Testing for Neglected Nonlinearity Using Twofold Unidentified Models under the Null and Hexic Expansions

JIN SEO CHO	ISAO ISHIDA	HALBERT WHITE
School of Economics	CSFI	Department of Economics
Yonsei University	Osaka University	University of California, San Diego
jinseocho@yonsei.ac.kr	i-ishida@sigmath.es.osaka-u.ac.jp	hwhite@weber.ucsd.edu

First version: March 2012. This version: February 2013

Abstract

We revisit the twofold identification problem discussed by Cho, Ishida, and White (*Neural Computation*, 2011), which arises when testing for neglected nonlinearity by artificial neural networks. We do not use the so-called "no-zero" condition and employ a sixth-order expansion to obtain the asymptotic null distribution of the quasi-likelihood ratio (QLR) test. In particular, we avoid restricting the number of explanatory variables in the activation function by using the distance and direction method discussed in Cho and White (*Neural Computation*, 2012). We find that the QLR test statistic can still be used to handle the twofold identification problem appropriately under the set of mild regularity conditions provided here, so that the asymptotic null distribution can be obtained in a manner similar to that in Cho, Ishida, and White (*Neural Computation*, 2011). This also implies that the weighted bootstrap in Hansen (*Econometrica*, 1996) can be successfully exploited when testing the linearity hypothesis using the QLR test.

Key Words: Artificial neural networks; Neglected nonlinearity; Quasi-likelihood ratio test; Distance and direction approach; Gaussian stochastic process; Twofold identification problem; Sixth-order (hexic) expansion; Weighted bootstrap.

JEL Classification: C12, C22, C45, C52.

Acknowledgements: The authors are most grateful to the editors, Niels Haldrup, Mika Meitz, Pentti Saikkonen, and two anonymous referees. We also acknowledge the helpful discussions with Timo Teräsvirta and the participants at the Nonlinear Time Series Analysis Conference held at Aarhus in June 2012. Jin Seo Cho thanks the research support by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-332-B00025), and Isao Ishida thanks the Japan Society for the Promotion of Science (Grants-in-Aid for Scientific Research No. 22243021).

1 Introduction

Testing for neglected nonlinearity is central to testing model specifications. Linear models are the simplest of all functional forms for the conditional mean, and econometric specification popularly assumes linear models. Thus, testing for neglected nonlinearity is the first step toward a deeper understanding of models.

Artificial neural networks (ANNs) are popularly used for testing for neglected nonlinearity because they are quite powerful, owing to their omnibus power. For example, the testing methodology introduced by Bierens (1990) is one of many examples in which ANNs are applied. Moreover, Stinchcombe and White (1998) show that the testing methodology of Bierens (1990) can be further extended by using other non-polynomial analytic functions.

Numerous testing methodologies using ANNs are now available in the literature, and their interrelationships remain unclear. Testing methodologies are traditionally classified into two main groups. The first group, which we call the type I methodology, includes the testing procedures proposed by Bierens (1987, 1990), Bierens and Hartog (1988), Hansen (1996), and Stinchcombe and White (1998), among others. These testing procedures attempt to test whether the coefficient of the activation function is zero. The second group, which we call the type II methodology, includes the testing procedures proposed by Luukkonen, Saikkonen, and Teräsvirta (1988), Teräsvirta, Lin, and Granger (1993), and Teräsvirta (1994), among others. They attempt to test whether the coefficients of predictors in the activation function are zero. Although they provide two different test statistics for testing the linearity, the interrelationships between the two testing methodologies still remain unclear, and therefore, while the two different sets of tests continue to be used widely, they are used independently.

These two methodologies are different mainly because they assume different alternative models, as a result of which the model expansions around the null model are also different. The asymptotic null distribution obtained by the type I methodology is closely associated with a Gaussian stochastic process. This is mainly because of the so-called Davies's (1977, 1987) identification problem, in which the null model cannot be obtained without the identification problem. On the other hand, this problem is trivially resolved for the type II methodology. The standard quadratic approximation (e.g. White 1994) cannot be applied instead, and quartic, hexic, or further higher-order approximations are required. This complicates the model analysis when there is more than a single predictor in the activation function. This is mainly

because standard matrix algebra is too restrictive to represent model expansions for multiple-predictor cases.

Recently, some studies attempt to unify these two different methodologies. Cho, Ishida, and White (2011) (hereafter, CIW) examine a quasi-likelihood ratio (QLR) statistic for the same goal and show that it can be used to test the two different hypotheses using type I and II methodologies simultaneously. In addition, they show that the weighted bootstrap suggested by Hansen (1996) can be successfully exploited when their regularity conditions hold. We call this the type III methodology.

Nevertheless, CIW focus on a particular set of activation functions and obtain the desired results. More specifically, they require that the so-called no-zero condition should hold for the activation functions; otherwise, their theory may not work, and testing the linearity using ANNs may result in a failure. Indeed, many analytic functions such as a logistic cumulative distribution function (CDF) and sine function do not satisfy the no-zero condition. This implies that the type III methodology has to be applied in a restrictive manner.

The main goal of this study is to replace this restriction with another and make the use of ANNs more promising for testing the linearity. Toward this goal, we employ CIW's methodology but modify their regularity conditions. Under these modified regularity conditions, we analyze the QLR test and obtain its asymptotic null distribution.

We achieve our goal in a conservative manner. In this study, our theory handles another subset of activation functions, which do not satisfy the no-zero condition. As detailed below, our analysis requires another condition called "zero-condition" to hold for activation functions. Although the no-zero condition is replaced by the zero-condition, many activation functions including those mentioned above can be handled by our theory.

The remainder of this paper is organized as follows. Section 2 describes the motivations of the current study and introduces the QLR test statistic. Section 3 examines the QLR test as defined in Section 2. In Section 3.1 and 3.2, we respectively examine the asymptotic null distribution using type I and II methodologies, respectively; in Section 3.3, we combine these. In Section 4, we describe Monte Carlo simulations carried out to validate the theoretical results presented in Section 3. Finally, we present the conclusions in Section 5. Mathematical proofs are separately provided. Interested readers can refer to the following URL: `http://web.yonsei.ac.kr/jinseocho/research.htm`.

Note that hereafter, $(\partial/\partial x)f(0)$ is used to indicate $(\partial/\partial x)f(x)|_{x=0}$. We also let "$\xrightarrow{\mathbb{P}}$" and "$\Rightarrow$"

denote convergence in probability and weak convergence as the sample size tends to infinity.

2 ANN Models with Neglected Nonlinearity

We proceed with our discussions by exploiting the regularity conditions in CIW. For this purpose, we first assume the following data generating process (DGP) condition.

Assumption 1 (DGP) $\{(Y_t, \mathbf{X}_t')' \in \mathbb{R}^{1+k}(k \in \mathbb{N}) : t = 1, 2, \cdots\}$ *is a strictly stationary and absolutely regular process defined on the complete probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, *with* $E[|Y_t|] < \infty$ *and mixing coefficient* β_τ *such that for some* $\rho > 1$, $\sum_{\tau=1}^{\infty} \tau^{1/(\rho-1)}\beta_\tau < \infty$. □

Here, Y_t and $\mathbf{X}_t$ are the target and the explanatory variables (or predictors), respectively, such that $\mathbf{X}_t$ does not contain the constant. This condition is often assumed when analyzing stationary time-series data. The mixing coefficient β_τ is as defined in Doukhan (1994). In particular, the mixing coefficient condition is imposed to apply the functional central limit theorem (FCLT) to our statistics defined below.

The main aim of this study is identical to that of CIW's study: we test $E[Y_t|\mathbf{X}_t]$ is a linear function of $\mathbf{Z}_t := (1, \mathbf{X}_t')'$. Formally, we state our interests as follows.

$$\mathcal{H}_0 : \text{ for some } (\alpha, \boldsymbol{\beta}')' \in \mathbb{R}^{1+k},\ E[Y_t|\mathbf{X}_t] = \alpha + \mathbf{X}_t'\boldsymbol{\beta} \text{ with probability 1;}$$

$$\mathcal{H}_1 : \text{ for any } (\alpha, \boldsymbol{\beta}')' \in \mathbb{R}^{1+k},\ E[Y_t|\mathbf{X}_t] = \alpha + \mathbf{X}_t'\boldsymbol{\beta} \text{ with probability less than 1.}$$

These hypotheses and the model provided below have been a popular research topic in the literature; and the most popular approach is thus far testing by parametric specifications. The first and notable study in terms of our focus, here, is by Ramsey (1969), who tests for neglected nonlinearity by adding powers of predicted values to the linear model. When the estimated coefficients of the powers do not converge to zero, this can be used as a testing basis for detecting neglected nonlinearity. Nevertheless, this test is severely influenced by outliers, and other testing procedures are introduced to improve the testing results. Before examining them, we assume the following model.

