

# Supplements to “Directionally Differentiable Econometric Models”

JIN SEO CHO

School of Economics

Yonsei University

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

Email: jinseocho@yonsei.ac.kr

HALBERT WHITE

Department of Economics

University of California, San Diego

9500 Gilman Dr., La Jolla, CA, 92093-0508, U.S.A.

Email: hwhite@weber.ucsd.edu

This version: October, 2016

## Abstract

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

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

**JEL Classification:** C12, C13, C22, C32.

**Acknowledgements:** The co-editor, Yoon-Jae Whang, and two anonymous referees provided very helpful comments for which we are most grateful. The second author (Halbert White) passed away while the final version was written. He formed the outline of the paper and also showed an exemplary guide for writing a quality research paper. The authors are benefited from discussions with Yoichi Arai, Seung Chan Ahn, In Choi, Horag Choi, Robert Davies, Graham Elliott, Chirok Han, John Hillas, Jung Hur, Hide Ichimura, Isao Ishida, Yongho Jeon, Estate Khmaladze, Chang-Jin Kim, Chang Sik Kim, Tae-Hwan Kim, Naoto Kunitomo, Taesuk Lee, Bruce Lehmann, Mark Machina, Jaesun Noh, Kosuke Oya, Taeyoung Park, Peter Phillips, Erwann Sbair, Donggyu Sul, Hung-Jen Wang, Yoshihiro Yajima, Byung Sam Yoo and other seminar participants at Sogang University, the University of Auckland, the University of Tokyo, Osaka University, the Econometrics Study Group of the Korean Econometric Society, VUW, Yonsei University, and other conference participants at NZESG (Auckland, 2013).

# 1 Introduction

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

## 2 Examples

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

### 2.1 Example 1: Stochastic Frontier Production Function Models

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

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

where  $Y_t \in \mathbb{R}$  is the output produced by inputs  $\mathbf{X}_t \in \mathbb{R}^k$  such that  $\boldsymbol{\beta}_*$  is an interior element of  $\mathbf{B} \subset \mathbb{R}^k$ ,  $E[U_t^2] < \infty$ ,  $E[X_{t,j}^2] < \infty$  for  $j = 1, 2, \dots, k$ , and  $E[\mathbf{X}_t \mathbf{X}_t']$  is positive-definite. Here,  $U_t$  stands for an error that is independent of  $\mathbf{X}_t$ . This model is first introduced by Aigner, Lovell, and Schmidt (1977).

One of the early uses of this specification is in identifying inefficiently produced outputs. Given output levels subject to the production function and inputs, if  $E[U_t] < 0$ , outputs are inefficiently produced. Aigner, Lovell, and Schmidt (1977) capture this inefficiency by decomposing  $U_t$  into  $U_t \equiv V_t - W_t$ , where  $V_t \sim N(0, \tau_*^2)$ ,  $W_t := \max[0, Q_t]$ ,  $Q_t \sim N(\mu_*, \sigma_*^2)$ , and  $V_t$  is independent of  $W_t$ . Here, it is assumed that  $\tau_* > 0$ ,  $\sigma_* \geq 0$ , and  $\mu_* \geq 0$ , and  $W_t$  is employed to capture inefficiently produced outputs. If  $\mu_* = 0$  and  $\sigma_*^2 = 0$ , this model reduces to Zellner, Kmenta, and Drèze's (1966) stochastic production function model, implying that outputs are efficiently produced. The key to identifying the inefficiency is, therefore, in testing whether  $\mu_* = 0$  and  $\sigma_*^2 = 0$ .

The original model introduced by Aigner, Lovell, and Schmidt (1977) assumes  $\mu_* = 0$ , so that the mode of  $W_t$  is always achieved at zero. Stevenson (1980) suggests to extend the model scope by letting  $\mu_*$  be different from zero, and the model with unknown  $\mu_*$  has been popularly specified for empirical data

analysis since then (e.g., Dutta, Narasimhan, and Rajiv (1999), Habib and Ljungqvist (2005), and etc.).

Nevertheless, we do not find a methodology testing whether  $\mu_* = 0$  and  $\sigma_*^2 = 0$  in prior literature to the best of our knowledge. This is mainly because the likelihood value is not identified under the null. Note that for each  $(\beta, \sigma, \mu, \tau)$ , the log-likelihood is given as

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

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

$$\tilde{\mu}_t := \frac{\tau^2 \mu - \sigma^2 (Y_t - \mathbf{X}_t' \beta)}{\tau^2 + \sigma^2} \quad \text{and} \quad \tilde{\sigma}^2 := \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2}.$$

Here, the log-likelihood is not identified if  $\theta_* := (\beta'_*, \mu_*, \sigma_*, \tau_*)' = (\beta'_*, 0, 0, \tau_*)$  because  $\mu_*/\sqrt{\sigma_*^2} = 0/0$ , so that  $\ln[\Phi(\mu_*/\sqrt{\sigma_*^2})]$  is not properly identified. Furthermore, if we let

$$\tilde{\mu}_{*t} := \frac{\tau_*^2 \mu_* - \sigma_*^2 U_t}{\tau_*^2 + \sigma_*^2} \quad \text{and} \quad \tilde{\sigma}_*^2 := \frac{\tau_*^2 \sigma_*^2}{\tau_*^2 + \sigma_*^2},$$

$\tilde{\mu}_{*t}/\sqrt{\tilde{\sigma}_*^2} = 0/0$ , so that  $\ln[\Phi(\tilde{\mu}_{*t}/\sqrt{\tilde{\sigma}_*^2})]$  is not also identified by the model. Even further, this model is not differentiable (D). This can be verified by examining the first-order directional derivative. Some tedious algebra shows that for given  $\mathbf{d} := (\mathbf{d}'_\beta, d_\mu, d_\sigma, d_\tau)'$ ,

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

which is the log-likelihood desired by the null condition. This limit is obtained by particularly using the fact that

$$\lim_{h \downarrow 0} \Phi \left( \frac{hd_\mu}{\sqrt{(hd_\sigma)^2}} \right) = \Phi \left( \frac{d_\mu}{\sqrt{d_\sigma^2}} \right) \quad \text{and} \quad \lim_{h \downarrow 0} \Phi \left( \frac{\tilde{\mu}_{*t}(h; \mathbf{d})}{\sqrt{\tilde{\sigma}(h; \mathbf{d})^2}} \right) = \Phi \left( \frac{d_\mu}{\sqrt{d_\sigma^2}} \right),$$

where

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

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

Using this directional limit, the first- and second-order directional derivatives of  $L_n(\cdot)$  at  $(\beta_*, 0, 0, \tau_*)$  are

