\mathcal{R}_n \Rightarrow (-\bar{A}_{s}^{(\delta, \delta)})^{-1}\max[0, \bar{Z}^{(\delta)}]^2 \quad (22)$$

under $H'_0$ because

$$\sup_{s_\phi \in \Delta(\phi_s)} \left\{ \frac{\max[0, s_\phi'Z^{(\phi)}]^2}{s_\phi'(-A_{s}^{(\phi, \phi)}_s)s_\phi} \right\} = (Z^{(\phi)})'(-A_{s}^{(\phi, \phi)})^{-1}(Z^{(\phi)})$$
by the proof of Corollary 1(ivii). As another example, if $\mu = \xi$ and $r_\psi = 0$ then different asymptotic null distribution is obtained. In this case, $\lambda = \omega$ and $\pi = (\lambda', \xi')'$. Thus,

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

and

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

by applying (20). Therefore, $\mathcal{H}_1 - \mathcal{H}_0$ becomes the asymptotic null distribution by Theorem 7. Further, this can be more simplified if $\Theta$ is more restricted:

A9 (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 $C(\xi_*) := \{ x \in \mathbb{R}^{r_\lambda} : \exists \xi' \in \Xi, x := \xi_* + \delta \xi', \delta \in \mathbb{R}^+ \}$.

Note that the null model is different (A9(i)) and does not contain a boundary parameter (A9(ii)). Further, A9(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 (A9(ii)). Then, we obtain the following:

**Corollary 2:** Given Assumptions A1 to A9,

$$
\mathcal{L}\mathcal{R}_n(1) \Rightarrow \sup_{s_\xi \in \Delta(\xi_*)} \max[0, \tilde{\mathcal{Y}}^{(\xi)}(s_\xi)]^2 + (Z^{(\lambda)})'(-A_s^{(\lambda, \lambda)})^{-1}(Z^{(\lambda)}) + \mathcal{H}_2,
$$

where for each $s_\xi \in \Delta(\xi_*) := \{ x \in \mathbb{R}^{r_\xi} : x_* + x \in \text{cl}\{C(\xi_*)\}, \| x \| = 1 \}$, $\tilde{\mathcal{Y}}^{(\xi)}(s_\xi) := (-A_s^{(\xi, \xi)}(s_\xi))^{-1/2} \tilde{Z}^{(\xi)}(s_\xi)$, $\tilde{Z}^{(\xi)}(s_\xi) := A_s^{(\xi, \xi)}(s_\xi)$, and also $\tilde{Z}^{(\xi)}(s_\xi) := Z^{(\xi)}(s_\xi) - A_s^{(\xi, \lambda)}(s_\xi) \lambda' A_s^{(\lambda, \lambda)}(s_\xi)$.

Note that $\tilde{Z}^{(\xi)}$ is obtained by projecting $Z^{(\xi)}$ on $Z^{(\lambda)}$. This follows from the fact the unknown nuisance parameter $\lambda_*$ needs to be estimated. Thus, combining Theorem 7, (23), (24), and Corollary 2 yields that under $H_0''$

$$
\mathcal{L}\mathcal{R}_n \Rightarrow \sup_{s_\xi \in \Delta(\xi_*)} \max[0, \tilde{\mathcal{Y}}^{(\xi)}(s_\xi)]^2.
$$
3.2 Wald Statistic

Given the $d$–diffle model specification, we can refine the standard Wald statistic for the same hypothesis. 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.

Before refining Wald statistic, we first consider the asymptotic distribution of the distance between $\hat{\mu}_n$ and $\mu_0$. Note that the distance between $\hat{\theta}_n$ and $\theta_*$ captured by $h_n$ (according to Theorem 4) cannot be used to test the null hypothesis because the inference on $\mu_*$ can be associated with that of other nuisance parameters $\omega_*$ and $\tau_*$. The distance $h_n$ needs to be broken down into pieces for $\mu$, $\omega$, and $\tau$ respectively. This can be achieved easily. We note that for any $hd$ and $d \in \Delta(\theta_*)$, there are $h_1(\mu)$, $h_1(\omega)$, and $h_1(\tau)$, and $(\hat{s}_\mu, \hat{s}_\omega, \hat{s}_\tau) \in \Delta(\mu_*) \times \Delta(\omega_*) \times \Delta(\tau_*)$ such that

$$h_d = (h_1(\mu), h_1(\omega), h_1(\tau))$$

if each parameter space for $\mu$, $\omega$, and $\tau$ can be approximated by a cone, and the parameter space for $\theta$ can be also approximated by a Cartesian product of these cones, which is given by A7 and A8. Thus,

$$\sup_d \sup_h L_n(\theta_* + hd) = \sup_{\{s_\mu, s_\omega, s_\tau\}} \sup_{\{h(\mu), h(\omega), h(\tau)\}} L_n(\theta_* + (h(\mu)s_\mu', h(\omega)s_\omega', h(\tau)s_\tau')),$$

and

$$\hat{h}_n(\hat{d}_n) = (\hat{h}_n(\mu), \hat{h}_n(\omega), \hat{h}_n(\tau))$$

and

$$(\hat{s}_\mu, \hat{s}_\omega, \hat{s}_\tau) \in \Delta(\mu_*) \times \Delta(\omega_*) \times \Delta(\tau_*)$$

where $(\hat{d}_n, \hat{h}_n)$ and $(\hat{s}_\mu, \hat{s}_\omega, \hat{s}_\tau)$ solve the LHS and RHS of (25) respectively.

Wald statistic can be refined by relating it to $\hat{h}_n(\mu)$. For this, 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 Wald statistic.

First, for each $(s_\mu, s_\omega, s_\tau) \in \Delta(\mu_*) \times \Delta(\omega_*) \times \Delta(\tau_*)$, we let

$$G^{(\mu)}(s_\mu, s_\omega), G^{(\omega)}(s_\mu, s_\omega), G^{(\tau)}(s_\tau) : \quad \left[ \begin{array}{c} G^{(\mu)}(s_\mu, s_\omega) \\ G^{(\omega)}(s_\mu, s_\omega) \\ G^{(\tau)}(s_\tau) \end{array} \right] : = \left[ \begin{array}{c} G^{(\mu)}(s_\mu, s_\omega) \\ G^{(\omega)}(s_\mu, s_\omega) \\ G^{(\tau)}(s_\tau) \end{array} \right] : = \left[ \begin{array}{c} -A_*(\eta, \eta)(s_\mu, s_\omega)^{-1}Z^{(\eta)}(s_\mu, s_\omega) \\ -A_*(\tau, \tau)(s_\tau)^{-1}Z^{(\tau)}(s_\tau) \end{array} \right] ,$$

where for each $(s_\mu, s_\omega)$, $Z^{(\mu)}(s_\mu, s_\omega) := (Z^{(\mu)}(s_\mu), Z^{(\omega)}(s_\omega))'$. Next, for each $(s_\mu, s_\omega) \in \Delta(\mu_*) \times \Delta(\omega_*)$, we also let

$$\hat{G}^{(\mu)}(s_\mu), \hat{G}^{(\omega)}(s_\omega) : = \left[ \begin{array}{c} -A_*(\mu, \mu)(s_\mu)^{-1}Z^{(\mu)}(s_\mu) \\ -A_*(\omega, \omega)(s_\omega)^{-1}Z^{(\omega)}(s_\omega) \end{array} \right] .$$
These constitute the asymptotic behavior of \( \hat{h}_n \). Following Lemma 4 establishes this.

**Lemma 4:** Given Assumptions A1 to A8,

\[
\sqrt{n} \hat{h}_n = \begin{bmatrix} G^{(\mu)} \\ G^{(\omega)} \\ 0 \end{bmatrix} \times \mathbf{1}_{\{\min[G^{(\mu)}, G^{(\omega)}] > 0\}} + \begin{bmatrix} \max[0, \hat{G}^{(\mu)}] \times \mathbf{1}_{\{G^{(\mu)} \geq 0 > G^{(\omega)}\}} \\ \max[0, \hat{G}^{(\omega)}] \times \mathbf{1}_{\{G^{(\omega)} \geq 0 > G^{(\mu)}\}} \\ \max[0, \hat{G}^{(\tau)}] \end{bmatrix}. \tag{26}
\]