Assumption 2 (Model) *For a non-polynomial analytic function* $\Psi : \mathbb{R} \mapsto \mathbb{R}$ *such that* $\Psi(0) \neq 0$, *we let*

$$f(\mathbf{X}_t\,; \alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta}) := \alpha + \boldsymbol{X}_t'\boldsymbol{\beta} + \lambda\Phi(\boldsymbol{X}_t'\boldsymbol{\delta}), \tag{1}$$

where $\Phi(\cdot) := \Psi(\cdot) - \Psi(0)$. *We define a model as* $\mathcal{M} := \{f(\,\cdot\,;\alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta}) : (\alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta}) \in \mathbf{A} \times \mathbf{B} \times \boldsymbol{\Lambda} \times \boldsymbol{\Delta}\}$, *where* $\mathbf{A} \subset \mathbb{R}$, $\mathbf{B} \subset \mathbb{R}^k$, $\boldsymbol{\Lambda} \subset \mathbb{R}$, *and* $\boldsymbol{\Delta} \subset \mathbb{R}^k$ *are non-empty compact and convex sets such that* $0 \in \mathrm{int}(\boldsymbol{\Lambda})$ *and* $\mathbf{0} \in \mathrm{int}(\boldsymbol{\Delta})$. □

Note that the final term in the right-hand side (RHS) of eq. (1) plays a similar role to the powers given by Ramsey (1969). The main difference from Ramsey's specification is that it contains unknown parameters in $\Psi(\cdot)$, and this causes the identification problem under the null. More specifically, when we let $(\alpha_*, \boldsymbol{\beta}_*', \lambda_*, \boldsymbol{\delta}_*')'$ be the parameter such that $E[Y_t|\mathbf{X}_*] = f(\mathbf{X}_t; \alpha_*, \boldsymbol{\beta}_*, \lambda_*, \boldsymbol{\delta}_*)$, the null hypothesis can be stated as follows.

$$\mathcal{H}_{01} : \lambda_* = 0 \ \text{ or } \ \mathcal{H}_{02} : \boldsymbol{\delta}_* = \mathbf{0}.$$

The literature can be divided into three subgroups in terms of the methodologies used to handle these hypotheses. The first group handles $\mathcal{H}_{01} : \lambda_* = 0$, and we call this approach the type I methodology. Bierens (1990) lets[1] $\Psi(\cdot) = \exp(\cdot)$ and tests $\mathcal{H}_{01} : \lambda_* = 0$. Bierens (1990) shows that any departure from the linearity can be consistently detected using his test statistic. Nevertheless, this approach suffers from the so-called Davies's (1977, 1987) identification problem. When the coefficient of $\Psi(\cdot)$ is zero, the coefficient $\boldsymbol{\delta}_*$ cannot be identified. This yields a null distribution that is different from the conventional chi-squared distribution even asymptotically. In general, test statistics constructed by the type I methodology weakly converge to a function of a Gaussian stochastic process under the null. Due to this, many studies try to find asymptotic null distributions directly or to introduce alternative testing procedures that do not suffer from these issues. For example, Hansen (1996) provides the weighted bootstrap procedure, and Lee, Granger, and White (1993) introduce another test statistic to apply the conventional chi-squared distribution as its asymptotic null distribution. As another extension of a type I methodology, Stinchcombe and White (1998) show that the omnibus power property suggested by Bierens (1990) is not only exhibited by the exponential function but also by any non-polynomial analytic function. For example, logistic CDF, sine, cosine, tangent, and other analytic functions can also be used for the same goal.

The second group handles $\mathcal{H}_{02} : \boldsymbol{\delta}_* = 0$, and we call this approach the type II methodology. Luukkonen, Saikkonen, and Teräsvirta (1988), Teräsvirta, Lin, and Granger (1993), and

[1] More precisely, Bierens's (1990) model is slightly different from the model in Assumption 2. He transforms the predictors in the activation function into bounded variables by $\tan^{-1}(\cdot)$. Nevertheless, Stinchcombe and White (1998) show that the omnibus power of Bierens's (1990) test is attributable to the exponential function and not the transformation.

Teräsvirta (1994), among others, examine this hypothesis and introduce relevant test statistics. These are mostly constructed using the principle of the Lagrange multiplier (LM) test. By assuming $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \mathbf{0}$, λ_* is not identified as before. Nevertheless, this problem turns out to be trivial because their asymptotic null distributions are not affected by this. One interesting aspect of this, though, is that it needs higher-order approximations. When the score is computed, the first-order derivative is identically zero, so that it cannot be used to define LM tests; then next-order derivatives have to be exploited as an effective score. For example, if Bierens's (1990) exponential function is used as the activation function, the second-order derivatives are relevant scores for defining LM tests. In another way, this also implies that the testing procedure using the LM testing principle can be complicated. When the number of explanatory variables exceeds one, the definition of the LM test needs to be re-considered, and deriving its asymptotic null distribution accordingly becomes complicated.

The third group attempts to combine type I and II methodologies and examine a testing procedure under both $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$. We call this the type III methodology. CIW examine the QLR test statistic defined as

$$QLR_n := n(1 - \hat{\sigma}^2_{n,A}/\hat{\sigma}^2_{n,0}),$$

where $\hat{\sigma}^2_{n,0} := \min_{\alpha,\boldsymbol{\beta}} n^{-1} \sum_{t=1}^n (Y_t - f(\mathbf{X}_t; \alpha, \boldsymbol{\beta}, 0, \boldsymbol{\delta}))^2$ with $\boldsymbol{\delta}$ being a placeholder, and $\hat{\sigma}^2_{n,A} := \min_{\alpha,\boldsymbol{\beta},\lambda,\boldsymbol{\delta}} n^{-1} \sum_{t=1}^n (Y_t - f(\mathbf{X}_t; \alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta}))^2$. As two different identification problems lurk under the null, CIW call this the twofold identification problem. They show that the QLR test has the capability of testing both $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$ simultaneously without necessarily focusing on a particular null hypothesis. More specifically, they obtain the asymptotic distributions of the QLR test under $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$ separately and then combine them to obtain the asymptotic distribution under $\mathcal{H}_0$. In this process, they attempt to acquire the omnibus power property targeted by the type I methodology and the power desired by the type II methodology.

Nevertheless, the analysis given by the type III methodology restricts the scope of models. They assume the "no-zero" condition, which means that the second-order derivative of $\Psi(\cdot)$ at zero is not zero. Many analytic functions such as the exponential function advocated by Bierens (1990) and the cosine function satisfy the no-zero condition. On the other hand, many other functions such as the logistic function advocated by White (1992) and the sine function do not satisfy this condition.

In this study, we eliminate this restriction and examine the asymptotic null distribution of the QLR test. Specifically, we instead assume the following condition.

Assumption 3 (Zero) *When we let* $c_j := (\partial^j/\partial x^j)\Psi(0)$ *and* $j = 0, 1, 2, \ldots$, $c_2 = 0$, $c_3 \neq 0$, *and* c_i *is finite for every* $i \geq 3$. □

CIW assume that $c_2 \neq 0$ and approximate their quasi-likelihood function by a quartic expansion. As Assumption 3 supposes that $c_2 = 0$, the no-zero condition does not hold any longer, and the quartic approximation cannot be applied either. As detailed below, it turns out that a sixth-order (hexic) approximation is effective under Assumption 3.

In achieving our goal, we relax a couple of limitations used in the previous studies. First, we do not restrict the number of variables in $\mathbf{X}_t$ to one. Many previous studies (e.g. White and Cho, 2012) assume that the number of predictors in the activation function to be one so that higher-order approximations can be easily obtained. Second, we do not restrict the parameter space so that the QLR test is equivalent to the LM tests under the null. White and Cho (2012) show that the QLR test is equivalent to the LM test when the space for λ is restricted not to include zero. We intentionally include zero in the space of λ and test both $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$ simultaneously.

Before moving to the next section, there are several relevant remarks. First, we note that the current model assumption is not identical to the model assumption in CIW, in which the activation function is not centered by $\Psi(0)$. In other words, their model is specified as $\alpha + \mathbf{X}_t'\boldsymbol{\beta} + \lambda\Psi(\mathbf{X}_t'\boldsymbol{\delta})$, and this specification leads to another identification problem. The intercept value is identified as $\alpha_* + \lambda_*\Psi(0)$ under $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \mathbf{0}$, but α_* and λ_* are not separately identified. This implies that the distribution of the QLR test has to be obtained in two different ways under $\mathcal{H}_{02}$ and the derivation process can be laborious. On the other hand, many difficulties disappear under our current model assumption. Only λ_* is not identified under $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \mathbf{0}$. Second, the twofold identification problem also exists in a different context. We observe that testing conditional heteroskedasticity using the model introduced in Rosenberg (1973) and King and Shively (1993) also suffers from the twofold identification problem. Third, different approaches with similar motivations are also found in the literature. Cho and Ishida (2012) and Baek and Cho (2012) approach this in a different way. They let $\Phi(\cdot)$ be a power function with unknown power coefficient and test the effect of omitted power transformation and neglected nonlinearity, respectively. Under their model conditions, the conven-

tional second-order expansion is sufficient for model approximations although identification problems have to be handled in more complicated ways. In this sense, their approaches are different from ours. This also implies that model analysis depends upon the functional form of the nonlinear component.

3 Asymptotic Null behavior of the QLR Test

We obtain the asymptotic null distributions of the QLR test by type I and II methodologies separately and identify their interrelationship in the final stage. More specifically, we first assume $\mathcal{H}_{01} : \lambda_* = 0$ and derive the asymptotic null distribution in Section 3.1 by following CIW. As its derivation is similar to CIW, we present only the key to the current study unless confusion would otherwise arise. Next, we assume $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \mathbf{0}$ and obtain the asymptotic null distribution by a further higher-order expansion in Section 3.2. Finally, we combine the separate results and derive the asymptotic null distribution under $\mathcal{H}_0 : \lambda_* = 0$ or $\boldsymbol{\delta}_* = \mathbf{0}$ in Section 3.3.

Here, we define some notations that will be used throughout this paper. For simplicity, we let the quasi-likelihood (QL) function be

$$L_n(\alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta}) := -\sum_{t=1}^{n} (Y_t - \alpha - \mathbf{X}_t'\boldsymbol{\beta} - \lambda\Phi_t(\boldsymbol{\delta}))^2,$$

where $\Phi_t(\boldsymbol{\delta}) := \Phi(\mathbf{X}_t'\boldsymbol{\delta})$. Thus, it also holds that $\hat{\sigma}_{n,0}^2 \equiv \max_{\alpha,\boldsymbol{\beta}} -n^{-1}L_n(\alpha, \boldsymbol{\beta}, 0, \boldsymbol{\delta})$ and $\hat{\sigma}_{n,A}^2 \equiv \max_{\alpha,\boldsymbol{\beta},\lambda,\boldsymbol{\delta}} -n^{-1}L_n(\alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta})$.