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

and

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

respectively, where  $\psi(d_\mu, d_\sigma) := |d_\sigma| \phi(d_\mu/|d_\sigma|) / \Phi(d_\mu/|d_\sigma|)$ . Here, if  $\theta_* = (\beta'_*, 0, 0, \tau_*)$ ,  $U_t \sim N(0, \tau_*^2)$ . These directional derivatives are neither linear nor quadratic with respect to  $\mathbf{d}$ , respectively, so that  $L_n(\cdot)$  is not twice D. Therefore, this model cannot be analyzed as for the standard D likelihood function. We examine this model by letting

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

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

It is not hard to identify the asymptotic behaviors of the first and second-order directional derivatives. Note that  $DL_n(\theta_*; \mathbf{d}) = Z_{1,n}(\mathbf{d}) + Z_{2,n}(\mathbf{d})$ , where for each  $\mathbf{d}$ ,

$$Z_{1,n}(\mathbf{d}) := \frac{d_\tau}{\tau_*^3} \sum_{t=1}^n (U_t^2 - \tau_*^2), \quad Z_{2,n}(\mathbf{d}) := \frac{1}{\tau_*^2} \sum_{t=1}^n [\mathbf{X}_t' \mathbf{d}_\beta + m(d_\mu, d_\sigma)] U_t,$$

and  $m(d_\mu, d_\sigma) := -[d_\mu + \psi(d_\mu, d_\sigma)]$ . Note that  $\psi(\cdot, \cdot)$  is Lipschitz continuous, so that Assumption 5(iii) holds with respect to the first-order directional derivative. Furthermore, for each  $\mathbf{d}$ , McLeish's (1974, Theorem 2.3) central limit theorem (CLT) can be applied to  $Z_{1,n}(\mathbf{d})$  and  $Z_{2,n}(\mathbf{d})$ : for each  $\mathbf{d}$ ,

$$n^{-1/2} \begin{bmatrix} Z_{n,1}(\mathbf{d}) \\ Z_{n,2}(\mathbf{d}) \end{bmatrix} \Rightarrow \begin{bmatrix} Z_1(\mathbf{d}) \\ Z_2(\mathbf{d}) \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{1}{\tau_*^2} \begin{bmatrix} 2d_\tau^2 & 0 \\ 0 & E[(\mathbf{X}_t' \mathbf{d}_\beta + m(d_\mu, d_\sigma))^2] \end{bmatrix} \right).$$

It also follows that for each  $\mathbf{d}$  and  $\tilde{\mathbf{d}}$ ,

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

$$E[\mathcal{Z}_2(\mathbf{d})\mathcal{Z}_2(\tilde{\mathbf{d}})] = \frac{1}{\tau_*^2} \begin{bmatrix} m(d_\mu, d_\sigma) \\ \mathbf{d}_\beta \end{bmatrix}' \begin{bmatrix} 1 & E[\mathbf{X}_t'] \\ E[\mathbf{X}_t] & E[\mathbf{X}_t\mathbf{X}_t'] \end{bmatrix} \begin{bmatrix} m(\tilde{d}_\mu, \tilde{d}_\sigma) \\ \tilde{\mathbf{d}}_\beta \end{bmatrix}.$$

Here,  $\mathcal{Z}_{n,1}(\mathbf{d})$  and  $\mathcal{Z}_{n,2}(\mathbf{d})$  are linear with respect  $d_\tau$  and  $[m(d_\mu, d_\sigma), \mathbf{d}_\beta]'$ , respectively. From this fact, their tightness trivially follows, so that  $n^{-1/2}DL_n(\boldsymbol{\theta}_*; \cdot) \Rightarrow \mathcal{Z}(\cdot)$ , where  $\mathcal{Z}(\cdot)$  is a zero-mean Gaussian stochastic process such that for each  $\mathbf{d}$  and  $\tilde{\mathbf{d}}$ ,  $E[\mathcal{Z}(\mathbf{d})\mathcal{Z}(\tilde{\mathbf{d}})] = B_*(\mathbf{d}, \tilde{\mathbf{d}})$  and

$$B_*(\mathbf{d}, \tilde{\mathbf{d}}) := \frac{1}{\tau_*^2} \begin{bmatrix} \mathbf{d}_\beta \\ m(d_\mu, d_\sigma) \\ d_\tau \end{bmatrix}' \begin{bmatrix} E[\mathbf{X}_t\mathbf{X}_t'] & E[\mathbf{X}_t] & 0 \\ E[\mathbf{X}_t'] & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{d}}_\beta \\ m(\tilde{d}_\mu, \tilde{d}_\sigma) \\ \tilde{d}_\tau \end{bmatrix}.$$

It is also possible to define  $\mathcal{Z}(\cdot)$  as  $\mathcal{Z}_1(\cdot) + \mathcal{Z}_2(\cdot)$ .

We provide another Gaussian stochastic process with the same covariance structure as that of  $\mathcal{Z}(\cdot)$ . If we let  $\tilde{\mathcal{Z}}(\mathbf{d}) := \boldsymbol{\delta}(\mathbf{d})'\boldsymbol{\Omega}_*^{1/2}\mathbf{W}$  such that for each  $\mathbf{d}$ ,

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

and  $\mathbf{W} \sim N(\mathbf{0}_{k+2}, \mathbf{I}_{k+2})$ , it follows that  $E[\tilde{\mathcal{Z}}(\mathbf{d})\tilde{\mathcal{Z}}(\tilde{\mathbf{d}})] = \boldsymbol{\delta}(\mathbf{d})'\boldsymbol{\Omega}_*\boldsymbol{\delta}(\tilde{\mathbf{d}})$  that is identical to  $B_*(\mathbf{d}, \tilde{\mathbf{d}})$ , so that  $\tilde{\mathcal{Z}}(\cdot) \stackrel{d}{=} \mathcal{Z}(\cdot)$ . Furthermore,  $\tilde{\mathcal{Z}}(\cdot)$  is linear with respect to  $\mathbf{W}$ . This feature makes it convenient to analyze the asymptotic distribution of the first-order directional derivative.