Heuristic explanation of Lemma 4 can be given as follows. First, note that both \((\hat{h}_n^{(\mu)}, \hat{h}_n^{(\omega)})'\) and \(\hat{h}_n^{(\tau)}\) are initially defined on \(\Delta(\mu) \times \Delta(\omega) \times \Delta(\tau)\), but their asymptotic counterparts in the RHS of (26) are respectively defined on \(\Delta(\mu) \times \Delta(\omega)\) and \(\Delta(\tau)\). This is mainly due to assumption A7(iv and v). By supposing that \(A^{(\eta, \tau)}_s(d_\eta, d_\tau) = 0\), the maximization in the RHS of (25) is asymptotically separated into two independent maximization procedures by Theorem 5, resulting in the different domains at the limit. Second, note that \(\hat{h}_n^{(\mu)}\) and \(\hat{h}_n^{(\omega)}\) cannot be less than zero though they can be zero. Thus, for each \((s_\mu, s_\omega, s_\tau)\) we obtain possibly either (i) both \(\hat{h}_n^{(\mu)}(s_\mu, s_\omega, s_\tau) > 0\) and \(\hat{h}_n^{(\omega)}(s_\mu, s_\omega, s_\tau) > 0\); (ii) \(\hat{h}_n^{(\mu)}(s_\mu, s_\omega, s_\tau) > 0\), \(\hat{h}_n^{(\omega)}(s_\mu, s_\omega, s_\tau) = 0\); (iii) \(\hat{h}_n^{(\mu)}(s_\mu, s_\omega, s_\tau) = 0\), \(\hat{h}_n^{(\omega)}(s_\mu, s_\omega, s_\tau) > 0\); or (iv) \(\hat{h}_n^{(\mu)}(s_\mu, s_\omega, s_\tau) = 0\), \(\hat{h}_n^{(\omega)}(s_\mu, s_\omega, s_\tau) = 0\). These four different probability events are determined by the sign of \(G^{(\eta)}\), and Lemma 4 distinguishes them using separate indication functions to identify the asymptotic weak limit of \(\sqrt{n} \hat{h}_n\). In particular, for a given direction \((s_\mu, s_\omega, s_\tau)\), if say \(\hat{h}_n^{(\omega)}(s_\mu, s_\omega, s_\tau) = 0\) but \(\hat{h}_n^{(\mu)}(s_\mu, s_\omega, s_\tau) > 0\), then the nuisance parameter \(\omega\) does not affect the asymptotic distribution of \(\hat{h}_n^{(\mu)}(s_\mu, s_\omega, s_\tau)\), so that \(\sqrt{n} \hat{h}_n^{(\mu)}(s_\mu, s_\omega, s_\tau) \Rightarrow \hat{G}^{(\mu)}(s_\mu)\). This mainly follows because \(\hat{h}_n^{(\omega)}(s_\mu, s_\omega, s_\tau) = 0\) implies that \(\hat{\omega}_n(s_\mu, s_\omega, s_\tau) = \omega_*\), so that the nuisance parameter estimation error is absent. Finally, if both \(\omega_*\) and \(\mu_*\) are interior elements, then the second term in the RHS of (26) can be virtually ignored while maximizing \(L_n\) over the directions.

Given these, we define Wald statistic as

\[
W_n := \sup_{s_\mu \in \Delta(\mu_0)} n\{\hat{h}_n^{(\mu)}(s_\mu)\}\{\hat{\omega}_n(s_\mu)\}\{\hat{h}_n^{(\mu)}(s_\mu)\},
\]

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

\[
L_n(\mu_0 + \hat{h}_n^{(\mu)}(s_\mu)s_\mu, \hat{\omega}_n(s_\mu), \hat{\tau}_n(s_\mu)) := \sup_{\{h^{(\mu)}, \omega, \tau\}} L_n(\mu_0 + h^{(\mu)}s_\mu, \omega, \tau),
\]

29
and $\hat{W}_n$ is a weight function, which can estimate a non-random positive function $W_\ast$ uniformly on $\Delta(\mu_0)$. As this assumption is important, we formalize this as follows.

**A10 (Weight Function I):** A weight function $\hat{W}_n$ strictly positive uniformly on $\Delta(\mu_0)$ and $n$ converges to $W_\ast$ strictly positive and bounded from above uniformly on $\Delta(\mu_0)$ a.s.--$\mathbb{P}$ as $n$ tends to infinity. That is, $\sup_{s_\mu} |\hat{W}_n(s_\mu) - W_\ast(s_\mu)| \to 0$ a.s.--$\mathbb{P}$.

In terms of this definition, the weight function $W_\ast$ is assumed to be identical to the asymptotic variance of $\sqrt{n}h_n^{(\mu)}$ if the standard Wald statistic principle is applied. If the parameter of interest is on the boundary, nevertheless, this needs to be chosen carefully because the asymptotic variance of weak limit of $\sqrt{n}h_n^{(\mu)}$ can have different forms whether a boundary parameter is involved or not.

From these aspects, the asymptotic null distribution of the Wald statistic can be derived as follows:

**Theorem 8:** Given Assumptions A1 to A8, A10, and $H_0$,

$$W_n \Rightarrow \sup_{s_\mu \in \Delta(\mu_0)} G^{(\mu)}(s_\mu, \bar{\omega}(s_\mu)) W_\ast(s_\mu) g^{(\mu)}(s_\mu, \bar{\omega}(s_\mu)) I_{\{\min[g^{(\mu)}(s_\mu, \bar{\omega}(s_\mu)), g^{(\nu)}(s_\mu, \bar{\omega}(s_\mu))] > 0\}}$$

$$+ \max[0, \hat{g}^{(\mu)}(s_\mu)] W_\ast(s_\mu) \max[0, \hat{g}^{(\mu)}(s_\mu)] I_{\{g^{(\mu)}(s_\mu, \bar{\omega}(s_\mu)) > 0, g^{(\nu)}(s_\mu, \bar{\omega}(s_\mu)) > 0\}},$$

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

$$\sup_{s_\omega \in \Delta(\omega_\ast)} G^{(\eta)}(s_\mu, s_\omega) = g^{(\mu)}(s_\mu, s_\omega) I_{\{\min[g^{(\mu)}(s_\mu, s_\omega), g^{(\nu)}(s_\mu, s_\omega)] > 0\}}$$

and

$$\max[0, \hat{g}^{(\mu)}(s_\mu)](s_\mu) (s_\mu) \max[0, \hat{g}^{(\mu)}(s_\mu)] I_{\{g^{(\mu)}(s_\mu, s_\omega) > 0, g^{(\nu)}(s_\mu, s_\omega) > 0\}},$$

respectively.

Briefly, the proof of Theorem 8 follows from the note that for given $s_\mu$ the arguments maximizing $L_n(\mu_0 + h^{(\mu)} s_\mu, \omega_\ast, \tau_\ast)$ is also obtained by

$$\sup_{\{s_\omega, s_\tau\} \in \{h^{(\mu)}, h^{(\nu)}, h^{(\tau)}\}} 2\{L_n(\mu_0 + h^{(\mu)} s_\mu, \omega_\ast + h^{(\nu)} s_\omega, \tau_\ast + h^{(\tau)} s_\tau) - L_n(\mu_0, \omega_\ast, \tau_\ast)\}. \quad (29)$$
For each \((s_\mu, s_\omega, s_\tau)\), we first approximate (29) by a quadratic function, noting that the presence of boundary parameter can result in different approximation as emphasized in Lemma 4. Thus, different approximations combined with different indication functions of Lemma 4 yield the separate maximization processes in Theorem 8. Equations (27) and (28) specifically show these different approximation and maximization processes. That is, for each \(s_\mu\), they show the maximization of (29) with respect to \((s_\omega, s_\tau)\) can result in different maximization whether or not to involve the boundary parameter. Apparently, the asymptotic null distribution of \(\sqrt{n}h_n^{(\mu)}(s_\mu)\) is captured by \(G^{(\mu)}(s_\mu, \bar{s}_\omega(s_\mu))\) and \(\max[0, \hat{G}^{(\mu)}(s_\mu)]\) under each approximation, so that the desired result follows by applying the CMT to the definition of Wald statistic combined with the weight function in A10.

The given result in Theorem 8 can be further simplified by considering the benchmark model in A9.

**Corollary 3:** Given Assumptions A1 to A9 and \(H''_0\), if \(\sup_{s_\xi \in \Delta(\xi_0)} |\hat{W}_n(s_\xi) + A^{(\xi)}(s_\xi)| \to 0\) a.s.−\(\mathbb{P}\), then \(W_n \Rightarrow \sup_{s_\xi \in \Delta(\xi_0)} \max[0, \hat{Y}^{(\xi)}(s_\xi)]^2\).

Note that the weak limit of the Wald test statistic in Corollary 3 is the same weak limit obtained by the LR test statistic.

### 3.3 Lagrange Multiplier Statistic

LM statistics for standard econometric models can be appropriately refined for d–diffle models in a way that the slope of the model can be distributed around zero for every direction under the null.

We define LM test statistic as

\[
LM_n := \sup_{(s_\mu, s_\omega) \in \Delta(\mu_0) \times \Delta(\omega_0)} n\bar{W}_n(s_\mu, s_\omega) \max \left[ 0, -\frac{DL_n(\hat{\theta}_n; s_\mu)}{D^2L_n(\hat{\theta}_n; s_\mu; s_\omega)} \right],
\]

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

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

\[
DL_n(\hat{\theta}_n; s_\mu; s_\omega) := \lim_{h \downarrow 0} h^{-1}\{DL_n(\mu_0, \omega_n + hs_\omega, \tau_n; s_\mu) - DL_n(\hat{\theta}_n; s_\mu)\},
\]

31
\[
DL_n(\bar{\theta}_n; s_\omega; s_\mu) := \lim_{h \to 0} h^{-1}\{DL_n(\mu_0 + h s_\mu, \bar{\omega}_n, \bar{\tau}_n; s_\omega) - DL_n(\bar{\theta}_n; s_\omega)\},
\]

and \(\bar{W}_n\) is a weight function satisfying the following condition.