3.1 Asymptotic Null Distribution under $\mathcal{H}_{01} : \lambda_* = 0$

We now consider the asymptotic distribution of the QLR test under $\mathcal{H}_{01} : \lambda_* = 0$. As $\boldsymbol{\delta}_*$ is not identified, we first concentrate the QL with respect to $(\alpha, \boldsymbol{\beta}, \lambda)$ and operate the unidentified $\boldsymbol{\delta}_*$ in the last. In other words, we let

$$QLR_n^{(1)} := \left\{ n - \frac{1}{\hat{\sigma}_{n,0}^2} \min_{\boldsymbol{\delta}} \min_{\lambda} \min_{\alpha,\boldsymbol{\beta}} \sum_{t=1}^{n} (Y_t - f(\mathbf{X}_t; \alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta}))^2 \right\}$$

and examine its asymptotic behavior under $\mathcal{H}_{01}$. Note that $QLR_n^{(1)}$ is another representation of QLR_n.

We obtain the asymptotic null behavior of $QLR_n^{(1)}$ by following CIW's approach. If we let $L_n^{(1)}(\lambda; \boldsymbol{\delta}) := \max_{\alpha,\boldsymbol{\beta}} L_n(\alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta})$ and note that $L_n^{(1)}(\lambda; \boldsymbol{\delta}) = -[\mathbf{Y} - \lambda\boldsymbol{\Phi}(\boldsymbol{\delta})]'\mathbf{M}[\mathbf{Y} - \lambda\boldsymbol{\Phi}(\boldsymbol{\delta})]$, where

$\mathbf{Y} := [Y_1, \ldots, Y_n]'$, $\mathbf{\Phi}(\boldsymbol{\delta}) := [\Phi(\mathbf{X}_1'\boldsymbol{\delta}), \ldots, \Phi(\mathbf{X}_n'\boldsymbol{\delta})]'$, $\mathbf{M} := \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, and $\mathbf{Z} := [\mathbf{Z}_1, \ldots, \mathbf{Z}_n]'$, it is not difficult to obtain

$$QLR_n^{(1)} \equiv \max_{\boldsymbol{\delta}} \frac{\{\mathbf{\Phi}(\boldsymbol{\delta})'\mathbf{MU}\}^2}{\hat{\sigma}_{n,0}^2 \mathbf{\Phi}(\boldsymbol{\delta})'\mathbf{M}\mathbf{\Phi}(\boldsymbol{\delta})},$$

where $\mathbf{U} := [U_1, \ldots, U_n]'$ and for each t, $U_t = Y_t - E[Y_t|\mathbf{X}_t]$. Furthermore, if we let $\mathbf{\Psi}(\boldsymbol{\delta}) := [\Psi(\mathbf{X}_1'\boldsymbol{\delta}), \ldots, \Psi(\mathbf{X}_n'\boldsymbol{\delta})]'$, it also follows that

$$\text{(2)} \quad QLR_n^{(1)} \equiv \max_{\boldsymbol{\delta}} \frac{\{\mathbf{\Psi}(\boldsymbol{\delta})'\mathbf{MU}\}^2}{\hat{\sigma}_{n,0}^2 \mathbf{\Psi}(\boldsymbol{\delta})'\mathbf{M}\mathbf{\Psi}(\boldsymbol{\delta})}.$$

It is mainly because $\mathbf{\Phi}(\boldsymbol{\delta}) = \mathbf{\Psi}(\boldsymbol{\delta}) - c_0\boldsymbol{\iota}$, so that $\mathbf{\Phi}(\boldsymbol{\delta})'\mathbf{MU} = \mathbf{\Psi}(\boldsymbol{\delta})'\mathbf{MU}$ and $\mathbf{\Phi}(\boldsymbol{\delta})'\mathbf{M}\mathbf{\Phi}(\boldsymbol{\delta}) = \mathbf{\Psi}(\boldsymbol{\delta})'\mathbf{M}\mathbf{\Psi}(\boldsymbol{\delta})$, where $\boldsymbol{\iota}$ is an $n \times 1$ vector of ones. We now note that this form is identical to eq. (5) in CIW, so that we can borrow their results for our derivation. For this purpose, the following assumptions are imposed.

Assumption 4 (Regularity I) *(i) For a sequence of stationary and ergodic random variables $\{M_t\}$, $|U_t| \leq M_t$, for $j = 1, 2, \ldots, k$, $|X_{t,j}| \leq M_t$, and for some $\kappa \geq 2(\rho - 1)$, $E[M_t^{4+2\kappa}] < \infty$;*

(ii) $\sup_{\boldsymbol{\delta} \in \mathbf{\Delta}} |\Psi_t(\boldsymbol{\delta})| \leq M_t$ and for $j = 1, 2, \ldots, k$, $\sup_{\boldsymbol{\delta} \in \mathbf{\Delta}} |(\partial/\partial\delta_j)\Psi_t(\boldsymbol{\delta})| \leq M_t$, where for $t = 1, 2, \ldots, n$, $\Psi_t(\boldsymbol{\delta}) := \Psi(\boldsymbol{X}_t'\boldsymbol{\delta})$;

(iii) $E[U_t|\boldsymbol{X}_t, U_{t-1}, \boldsymbol{X}_{t-1}, \ldots] = 0$; and

(iv) For each $\epsilon > 0$ and $\boldsymbol{\delta} \in \mathbf{\Delta}(\epsilon)$, $\bar{\boldsymbol{V}}_1(\boldsymbol{\delta})$ and $\bar{\boldsymbol{V}}_2(\boldsymbol{\delta})$ are positive definite, where for given $\epsilon > 0$, $\mathbf{\Delta}(\epsilon) := \{\boldsymbol{\delta} \in \mathbf{\Delta} : \sum_{j=1}^k |\delta_j| \geq \epsilon\}$,

$$\bar{\boldsymbol{V}}_1(\boldsymbol{\delta}) := \begin{bmatrix} E[U_t^2\Psi_t(\boldsymbol{\delta})^2] & E[U_t^2\Psi(\boldsymbol{\delta})\boldsymbol{Z}_t'] \\ E[U_t^2\boldsymbol{Z}_t\Psi(\boldsymbol{\delta})] & E[U_t^2\boldsymbol{Z}_t\boldsymbol{Z}_t'] \end{bmatrix} \quad \textit{and} \quad \bar{\boldsymbol{V}}_2(\boldsymbol{\delta}) := \begin{bmatrix} E[\Psi_t(\boldsymbol{\delta})^2] & E[\Psi(\boldsymbol{\delta})\boldsymbol{Z}_t'] \\ E[\boldsymbol{Z}_t\Psi(\boldsymbol{\delta})] & E[\boldsymbol{Z}_t\boldsymbol{Z}_t'] \end{bmatrix}. \ \square$$

These assumptions are sufficient for deriving the asymptotic null distribution. The motivations of these conditions are already given in CIW. In particular, Assumption 4(*i*) is imposed to apply the FCLT of Doukhan, Massart, and Rio (1995). With this condition holding along with the mixing coefficient condition in Assumption 1, it is not difficult to apply the FCLT. We also modify our parameter space condition into $\mathbf{\Delta}(\epsilon)$. As noted by CIW, if $\boldsymbol{\delta} = \mathbf{0}$, $\mathbf{\Phi}(\mathbf{0}) = \mathbf{0}$, implying that $QLR_n^{(1)}$ is not appropriately defined. We avoid this by removing $\mathbf{0}$ from $\mathbf{\Delta}$, and this is given in the form of $\mathbf{\Delta}(\epsilon)$. We also let

$$QLR_n^{(1)}(\epsilon) := \max_{\boldsymbol{\delta} \in \mathbf{\Delta}(\epsilon)} \frac{\{\mathbf{\Psi}(\boldsymbol{\delta})'\mathbf{MU}\}^2}{\hat{\sigma}_{n,0}^2 \mathbf{\Psi}(\boldsymbol{\delta})'\mathbf{M}\mathbf{\Psi}(\boldsymbol{\delta})}$$

to accommodate the influence of this modification. Given this, the following result is derived based on theorem 1 of CIW.

Theorem 1 (CIW (Theorem 1)) *Given Assumptions 1, 2, 4, and $\mathcal{H}_{01}$, for each $\epsilon > 0$,*

$$QLR_n^{(1)}(\epsilon) \Rightarrow \sup_{\boldsymbol{\delta} \in \boldsymbol{\Delta}(\epsilon)} \bar{\mathcal{G}}_0(\boldsymbol{\delta})^2,$$

where $\bar{\mathcal{G}}_0(\cdot)$ is a mean-zero Gaussian stochastic process such that for each $\boldsymbol{\delta}$ and $\widetilde{\boldsymbol{\delta}} \in \boldsymbol{\Delta}(\epsilon)$,

$$\bar{\rho}(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) := E[\bar{\mathcal{G}}_0(\boldsymbol{\delta})\bar{\mathcal{G}}_0(\widetilde{\boldsymbol{\delta}})] = \frac{\mathcal{T}_1(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}})}{\{\sigma_*^2 \mathcal{J}_1(\boldsymbol{\delta}, \boldsymbol{\delta})\}^{1/2}\{\sigma_*^2 \mathcal{J}_1(\widetilde{\boldsymbol{\delta}}, \widetilde{\boldsymbol{\delta}})\}^{1/2}}$$

with $\mathcal{T}_1(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) := E[U_t^2 \Psi_t^*(\boldsymbol{\delta})\Psi_t^*(\widetilde{\boldsymbol{\delta}})]$ and $\mathcal{J}_1(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) := E[\Psi_t^*(\boldsymbol{\delta})\Psi_t^*(\widetilde{\boldsymbol{\delta}})]$, where $\sigma_*^2 := E[U_t^2]$ and for each $\boldsymbol{\delta} \in \boldsymbol{\Delta}(\epsilon)$ and $t = 1, 2, \ldots, n$, $\Psi_t^*(\boldsymbol{\delta}) := \Psi_t(\boldsymbol{\delta}) - E[\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t']E[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbf{Z}_t$. □

The proof of Theorem 1 is already given in CIW and is not restated here.