The probability limit of the second-order directional derivative can also be similarly obtained. Note that  $D^2L_n(\boldsymbol{\theta}_*; \cdot)$  is Lipschitz continuous on  $\Delta(\boldsymbol{\theta}_*)$ , so that Assumption 5(iii) holds, and we can apply the law of large numbers (LLN):

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

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

This implies that

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

and this is identical to  $-B(\mathbf{d}, \mathbf{d})$ . Thus,  $2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \sup_{\mathbf{d} \in \Delta(\boldsymbol{\theta}_*)} [0, \mathcal{Y}(\mathbf{d})]^2$  by Theorem 1(iii), where

$$\mathcal{Y}(\mathbf{d}) := \frac{\boldsymbol{\delta}(\mathbf{d})' \boldsymbol{\Omega}_*^{1/2} \mathbf{W}}{\{\boldsymbol{\delta}(\mathbf{d})' \boldsymbol{\Omega}_* \boldsymbol{\delta}(\mathbf{d})\}^{1/2}},$$

and for each  $\mathbf{d}$  and  $\tilde{\mathbf{d}}$ ,

$$E[\mathcal{Y}(\mathbf{d})\mathcal{Y}(\tilde{\mathbf{d}})] = \frac{\boldsymbol{\delta}(\mathbf{d})' \boldsymbol{\Omega}_* \boldsymbol{\delta}(\tilde{\mathbf{d}})}{\{\boldsymbol{\delta}(\mathbf{d})' \boldsymbol{\Omega}_* \boldsymbol{\delta}(\mathbf{d})\}^{1/2} \{\boldsymbol{\delta}(\tilde{\mathbf{d}})' \boldsymbol{\Omega}_* \boldsymbol{\delta}(\tilde{\mathbf{d}})\}^{1/2}}.$$

This result shows that the directional limit of the likelihood is well defined under the null, and the null limit distribution can be obtained using this, although the log-likelihood is not properly identified under the null.

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

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

Then, for each  $\mathbf{d}$  and  $\tilde{\mathbf{d}}$ ,

$$\mathbf{B}_*(\mathbf{d}, \tilde{\mathbf{d}}) = \begin{bmatrix} \mathbf{B}_*^{(\boldsymbol{\pi}, \boldsymbol{\pi})}(\mathbf{d}_\pi, \tilde{\mathbf{d}}_\pi) & \mathbf{0}' \\ \mathbf{0} & \frac{2}{\tau_*^2} d_\tau' \tilde{d}_\tau \end{bmatrix},$$

and

$$\mathbf{B}_*^{(\boldsymbol{\pi}, \boldsymbol{\pi})}(\mathbf{d}_\pi, \tilde{\mathbf{d}}_\pi) = \frac{1}{\tau_*^2} \begin{bmatrix} \mathbf{d}_\beta' E[\mathbf{X}_t \mathbf{X}_t'] \tilde{\mathbf{d}}_\beta & \mathbf{d}_\beta' E[\mathbf{X}_t'] m(d_\mu, d_\sigma) \\ m(\tilde{d}_\mu, \tilde{d}_\sigma) E[\mathbf{X}_t] \tilde{\mathbf{d}}_\beta & m(d_\mu, d_\sigma) m(\tilde{d}_\mu, \tilde{d}_\sigma) \end{bmatrix}.$$

By the information matrix equality, for each  $\mathbf{d}$ ,  $\mathbf{B}_*(\mathbf{d})$  is identical to  $-\mathbf{A}_*(\mathbf{d})$ .

The null limit distributions of the test statistics are identified by the theorems in Cho and White (2016). First, we apply the QLR test. Applying Theorem 2 shows that

$$\mathcal{LR}_n^{(1)} := 2\{L_n(\widehat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \sup_{\mathbf{s}_\pi \in \Delta(\boldsymbol{\pi}_*)} \max[0, \mathcal{Y}^{(\boldsymbol{\pi})}(\mathbf{s}_\pi)]^2 + \mathcal{H}_2,$$

where for each  $\mathbf{s}_\pi \in \Delta(\boldsymbol{\pi}_*) := \{(\mathbf{s}_\beta', s_\mu, s_\sigma)' \in \mathbb{R}^{k+2} : \mathbf{s}_\beta' \mathbf{s}_\beta + s_\mu^2 + s_\sigma^2 = 1, s_\mu > 0, \text{ and } s_\sigma > 0\}$ ,

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

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

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

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

$$\tilde{\mathcal{Y}}^{(v)}(\mathbf{s}_v) := (\tilde{B}_*^{(v,v)}(\mathbf{s}_v))^{-1/2} \tilde{\mathcal{Z}}^{(v)}(\mathbf{s}_v),$$

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

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

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

where for each  $\mathbf{s}_\beta \in \Delta(\boldsymbol{\beta}_*) := \{\mathbf{s}_\beta \in \mathbb{R}^k : \mathbf{s}'_\beta \mathbf{s}_\beta = 1\}$ ,  $\mathcal{Y}^{(\beta)}(\mathbf{s}_\beta) := \{\mathbf{s}'_\beta E[\mathbf{X}_t \mathbf{X}'_t] \mathbf{s}_\beta\}^{-1/2} \mathbf{Z}^{(\beta)'} \mathbf{s}_\beta$ , and applying Theorem 2(iii) implies that  $\mathcal{LR}_n^{(2)} := 2\{L_n(\ddot{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} \Rightarrow \mathbf{Z}^{(\beta)'} E[\mathbf{X}_t \mathbf{X}'_t]^{-1} \mathbf{Z}^{(\beta)} + \mathcal{H}_2$ .

Therefore, Theorem 2(iv) now yields that

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

under  $H_0$ , where  $Z := \{1 - E[\mathbf{X}_t]' E[\mathbf{X}_t \mathbf{X}'_t]^{-1} E[\mathbf{X}_t]\}^{-1/2} \{Z^{(v)} - E[\mathbf{X}_t]' E[\mathbf{X}_t \mathbf{X}'_t]^{-1} \mathbf{Z}^{(\beta)}\} \sim N(0, 1)$ .

If we let  $r(x) := \phi(x)/[x\Phi(x)]$ ,

$$\frac{m(s_\mu, s_\sigma)}{|m(s_\mu, s_\sigma)|} = -\frac{s_\mu}{|s_\mu|} \left( \frac{1 + r(s_\mu/|s_\sigma|)}{|1 + r(s_\mu/|s_\sigma|)|} \right),$$

which is  $-1$  uniformly on  $\Delta(\mathbf{v}_0)$ . Thus, the null limit distribution reduces to  $\max[0, -Z]^2$ , and this implies that  $\mathcal{LR}_n \overset{A}{\rightsquigarrow} \max[0, -Z]^2$  under  $H_0$ .