**A11 (Weight Function II):** (i) The unknown nuisance parameter \(\omega_*\) is interior to \(\Omega\);

(ii) A weight function \(\bar{W}_n\) strictly positive uniformly on \(\Delta(\mu_0) \times \Delta(\omega_n)\) and \(n\) converges to \(\bar{W}\) strictly positive and bounded from above uniformly on \(\Delta(\mu_0) \times \Delta(\omega_*)\) a.s. as \(n\) tends to infinity, i.e.,

\[
\sup_{(s_\mu, s_\omega) \in \Delta(\mu_0) \times \Delta(\omega_*)} |\bar{W}_n(s_\mu, s_\omega) - \bar{W}(s_\mu, s_\omega)| \to 0 \text{ a.s. as } n \to \infty.
\]

There are several comments in terms of this definition. First, the asymptotic structure of the LM statistic is designed to be close to those for the LR and Wald statistics. Each component of the LM 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 \((\mu_0', \bar{\omega}_n', \bar{\tau}_n') = \bar{\theta}_n\). Also, the 'max' operator in the LM statistic is related to the behavior of \(n\bar{\omega}_n(\mu_0(s_\mu, s_\omega, s_\tau))\), which cannot be less than zero. Second, the LM statistic can be represented as the supremum of a squared random score function when \(\bar{W}_n(s_\mu, s_\omega)\) is asymptotically equivalent to \(-n^{-1}\bar{D}^2L_n(\bar{\theta}_n; s_\mu, s_\omega)\). This random function is also designed to have a statistic asymptotically equivalent to the LR statistic particularly for the popular models given in A9. Third, we emphasize that \(\bar{W}_n\) is defined on \(\Delta(\mu_0) \times \Delta(\omega_*)\), and \(\omega_*\) is interior to \(\Omega\). The domain \(\Delta(\omega_*)\) estimates \(\Delta(\omega_*)\), and the interiority condition lets \(\Delta(\omega_*)\) be identical to \(\Delta(\omega_*)\) for sufficiently large \(n\). If \(\omega_*\) is on the boundary, then \(\Delta(\omega_*)\) can be different from \(\Delta(\omega_*)\). Condition A11(i) precludes this occasion.

Given the definition of the LM statistic, its asymptotic null distribution can be obtained straightforwardly. The following Theorem 9 establishes this.

**Theorem 9:** Given Assumptions A1 to A8, A11, and \(H_0\),

\[
\mathcal{L} M_n \Rightarrow \sup_{(s_\mu, s_\omega) \in \Delta(\mu_0) \times \Delta(\omega_*)} \max[0, \bar{G}(\mu)(s_\mu, s_\omega; s_\omega)] \bar{W}(s_\mu, s_\omega) \max[0, \bar{G}(\mu)(s_\mu, s_\omega; s_\omega)],
\]

where for each \((s_\mu, s_\omega) \in \Delta(\mu_0) \times \Delta(\omega_*)\),

\[
\bar{G}(\mu)(s_\mu, s_\omega; s_\omega) := \{-\bar{A}_\mu(\mu,\mu)(s_\mu, s_\omega)\}^{-1}\bar{Z}(\mu)(s_\mu; s_\omega),
\]

\[
\bar{Z}(\mu)(s_\mu; s_\omega) := Z(\mu)(s_\mu) - (-\bar{A}_\mu^{(\mu,\omega)}(s_\mu, s_\omega))(-\bar{A}_\mu^{(\omega,\omega)}(s_\omega))^{-1}Z(\omega)(s_\omega),
\]

32
\[ \tilde{A}_s^{(\mu,\mu)}(s_\mu, s_\omega) := A_s^{(\mu,\mu)}(s_\mu) - A_s^{(\mu,\omega)}(s_\mu, s_\omega)(A_s^{(\omega,\omega)}(s_\omega))^{-1} A_s^{(\omega,\mu)}(s_\omega, s_\mu), \]

and \( \tilde{s}_\omega \) is such that \( \max[0, \gamma'(\omega)(\tilde{s}_\omega)]^2 = \sup_{s_\omega \in \Delta(\omega_*)} \max[0, \gamma'(\omega)(s_\omega)]^2. \)

Note that the asymptotic null distribution of the LM statistic depends upon that of \( \tilde{s}_\omega \). This is mainly because the weak limit of \( n^{-1/2} DL_n(\tilde{\theta}_n; \cdot \cdot) \) is captured by \( \tilde{Z}(\mu)(\cdot; \tilde{s}_\omega) \), which is obtained by the weak limit of Taylor expansion of \( n^{-1/2} DL_n(\tilde{\theta}_n; \cdot \cdot) \) at \( \omega_* \).

Theorem 9 can be further simplified for the benchmark model. We provide its asymptotic null distribution in Corollary 4.

**Corollary 4:** Given Assumptions A1 to A9 and \( H_0^n \), if \( \sup_{(s_\xi, s_\lambda) \in \Delta(\xi_0) \times \Delta(\lambda_n)} |\tilde{W}_n(s_\xi, s_\lambda) + \tilde{A}_s^{(\xi,\xi)}(s_\xi)| \to 0 \ \text{a.s.} - \mathbb{P} \), then \( \mathcal{LM}_n \Rightarrow \sup_{s_\xi \in \Delta(\xi_0)} \max[0, \tilde{\gamma}'(\xi)(s_\xi)]^2. \)

The LR, Wald, and LM 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 we reconcile the parameters in the model in terms of 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 \( \tilde{d} \),

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

and

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

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

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

33
and
\[ k(d_2/d_1, \tilde{d}_2/\tilde{d}_1) := \frac{m(d_2/d_1)m(\tilde{d}_2/\tilde{d}_1)}{1 - m(d_2/d_1)m(\tilde{d}_2/\tilde{d}_1)}. \]

Due to the block diagonal aspect of \( B_s(d, \tilde{d}) \) and the information matrix equality, we can focus separately to the asymptotic null distribution associated with each block diagonal submatrix. Further, \( \sigma_s^{-2} d_s E[Q_t Q_t'] \tilde{d}_s \) is associated with only \( \gamma \), so that it can be ignored. We also note that \( \nu_3' B_s(d, \tilde{d}) \nu_3 = B_s^{(1)}(d, \tilde{d}) + B_s^{(2)}(d, \tilde{d}) + B_s^{(3)}(d, \tilde{d}) \), where each \( B_s^{(i)}(d, \tilde{d}) \) \( (i = 1, 2, 3) \) is the covariance constituting the independent Gaussian process shown before. Further, the relevant asymptotic null distribution can be more easily obtained by rephrasing \( B_s^{(\pi, \pi)}(d_\pi, \tilde{d}_\pi) \). For this examination, we let
\[
\tilde{B}_s^{(\pi, \pi)}(d_\pi, \tilde{d}_\pi) := \begin{bmatrix}
\tilde{B}_s^{(\pi, \pi)}(d_\pi, \tilde{d}_\pi) & 0 \\
0' & q(d_1, d_2, \tilde{d}_1, \tilde{d}_2) E[W_t^2]^2
\end{bmatrix},
\]
where
\[
\tilde{B}_s^{(\pi, \pi)}(d_\pi, \tilde{d}_\pi) := \frac{1}{2} \begin{bmatrix}
    h(d_1, d_2)h(\tilde{d}_1, \tilde{d}_2) E[W_t^4] & \frac{1}{\sigma_s^2} d_s h(d_1, d_2) E[W_t^2] \\
    \frac{1}{\sigma_s^2} d_s h(\tilde{d}_1, \tilde{d}_2) E[W_t^2] & \frac{1}{\sigma_s^2} d_s \tilde{d}_s \tilde{d}_s
\end{bmatrix},
\]
and
\[
q(d_1, d_2, \tilde{d}_1, \tilde{d}_2) := h(d_1, d_2)h(\tilde{d}_1, \tilde{d}_2)k(d_2/d_1, \tilde{d}_2/\tilde{d}_1).
\]

Note that \( \nu_3' \tilde{B}_s^{(\pi, \pi)}(d_\pi, \tilde{d}_\pi) \nu_3 = B_s^{(2)}(d, \tilde{d}) + B_s^{(3)}(d, \tilde{d}) \) and \( \nu_2' \tilde{B}_s^{(\pi, \pi)}(d_\pi, \tilde{d}_\pi) \nu_2 = B_s^{(2)}(d, \tilde{d}) \).

Also, \( B_s^{(\pi, \pi)}(d_\pi, \tilde{d}_\pi) \) is block diagonal, implying that the Gaussian process associated with \( B_s^{(\pi, \pi)}(d_\pi, \tilde{d}_\pi) \) is independent of that associated with \( q(d_1, d_2, \tilde{d}_1, \tilde{d}_2) E[W_t^2]^2 \). In addition to this, \( B_s^{(\pi, \pi)}(d_\pi, \tilde{d}_\pi) \) is bilinear with respect to \( h(d_1, d_2) \) and \( d_s \).