3.2 Asymptotic Null Distribution under $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \mathbf{0}$

We now examine the asymptotic distribution of the QLR test under $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \mathbf{0}$. This lets us first maximize the QL with respect to $(\alpha, \boldsymbol{\beta}, \boldsymbol{\delta})$ and finally with respect to λ. In other words, our QLR test is now analyzed by examining

$$QLR_n^{(2)} := \left\{ n - \min_{\lambda} \min_{\boldsymbol{\delta}} \min_{\alpha, \boldsymbol{\beta}} \frac{1}{\hat{\sigma}_{n,0}^2} \sum_{t=1}^{n} (Y_t - f(\mathbf{X}_t; \alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta}))^2 \right\}.$$

We note that $QLR_n^{(2)}$ is another representation of QLR_n. We simply minimize QL with respect to λ in the last as it is unidentified under $\mathcal{H}_{02}$.

Our analysis of $QLR_n^{(2)}$ here is quite different from that in CIW because we replace their no-zero condition with Assumption 3.

We manage these different results in the following manner. We first concentrate the QL by maximizing it with respect to $(\alpha, \boldsymbol{\beta})$. The concentrated QL is now obtained as $L_n^{(2)}(\boldsymbol{\delta}; \lambda) := -[\mathbf{Y} - \lambda\boldsymbol{\Phi}(\boldsymbol{\delta})]'\mathbf{M}[\mathbf{Y} - \lambda\boldsymbol{\Phi}(\boldsymbol{\delta})]$, and we next approximate the concentrated QL using Taylor's expansion with respect to $\boldsymbol{\delta}$. As pointed out by CIW, the first-order derivative of this does not play any role, and this is the same for our case, too. We thus approximate it using a higher-order expansion. Nevertheless, a further higher-order approximation is needed under Assumption 3. It is mainly because the second-order derivative is identical to zero as detailed below. It turns out that a sixth-order expansion is appropriate as in White and Cho (2012).

A higher-order approximation can be complicated because of the laborious differentiations involved. In particular, when the dimension of $\boldsymbol{\delta}$ exceeds one, it becomes more complicated.[2]

[2] White and Cho (2012) avoid this complication by letting the dimension of $\boldsymbol{\delta}$ be one.

In this study, we avoid this by exploiting the distance and direction method in CIW, Cho and White (2012), and Cho (2012). Specifically, for any two elements $\boldsymbol{\delta}_*$ and $\boldsymbol{\delta}$, there is $(h, \boldsymbol{d})$ such that

$$\boldsymbol{\delta} \equiv \boldsymbol{\delta}_* + h\boldsymbol{d}, \tag{3}$$

where $h \geq 0$ and $\boldsymbol{d} \in \mathbb{S}^{k-1} := \{\mathbf{x} \in \mathbb{R}^k : \mathbf{x}'\mathbf{x} = 1\}$. If $\boldsymbol{\delta}$ and $\boldsymbol{\delta}_*$ are identical, it is mainly because $h = 0$ uniformly in $\boldsymbol{d}$. Furthermore, for a particular direction, the distance between $\boldsymbol{\delta}$ and $\boldsymbol{\delta}_*$ is captured by h, a single number, so that for each $\boldsymbol{d}$, we can apply Taylor's expansion with respect to h and treat $\boldsymbol{d}$ as a nuisance parameter. Thus, for given $\boldsymbol{\delta}$, we now let $(h, \boldsymbol{d}) \in \mathbb{R}^+ \times \mathbb{S}^{k-1}$ capture the identity, $L_n^{(2)}(\boldsymbol{\delta}; \lambda) \equiv L_n^{(2)}(h\boldsymbol{d}; \lambda)$. Here, $\boldsymbol{\delta}_* = \mathbf{0}$ under $\mathcal{H}_{02}$. For a particular direction $\boldsymbol{d}$, applying Taylor's sixth-order expansion with respect to h yields the following expansion.

$$\begin{aligned} L_n^{(2)}(h\boldsymbol{d}; \lambda) =& L_n^{(2)}(\mathbf{0}; \lambda) + \frac{\partial}{\partial h} L_n^{(2)}(\mathbf{0}; \lambda) h + \frac{1}{2!} \frac{\partial^2}{\partial h^2} L_n^{(2)}(\mathbf{0}; \lambda) h^2 + \frac{1}{3!} \frac{\partial^3}{\partial h^3} L_n^{(2)}(\mathbf{0}; \lambda) h^3 \\ &+ \frac{1}{4!} \frac{\partial^4}{\partial h^4} L_n^{(2)}(\mathbf{0}; \lambda) h^4 + \frac{1}{5!} \frac{\partial^5}{\partial h^5} L_n^{(2)}(\mathbf{0}; \lambda) h^5 + \frac{1}{6!} \frac{\partial^6}{\partial h^6} L_n^{(2)}(\mathbf{0}; \lambda) h^6 + o_{\mathbb{P}}(h^6). \end{aligned} \tag{4}$$

We contain the partial derivatives constituting the RHS in the following lemma.

Lemma 1 *Given Assumptions 2 and 3, the following holds under $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \mathbf{0}$: for each $\boldsymbol{d} \in \mathbb{S}^{k-1}$,*

(i) $(\partial/\partial h) L_n(h\boldsymbol{d}; \lambda)|_{h=0} = c_1 \lambda \boldsymbol{\iota}' \boldsymbol{D}_1(\boldsymbol{d}) \boldsymbol{M}\boldsymbol{U} = 0$*;*

(ii) $(\partial^2/\partial h^2) L_n(h\boldsymbol{d}; \lambda)|_{h=0} = 2\lambda c_2 \boldsymbol{\iota}' \boldsymbol{D}_2(\boldsymbol{d}) \boldsymbol{M}\boldsymbol{U} - 2\lambda^2 c_1^2 \boldsymbol{\iota}' \boldsymbol{D}_1(\boldsymbol{d}) \boldsymbol{M}\boldsymbol{D}_1(\boldsymbol{d}) \boldsymbol{\iota} = 0$*;*

(iii) $(\partial^3/\partial h^3) L_n(h\boldsymbol{d}; \lambda)|_{h=0} = 2\lambda c_3 \boldsymbol{\iota}' \boldsymbol{D}_3(\boldsymbol{d}) \boldsymbol{M}\boldsymbol{U}$*;*

(iv) $(\partial^4/\partial h^4) L_n(h\boldsymbol{d}; \lambda)|_{h=0} = 2\lambda c_4 \boldsymbol{\iota}' \boldsymbol{D}_4(\boldsymbol{d}) \boldsymbol{M}\boldsymbol{U}$*;*

(v) $(\partial^5/\partial h^5) L_n(h\boldsymbol{d}; \lambda)|_{h=0} = 2\lambda c_5 \boldsymbol{\iota}' \boldsymbol{D}_5(\boldsymbol{d}) \boldsymbol{M}\boldsymbol{U}$*; and*

(vi) $(\partial^6/\partial h^6) L_n(h\boldsymbol{d}; \lambda)|_{h=0} = 2\lambda c_6 \boldsymbol{\iota}' \boldsymbol{D}_6(\boldsymbol{d}) \boldsymbol{M}\boldsymbol{U} - 20\lambda^2 c_3^2 \boldsymbol{\iota}' \boldsymbol{D}_3(\boldsymbol{d}) \boldsymbol{M}\boldsymbol{D}_3(\boldsymbol{d}) \boldsymbol{\iota}$*, where for each $\boldsymbol{d} \in \mathbb{S}^{k-1}$ and $m = 1, 2, \ldots$, we let*

$$\boldsymbol{D}_m(\boldsymbol{d}) := \begin{bmatrix} (\boldsymbol{X}_1'\boldsymbol{d})^m & 0 & \cdots & 0 \\ 0 & (\boldsymbol{X}_2'\boldsymbol{d})^m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\boldsymbol{X}_n'\boldsymbol{d})^m \end{bmatrix},$$

and $\boldsymbol{D}_0(\boldsymbol{d}) := \boldsymbol{I}$. □

The partial derivatives in Lemma 1 are derived relatively easily because the dimension of h is one. Obtaining the partial derivatives can also be simplified by the fact that $\mathbf{MD}_1(\boldsymbol{d})\boldsymbol{\iota} = \mathbf{MX}\boldsymbol{d} = \mathbf{0}$. The proof of Lemma 1 is elementary.

Lemma 1 explains the necessity of higher-order expansions. We note that the second-order derivative is identical to zero under $\mathcal{H}_{02}$, and therefore, we cannot apply the central limit theorem (CLT) to this. On the other hand, we can apply CLT to the next-order derivative. This explains why a further higher-order expansion is needed for the QLR test. This result is different from that in CIW. Under their no-zero condition, the second-order derivative is obtained as $(\partial^2/\partial h^2)L_n(h\boldsymbol{d};\lambda)|_{h=0} = 2\lambda c_2\boldsymbol{\iota}'\mathbf{D}_2(\boldsymbol{d})\mathbf{MU}$, and we can apply the CLT to this, leading to a quartic expansion as a relevant order of expansion.

A sixth-order expansion is relevant to our model, and the asymptotic behavior of this expansion is determined by the terms constituting the partial derivatives. Before examining them, we first provide the following regularity conditions.