We conduct simulations to verify this. We let  $(\mathbf{X}'_t, U_t)' \sim \text{IID } N(\mathbf{0}_2, \mathbf{I}_2)$  and obtain the null limit distribution of the QLR test statistic by repeating the same independent experiments 2,000 times for  $n = 50, 100$ , and  $200$ . Simulation results are summarized in Figure 1. Note that the null limit distributions of the QLR test statistics exactly overlap with that of  $\max[0, -Z]^2$ .

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

Second, we apply the Wald test. For this, if we let

$$\widehat{W}_n(s_\mu, s_\sigma) := m(s_\mu, s_\sigma)^2 \left\{ 1 - n^{-1} \sum_{t=1}^n \mathbf{X}_t' (n^{-1} \sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t')^{-1} n^{-1} \sum_{t=1}^n \mathbf{X}_t \right\},$$

the LLN implies that  $\sup_{s_\mu, s_\sigma} |\widehat{W}_n(s_\mu, s_\sigma) - \widetilde{B}_*^{(v,v)}(s_\mu, s_\sigma)| \rightarrow 0$  a.s.- $\mathbb{P}$ . In particular,  $m(\cdot, \cdot)^2$  is bounded by 1 and  $2/\pi$  from above and below, respectively. Using  $\widehat{W}_n(s_\mu, s_\sigma)$ , we let the Wald test statistic be defined as

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

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

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

Theorem 3 now implies that  $\mathcal{W}_n \Rightarrow \sup_{\mathbf{s}_v \in \Delta(\mathbf{v}_0)} \max[0, \widetilde{\mathcal{Y}}^{(v)}(\mathbf{s}_v)]^2$ , and this is the weak limit identical to that of the QLR test statistic. Thus,  $\mathcal{W}_n \overset{A}{\rightsquigarrow} \max[0, -Z]^2$  under  $H_0$ .

Finally, we investigate the LM test statistic. We let the LM test statistic be defined as

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

where  $\ddot{\theta}_n = (\ddot{\beta}_n, 0, 0, \ddot{\tau}_n)$  with  $\ddot{\beta}_n = (\sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t')^{-1} \sum_{t=1}^n \mathbf{X}_t Y_t$ ,  $\ddot{\tau}_n = (n^{-1} \sum_{t=1}^n \ddot{U}_t^2)^{1/2}$ ,  $\ddot{U}_t := Y_t - \mathbf{X}_t' \ddot{\beta}_n$ ,  $\Delta(\ddot{\beta}_n) := \{\mathbf{x} \in \mathbb{R}^k : \mathbf{x}' \mathbf{x} = 1\}$ ,  $DL_n(\ddot{\theta}_n; s_\mu, s_\sigma) = \{m(s_\mu, s_\sigma)/\ddot{\tau}_n^2\} \sum_{t=1}^n \ddot{U}_t$ , and

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



In particular, applying the LLN implies that for each  $(s_\mu, s_\sigma)$ ,

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

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

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

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

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

by optimizing the objective function with respect to  $\mathbf{s}_\beta$ , so that

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

under  $H_0$ . Therefore,  $\mathcal{LM}_n \overset{A}{\sim} \max[0, -Z]^2$  by noting that

$$\frac{m(\cdot, \cdot)}{|m(\cdot, \cdot)|} = -1$$

on  $\Delta(\mathbf{v}_0)$  and  $n^{-1/2} \sum_{t=1}^n \ddot{U}_t \sim N[0, \tau_*^2 (1 - E[\mathbf{X}_t]' E[\mathbf{X}_t \mathbf{X}'_t]^{-1} E[\mathbf{X}_t])]$ . This is exactly what Theorem 4 asserts.

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

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

where  $\Delta(\mathbf{v}_0)' := \{(s_\mu, s_\sigma) \in \mathbb{R}^2 : s_\mu^2 + s_\sigma^2 = 1 \text{ and } s_\sigma > 0\}$ . Furthermore, it analytically follows that  $m(s_\mu, s_\sigma)/|m(s_\mu, s_\sigma)| = -1$  uniformly on  $\Delta(\mathbf{v}_0)'$ , so that  $\mathcal{LR}_n \Rightarrow \max[0, -Z]^2$ , which is the same as for

the previous case. Nevertheless, our Monte Carlo experiments assuming the same condition showed that the empirical distribution of  $\mathcal{LR}_n$  exactly overlaps with that of  $Z^2$  under the null.

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

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

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

## 2.2 Example 2: Box-Cox's (1964) Transformation

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

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

where  $\{(Y_t, X_t, \mathbf{Z}_t') \in \mathbb{R}^{2+k} : t = 1, 2, \dots\}$  is assumed to be IID,  $X_t$  is strictly greater than zero almost surely, and  $U_t := Y_t - E[Y_t | \mathbf{Z}_t, X_t]$ . Furthermore,  $\boldsymbol{\theta} := (\boldsymbol{\theta}_0', \theta_1, \theta_2)' \in \boldsymbol{\Theta}_0 \times \boldsymbol{\Theta}_{12}$ ,  $\boldsymbol{\Theta}_0$  is a convex and compact set in  $\mathbb{R}^k$ , and

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

Our interests are in testing whether  $X_t$  influences  $E[Y_t | \mathbf{Z}_t, X_t]$  or not by testing that the second term of (1) vanishes.

This model is introduced to avoid Davies's (1977, 1987) identification problem. If the Box-Cox transformation is specified in the conventional way as in Hansen (1996), so that

$$Y_t = \mathbf{Z}_t' \boldsymbol{\theta}_0 + \beta_1 (X_t^\gamma - 1) + U_t$$

is assumed, then  $\gamma_*$  is not identified when  $\beta_{1*} = 0$ , where the subscript ‘\*’ indicates the limit of the nonlinear least squares (NLS) estimator. We may instead examine another null hypothesis:  $\gamma_* = 0$ . Note that letting  $\gamma_* = 0$  also renders  $\beta_{1*}$  be unidentified.

We avoid the identification problem by re-parameterizing the model using  $\theta_1$  and  $\theta_2$  as given in (1). If  $\theta_{2*} = 0$ ,  $\theta_{1*}$  has to be zero by the model condition on  $\Theta_{12}$ , and the identification problem does not arise any longer.