From these, the asymptotic null distribution of the test statistics are derived straightforwardly. The asymptotic null distribution of the LR statistic is obtained as follows: First,
\[
\mathcal{L}R_n^{(1)} \Rightarrow \sup_{s_\pi \in \Delta(\pi_s)} \max[0, \mathcal{Y}^{(\pi)}(s_\pi)]^2 + \mathcal{H}_2
\]
by Theorem 5, where \( \mathcal{Y}^{(\pi)}(s_\pi) \) is a standard Gaussian process with covariance structure
\[
E[\mathcal{Y}^{(\pi)}(s_\pi) \mathcal{Y}^{(\pi)}(s_\pi')] = \frac{\nu_3' \tilde{B}_s^{(\pi, \pi)}(s_\pi, \tilde{s}_\pi) \nu_3}{\{\nu_3' \tilde{B}_s^{(\pi, \pi)}(s_\pi, s_\pi) \nu_3\}^{1/2}} \{\nu_3' \tilde{B}_s^{(\pi, \pi)}(\tilde{s}_\pi, \tilde{s}_\pi) \nu_3\}^{1/2}.
\]
and $H_2$ is a chi-square random variable with degree of freedom $k + 1$. Second,

$$\mathcal{L}R_n^{(2)} \Rightarrow \sup_{s_2 \in \{-1, 1\}} \max[0, \mathcal{Y}^{(\sigma^2)}(s_2)]^2 + H_2$$

by Theorem 6, where $\mathcal{Y}^{(\sigma^2)}$ is a standard Gaussian process having $E[\mathcal{Y}^{(\sigma^2)}(s_2)] = 1$, implying that it is free of the direction $s_2$, so that $\sup_{s_2 \in \{-1, 1\}} \max[0, \mathcal{Y}^{(\sigma^2)}(s_2)]^2$ is a chi-square random variable with degree of freedom 1 by the proof of Corollary 1(vi). Thus,

$$\mathcal{L}R_n \Rightarrow \sup_{s_2/s_1 \in [\xi, \bar{c}]} \max[0, \tilde{Y}^{(\theta)}(s_1, s_2)]^2$$

by Corollary 2, where $\tilde{Y}^{(\theta)}$ is a standard Gaussian process with covariance structure

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

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

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

This structure is homogenous of degree zero with respect to $s_1$ and $s_2$, so that $\tilde{Y}^{(\theta)}$ can be represented equivalently as a function of $s_2/s_1$. That’s why the maximization of (30) is taken over $[\xi, \bar{c}]$.

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

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

where $\hat{E}_n[W_t^2] := n^{-1} \sum_{t=1}^n W_t^2$ and $\var_n[W_t^2] := n^{-1} \sum_{t=1}^n W_t^4 - (n^{-1} \sum_{t=1}^n W_t^2)^2$. It is trivial to show that this estimator satisfies A10. Then the Wald statistic defined as

$$W_n = n \{\tilde{h}_n^{(\theta)}(s_2/s_1)\} \hat{W}_n(s_2/s_1, s_2/s_1) \{\tilde{h}_n^{(\theta)}(s_2/s_1)\}$$

has the following asymptotic null behavior:

$$W_n \Rightarrow \sup_{s_2/s_1 \in [\xi, \bar{c}]} \max[0, \tilde{Y}^{(\theta)}(s_1, s_2)]^2$$

by Corollary 3, where $\tilde{h}_n^{(\theta)}(s_2/s_1)$ is such that

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

35
and \( s_1^2 + s_2^2 = 1 \).

The LM statistic can be also applied, as well. Following the definition,

\[
\mathcal{LM}_n = \sup_{s_2/s_1 \in [\varepsilon, \bar{c}]} \left\{ \frac{\max \left[ 0, DL_n(\hat{\gamma}_n, \hat{\sigma}_n^2; 0; s_1, s_2) \right]}{-D^2 L_n(\hat{\gamma}_n, \hat{\sigma}_n^2; 0; s_1, s_2, s_{\sigma^2})} \right\}^2 \hat{W}_n(s_2/s_1, s_2/s_1),
\]

where \( DL_n(\hat{\gamma}_n, \hat{\sigma}_n^2; 0; s_1, s_2) = \{2\hat{\sigma}_n^2\}^{-1} \left\{ U^n(\hat{\gamma}_n)\Omega^n(m(s_2/s_1)) U^n(\hat{\gamma}_n) - \hat{\sigma}_n^2 \text{tr}[\Omega^n(m(s_2/s_1))] \right\} \); 

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

\[
- U^n(\hat{\gamma}_n) \Omega^n(m(s_2/s_1)) U^n(\hat{\gamma}_n) \left[ \frac{n}{2\hat{\sigma}_n^4} - \frac{1}{\hat{\sigma}_n^4} U^n(\hat{\gamma}_n) U^n(\hat{\gamma}_n) \right]^{-1}
\times U^n(\hat{\gamma}_n) \Omega^n(m(s_2/s_1)) U^n(\hat{\gamma}_n);
\]

\((\hat{\gamma}_n, \hat{\sigma}_n^2)\) is such that \( L_n(\hat{\gamma}_n, \hat{\sigma}_n^2, 0) = \sup_{(\gamma, \sigma^2)} L_n(\gamma, \sigma^2, 0) \); and the same weight matrix is employed as before. Then by Corollary 4,

\[
\mathcal{LM}_n \Rightarrow \sup_{s_2/s_1 \in [\varepsilon, \bar{c}]} \max[0, \hat{\gamma}^{(\theta)}(s_1, s_2)]^2,
\]

so that all three test statistics are asymptotically equivalent.

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

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

where \( Z_j \sim \text{i.i.d. } N(0, 1) \). Due to the i.i.d. condition, it’s not hard to generate \( \hat{\gamma}^{(\theta)} \) by a simulation method.

There are a couple of caveats in terms of this method. First, \( \text{var}(W_t^2) \) and \( E[W_t^2] \) are unknown. We should instead use their consistent estimators. Second, the index \( j \) needs to be truncated at a moderately large number in a way it does not affect the null distribution significantly. For this examination, we implement a Monte Carlo simulations using these. The DGP for \( Y_t = U_t \sim \text{i.i.d. } N(0, 1) \) and \( W_t \sim \text{i.i.d. } N(0, 1) \), which is independent of \( U_t \). We assume the other parameters than \( \alpha_s, \sigma_0^2, \theta_1 s, \) and \( \theta_2 s \) are known and also let \( \varepsilon = 0.5, \bar{c} = 1.5 \). The total number of repetitions is 2,000. Figure 1 shows the Q-Q plot between \( \sup_{(s_1, s_2)} \max[0, \hat{\gamma}^{(\theta)}(s_1, s_2)]^2 \) and \( \sup_{(s_1, s_2)} \max[0, \hat{\gamma}^{(\theta)}(s_1, s_2)]^2 \), where for each \((s_1, s_2)\),

\[
\hat{\gamma}^{(\theta)}(s_1, s_2) := \frac{1}{c_n(s_2/s_1, s_2/s_1)^{1/2}} \left[ \left\{ \text{var}_n(W_t^2) \right\}^{1/2} Z_0 + \hat{E}_n[W_t^2] \sum_{j=1}^{150} m(s_2/s_1)^j Z_j \right],
\]
and \( \text{var}(W_t^2) \) and \( \hat{E}[W_t^2] \) are the method of moments estimators for \( \text{var}(W_t^2) \) and \( E[W_t^2] \) when \( n = 500 \). Note that the Q-Q line in Figure 1 is 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 LR statistic for various sample sizes and the asymptotic null distribution. As \( n \) gets large, the empirical distribution of the LR statistic approaches the asymptotic distribution.

This aspect becomes more apparent if the number of nuisance parameters reduces. As another simulation, we assume that \( \alpha_* \) is also known. Figure 3 shows that the empirical distribution of the LR statistic is closer to the asymptotic distribution even when the sample size is much smaller than the prior case.

### 3.4.2 Example 2 (continued)

As the model is now different, \( r_\nu = 0 \), and further supposing that \( r_\tau = 0 \) does not matter much by Theorem 5. Thus, we can let \( \theta = \pi = \lambda = (\delta', \phi')' = (\mu', \omega')' \).

Given that the objective function \( Q_n \) does not satisfy the model condition A2, we cannot exactly apply the definition of the LR test statistic to \( Q_n \). Nevertheless, we can develop the LR-test like test statistic using \( Q_n \). We let

\[
QLR_n := \left\{ \sup_{\delta, \phi} Q_n(\delta, \phi) - \sup_{\phi} Q_n(\delta_0, \phi) \right\}
\]

and rephrase \( C_*'=\left[-M_*\right]^{-1}W \) and \( C_*'=\left[-M_*\right]^{-1}C_* \) into \( Z^{(\lambda)} = (Z^{(\delta)'}, Z^{(\phi)'})' \) and \( A_*^{(\lambda \lambda)} \) of Section 3.1 respectively. The asymptotic null distribution of the QLR test statistic is already given in (21). That is, \( QLR_n \Rightarrow (\tilde{Z}^{(\delta)})'(-\tilde{A}_*^{(\delta \delta)})^{-1}(\tilde{Z}^{(\delta)}) \).

The asymptotic null distribution of the Wald test statistic defined by the GMM estimator can be obtained by applying Corollary 3. For this, we define

\[
QW_n := \sup_{\delta_0 \in \Delta(\delta_0)} \left\{ \sum_{s_\delta \in \Delta(\delta_0)} n\{h_n^{(\delta)}(s_\delta)\}\{\hat{W}_n(s_\delta)\}\{\tilde{h}_n^{(\delta)}(s_\delta)\} \right\},
\]

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

\[
Q_n(\delta_0 + \tilde{h}_n^{(\delta)}(s_\delta) s_\delta, \tilde{\phi}_n(s_\delta)) := \sup_{\{h^{(\delta)}, \phi\}} Q_n(\delta_0 + h^{(\delta)} s_\delta, \phi).
\]
Note that the definition of $QW_n$ is exactly the same as $W_n$ except that $\tilde{h}^{(\delta)}(s_\delta)$ is defined by $Q_n$ instead of $L_n$. If we further let the weight function $\hat{W}_n(s_\delta)$ be $s_\delta'\hat{W}_n s_\delta$ such that $\hat{W}_n$ converges to $-\tilde{A}^{(\delta)}_s$ a.s. $\Rightarrow$, then

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

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

Finally, the asymptotic null distribution of the LM statistic can be similarly obtained. For this we let

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

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

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

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

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

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

### 3.4.3 Example 3 (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. Note that $\theta_1 = 0$ if and only if $\theta_2 = 0$ from the model assumption. Then, for given $r := d_1/d_2 \in [c, \tilde{c}]$, we can apply Corollary 2 to obtain the asymptotic null distribution of the LR test statistic. That is,

$$LR_n \Rightarrow \sup_{s_\xi \in \Delta(\xi_0)} \max[0, \tilde{Y}^{(\xi)}(s_\xi)]^2,$$
where \( s_\xi := s_2, \Delta(\xi_0) := \{-1, 1\}, \) and

\[
\tilde{Y}(\xi)(s_\xi) := \frac{s_2 \tilde{Z}(\xi)}{(\tilde{A}_s(\xi, \xi))^{1/2}} := \frac{r}{\sqrt{r^2}} \frac{s_2(W_1 - (-A_s^{(0,1)})(-A_s^{(0,0)})^{-1}W_0)}{((-A_s^{(1,1)})) - (-A_s^{(1,0)})(-A_s^{(0,0)})^{-1}(-A_s^{(0,1)}))^{1/2}}.
\]

Note that \( d_1/d_2 \) is cancelled out. That’s why the asymptotic null distribution is obtained without considering \( r \). From the structure of \( \Delta(\xi_0) \), it also follows that \( \mathcal{L}R_n \Rightarrow \tilde{Z}(\xi)(\tilde{A}_s(\xi, \xi))^{-1}\tilde{Z}(\xi) \).