Assumption 5 (Regularity II) *(i)* $E[|U_t|^4] < \infty$ *and for* $j = 1, 2, \ldots, k$, $E[|X_{t,j}|^{32}] < \infty$*; or* $E[|U_t|^8] < \infty$ *and for* $j = 1, 2, \ldots, k$, $E[|X_{t,j}|^{16}] < \infty$*;*

(ii) $E[U_t|\boldsymbol{X}_t, U_{t-1}, \boldsymbol{X}_{t-1}, \ldots] = 0$*; and*

(iii) For each $j = 1, 2, \ldots, k$*, we let* $\boldsymbol{X}_{t(j:k)} := [X_{t,j}, X_{t,j+1}, \ldots, X_{t,k}]'$ *and also let*

$$\boldsymbol{C}_t := [X_{t,1}\text{vech}[\boldsymbol{X}_{t(1:k)}\boldsymbol{X}'_{t(1:k)}]', X_{t,2}\text{vech}[\boldsymbol{X}_{t(2:k)}\boldsymbol{X}'_{t(1:k-1)}]', \ldots, X_{t,k}\text{vech}[\boldsymbol{X}_{t(k:k)}\boldsymbol{X}'_{t(1:1)}]']'.$$

Given this, $\widetilde{\boldsymbol{V}}_1$ *and* $\widetilde{\boldsymbol{V}}_2$ *are positive definite, where*

$$\widetilde{\boldsymbol{V}}_1 := \begin{bmatrix} E[U_t^2\boldsymbol{C}_t\boldsymbol{C}'_t] & E[U_t^2\boldsymbol{C}_t\boldsymbol{Z}'_t] \\ E[U_t^2\boldsymbol{Z}_t\boldsymbol{C}'_t] & E[U_t^2\boldsymbol{Z}_t\boldsymbol{Z}'_t] \end{bmatrix} \quad \textit{and} \quad \widetilde{\boldsymbol{V}}_2 := \begin{bmatrix} E[\boldsymbol{C}_t\boldsymbol{C}'_t] & E[\boldsymbol{C}_t\boldsymbol{Z}'_t] \\ E[\boldsymbol{Z}_t\boldsymbol{C}'_t] & E[\boldsymbol{Z}_t\boldsymbol{Z}'_t] \end{bmatrix}. \qquad \square$$

Assumption 5 is provided for regular asymptotic behaviors of relevant statistics. Specifically, Assumption 5(*i*) provides higher-order finite moment conditions. They are necessary when applying the CLT and the ergodic theorem (e.g. White 2001). Assumption 5(*ii*) is the same MDA condition as that given by Assumption 4(*iii*). We repeat this condition to avoid referring to Assumption 4. Assumption 5(*iii*) is required to obtain a non-degenerate null distribution, and it corresponds to the absence of perfect multicolinearity in a conventional linear model. When a quartic approximation applies, CIW provide a similar condition using $\text{vech}[\mathbf{X}'_t\mathbf{X}_t]$. We extend this to the vector $\mathbf{C}_t$, which extends the notion of $\text{vech}[\mathbf{X}'_t\mathbf{X}_t]$ to the half-vectorization of the three-dimensional Cartesian product of $\mathbf{X}_t$, $\mathbf{X}_t \otimes \mathbf{X}_t \otimes \mathbf{X}_t$. This implies that Assumption

5(*iii*) does not hold if one of the elements in $\mathbf{X}_t$ is obtained by squaring another element or by multiplying two different elements in $\mathbf{X}_t$.

If the asymptotic behavior of each derivative is available, we can combine them to obtain the asymptotic null distribution of the test. We first contain the asymptotic behaviors in the following lemma.

Lemma 2 *Given Assumptions 1, 2, 3, 5, and* $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \boldsymbol{0}$*,*

(i) $\{n^{-1/2}\boldsymbol{\iota}'\boldsymbol{D}_3(\cdot)\boldsymbol{MU}, n^{-1}\boldsymbol{\iota}'\boldsymbol{D}_3(\cdot)\boldsymbol{MD}_3(\cdot)\boldsymbol{\iota}\} \Rightarrow \{\mathcal{G}_2(\cdot), \mathcal{J}_2(\cdot,\cdot)\}$*, where* $\mathcal{G}_2(\cdot)$ *is a mean-zero Gaussian stochastic process such that for each* $\boldsymbol{d}$ *and* $\widetilde{\boldsymbol{d}} \in \mathbb{S}^{k-1}$*,*

$$E[\mathcal{G}_2(\boldsymbol{d})\mathcal{G}_2(\widetilde{\boldsymbol{d}})] = \mathcal{T}_2(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) := \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{\ell=1}^{k}\sum_{i'=1}^{k}\sum_{j'=1}^{k}\sum_{\ell'=1}^{k} d_i d_j d_\ell \widetilde{d}_{i'}\widetilde{d}_{j'}\widetilde{d}_{\ell'} E[U_t^2 V_{t,ji\ell}^* V_{t,j'i'\ell'}^*],$$

$$\mathcal{J}_2(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) := \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{\ell=1}^{k}\sum_{i'=1}^{k}\sum_{j'=1}^{k}\sum_{\ell'=1}^{k} d_i d_j d_\ell \widetilde{d}_{i'}\widetilde{d}_{j'}\widetilde{d}_{\ell'} E[U_t^2 V_{t,ji\ell}^* V_{t,j'i'\ell'}^*],$$

and for each $i, j, \ell = 1, 2, \ldots, k$*,* $V_{t,ji\ell}^* := X_{t,j}X_{t,i}X_{t,\ell} - E[X_{t,j}X_{t,i}X_{t,\ell}\boldsymbol{Z}_t']E[\boldsymbol{Z}_t\boldsymbol{Z}_t']^{-1}\boldsymbol{Z}_t$*; and*

(ii) $\boldsymbol{\iota}'\boldsymbol{D}_4(\cdot)\boldsymbol{MU} = O_{\mathbb{P}}(n^{1/2})$*,* $\boldsymbol{\iota}'\boldsymbol{D}_5(\cdot)\boldsymbol{MU} = O_{\mathbb{P}}(n^{1/2})$*, and* $\boldsymbol{\iota}'\boldsymbol{D}_6(\cdot)\boldsymbol{MU} = O_{\mathbb{P}}(n^{1/2})$*.* □

We now combine all our derivations obtained so far into eq. (4) and obtain the asymptotic null distribution. We first rewrite eq. (4) as

$$\begin{aligned}(5)\quad L_n^{(2)}(h\boldsymbol{d};\lambda) - L_n^{(2)}(\boldsymbol{0};\lambda) &= \frac{2\lambda c_3}{3!n^{3/6}}\boldsymbol{\iota}'\mathbf{D}_3(\boldsymbol{d})\mathbf{MU}\delta_n^3 + \frac{2\lambda c_4}{4!n^{4/6}}\boldsymbol{\iota}'\mathbf{D}_4(\boldsymbol{d})\mathbf{MU}\delta_n^4 + \frac{2\lambda c_5}{5!n^{5/6}}\boldsymbol{\iota}'\mathbf{D}_5(\boldsymbol{d})\mathbf{MU}\delta_n^5 \\ &+ \frac{2\lambda c_6}{6!n}\boldsymbol{\iota}'\mathbf{D}_6(\boldsymbol{d})\mathbf{MU}\delta_n^6 - \frac{20\lambda^2 c_3^2}{6!n}\boldsymbol{\iota}'\mathbf{D}_3(\boldsymbol{d})\mathbf{MD}_3(\boldsymbol{d})\boldsymbol{\iota}\delta_n^6 + O_{\mathbb{P}}(n^{-1/6}),\end{aligned}$$

where $\delta_n := n^{1/6}h$, and we note that maximizing the LHS with respect to h is asymptotically equivalent to maximizing the RHS with respect to δ_n. We also note that the fourth-, fifth-, and sixth-order derivatives are uniformly negligible in $\boldsymbol{d}$ by Lemma 2(*ii*). That is, $n^{-4/6}\boldsymbol{\iota}'\mathbf{D}_4(\cdot)\mathbf{MU} = o_{\mathbb{P}}(1)$, $n^{-5/6}\boldsymbol{\iota}'\mathbf{D}_5(\cdot)\mathbf{MU} = o_{\mathbb{P}}(1)$, and $n^{-1}\boldsymbol{\iota}'\mathbf{D}_6(\cdot)\mathbf{MU} = o_{\mathbb{P}}(1)$. Therefore, eq. (5) can also be written as

$$(6)\quad L_n^{(2)}(h\boldsymbol{d};\lambda) - L_n^{(2)}(\boldsymbol{0};\lambda) = \frac{2\lambda c_3}{3!n^{1/2}}\boldsymbol{\iota}'\mathbf{D}_3(\boldsymbol{d})\mathbf{MU}\delta_n^3 - \frac{20\lambda^2 c_3^2}{6!n}\boldsymbol{\iota}'\mathbf{D}_3(\boldsymbol{d})\mathbf{MD}_3(\boldsymbol{d})\boldsymbol{\iota}\delta_n^6 + o_{\mathbb{P}}(1).$$

Furthermore, the asymptotic behavior of $\{n^{-1/2}\boldsymbol{\iota}'\mathbf{D}_3(\cdot)\mathbf{MU}, n^{-1}\boldsymbol{\iota}'\mathbf{D}_3(\cdot)\mathbf{MD}_3(\cdot)\boldsymbol{\iota}\}$ is given by Lemma 2(*i*). We now combine all these and obtain the asymptotic null distribution as follows.