Nevertheless, the re-parameterized model becomes obscure by the null condition:  $\theta_{1*} = 0$  and  $\theta_{2*} = 0$ . If so, the null model is not properly obtained from the model. Note that  $\theta_{1*}(X_t^{\theta_{2*}} - 1)/\theta_{2*} = 0 \times 0/0$ , implying that the standard tests cannot be applied.

On the other hand, the directional limits are well defined, and they can be used to analyze the asymptotic behavior of the quasi-likelihood. For this purpose, we let  $\mathbf{d} = (\mathbf{d}'_0, d_1, d_2)'$  and  $\boldsymbol{\theta}_* = (\boldsymbol{\theta}'_{0*}, 0, 0)'$  with  $\boldsymbol{\theta}_{0*}$  interior to  $\Theta_0$ . The following quasi-likelihood function is obtained from this:

$$L_n(\boldsymbol{\theta}_* + h\mathbf{d}) = -\frac{1}{2} \sum_{t=1}^n \left\{ Y_t - \mathbf{Z}_t'(\boldsymbol{\theta}_{0*} + \mathbf{d}_0 h) - \frac{d_1}{d_2} (X_t^{d_2 h} - 1) \right\}^2,$$

which is now D with respect to  $h$  at 0. Therefore, for each  $\mathbf{d}$ ,  $\lim_{h \downarrow 0} L_n(\boldsymbol{\theta}_* + h\mathbf{d}) = -\frac{1}{2} \sum_{t=1}^n \{Y_t - \mathbf{Z}_t' \boldsymbol{\theta}_{0*}\}^2$ . The directional derivatives are also derived as

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

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

that are linear and quadratic in  $(\mathbf{d}_0, d_2, d_2)$ , respectively. Therefore, the model may be analyzed as if it is D, although the null model is not properly obtained from the model.

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

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

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

a.s.- $\mathbb{P}$ , where  $\ddot{\mathbf{d}} \in \ddot{\Delta}(\boldsymbol{\theta}_*) := \{\mathbf{x} \in \mathbb{R}^{k+1} : \|\mathbf{x}\| = 1\}$ ,  $\mathbf{W}$  is a multivariate normal:

$$\begin{bmatrix} n^{-1/2} \sum U_t \mathbf{Z}_t' \\ n^{-1/2} \sum U_t \log(X_t) \end{bmatrix} \Rightarrow \mathbf{W} := \begin{bmatrix} \mathbf{W}_0' \\ W_1 \end{bmatrix} \sim N(\mathbf{0}, \mathbf{B}_*)$$

with  $\mathbf{B}_*$  being a  $(k+1) \times (k+1)$  positive definite matrix with a finite maximum eigenvalue, and

$$\mathbf{A}_* := \begin{bmatrix} \mathbf{A}_*^{(0,0)} & \mathbf{A}_*^{(0,1)} \\ \mathbf{A}_*^{(1,0)} & \mathbf{A}_*^{(1,1)} \end{bmatrix} := \begin{bmatrix} -E[\mathbf{Z}_t \mathbf{Z}_t'] & -E[\mathbf{Z}_t \log(X_t)] \\ -E[\log(X_t) \mathbf{Z}_t'] & -E[\log(X_t)^2] \end{bmatrix}.$$

Here, we assume  $E[\log(X_t)^2] < \infty$  and for each  $j$ ,  $E[Z_{t,j}^2] < \infty$  to obtain these limits. We also separate the set of directions into  $\ddot{\Delta}(\boldsymbol{\theta}_*)$  and the set for  $d_2$  to derive the asymptotic distribution more efficiently. By this separation, the maximization process can also be separated into a two-step maximization process:

$$\begin{aligned} 2\{L_n(\hat{\boldsymbol{\theta}}_n) - L_n(\boldsymbol{\theta}_*)\} &\Rightarrow \sup_{d_2} \sup_{\ddot{\mathbf{d}} \in \ddot{\Delta}(\boldsymbol{\theta}_*)} \max[0, \mathbf{W}' \ddot{\mathbf{d}}]^2 \{-\ddot{\mathbf{d}}' \mathbf{A}_* \ddot{\mathbf{d}}\}^{-1} \\ &= \sup_{\ddot{\mathbf{d}} \in \ddot{\Delta}(\boldsymbol{\theta}_*)} \max[0, \mathbf{W}' \ddot{\mathbf{d}}]^2 \{-\ddot{\mathbf{d}}' \mathbf{A}_* \ddot{\mathbf{d}}\}^{-1} = \mathbf{W}' (-\mathbf{A}_*)^{-1} \mathbf{W} \end{aligned}$$

by Theorem 1(iii), where  $\hat{\boldsymbol{\theta}}_n$  is the NLS estimator, and applying the proof of Corollary 1(vii) obtains the last equality. Note that maximizing the limit with respect to  $d_2$  is an innocuous process to obtaining the null limit distribution because  $d_2$  vanishes at the limit. We also note that the limit result is the same as what is obtained when an identified model is D.

For the model inference, we let  $\boldsymbol{\pi} = (\boldsymbol{\lambda}', \boldsymbol{v}')'$  such that  $\boldsymbol{\lambda} = \boldsymbol{\theta}_0$  and  $\boldsymbol{v} = \boldsymbol{\theta}_2$ , so that  $\boldsymbol{\Omega} = \boldsymbol{\Theta}_0$ , and  $\mathbf{M}$  is a closed interval with zero as an interior element. Note that  $\theta_{1*} = 0$  if and only if  $\theta_{2*} = 0$  from the model assumption. Using these conditions, Theorem 2(iv) can be applied. That is,

$$\mathcal{LR}_n \Rightarrow \sup_{\mathbf{s}_v \in \Delta(\mathbf{v}_0)} \max[0, \tilde{\mathcal{Y}}^{(\mathbf{v})}(\mathbf{s}_v)]^2,$$

where  $\mathbf{s}_v := s_1$ ,  $\Delta(\mathbf{v}_0) := \{-1, 1\}$ , and

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

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

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

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

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

then

$$\mathcal{W}_n := n \{\tilde{h}_n^{(v)}(\mathbf{s}_v)\} \{\widehat{W}_n\} \{\tilde{h}_n^{(v)}(\mathbf{s}_v)\} \Rightarrow \tilde{\mathcal{Z}}^{(v)} (\tilde{A}_*^{(v,v)})^{-1} \tilde{\mathcal{Z}}^{(v)}$$