In a similar way, we can apply Corollary 3 for the Wald statistic. We note that for given \( r \),

\[
\sqrt{n}\tilde{h}_n^{(\mu)}(s_\xi) \Rightarrow (r\tilde{A}_s(\xi, \xi))^{-1}\max[0, s_2\tilde{Z}(\xi)].
\]

In order to have a statistic ineffective to \( r \), we choose \( \hat{W}_n \) to be a consistent estimator for \( (r\tilde{A}_s(\xi, \xi))^{-1} \). For example, we can let

\[
\hat{W}_n := r^{-1}\{(n^{-1} \sum \log(X_t)^2) - (n^{-1} \sum \log(X_t)Z_t') (n^{-1} \sum Z_t Z_t')^{-1} (n^{-1} \sum Z_t \log(X_t))\}.
\]

Then, \( W_n := n\{(\tilde{h}_n^{(\mu)}(s_\xi)\{\tilde{W}_n\}\hat{h}_n^{(\mu)}(s_\xi)\} \Rightarrow \tilde{Z}(\xi)(\tilde{A}_s(\xi, \xi))^{-1}\tilde{Z}(\xi) \) by Corollary 3. Also, we can apply Corollary 4 to obtain the same asymptotic null distribution of LM test statistic by employing the same weight function.

### 4 Conclusion

In this paper we examine \( d \)-diffle econometric models and provide conditions for these to behave regularly when the number of observations is large. Specifically, given our regularity conditions, we show that extremum estimator has a different distribution from that for standard diffle models, and that it can be represented as a functional of a Gaussian process indexed by direction. Also, our analysis yields the same results as in the literature by treating the extremum estimator for diffle models as a special case of \( d \)-diffle models.

Further, we refine standard LR, Wald, and LM test statistics appropriately for \( d \)-diffle models. These definitions are provided for general \( d \)-diffle models in the sense that their null models can be \( d \)-diffle with respect to other nuisance parameters. Nevertheless, they turn out to have asymptotically equivalent null distributions under the benchmark model assumption mainly motivated by King and Shively (1993), diffle models for GMM estimation, and Box-Cox transformation.
5 Appendix: Proofs

Proof of Lemma 1: To show the given claim, we show that $f$ is twice continuously diffle at $\theta_0$. The same proof can be applied to other parameter values as well.

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

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

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

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

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

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

Therefore,

$$\lim_{\theta \to \theta_0} \sup_{d} \left| \frac{1}{\|\theta - \theta_0\|^2} \left\{ f(\theta) - f(\theta_0) - A(\theta_0)(\theta - \theta_0) - \frac{1}{2}(\theta - \theta_0)^{\prime}B(\theta_0)(\theta - \theta_0) \right\} \right|$$

$$= \lim_{\theta \to \theta_0} \sup_{d} \left| \frac{1}{2\|\theta - \theta_0\|^2} (D^2f(\bar{\theta}; d) - D^2f(\theta_0; d))h^2 \right|$$

$$= \lim_{\theta \to \theta_0} \sup_{d} \left| \frac{1}{2} (D^2f(\bar{\theta}; d) - D^2f(\theta_0; d)) \right|$$

$$\leq \lim_{\theta \to \theta_0} \frac{M}{2} \|\bar{\theta} - \theta_0\| = 0,$$

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

Proof of Lemma 2: (i) To show the given claim, we verify the conditions of Wooldridge and White (1988). First, AC1 of Wooldridge and White (1988) is satisfied by A6(iii) because we can let $n^{-1/2} \sum \ell_t(\theta; d)$ be their $\sum Z_m$. Second, the conditions (i, ii, iii) of AC2 in Wooldridge and White (1988) trivially hold by our assumptions that $\|D\ell_t(\theta; d)\|_s < \Delta$ uniformly in $t$, that $\nu_r$ 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 that $\|\ell_t(\theta, d)\|_s < \Delta < \infty$ uniformly $t$ and $d$. Finally, their condition A5 needs not to be proved as our goal is not obtain the standard normal distribution.

(ii) We can apply the ergodic theorem given assumptions A1 and A6(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 that $\sqrt{n}\hat{h}_n(d) = \max[0, \{D^2L_n(\theta_n(d); d)\}^{-1}DL_n(\theta_s; d)]$ by (9) and Khun–Tucker theorem, and that $\hat{\theta}_n(d) \to \theta_s$ a.s.$-P$ by Theorem 1, it follows that $\sqrt{n}\hat{h}_n(d) \to \max[0, \{A_s(d)\}^{-1}Z(d)]$ by Lemma 2(iii). The desired result follows from the definition of $G(d)$.

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

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

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

Thus, the desired result follows from (10).

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

The tightness can be proved by verifying the conditions of theorem 4 in Hansen (1996). First, from the fact that $\{Y_t\}$ is a strong mixing sequence of $-sq/(s-q)$, for some $\epsilon > 0$, $\alpha^{-(s-q)/(sq)} = O(\tau^{-1-\epsilon})$, so that $\sum_{t=1}^\infty \alpha^{-(s-q)/(sq)} < \infty$. Second, $\|M_t\|_s < \infty$ uniformly in $t$ from the stationarity assumption of $\{M_t\}$ in A6(iv). Third, $\|D\ell_t(\theta_s; d)\|_s < \infty$ uniformly in $t$ and $d$ from A6(iv). Fourth, from that $\nu_\tau$ is of size $-1/(1-\gamma)$, for some $\epsilon > 0$, $\nu_\tau = O(\tau^{-1/(1-\gamma)-\epsilon})$, implying that $\sum_{t=1}^\infty \nu_{\tau-1} < \infty$. Finally, it is already assumed in A6(iv) that $q > (r-1)/(\gamma\lambda)$. All these verify the conditions in theorem 4 of Hansen (1996), and the tightness of $\{n^{-1/2} \sum D\ell_t(\theta_s; \cdot)\}$ follows.

(ii) By assumption A5(iii), $|n^{-1}D^2L_n(\theta_d) - n^{-1}D^2L_n(\theta; d2)| \leq n^{-1} \sum M_t \|d_1 - d_2\|^\lambda$. Further, we can apply the ergodic theorem to $\{n^{-1} \sum M_t\}$, so that for any $\omega \in F$, $\mathbb{P}(F) = 1$, and $\epsilon > 0$, there is an $n^*(\omega, \epsilon)$ such that if $n \geq n^*(\omega, \epsilon)$, then $|n^{-1} \sum M_t - E[M_t]| \leq \epsilon$, 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
\[ n^{-1} \sum M_t \delta = n^{-1} \sum M_t \varepsilon / (\varepsilon + E[M_t]) \leq \varepsilon. \] That is, for any \( \omega \in F \), \( \mathbb{P}(F) = 1 \) and \( \varepsilon > 0 \), there is \( n^* (\omega, \varepsilon) \) and \( \delta \) such that if \( n \geq n^* (\omega, \varepsilon) \) and \( \| d_1 - d_2 \|^\lambda \leq \delta \), then \[ | n^{-1} D^2 L_n(\theta; d_1) - n^{-1} D^2 L_n(\theta; d_2) | < \varepsilon, \] which means that \( \{ n^{-1} D^2 L_n(\theta; \cdot) \}_n \) is equicontinuous. Therefore, it follows that \( n^{-1} D^2 L_n(\theta; \cdot) \) converges to \( A_* \) uniformly on \( \Delta(\theta_* \varepsilon) \) a.s. - \( \mathbb{P} \) by Rudin (1976, p.168).

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

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

(iii) We can apply the CMT to (11).

(iv) From the definition of \( \hat{h}_n \), for each \( d \), \( \sqrt{n} (\hat{\theta}_n(d) - \theta_*) = \sqrt{n} \hat{h}_n(d) d \). Theorems 3(ii), 4(i to iii), (11), and the CMT yield the given result.

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

(i) As the weak convergence is proven for a general function, we verify only the pointwise weak convergence for this case. From the definition of \( DL_n(\theta_*; d) = \nabla_{\theta} L_n(\theta)^T d \), and \( n^{-1/2} \nabla_{\theta} L_n(\theta) \)
\( \Rightarrow Z \) by theorem 1 of DMR (1995). Therefore, \( n^{-1/2} DL_n(\theta_*; d) \Rightarrow Z'd \) for every \( d \in \Delta(\theta_*) \).