Theorem 2 *Given Assumptions 1, 2, 3, 5, and* $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \boldsymbol{0}$, $QLR_n^{(2)} \Rightarrow \sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} \max[0, \widetilde{\mathcal{G}}_0(\boldsymbol{d})]^2$, *where* $\widetilde{\mathcal{G}}_0(\cdot)$ *to be a continuous mean-zero Gaussian stochastic process defined on* $\mathbb{S}^{k-1}$ *such that for each* $\boldsymbol{d}$ *and* $\widetilde{\boldsymbol{d}} \in \mathbb{S}^{k-1}$,

$$E[\widetilde{\mathcal{G}}_0(\boldsymbol{d})\widetilde{\mathcal{G}}_0(\widetilde{\boldsymbol{d}})] = \widetilde{\rho}(\boldsymbol{d}, \widetilde{\boldsymbol{d}}) := \frac{\mathcal{T}_2(\boldsymbol{d}, \widetilde{\boldsymbol{d}})}{\{\sigma_*^2 \mathcal{J}_2(\boldsymbol{d}, \boldsymbol{d})\}^{1/2}\{\sigma_*^2 \mathcal{J}_2(\widetilde{\boldsymbol{d}}, \widetilde{\boldsymbol{d}})\}^{1/2}}. \qquad \square$$

In fact, the given consequence in Theorem 2 comes from the notice that

$$(7) \qquad QLR_n^{(2)} = \sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} \frac{1}{\hat{\sigma}_{n,0}^2} \max\left[0, \frac{\boldsymbol{\iota}'\mathbf{D}_3(\boldsymbol{d})\mathbf{MU}}{\sqrt{\boldsymbol{\iota}'\mathbf{D}_3(\boldsymbol{d})\mathbf{M}\mathbf{D}_3(\boldsymbol{d})\boldsymbol{\iota}}}\right]^2 + o_{\mathbb{P}}(1),$$

and applying Lemma 2(i) to this delivers the desired result. Here, we note that λ is not associated with the asymptotic null distribution. Although λ is not identified under $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \boldsymbol{0}$, its role is asymptotically negligible under $\mathcal{H}_{02}$. This follows from the fact that λ is canceled off while maximizing the RHS of eq. (6) with respect to δ_n. Thus, maximizing the LHS of eq. (6) with respect to λ is asymptotically innocuous. The asymptotic weak limit is also associated with "$\max[0, \cdot]$". By definition, the distance δ_n cannot be less than zero. This lets "$\max[0, \cdot]$" be involved with the weak limit.

Before moving to the next subsection, we note that model approximations using higher-order expansions are also found in the literature. Cho and White (2007, 2010) provide relevant conditions under which higher-order approximations are necessary for obtaining the asymptotic null distribution of their likelihood-ratio tests. They examine fourth-, sixth-, and eighth-order expansions and derive the asymptotic null distribution of the test.

3.3 Asymptotic Null Distribution under $\mathcal{H}_0 : \lambda_* = 0$ or $\boldsymbol{\delta}_* = \boldsymbol{0}$

We now derive the asymptotic null distribution of the QLR test under $\mathcal{H}_0 : \lambda_* = 0$ or $\boldsymbol{\delta}_* = \boldsymbol{0}$. In fact, this derivation requires the results in Theorems 1 and 2 to be combined. We start from the approximation of $QLR_n^{(1)}$ and derive its interrelationship with $QLR_n^{(2)}$. We note that eq. (2) can also be represented using $(h, \boldsymbol{d})$. That is,

$$QLR_n^{(1)} = \sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} \sup_h \frac{\{\boldsymbol{\Psi}(h\boldsymbol{d})'\mathbf{MU}\}^2}{\hat{\sigma}_{n,0}^2 \boldsymbol{\Psi}(h\boldsymbol{d})'\mathbf{M}\boldsymbol{\Psi}(h\boldsymbol{d})}.$$

We now examine what happens to

$$\frac{\{\boldsymbol{\Psi}(h\boldsymbol{d})'\mathbf{MU}\}^2}{\hat{\sigma}_{n,0}^2 \boldsymbol{\Psi}(h\boldsymbol{d})'\mathbf{M}\boldsymbol{\Psi}(h\boldsymbol{d})},$$

as h converges to zero. As pointed out by Theorem 1, if $h = 0$ then $\mathbf{\Psi}(\mathbf{0})'\mathbf{MU} = c_0\boldsymbol{\iota}'\mathbf{MU} = 0$ and $\mathbf{\Psi}(\mathbf{0})'\mathbf{M}\mathbf{\Psi}(0) = c_0^2\boldsymbol{\iota}'\mathbf{M}\boldsymbol{\iota} = 0$. We therefore apply L'Hôpital's rule to this ratio. For notational simplicity, for each $(h, \boldsymbol{d})$, we let

$$N_n(h, \boldsymbol{d}) := \{\mathbf{\Psi}(h\boldsymbol{d})'\mathbf{MU}\}^2 \quad \text{and} \quad D_n(h, \boldsymbol{d}) := \mathbf{\Psi}(h\boldsymbol{d})'\mathbf{M}\mathbf{\Psi}(h\boldsymbol{d}).$$

The following lemma provides the partial derivatives of $N_n(h, \boldsymbol{d})$ and $D_n(h, \boldsymbol{d})$ with respect to h.

Lemma 3 *Given Assumptions 2 and 3, for each $\boldsymbol{d} \in \mathbb{S}^{k-1}$,*

(i) For $\ell = 0, 1, 2, 3, 4,$ and 5, $\lim_{h\downarrow 0} N_n^{(\ell)}(h, \boldsymbol{d}) = 0$ a.s.$-\mathbb{P}$ and $\lim_{h\downarrow 0} D_n^{(\ell)}(h, \boldsymbol{d}) = 0$ a.s.$-\mathbb{P}$;

(ii) $\lim_{h\downarrow 0} N_n^{(6)}(h, \boldsymbol{d}) = 20c_3^2\{\boldsymbol{\iota}'\boldsymbol{D}_3(\boldsymbol{d})\boldsymbol{MU}\}^2$ a.s.$-\mathbb{P}$; and

(iii) $\lim_{h\downarrow 0} D_n^{(6)}(h, \boldsymbol{d}) = 20c_3^2\boldsymbol{\iota}'\boldsymbol{D}_3(\boldsymbol{d})\boldsymbol{MD}_3(\boldsymbol{d})\boldsymbol{\iota}$ a.s.$-\mathbb{P}$. □

Indeed, for $\ell = 0, 1, 2, 3,$ and 4, Lemma 3(*i*) holds as a corollary of lemma 7 of CIW. Lemmas 3 provides additional derivatives. Readers can find the proofs of Lemmas 3(*i*) with $\ell = 5$ and 3(*ii* and *iii*) from the URL mentioned above.

This now implies that the QLR test under $\mathcal{H}_{01} : \lambda_* = 0$ can be used to treat the QLR test under $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \mathbf{0}$ as a special case. More specifically, we can claim the following.

Lemma 4 *Given Assumptions 2 and 3, $QLR_n^{(1)} \geq QLR_n^{(2)} + o_{\mathbb{P}}(1)$.*

Thus, $QLR_n^{(1)}$ asymptotically dominates $QLR_n^{(2)}$. Proving Lemma 4 is almost identical to that in CIW. Lemma 4 implies that the asymptotic distribution of the QLR test can be obtained under both $\mathcal{H}_{01} : \lambda_* = 0$ and $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \mathbf{0}$ by combining the regularity conditions for $QLR_n^{(1)}$ with those for $QLR_n^{(2)}$. The following set of conditions is provided for this goal.

Assumption 6 (Regularity III) *(i) For some $\kappa \geq 2(\rho - 1)$, $E[|U_t|^{4+2\kappa}] < \infty$ and for $j = 1, 2, \ldots, k$, $E[|X_{t,j}|^{32}] < \infty$; or $E[|U_t|^8] < \infty$ and for $j = 1, 2, \ldots, k$, $E[|X_{t,j}|^{16}] < \infty$;*

(ii) $\sup_{\boldsymbol{\delta}\in\boldsymbol{\Delta}} |\Psi_t(\boldsymbol{\delta})| \leq M_t$ and for $j = 1, 2, \ldots, k$, $\sup_{\boldsymbol{\delta}\in\boldsymbol{\Delta}} |(\partial/\partial\delta_j)\Psi_t(\boldsymbol{\delta})| \leq M_t$;

(iii) $E[U_t|\boldsymbol{X}_t, U_{t-1}, \boldsymbol{X}_{t-1}, \ldots] = 0$; and

(iv) For each $\epsilon > 0$ and $\boldsymbol{\delta} \in \boldsymbol{\Delta}(\epsilon)$, $\bar{\boldsymbol{V}}_1(\boldsymbol{\delta})$ and $\bar{\boldsymbol{V}}_2(\boldsymbol{\delta})$ are positive definite, where for given ϵ, $\boldsymbol{\Delta}(\epsilon) := \{\boldsymbol{\delta} \in \boldsymbol{\Delta} : \sum_{j=1}^k |\delta_j| \geq \epsilon\}$,

$$\ddot{\boldsymbol{V}}_1(\boldsymbol{\delta}) := \begin{bmatrix} E[U_t^2\Psi_t(\boldsymbol{\delta})^2] & E[U_t^2\Psi(\boldsymbol{\delta})\boldsymbol{Z}_t'] & E[U_t^2\Psi(\boldsymbol{\delta})\boldsymbol{C}_t'] \\ E[U_t^2\boldsymbol{Z}_t\Psi(\boldsymbol{\delta})] & E[U_t^2\boldsymbol{Z}_t\boldsymbol{Z}_t'] & E[U_t^2\boldsymbol{Z}_t\boldsymbol{C}_t'] \\ E[U_t^2\boldsymbol{C}_t\Psi(\boldsymbol{\delta})] & E[U_t^2\boldsymbol{C}_t\boldsymbol{Z}_t'] & E[U_t^2\boldsymbol{C}_t\boldsymbol{C}_t'] \end{bmatrix} \quad \textit{and}$$

$$\ddot{\boldsymbol{V}}_2(\boldsymbol{\delta}) := \begin{bmatrix} E[\Psi_t(\boldsymbol{\delta})^2] & E[\Psi(\boldsymbol{\delta})\boldsymbol{Z}_t'] & E[\Psi(\boldsymbol{\delta})\boldsymbol{C}_t'] \\ E[\boldsymbol{Z}_t\Psi(\boldsymbol{\delta})] & E[\boldsymbol{Z}_t\boldsymbol{Z}_t'] & E[\boldsymbol{Z}_t\boldsymbol{C}_t'] \\ E[\boldsymbol{C}_t\Psi(\boldsymbol{\delta})] & E[\boldsymbol{C}_t\boldsymbol{Z}_t'] & E[\boldsymbol{C}_t\boldsymbol{C}_t'] \end{bmatrix}. \qquad \square$$

Using this set of assumptions, we can obtain the asymptotic distribution under $\mathcal{H}_{01}$ or $\mathcal{H}_{02}$. When Assumption 6 is imposed, Assumptions 4 and 5 hold. The following theorem states the synthesis of the results that we have separately obtained above.