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

### 2.3 Example 3: Generalized Method of Moments (GMM)

Hansen (1982) examines an estimation method by generalizing the method of moments estimation that requires differentiability as one of the regularity conditions. We consider the GMM estimator  $\hat{\theta}_n$  obtained by maximizing

$$Q_n(\theta) := \mathbf{g}_n(\mathbf{X}^n; \theta)' \{-\mathbf{M}_n\}^{-1} \mathbf{g}_n(\mathbf{X}^n; \theta)$$

with respect to  $\theta$ , where  $\{\mathbf{X}_t : t = 1, 2, \dots\}$  is a sequence of strictly stationary and ergodic random variables,  $\mathbf{g}_n(\mathbf{X}^n; \theta) := n^{-1} \sum_{t=1}^n \mathbf{q}(\mathbf{X}_t; \theta)$  with  $\mathbf{q}_t := \mathbf{q}(\mathbf{X}_t; \cdot) : \Theta \mapsto \mathbb{R}^k$  being D a.s.- $\mathbb{P}$  on  $\Theta$  given in Assumption 2 ( $r \leq k$ ) such that for each  $\theta \in \Theta$ ,  $\mathbf{q}(\cdot; \theta)$  is measurable, and  $\mathbf{M}_n$  is a symmetric and positive definite random matrix a.s.- $\mathbb{P}$  uniformly in  $n$  that converges to a symmetric and positive definite  $\mathbf{M}_*$  a.s.- $\mathbb{P}$ . Furthermore, for some integrable  $m(\mathbf{X}_t)$ ,  $\|\mathbf{q}_t(\cdot)\|_\infty \leq m(\mathbf{X}_t)$  and  $\|\nabla_\theta \mathbf{q}_t(\cdot)\|_\infty \leq m(\mathbf{X}_t)$ , and there is a unique  $\theta_*$  that maximizes  $E[\mathbf{q}_t(\cdot)]' \{-\mathbf{M}_*\}^{-1} E[\mathbf{q}_t(\cdot)]$  on the interior part of  $\Theta$ . We denote the uniform matrix norm by  $\|\cdot\|_\infty$ . We further suppose that  $n^{1/2} \mathbf{g}_n(\mathbf{X}^n; \theta_*) \Rightarrow \mathbf{W} \sim N(\mathbf{0}, \mathbf{S}_*)$  for some positive definite matrix  $\mathbf{S}_*$ . The GMM estimator is widely applied for empirical data.

The given conditions for  $Q_n(\cdot)$  do not exactly satisfy the conditions in Assumption 2. Even so, our D-D analysis can be easily adapted to the GMM estimation framework. Directional derivatives play a key role as before. We note that the first-order directional derivative of  $\mathbf{g}_n(\cdot) := \mathbf{g}_n(\mathbf{X}^n; \cdot)$  is

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

where  $\nabla_{\theta} \mathbf{g}_n(\mathbf{X}^n; \theta) := [\nabla_{\theta_1} g_{1,n}(\mathbf{X}^n; \theta), \dots, \nabla_{\theta_r} g_{k,n}(\mathbf{X}^n; \theta)]'$  and  $g_{j,n}(\mathbf{X}^n; \theta)$  is the  $j$ -th element of  $\mathbf{g}_n(\mathbf{X}^n; \theta)$ . As (4) makes it clear,  $D\mathbf{g}_n(\theta; \mathbf{d})$  is now linear with respect to  $\mathbf{d}$ . Applying the mean-value theorem implies that for each  $\mathbf{d}$ ,

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

Here,  $\bar{\theta} := [\bar{\theta}_1, \dots, \bar{\theta}_r]$  is the collection of the parameter values between  $\theta$  and  $\theta_*$ , and  $D\mathbf{g}_n(\bar{\theta}; \mathbf{d})$  denotes  $[\nabla_{\theta_1} g_{1,n}(\mathbf{X}^n; \bar{\theta}_1), \dots, \nabla_{\theta_r} g_{k,n}(\mathbf{X}^n; \bar{\theta}_r)]' \mathbf{d}$ . Furthermore,  $DQ_n(\theta; \mathbf{d}) = -2\mathbf{d}' \nabla_{\theta} \mathbf{g}_n(\theta)' \mathbf{M}_n^{-1} \mathbf{g}_n(\theta)$ . This implies that for each  $\mathbf{d}$ ,  $n^{1/2} DQ_n(\theta_*; \mathbf{d}) \Rightarrow -2\mathbf{d}' \mathbf{C}_*' \mathbf{M}_*^{-1} \mathbf{W}$  by the CLT. Here, we applied the LLN to obtain that  $\nabla_{\theta} \mathbf{g}_n(\theta_*)$  converges to  $\mathbf{C}_* := E[\nabla_{\theta} \mathbf{q}_t(\theta_*)]$  a.s.- $\mathbb{P}$  by the fact that  $\|\nabla_{\theta} \mathbf{q}_t(\cdot)\|_{\infty} \leq m(\mathbf{X}_t)$ . We below use these facts and the vehicles for D-D analysis to obtain the asymptotic behavior of the GMM estimator.

Given (4), it is trivial to show that  $\{n^{1/2} DQ_n(\theta_*; \cdot)\}$  is asymptotically tight by the fact that it is linear with respect to  $\mathbf{d}$ . Next, we obtain that for some  $\bar{\theta}$  between  $\theta$  and  $\theta_*$ ,

$$n\{Q_n(\theta) - Q_n(\theta_*)\} = -2\mathbf{d}' \nabla_{\theta} \mathbf{g}_n(\bar{\theta})' \mathbf{M}_n^{-1} \sqrt{n} \mathbf{g}_n(\theta_*) \sqrt{n} h - \mathbf{d}' \nabla_{\theta} \mathbf{g}_n(\bar{\theta})' \mathbf{M}_n^{-1} \nabla_{\theta} \mathbf{g}_n(\bar{\theta}) \mathbf{d} (\sqrt{n} h)^2$$

by substituting  $\mathbf{g}_n$  in (5) into  $Q_n(\cdot)$ , and so

$$\{Q_n(\hat{\theta}_n) - Q_n(\theta_*)\} \Rightarrow \sup_{\mathbf{d}} \sup_h -2\mathbf{d}' \mathbf{C}_*' \mathbf{M}_*^{-1} \mathbf{W} h - \mathbf{d}' \mathbf{C}_*' \mathbf{M}_*^{-1} \mathbf{C}_* \mathbf{d} h^2.$$