(ii) Note that \( D^2 L_n(\theta; d) = d' \nabla^2_{\theta} L_n(\theta_*) d \), 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.

(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).

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

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

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

For notational simplicity, let
\[ v := \frac{(-A_*)^{1/2} d}{\{ d'(-A_*) d \}^{1/2}}. \]
then it follows that \( v'v = 1 \) and \( v'(-A_s)^{-1/2}ZZ(-A_s)^{-1/2}v = d'ZZ'd\{d'(-A_s)d\}^{-1} \). Given this, note that

\[
\max_v v'(-A_s)^{-1/2}ZZ'(-A_s)^{-1/2}v = \max_d d'ZZ'd\{d'(-A_s)d\}^{-1},
\]

and it is equal to the maximum eigenvalue of \((-A_s)^{-1/2}ZZ'(-A_s)^{-1/2}\), which is computed as \(Z'(-A_s)^{-1}Z\). It is because \( \text{rank}((-A_s)^{-1/2}ZZ'(-A_s)^{-1/2}) = 1 \) (so that there is one positive eigenvalue, and other eigenvalues are zero), and the sum of eigenvalues is \( \text{tr}((-A_s)^{-1/2}ZZ'(-A_s)^{-1/2}) = Z'(-A_s)^{-1}Z \). 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

\[
\sqrt{n}(\hat{\theta}_n - \theta_s) \Rightarrow Z'd_s\{d_s'(-A_s)d_s\}^{-1}d_s.
\]

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 \( Z'(-A_s)^{-1}Z \), so that

\[
(-A_s)^{-1/2}ZZ'(-A_s)^{-1/2}v_* = Z'(-A_s)^{-1}Zv_* \tag{33}
\]

by the definition of eigenvector. This implies that

\[
v_*'(-A_s)^{-1/2}ZZ'(-A_s)^{-1/2}v_* = Z'(-A_s)^{-1}Zv_*'v_* = Z'(-A_s)^{-1}Z \tag{34}
\]

because \( v_*'v_* = 1 \). Plugging the definition of \( v_* \) to the LHS of (34) leads to that

\[
Z'd_s\{d_s'(-A_s)d_s\}^{-1} = (d_s'Z)^{-1}Z'(-A_s)^{-1}Z. \tag{35}
\]

Thus, \( Z'd_s\{d_s'(-A_s)d_s\}^{-1}d_s = (d_s'Z)^{-1}Z'(-A_s)^{-1}Zd_s \). Also, plugging the definition of \( v_* \) to (33) yields that \( (d_s'Z)^{-1}Z'(-A_s)^{-1}Zd_s = (-A)^{-1}Z \). Therefore, \( \sqrt{n}(\hat{\theta}_n - \theta_s) \Rightarrow Z'd_s\{d_s'(-A_s)d_s\}^{-1}d_s = (-A_s)^{-1}Z \). This completes the proof.

Proof of Theorem 5: Note that for any \( hd \) such that \( h \in \mathbb{R}^+ \) and \( d \in \Delta(\pi_s) \), there are \( h(\pi) \in \mathbb{R}^+ \), \( \tau(\pi) \in \mathbb{R}^+ \), and \( s_{\pi} \in \Delta(\pi_s) \) such that \( \tau d = [h(\pi)s_{\pi}', \tau(\pi)s_{\pi}'] \) by A7. Thus,
\[ L_n(\theta + h d) = L_n(\pi + h(\pi) s_\pi, \tau + h(\tau) s_\tau), \] implying that

\[ 2\{L_n(\pi + h(\pi) s_\pi, \tau + h(\tau) s_\tau) - L_n(\pi, \tau)\} = 2D L_n(\pi, \tau; s_\pi) h(\pi) + 2D L_n(\pi, \tau; s_\tau) h(\tau) + D^2 L_n(\pi, \tau; s_\pi)(h(\pi))^2 + D^2 L_n(\pi, \tau; s_\tau)(h(\tau))^2 + 2D L_n(\pi, \tau; s_\pi; s_\tau) h(\pi) h(\tau) + o_{\mathbb{P}_{d\pi}}(1) + o_{\mathbb{P}_{d\tau}}(1), \tag{36} \]

where \( D L_n(\pi, \tau; s_\pi; s_\tau) \) is the directional derivative of \( D L_n(\cdot, \\cdot; s_\pi) \) with respect to \( s_\tau \) evaluated at \( (\pi, \tau) \), and

\[
\sup_d \sup_h L_n(\theta + h d) = \sup_{\{s_\pi, s_\tau\}} \{h(\pi), h(\tau)\} L_n(\pi + h(\pi) s_\pi, \tau + h(\tau) s_\tau).
\]

Therefore,

\[ 2\{L_n(\hat{\theta}_n) - L_n(\theta)\} = \sup_d \sup_h 2\{L_n(\theta + h d) - L_n(\theta)\} = \sup_{s_\pi} \sup_{h(\pi)} \{2D L_n(\theta; s_\pi) h(\pi) + D^2 L_n(\theta; s_\pi)(h(\pi))^2 + o_{\mathbb{P}_{d\pi}}(1)\} + \sup_{s_\tau} \sup_{h(\tau)} \{2D L_n(\theta; s_\tau) h(\tau) + D^2 L_n(\theta; s_\tau)(h(\tau))^2 + o_{\mathbb{P}_{d\tau}}(1)\}, \tag{37} \]

where we exploited the facts that \( n^{-1} D L_n(\pi, \tau; s_\pi; s_\pi) \) has probability limit \( A^{(\pi, \pi)}(s_\pi, s_\pi) = 0 \) by assumption A7(iv), and that \( D L_n(\theta; \\cdot) \) and \( D L_n(\theta; \\cdot) \) are \( O_P(n^{1/2}) \) 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 (37).

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

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

**Proof of Corollary 2:** To show the given claim, we use (37). As it is trivial from the proof of Theorem 5 that \( H_{2,n} := \sup_{s_\tau} \sup_{h(\tau)} \{2D L_n(\theta; s_\tau) h(\tau) + D^2 L_n(\theta; s_\tau)(h(\tau))^2\} \Rightarrow H_2 \), we here focus to the weak limit of

\[ H_{1,n} := \sup_{s_\pi} \sup_{h(\pi)} \{2D L_n(\theta; s_\pi) h(\pi) + D^2 L_n(\theta; s_\pi)(h(\pi))^2\}. \]
From the fact that for any \( h \in \mathbb{R}^+ \) and \( d_\pi \in \Delta(\pi_*) \), there are \( h^{(\lambda)} \in \mathbb{R}^+ \), \( h^{(\xi)} \in \mathbb{R}^+ \), \( s_\lambda \in \Delta(\lambda_*) \), and \( s_\pi \in \Delta(\pi_*) \) such that \( h d_\pi = [h^{(\lambda)} s_\lambda', h^{(\xi)} s_\xi']' \) and

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

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,

\[
H_{1,n} \Rightarrow \mathcal{H}_1 = \sup_{\{s_\xi, s_\lambda\}} \sup_{\{h^{(\xi)}, h^{(\lambda)}\}} 2\mathcal{Z}(\xi)(s_\xi) h^{(\xi)} + 2s_\lambda' \mathcal{Z}(\lambda) h^{(\lambda)} + 2s_\lambda' A^{(\lambda, \xi)}(s_\xi) h^{(\lambda)} h^{(\xi)} \tag{38}
+ A^{(\xi, \xi)}(s_\xi)(h^{(\xi)})^2 + s_\lambda' A^{(\lambda, \lambda)}(s_\lambda)(h^{(\lambda)})^2,
\]

and there can be possibly four different cases for the solutions of the RHS of (38) 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 (38) then 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 \( \hat{h}^{(\lambda)}(s_\xi, s_\lambda) > 0 \), then the RHS of (38) is identical to

\[
\sup_{\{s_\xi, s_\lambda\}} [\mathcal{Z}(\xi)(s_\xi) s_\lambda' \mathcal{Z}(\lambda)] \begin{bmatrix}
-A^{(\xi, \xi)}(s_\xi) & -s_\lambda' A^{(\lambda, \xi)}(s_\xi) \\
-s_\lambda' A^{(\lambda, \xi)}(s_\xi) & -s_\lambda' A^{(\lambda, \lambda)}(s_\lambda)
\end{bmatrix}^{-1} \begin{bmatrix}
\mathcal{Z}(\xi)(s_\xi) \\
s_\lambda' \mathcal{Z}(\lambda)
\end{bmatrix},
\]

and maximizing this with respect to \( s_\lambda \) for a given \( s_\xi \) yields \( \hat{\mathcal{Y}}(\xi)(s_\xi)^2 + (\mathcal{Z}(\lambda))' (A^{(\lambda, \lambda)} - 1 \mathcal{Z}(\lambda)) \).