Theorem 3 *Given Assumptions 1, 2, 3, 6, and the null* $\mathcal{H}_0 : \lambda_* = 0$ *or* $\boldsymbol{\delta}_* = \boldsymbol{0}$, $QLR_n \Rightarrow \sup_{h,\boldsymbol{d}} \ddot{\mathcal{G}}_0(h,\boldsymbol{d})^2$, *where*

$$\ddot{\mathcal{G}}_0(h,\boldsymbol{d}) := \begin{cases} \bar{\mathcal{G}}_0(h\boldsymbol{d}), & \textit{if } h \neq 0; \\ \widetilde{\mathcal{G}}_0(\boldsymbol{d}), & \textit{otherwise.} \end{cases} \qquad \square$$

As Theorem 3 is directly implied by Theorems 1 and 2. By the definition of $\ddot{\mathcal{G}}_0(\cdot)$, its covariance structure can be identified as follows: for each $(h,\boldsymbol{d})$ and $(\widetilde{h},\widetilde{\boldsymbol{d}})$,

$$\ddot{\rho}(\boldsymbol{d},\widetilde{\boldsymbol{d}}) := \begin{cases} \bar{\rho}(h\boldsymbol{d},\widetilde{h}\widetilde{\boldsymbol{d}}), & \text{if } h \neq 0 \text{ and } \widetilde{h} \neq 0; \\ \widetilde{\rho}(\boldsymbol{d},\widetilde{\boldsymbol{d}}), & \text{if } h = 0 \text{ and } \widetilde{h} = 0; \\ \widehat{\rho}(\boldsymbol{d},\widetilde{h}\widetilde{\boldsymbol{d}}), & \text{if } h = 0 \text{ and } \widetilde{h} \neq 0, \end{cases}$$

where for each $(\boldsymbol{d},\widetilde{h}\widetilde{\boldsymbol{d}})$,

$$\widehat{\rho}(\boldsymbol{d},\widetilde{h}\widetilde{\boldsymbol{d}}) := \frac{\mathcal{T}_3(\boldsymbol{d},\widetilde{h}\widetilde{\boldsymbol{d}})}{\{\sigma_*^2\mathcal{J}_2(\boldsymbol{d},\boldsymbol{d})\}^{1/2}\{\sigma_*^2\mathcal{J}_1(\widetilde{h}\widetilde{\boldsymbol{d}},\widetilde{h}\widetilde{\boldsymbol{d}})\}^{1/2}}$$

and $\mathcal{T}_3(\boldsymbol{d},\widetilde{h}\widetilde{\boldsymbol{d}}) := \sum_{j=1}^k\sum_{i=1}^k\sum_{\ell=1}^k d_jd_id_\ell E[U_t^2\Psi_t^*(\widetilde{h}\widetilde{\boldsymbol{d}})V_{t,ji\ell}^*]$. We define the covariance between $n^{-1/2}\boldsymbol{\Psi}(\widetilde{h}\widetilde{\boldsymbol{d}})'\mathbf{MU}$ and $n^{-1/2}\boldsymbol{\iota}'\mathbf{D}_3(\boldsymbol{d})\mathbf{MU}$ as $\mathcal{T}_3(\boldsymbol{d},\widetilde{h}\widetilde{\boldsymbol{d}})$, and they are the scores we could apply the CLT on MDA to in Sections 3.1 and 3.2, respectively. By Assumption 6, the separately obtained Gaussian stochastic processes in Sections 3.1 and 3.2 are not irrelevant processes any longer, and their dependence structure is now captured by $\widehat{\rho}(\cdot,\cdot)$.

We also note that the correlation functions used to define $\ddot{\rho}(\cdot,\cdot)$ are closely interrelated with each other. For identifying this interrelationship, we impose the following conditions.

Assumption 7 (Domination) *(i) For each* $\ell = 0,1,\ldots,6$ *and* $j = 1,2,\ldots,k+1$, $E[\sup_{h,\boldsymbol{d}}|(\partial^\ell/\partial h^\ell)\Xi_{t,j}(h\boldsymbol{d})|^2] < \infty$, *where for each* $j = 1,2,\ldots,k+1$, $\Xi_{t,j}(h\boldsymbol{d}) := \Psi_t(h\boldsymbol{d})Z_{t,j}$; *and*

(ii) For each ℓ *and* $m = 0,\ldots,6$ *such that* $\ell + m \leq 6$, $E[\sup_{h,\boldsymbol{d},\widetilde{h},\widetilde{\boldsymbol{d}}}|(\partial^{\ell+m}/\partial h^\ell\partial\widetilde{h}^m)\Lambda_t(h\boldsymbol{d},\widetilde{h}\widetilde{\boldsymbol{d}})|^2] < \infty$, *where* $\Lambda_t(h\boldsymbol{d},\widetilde{h}\widetilde{\boldsymbol{d}}) := \Psi_t(h\boldsymbol{d})\Psi_t(\widetilde{h}\widetilde{\boldsymbol{d}})$. $\square$

We provide the following lemma to formally state the interrelationship.

Lemma 5 *Given Assumptions 1, 2, 3, 6, 7, and the null hypothesis* $\mathcal{H}_0 : \lambda_* = 0$ *or* $\boldsymbol{\delta}_* = \boldsymbol{0}$*,*

(i) $\lim_{h\downarrow 0} \bar{\rho}(h\boldsymbol{d}, \widetilde{h}\widetilde{\boldsymbol{d}}) = \text{sgn}[c_3]\widehat{\rho}(\boldsymbol{d}, \widetilde{h}\widetilde{\boldsymbol{d}})$*; and*

(ii) $\lim_{\widetilde{h}\downarrow 0} \lim_{h\downarrow 0} \bar{\rho}(h\boldsymbol{d}, \widetilde{h}\widetilde{\boldsymbol{d}}) = \widetilde{\rho}(\boldsymbol{d}, \widetilde{\boldsymbol{d}})$. □

This lemma implies that the Gaussian stochastic process obtained under $\mathcal{H}_{01} : \lambda_* = 0$ can generate the asymptotic Gaussian stochastic process under $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \boldsymbol{0}$ as a special case, so that $\widetilde{\mathcal{G}}_0(\cdot)$ can be treated as a special case of $\bar{\mathcal{G}}_0(\cdot)$. Therefore, we do not have to separately derive the asymptotic null distribution under $\mathcal{H}_{02} : \boldsymbol{\delta}_* = \boldsymbol{0}$ when the conditions provided so far are satisfied. We formally state this result in the following theorem.

Theorem 4 *Given Assumptions 1, 2, 3, 6, 7, and* $\mathcal{H}_0 : \lambda_* = 0$ *or* $\boldsymbol{\delta}_* = \boldsymbol{0}$*,* $QLR_n \Rightarrow \sup_{\boldsymbol{\delta}\in\Delta} \bar{\mathcal{G}}_0(\boldsymbol{\delta})^2$. □

Note that the asymptotic null distribution is captured by $\bar{\mathcal{G}}_0(\cdot)$ and we do not restrict the parameter space as in Theorem 1. We adopt this idea from CIW although our interest lies in a different expansion.

Before moving to the next section, we provide additional comments on the QLR test. We note that the current examination provides a clue to the use of activation functions. Given that it is determined by the researcher, its selection influences the performances of the tests. If the researcher has a particular nonlinearity as an alternative in mind, the activation function can be easily determined. For example, if the concerned nonlinearity is better approximated by a cubic function, it is better to choose the activation function satisfying the zero-condition. On the other hand, if the researcher does not have a particular alternative, several aspects of the tests can be considered to determine the activation function. First, the activation functions satisfying the no-zero condition should perform better than those satisfying the zero condition in terms of asymptotic standard as their convergence rate is bigger than the current QLR test. Second, we can select better activation functions by comparing the powers of the tests under the Pitman-type local alternative. Under certain regularity conditions, the local powers of the QLR tests can be compared by using the optimality argument in Davies (1977). Third, we can also consider constructing a QLR test by adding two activation functions together to the RHS such that they respectively satisfies the zero and no-zero conditions. As their convergence rates are different, the QLR test has to be redefined to accommodate this,

and the local power of the test can be different from the QLR tests in CIW and the current paper. We leave examining these as future research topics.