We may let  $\mathcal{Z}(\mathbf{d}) := -\mathbf{d}' \mathbf{C}_*' \mathbf{M}_*^{-1} \mathbf{W}$  and  $A_*(\mathbf{d}) := -\mathbf{d}' \mathbf{C}_*' \mathbf{M}_*^{-1} \mathbf{C}_* \mathbf{d}$ . Note that these derivatives are linear and quadratic in  $\mathbf{d}$ , respectively. Therefore,

$$\{Q_n(\hat{\theta}_n) - Q_n(\theta_*)\} \Rightarrow \mathbf{W}' \mathbf{M}_*^{-1} \mathbf{C}_* \{-\mathbf{C}_*' \mathbf{M}_*^{-1} \mathbf{C}_*\}^{-1} \mathbf{C}_*' \mathbf{M}_*^{-1} \mathbf{W}$$

by Corollary 1(vii). Furthermore, by applying Corollary 1(v), we obtain that

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

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

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

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

$$\mathcal{QLR}_n := \left\{ \sup_{\mathbf{v}, \boldsymbol{\lambda}} Q_n(\mathbf{v}, \boldsymbol{\lambda}) - \sup_{\boldsymbol{\lambda}} Q_n(\mathbf{v}_0, \boldsymbol{\lambda}) \right\}$$

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

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

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

$$\mathcal{QW}_n := \sup_{\mathbf{s}_v \in \Delta(\mathbf{v}_0)} n \{ \tilde{h}_n^{(v)}(\mathbf{s}_v) \} \{ \widehat{W}_n(\mathbf{s}_v) \} \{ \tilde{h}_n^{(v)}(\mathbf{s}_v) \},$$

where  $\tilde{h}_n^{(v)}(\mathbf{s}_v)$  is such that for each  $\mathbf{s}_v \in \Delta(\mathbf{v}_0)$ ,

$$Q_n(\mathbf{v}_0 + \tilde{h}_n^{(v)}(\mathbf{s}_v)\mathbf{s}_v, \tilde{\boldsymbol{\lambda}}_n(\mathbf{s}_v)) := \sup_{\{h^{(v)}, \boldsymbol{\lambda}\}} Q_n(\mathbf{v}_0 + h^{(v)}\mathbf{s}_v, \boldsymbol{\lambda}),$$

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

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

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

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

where for each  $(s_v, s_\lambda)$ ,

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

and  $\ddot{\theta}_n := (v_0, \ddot{\lambda}_n)$  such that  $\ddot{\lambda}_n := \arg \max_{\lambda} Q_n(v_0, \lambda)$ . If we let  $\widetilde{W}_n(s_v, s_\lambda) = s_v' \widehat{W}_n s_v$  for each  $(s_v, s_\lambda) \in \Delta(v_0) \times \Delta(\lambda_*)$ ,

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

by Theorem 4, the interiority condition of  $v_0$ , and the proof of Corollary 1(vii), where  $\widehat{W}_n$  is the same weight matrix as used for the  $\mathcal{QW}_n$  test statistic.

Indeed, many nonlinear models share the similar features. For example, table 1 of Cheng, Evans, and Iles (1992) collects a number of nonlinear models with parameter instability problems. Many of them can be analyzed using the approach of the current study. Furthermore, D-D analysis tools simplify dimensional complexities that arise when higher-order approximations are needed for model analysis. Cho, Ishida, and White (2011, 2014) and White and Cho (2012) revisit testing neglected nonlinearity using artificial neural networks, and it requires higher-order model approximations. They resolve the relevant issues by applying the D-D analysis to their models.

### 3 Differentiable Model and Directionally Differentiable Model

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

**Theorem C1.** *If a function  $f : \Theta \mapsto \mathbb{R}$  is (i) D-D on  $\Theta$ ; (ii) for each  $\theta, \theta'$  and for some  $M < \infty$ ,  $|Df(\theta'; d) - Df(\theta; d)| \leq M\|\theta' - \theta\|$  uniformly on  $\Delta(\theta) \cap \Delta(\theta')$ ; and (iii) for each  $\theta \in \Theta$ ,  $Df(\theta; d)$  is linear in  $d \in \Delta(\theta)$ , then  $f : \Theta \mapsto \mathbb{R}$  is D on  $\Theta$ .*  $\square$

**Proof of Theorem C1:** Refer to Troutman (1996, p. 122).  $\blacksquare$

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



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

$$\lim_{\|\tilde{\theta}-\theta_0\|\rightarrow 0} \sup_{\|\theta-\theta_0\|=1} \frac{1}{\|\tilde{\theta}-\theta_0\|} \left| \mathbf{A}(\tilde{\theta})'(\theta-\theta_0) - \mathbf{A}(\theta_0)'(\theta-\theta_0) - (\tilde{\theta}-\theta_0)'\mathbf{B}(\theta_0)(\theta-\theta_0) \right| = 0.$$

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

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

implying that  $Df(\theta_0 + h\tilde{\mathbf{d}}; \mathbf{d}) = Df(\theta_0; \mathbf{d}) + D^2f(\theta_0; \tilde{\mathbf{d}}; \mathbf{d})h\bar{h}$ , where

$$D^2f(\theta_0; \tilde{\mathbf{d}}; \mathbf{d}) := \lim_{h \downarrow 0} \frac{Df(\theta_0 + h\tilde{\mathbf{d}}; \mathbf{d}) - Df(\theta_0; \mathbf{d})}{h}.$$