For \( \hat{h}^{(\xi)}(s_\xi, s_\lambda) \) to be greater than zero, it is necessary that \( \hat{\mathcal{Y}}(\xi)(s_\xi) \) is greater than zero, too. Second, if \( \hat{h}^{(\lambda)}(s_\xi, s_\lambda) > 0 \) and \( \hat{h}^{(\xi)}(s_\xi, s_\lambda) = 0 \) then the RHS of (38) is identical to \( 2s_\lambda' \mathcal{Z}(\lambda) + s_\lambda' A^{(\lambda, \lambda)}(s_\lambda) \), and maximizing this with respect to \( s_\lambda \) leads to \( (\mathcal{Z}(\lambda))' (A^{(\lambda, \lambda)}) (\mathcal{Z}(\lambda)) \) as its maximum. Also, \( \hat{\mathcal{Y}}(\xi)(s_\xi) \) 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 \( \hat{h}^{(\xi)}(s_\xi, s_\lambda) > 0 \) then we can consider \( -s_\lambda \) as an alternative to \( s_\lambda \) for the same \( s_\xi \) while maximizing the RHS of (38) with respect to \( s_\lambda \) from the fact that \( \lambda_* \) 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 \( \hat{h}^{(\xi)}(s_\xi, s_\lambda) = 0 \), then \( -s_\lambda \) can be considered as an alternative as well,
and this modifies the maximization to the second case. Therefore, combining all these leads to that

\[ \mathcal{H}_1 = \sup_{s \in \Delta(\xi)} \max[0, \hat{y}(\xi)(s_\xi)]^2 + (Z(\lambda))'(-A_s(\lambda,\lambda))(Z(\lambda)). \]

**Proof of Lemma 4:** (i) We exploit (37) further. First, applying the CMT to Theorem 4(i) shows that

\[ (n^{-1/2}D\lambda_{\alpha}^{(\tau)}, n^{-1}D^2\lambda_{\alpha}^{(\tau)}) = (\mathcal{Z}(\tau), A_s^{(\tau,\tau)}). \]

Thus,

\[ \sup_{n^{1/2}h(\tau)} 2D\lambda_n(\theta_s; s_\tau)h^{(\tau)} + D^2\lambda_n(\theta_s; s_\tau)(h^{(\tau)})^2 \Rightarrow \sup_{h(\tau) \in \mathbb{R}^+} 2\mathcal{Z}(\tau)(s_\tau)h^{(\tau)} + A_s^{(\tau,\tau)}(s_\tau)(h^{(\tau)})^2, \]

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

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

\[ \sup_{h^{(\eta)}} \{2D\lambda_n(\theta_s; d_\eta)h^{(\eta)} + D^2\lambda_n(\theta_s; d_\eta)(h^{(\eta)})^2 \} \]

\[ \Rightarrow \sup_{(h^{(\mu)}, h^{(\omega)})} 2\mathcal{Z}(s_\mu)h^{(\mu)} + 2\mathcal{Z}(s_\omega)h^{(\omega)} + 2A_s^{(\omega,\mu)}(s_\mu, s_\omega)h^{(\mu)}h^{(\omega)} \]

\[ + A_s^{(\mu,\mu)}(s_\mu)(h^{(\mu)})^2 + A_s^{(\omega,\omega)}(s_\omega)(h^{(\omega)})^2. \]

Given this, \( h^{(\mu)} \) and \( h^{(\omega)} \) on the RHS of (39) are present on the positive Euclidean line, so that there can be possibly four different inequality constraints. We examine each case one by one. First, if any inequality condition does not bind, then \( \sqrt{n}(\hat{h}_n^{(\mu)}(s_\mu, s_\omega), \hat{h}_n^{(\omega)}(s_\mu, s_\omega))' \Rightarrow \mathcal{G}^{(\eta)}(s_\mu, s_\omega) \) by the standard FOC and Lemma 2. This occurs if each component of \( \mathcal{G}^{(\eta)}(s_\mu, s_\omega) \) is strictly greater than zero. Second, if \( \mathcal{G}^{(\mu)}(s_\mu, s_\omega) < 0, \) \( \hat{h}_n^{(\omega)}(s_\mu, s_\omega) = \max[0, \hat{G}^{(\omega)}(s_\mu, s_\omega)]. \) Thus, \( \sqrt{n}(\hat{h}_n^{(\mu)}(s_\mu, s_\omega), \hat{h}_n^{(\omega)}(s_\mu, s_\omega), \hat{h}_n^{(\mu)}(s_\mu, s_\omega, s_\tau))' \Rightarrow (0, \max[0, \hat{G}^{(\omega)}(s_\mu, s_\omega)])'. \) Likewise, if \( \hat{h}_n^{(\omega)}(s_\mu, s_\omega) = 0 \) in the RHS of (39) because \( \mathcal{G}^{(\omega)}(s_\mu, s_\omega) < 0, \) then \( \hat{h}_n^{(\mu)}(s_\mu, s_\omega) = \max[0, \hat{G}^{(\mu)}(s_\mu, s_\omega)]. \) This implies that \( \sqrt{n}(\hat{h}_n^{(\mu)}(s_\mu, s_\omega), \hat{h}_n^{(\omega)}(s_\mu, s_\omega, s_\tau))' \Rightarrow (\max[0, \hat{G}^{(\mu)}(s_\mu, s_\omega)]', 0'). \) Fourth, it must
be the case that \( \sqrt{n}(\hat{h}^{(\mu)}(s_{\mu}, s_{\omega}, s_{\tau}), \hat{h}^{(\omega)}(s_{\mu}, s_{\omega}, s_{\tau}))' \Rightarrow (0, 0)' \) for any other case. Therefore,

\[
\sqrt{n} \begin{bmatrix} \hat{h}^{(\mu)}(s_{\mu}, s_{\omega}, s_{\tau}) \\ \hat{h}^{(\omega)}(s_{\mu}, s_{\omega}, s_{\tau}) \\ \hat{h}^{(\tau)}(s_{\mu}, s_{\omega}, s_{\tau}) \end{bmatrix} \Rightarrow \begin{bmatrix} G(\mu) \\ G(\omega) \\ 0 \end{bmatrix} \mathbf{1}_{\{\min(G(\mu), G(\omega)) \geq 0\}} + \begin{bmatrix} \max[0, \hat{G}(\mu)] \mathbf{1}_{\{G(\omega) < 0\}} \\ \max[0, \hat{G}(\omega)] \mathbf{1}_{\{G(\mu) < 0\}} \\ \max[0, G(\tau)] \end{bmatrix}
\]

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

**Proof of Theorem 8:** We can approximate (29) by a quadratic function and apply Lemma 4 to obtain that

\[
\sup_{\{h^{(\mu)}, \omega, \tau\}} 2\{L_{n}(\mu_{0} + h^{(\mu)} s_{\mu}, \omega, \tau) - L_{n}(\mu_{0}, \omega_{a}, \tau_{a})\}
\]

\[
= \sup_{\omega_{a}, \tau_{a}} \sup_{\{h^{(\mu)}, \omega, \tau\}} 2\{L_{n}(\mu_{0} + h^{(\mu)} s_{\mu}, \omega_{a} + h^{(\omega)} s_{\omega}, \tau_{a} + h^{(\tau)} s_{\tau}) - L_{n}(\mu_{0}, \omega_{a}, \tau_{a})\}
\]

\[
\Rightarrow \sup_{s_{\omega}} G^{(\eta)}(s_{\mu}, s_{\omega})'(-A^{(\eta, \eta)}(s_{\mu}, s_{\omega}))G^{(\eta)}(s_{\mu}, s_{\omega})\mathbf{1}_{\{\min(G^{(\mu)}(s_{\mu}, s_{\omega}), G^{(\omega)}(s_{\mu}, s_{\omega})) > 0\}}
\]

\[
+ \sup_{s_{\omega}} \max[0, \hat{G}^{(\mu)}(s_{\mu})](-A^{(\mu, \mu)}(s_{\mu})) \max[0, \hat{G}^{(\mu)}(s_{\mu})] \mathbf{1}_{\{G^{(\mu)}(s_{\mu}, s_{\omega}) \geq 0, \hat{G}^{(\mu)}(s_{\mu}, s_{\omega}) \geq 0\}}
\]

\[
+ \sup_{s_{\omega}} \max[0, \hat{G}^{(\omega)}(s_{\omega})](-A^{(\omega, \omega)}(s_{\omega})) \max[0, \hat{G}^{(\omega)}(s_{\omega})] \mathbf{1}_{\{G^{(\omega)}(s_{\mu}, s_{\omega}) \geq 0, \hat{G}^{(\mu)}(s_{\mu}, s_{\omega}) \geq 0\}}
\]

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

which is identical to

\[
G^{(\eta)}(s_{\mu}, \bar{s}_{\omega}(s_{\mu}))'(-A^{(\eta, \eta)}(s_{\mu}, \bar{s}_{\omega}(s_{\mu})))G^{(\eta)}(s_{\mu}, \bar{s}_{\omega}(s_{\mu}))\mathbf{1}_{\{\min(G^{(\mu)}(s_{\mu}, \bar{s}_{\omega}(s_{\mu})), G^{(\omega)}(s_{\mu}, \bar{s}_{\omega}(s_{\mu})) > 0\}}
\]

\[
+ \max[0, \hat{G}^{(\mu)}(s_{\mu})](-A^{(\mu, \mu)}(s_{\mu})) \max[0, \hat{G}^{(\mu)}(s_{\mu})] \mathbf{1}_{\{G^{(\mu)}(s_{\mu}, \bar{s}_{\omega}(s_{\mu})) \geq 0, \hat{G}^{(\mu)}(s_{\mu}, \bar{s}_{\omega}(s_{\mu})) \geq 0\}}
\]

\[
+ \max[0, \hat{G}^{(\omega)}(s_{\omega})](-A^{(\omega, \omega)}(s_{\omega})) \max[0, \hat{G}^{(\omega)}(s_{\omega})] \mathbf{1}_{\{G^{(\omega)}(s_{\mu}, \bar{s}_{\omega}(s_{\omega})) \geq 0, \hat{G}^{(\omega)}(s_{\mu}, \bar{s}_{\omega}(s_{\omega})) \geq 0\}}
\]

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

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

\[
\max[0, \hat{G}^{(\omega)}(s_{\omega})](-A^{(\omega, \omega)}(s_{\omega})) \max[0, \hat{G}^{(\omega)}(s_{\omega})] \mathbf{1}_{\{G^{(\omega)}(s_{\mu}, \bar{s}_{\omega}(s_{\omega})) \geq 0, \hat{G}^{(\omega)}(s_{\mu}, \bar{s}_{\omega}(s_{\omega})) \geq 0\}}
\]

\[
= \sup_{s_{\omega} \in \Delta(\omega_{a})} \max[0, \hat{G}^{(\omega)}(s_{\omega})](-A^{(\omega, \omega)}(s_{\omega})) \max[0, \hat{G}^{(\omega)}(s_{\omega})] \mathbf{1}_{\{G^{(\omega)}(s_{\mu}, \bar{s}_{\omega}(s_{\omega})) \geq 0, \hat{G}^{(\omega)}(s_{\mu}, \bar{s}_{\omega}(s_{\omega})) \geq 0\}}.
\]

Given this, the asymptotic behavior of \( \sqrt{n}h^{(\mu)}(s_{\mu}) \) can be captured by \( G^{(\mu)}(s_{\mu}, \bar{s}_{\omega}(s_{\mu})) \) and \( \max[0, \hat{G}^{(\mu)}(s_{\mu})] \) on the probability 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}, \tilde{s}_{\omega}(s_{\mu})) \geq 0 > G^\omega(s_{\mu}, \tilde{s}_{\omega}(s_{\mu}))\} \text{ respectively. Otherwise, } \sqrt{n} \tilde{h}^\mu_n(s_{\mu}) \Rightarrow 0. \text{ Thus, }

\sqrt{n} \tilde{h}^\mu_n(s_{\mu}) \Rightarrow G^\mu(s_{\mu}, \tilde{s}_{\omega}(s_{\mu})) \mathbb{1}_{\{\min[|G^\mu(s_{\mu}, \tilde{s}_{\omega}(s_{\mu})), |G^\omega(s_{\mu}, \tilde{s}_{\omega}(s_{\mu}))]| > 0\}} + \max[0, \dot{G}^\mu(s_{\mu})] \mathbb{1}_{\{|G^\mu(s_{\mu}, \tilde{s}_{\omega}(s_{\mu}))| \geq 0 > |G^\omega(s_{\mu}, \tilde{s}_{\omega}(s_{\mu}))|\}}.

This holds as a function of \( s_{\mu} \), too, so that Theorem 8 follows by applying again the CMT to the Wald test statistic.

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

\textbf{Proof of Theorem 9:} Before proceeding our proof, 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 find out the convergence limit of each component constituting the LM test statistic. First, there is \( n^* \) a.s.\(-\mathbb{P}\) such that if \( n > n^* \) then \( \Delta(\hat{\omega}_n) = \Delta(\omega_s) \). Note that \( \omega_s \) is an interior element by A11(i), so that \( \Delta(\omega_s) = \{x \in \mathbb{R}^{\tau_s} : ||x|| = 1\} \), and further for an open ball with radius \( \varepsilon > 0 \) such that \( B(\omega_s, \varepsilon) \subset \Omega \), there is \( n(\varepsilon) \) a.s.\(-\mathbb{P}\) if \( n > n(\varepsilon) \) then \( \omega_n \in B(\omega_s, \varepsilon) \) by Theorem 6(i), implying that \( \omega_n \) is an interior element, too. Thus, if we let \( n^* > n(\varepsilon) \) then \( \Delta(\hat{\omega}_n) = \{x \in \mathbb{R}^{\tau_s} : ||x|| = 1\} \), which is also \( \Delta(\omega_s) \). Second, \( n^{-1/2} DL_n(\tilde{\theta}_n; \cdot) \Rightarrow \tilde{Z}^\mu(\cdot; \tilde{s}_{\omega}) \).

Applying the mean–value theorem shows that for each \( s_{\mu} \) there is \( \omega(\cdot; s_{\mu}) \) such that

\[
DL_n(\tilde{\theta}_n; s_{\mu}) - DL_n(\theta_s; s_{\mu}) = DL_n(\mu_0, \omega_n(s_{\mu}), \tau_s; s_{\mu}, \tilde{s}_{\omega,n}) \{\tilde{h}^\omega_n(\tilde{s}_{\omega,n})\} = DL_n(\mu_0, \omega_n(s_{\mu}), \tau_s; s_{\mu}, \tilde{s}_{\omega,n}) \{\tilde{h}^\omega_n(\tilde{s}_{\omega,n})\} - DL_n(\theta_s; \tilde{s}_{\omega,n}),
\]

where we define \( (\tilde{h}^\omega_n(\tilde{s}_{\omega,n}), \tilde{s}_{\omega,n}) \) is such that \( L_n(\mu_0, \omega_s + \tilde{h}^\omega_n(\tilde{s}_{\omega,n}) \tilde{s}_{\omega,n}, \tau_s) = \sup_{\omega_n} \sup_{h(\omega)(\cdot; \tilde{s}_{\omega,n}, \tau_s)} L_n(\mu_0, \omega_s + h(\omega)(\cdot; \tilde{s}_{\omega,n}, \tau_s), \tau_s) \), and the last equality follows from the mean–value theorem: there is \( \omega_n(s_{\mu}) \) such that (9) holds. Given this and Theorem 6(i), the ULLN can be applied to \( n^{-1} DL_n(\mu_0, \omega_n(s_{\mu}), \tau_s; \cdot) \) and \( n^{-1} D^2 L_n(\mu_0, \omega_n(s_{\mu}), \tau_s; \cdot) \) to obtain \( A_s^\mu(\omega, \tau_s) \) and \( A_s^\omega(\omega, \tau_s) \) as their respective probability limits. Also, it trivially follows that \( n^{-1/2} (DL_n(\theta_s; \cdot), DL_n(\theta_s; \tilde{s}_{\omega,n})) \Rightarrow (Z^\mu, Z^\omega(\cdot; \tilde{s}_{\omega})) \) from that \( n^{-1/2} (DL_n(\theta_s; s_{\mu}), DL_n(\theta_s; s_{\omega})) \Rightarrow (Z^\mu, Z^\omega) \) as functions of \( (s_{\mu}, s_{\omega}) \), and that \( \max[0, DL_n(\theta_s; \tilde{s}_{\omega,n})]^2 \{ -D^2 L_n(\theta_s; \tilde{s}_{\omega,n}) \}^{-1} \Rightarrow \max[0, \gamma^\omega(\tilde{s}_{\omega})]^2 \). Thus, \( n^{-1/2} DL_n(\tilde{\theta}_n; \cdot) \Rightarrow \tilde{Z}^\mu(\cdot; \tilde{s}_{\omega}) \).
\[ \mathcal{L} \mathcal{M}_n \Rightarrow \sup_{s \in \Delta(\xi_0)} \left( \frac{\max\{0, \tilde{Z}(\xi)(s_\xi; \tilde{s}_\lambda)\}^2}{\inf_{s \in \Delta(\lambda)} \{ -\bar{A}_{\lambda}(s_\xi, s_\lambda) \} } \right), \]

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

\[
\inf_{s_\lambda \in \Delta(\lambda)} \{-\bar{A}_{\lambda}(s_\xi, s_\lambda)\} \\
= -\bar{A}_{\lambda}(s_\xi) - \sup_{s_\lambda \in \Delta(\lambda)} \{-\bar{A}_{\lambda}(s_\xi) \bar{s}_\lambda \} \{-s_\lambda' A_{\lambda}' A_{\lambda}^{-1} \{ -s_\lambda' A_{\lambda}(s_\xi) \} \} \\
= -\{-A_{\lambda}(s_\xi) - A_{\lambda}(s_\xi)' \{ A_{\lambda}^{-1} \} A_{\lambda}(s_\xi) \} = -\bar{A}_{\lambda}(s_\xi).
\]

Next, \( \tilde{Z}(\xi)(s_\xi; \tilde{s}_\lambda) = Z(\xi)(s_\xi) - \{-A_{\lambda}(s_\xi) \bar{s}_\lambda \} \{-s_\lambda' A_{\lambda}' A_{\lambda}(s_\xi) \} \bar{s}_\lambda \} \), and we already proved in the proof of Corollary 1(ii) that \( \bar{s}_\lambda \{-s_\lambda' A_{\lambda}' A_{\lambda}(s_\xi) \} \bar{s}_\lambda \} \bar{s}_\lambda \} = \{-A_{\lambda}(s_\xi) \} \bar{s}_\lambda \} \), implying that \( \tilde{Z}(\xi)(s_\xi; \tilde{s}_\lambda) = Z(\xi)(s_\xi) - \{-A_{\lambda}(s_\xi) \} \{-A_{\lambda}(s_\xi) \} \bar{s}_\lambda \} \). Therefore, the desired result follows.

\section*{References}


Figure 1: Q-Q Plot between \( \sup \max[0, \tilde{Y}^{(\theta)}(s_1, s_2)]^2 \) and \( \sup \max[0, \hat{Y}^{(\theta)}(s_1, s_2)]^2 \)

Figure 2: Empirical and Asymptotic Distributions of LR Statistic
Figure 3: Empirical and Asymptotic Distributions of LR Statistic