Given this, note that the given conditions imply that  $Df(\theta_0; \mathbf{d}) = \mathbf{A}(\theta_0)'\mathbf{d}$  and  $D^2f(\theta_0; \tilde{\mathbf{d}}; \mathbf{d}) = \tilde{\mathbf{d}}'\mathbf{B}(\theta_0)\mathbf{d}$ . Therefore, if we let  $\tilde{\theta} := \theta_0 + h\tilde{\mathbf{d}}$ , then  $\mathbf{A}(\tilde{\theta})'\mathbf{d} = \mathbf{A}(\theta_0)'\mathbf{d} + h\tilde{\mathbf{d}}'\mathbf{B}(\theta_0 + \bar{h}\tilde{\mathbf{d}})\mathbf{d}$ , so that

$$\mathbf{A}(\tilde{\theta})'\mathbf{d} - \mathbf{A}(\theta_0)'\mathbf{d} - h\tilde{\mathbf{d}}'\mathbf{B}(\theta_0)\mathbf{d} \leq h\tilde{\mathbf{d}}'\mathbf{B}(\theta_0 + \bar{h}\tilde{\mathbf{d}})\mathbf{d} - h\tilde{\mathbf{d}}'\mathbf{B}(\theta_0)\mathbf{d},$$

implying that

$$\frac{1}{h} |\mathbf{A}(\tilde{\theta})'\mathbf{d} - \mathbf{A}(\theta_0)'\mathbf{d} - h\tilde{\mathbf{d}}'\mathbf{B}(\theta_0)\mathbf{d}| \leq \frac{1}{h} |\tilde{\mathbf{d}}'[\mathbf{B}(\theta_0 + \bar{h}\tilde{\mathbf{d}}) - \mathbf{B}(\theta_0)]\mathbf{d}| \leq M \cdot \|\tilde{\theta} - \theta_0\|,$$

where the last inequality follows from the uniform bound condition. We further note that  $h = \|\tilde{\theta} - \theta_0\|$ . This implies that

$$\lim_{\|\tilde{\theta}-\theta_0\|\rightarrow 0} \frac{1}{\|\tilde{\theta}-\theta_0\|} |\mathbf{A}(\tilde{\theta})'\mathbf{d} - \mathbf{A}(\theta_0)'\mathbf{d} - h\tilde{\mathbf{d}}'\mathbf{B}(\theta_0)\mathbf{d}| \leq \lim_{\|\tilde{\theta}-\theta_0\|\rightarrow 0} M \cdot \|\tilde{\theta} - \theta_0\| = 0.$$

This completes the proof. ■

## References

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

- BOX, G. AND COX, D. (1964): “An Analysis of Transformations,” *Journal of the Royal Statistical Society, Series B*, 26, 211–252.
- CHENG, R., B. EVANS, AND T. ILES. (1992): “Embedded Models in Non-Linear Regression,” *Journal of the Royal Statistical Society, Series B*, 54, 877–888.
- CHO, J. AND WHITE, H. (2016): “Directionally Differentiable Econometric Models,” Discussion Paper, School of Economics, Yonsei University.
- CHO, J., ISHIDA, I., AND WHITE, H. (2011): “Revisiting Tests for Neglected Nonlinearity Using Artificial Neural Networks,” *Neural Computation*, 23, 1133–1186.
- CHO, J., ISHIDA, I., AND WHITE, H. (2014): “Testing for Neglected Nonlinearity Using Twofold Unidentified Models under the Null and Hexic Expansions,” *Essays in Nonlinear Time Series Econometrics*. 3–27. Oxford: Oxford University Press.
- DAVIES, R. (1977): “Hypothesis Testing When a Nuisance Parameter is Present only under the Alternative,” *Biometrika*, 64, 247–254.
- DAVIES, R. (1987): “Hypothesis Testing When a Nuisance Parameter is Present only under the Alternative,” *Biometrika*, 74, 33–43.
- DUOFUR, J.-M. (2006): “Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics,” *Journal of Econometrics*, 133, 443–477.
- DUTTA, S., NARASIMHAN, O., AND RAJIV, S. (1999): “Success in High-Technology Markets: Is Marketing Capability Critical?” *Marketing Science*, 18, 547–598.
- HABIB, M. AND LJUNGQVIST, A. (2005): “Firm Values and Managerial Incentives: A Stochastic Frontier Approach,” *Journal of Business*, 78, 2053–2094.
- HANSEN, L. (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50, 1029–1054.
- HANSEN, B. (1996): “Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis,” *Econometrica*, 64, 413–440.

- MCLEISH, D. (1974): “Dependent Central Limit Theorem and Invariance Principles,” *The Annals of Probability*, 2, 620–628.
- NEWY, W. AND WEST, K. (1987): “Hypothesis Testing with Efficient Method of Moment Estimation,” *International Economic Review*, 28, 777–787.
- STEVENSON, R. (1980): “Likelihood Functions for Generalized Stochastic Frontier Estimation,” *Journal of Econometrics*, 13, 57–66.
- TROUTMAN, J. (1996): *Variational Calculus and Optimal Control*. New York: Springer-Verlag.
- WHITE, H. AND J.S. CHO (2012): “Higher-Order Approximations for Testing Neglected Nonlinearity,” *Neural Computation*, 24, 273–287.
- ZELLNER, A., KMENTA, J., AND DRÈZE, J. (1966): “Specification and Estimation of Cobb-Douglas Production Functions,” *Econometrica*, 34, 784–795.

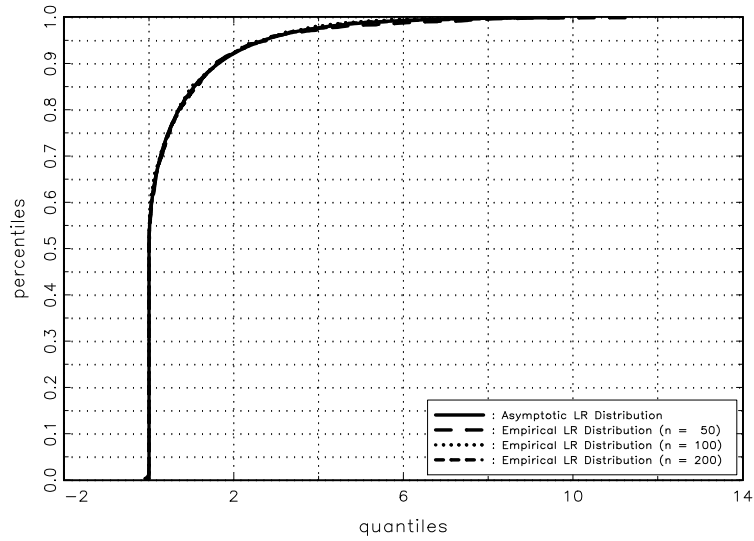


Figure 1: EMPIRICAL AND ASYMPTOTIC DISTRIBUTIONS OF THE QLR TEST STATISTIC. This figure shows the null limit distribution of the QLR test statistic, which is obtained as  $\max[0, -Z]^2$ , and the empirical distributions of the QLR test statistic for various sample sizes:  $n = 50, 100$ , and  $500$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