4 Model Exercise and Monte Carlo Experiments

4.1 Model Exercise

Before conducting our experiments, we first examine a simple model and affirm the theoretical results in the previous section. We assume that $Y_t = 0.5Y_{t-1} + U_t$ with $U_t \sim$ IID $N(0, \sigma_*^2)$ and also that the following model is specified for $E[Y_t|Y_{t-1}, Y_{t-2}, \ldots]$:

$$\begin{aligned}\mathcal{M} := \{f(Y_{t-1}; \cdot, \cdot\ \cdot\ \cdot) : f(Y_{t-1}; \alpha, \beta, \lambda, \delta) := \alpha + \beta Y_{t-1} + \lambda \sin(\delta Y_{t-1}),\\ \alpha \in A, \beta \in (-1, 1), \lambda \in [-\bar{c}, \bar{c}], \delta \in \Delta\},\end{aligned}$$

where A and Δ are compact sets in $\mathbb{R}$ containing zero as an interior element, and $\bar{c}$ is a finite positive number. Furthermore, we suppose that U_t exhibits conditional homoskedasticity, and we let $\mathbf{Z}_t$ be $(1, Y_{t-1})'$ for this exercise, implying that $\mathbf{X}_t = Y_{t-1}$. Thus, our null model is a simple AR(1) model. As there is a single predictor in the activation function, we can decompose δ into hd such that $d \in \{-1, 1\}$ and $h \geq 0$. In addition to this, $E[Y_t|Y_{t-1}] = 0.5Y_{t-1}$ from the fact that $\{Y_t\}$ is an AR(1) process. We further note that $(d/dx)\sin(x)|_{x=0} = 1$ and $(d^2/dx^2)\sin(x)|_{x=0} = 0$, so that the no-zero condition in CIW does not hold for this activation function, whereas Assumption 3 holds. By this fact, we analyze the asymptotic null distribution using the methodology that we provided in the previous section when testing $\mathcal{H}_{01} : \lambda_* = 0$ or $\mathcal{H}_{02} : \delta_* = 0$.

Given this DGP and model assumptions, we now obtain that for each δ and $\widetilde{\delta}$,

$$\begin{aligned}\mathcal{T}_1(\delta, \widetilde{\delta}) &= E[U_t^2]E[\Psi_t^*(\delta)\Psi_t^*(\widetilde{\delta})]\\ &= \sigma_*^2 \exp\left[-(\delta^2 + \widetilde{\delta}^2)\frac{\sigma_*^2}{2}\right]\left[\frac{1}{2}\exp(\delta\widetilde{\delta}\sigma_*^2) - \frac{1}{2}\exp(-\delta\widetilde{\delta}\sigma_*^2) - \delta\widetilde{\delta}\sigma_*^2\right],\end{aligned}$$

and

$$\mathcal{J}_1(\delta, \delta) = \exp(-\delta^2\sigma_*^2)\left\{\frac{1}{2}\exp(\delta^2\sigma_*^2) - \frac{1}{2}\exp(-\delta^2\sigma_*^2) - \delta^2\sigma_*^2\right\},$$

so that we can derive

$$\bar{\rho}(\delta, \widetilde{\delta}) = \frac{\sinh(\delta\widetilde{\delta}\sigma_*^2) - \delta\widetilde{\delta}\sigma_*^2}{\{\sinh(\delta^2\sigma_*^2) - \delta^2\sigma_*^2\}^{1/2}\{\sinh(\widetilde{\delta}^2\sigma_*^2) - \widetilde{\delta}^2\sigma_*^2\}^{1/2}} \tag{8}$$

by following the definition of $\bar{\rho}(\cdot,\cdot)$ and using the fact that $\sinh(x) = \exp(x)/2 - \exp(-x)/2$. We also note that $\mathcal{T}_2(d,\widetilde{d}) = d^3\widetilde{d}^3 E[U_t^2]E[V_t^{*2}] = 6d^3\widetilde{d}^3\sigma_*^8$, where $V_t := Y_{t-1}^3 - 3\sigma_*^2 Y_{t-1}$, and $\mathcal{J}_2(d,d) = 6d^6\sigma_*^6$. This implies that $\widetilde{\rho}(d,\widetilde{d}) = d^3\widetilde{d}^3/\{d^6\widetilde{d}^6\}^{1/2}$ by the definition of $\widetilde{\rho}(\cdot,\cdot)$. Finally, $\mathcal{T}_3(d,\widetilde{h}\widetilde{d}) = d^3 E[U_t^2]E[V_t^*\Psi_t^*(\widetilde{h}\widetilde{d})] = -d^3\widetilde{\delta}^3\sigma_*^8\exp(-\sigma_*^2\widetilde{\delta}^2/2)$, so that

$$\widehat{\rho}(d,\widetilde{h}\widetilde{d}) = -\frac{-\sigma_*^8 d^3\widetilde{\delta}^3\exp\left(-\frac{\sigma_*^2\widetilde{\delta}^2}{2}\right)}{\{6\sigma_*^8 d^6\}^{1/2}\{\sigma_*^2\exp(-\widetilde{h}^2\widetilde{d}^2\sigma_*^2)[\sinh(\widetilde{h}^2\widetilde{d}^2\sigma_*^2) - \widetilde{h}^2\widetilde{d}^2\sigma_*^2]\}^{1/2}}$$

by the definition of $\widehat{\rho}(\cdot,\cdot)$. From this, it simply follows that $\lim_{h\downarrow 0}\bar{\rho}(hd,\widetilde{h}\widetilde{d}) = -\widehat{\rho}(d,\widetilde{h}\widetilde{d})$ by applying L'Hôpital's rule repeatedly. The negative sign in the RHS is attributable to the fact that $(d^3/dx^3)\sin(x)|_{x=0} = -1$, implying that $\text{sgn}[c_3] = -1$. This is the result implied by Lemma 5(*i*). In addition, $\lim_{\widetilde{h}\downarrow 0}\lim_{h\downarrow 0}\bar{\rho}(hd,\widetilde{h}\widetilde{d}) = \widetilde{\rho}(d,\widetilde{d})$, so that the covariance structure of the Gaussian stochastic process obtained under $\mathcal{H}_{02} : \delta_* = 0$ is obtained from the Gaussian stochastic process derived under $\mathcal{H}_{01} : \lambda_* = 0$. This implies that the Gaussian stochastic process obtained under $\mathcal{H}_{01} : \lambda_* = 0$ can be used to deliver the asymptotic behavior of the Gaussian stochastic process derived under $\mathcal{H}_{02} : \delta_* = 0$.

4.2 Monte Carlo Simulations

In this subsection, we conduct Monte Carlo experiments using the model exercise examined in the previous subsection.

We first generate the asymptotic distribution of the QLR test under $\mathcal{H}_0$. The key element for this is generating a Gaussian stochastic process with the correlation structure in eq. (8). We note that $\sinh(x) = \sum_{j=1}^{\infty} x^{(2j-1)}/(2j-1)!$, so that if we let

$$\mathcal{B}(\delta;\sigma_*^2) := \frac{1}{\{\sinh(\delta^2\sigma_*^2) - \delta^2\sigma_*^2\}^{1/2}}\sum_{j=1}^{\infty}\frac{(\delta\sigma_*)^{2j+1}}{\sqrt{(2j+1)!}}Z_j,$$

where $Z_j \sim$ IID $N(0,1)$, for each δ and $\widetilde{\delta}$,

$$\begin{aligned} E[\mathcal{B}(\delta;\sigma_*^2)\mathcal{B}(\widetilde{\delta};\sigma_*^2)] &= \frac{1}{\{\sinh(\delta^2\sigma_*^2) - \delta^2\sigma_*^2\}^{1/2}\{\sinh(\widetilde{\delta}^2\sigma_*^2) - \widetilde{\delta}^2\sigma_*^2\}^{1/2}}\sum_{j=1}^{\infty}\frac{(\delta\widetilde{\delta}\sigma_*^2)^{2j+1}}{(2j+1)!} \\ &= \frac{\sinh(\delta\widetilde{\delta}\sigma_*^2) - \delta\widetilde{\delta}\sigma_*^2}{\{\sinh(\delta^2\sigma_*^2) - \delta^2\sigma_*^2\}^{1/2}\{\sinh(\widetilde{\delta}^2\sigma_*^2) - \widetilde{\delta}^2\sigma_*^2\}^{1/2}}. \end{aligned}$$

In other words, the covariance structure of $\bar{\mathcal{G}}_0(\cdot)$ is identical to that of $\mathcal{B}(\cdot;\sigma_*^2)$. Therefore, $\mathcal{B}(\cdot;\sigma_*^2) \overset{\text{d}}{=} \bar{\mathcal{G}}_0(\cdot)$. This also implies that the asymptotic null distribution can be obtained by generating $\sup_{\delta\in\Delta}\mathcal{B}(\delta;\sigma_*^2)^2$ iteratively.

For this goal, we instead consider the following Gaussian stochastic process: for each $\delta \in \Delta$,

$$\bar{\mathcal{B}}(\delta; \hat{\sigma}_n^2) := \frac{1}{\{\sinh(\delta^2 \hat{\sigma}_n^2) - \delta^2 \hat{\sigma}_n^2\}^{1/2}} \sum_{j=1}^{150} \frac{(\delta \hat{\sigma}_n)^{2j+1}}{\sqrt{(2j+1)!}} Z_j,$$

where $\hat{\sigma}_n^2 := (n-1)^{-1} \sum_{t=2}^{n} \hat{U}_t^2$ and $\hat{U}_t$ is the prediction error obtained by estimating the null model. Note that $\mathcal{B}(\cdot; \sigma_*^2)$ contains unknown σ_*^2, and we cannot draw independent standard normal random variables an infinite number of times. We instead estimate σ_*^2 and draw 150 standard normal random variables.

The environments for our Monte Carlo experiments are set up as follows. First, we consider four different parameter spaces. We let $\Delta_1 := [-0.5, 0.5]$, $\Delta_2 := [-1.0, 1.0]$, $\Delta_3 := [-1.5, 1.5]$, and $\Delta_4 := [-2.0, 2.0]$. Note that the asymptotic null distribution is influenced by the size of Δ, so that different critical values must be applied for different parameter spaces. Second, we gradually increase the sample size and investigate how the empirical null distribution varies under each parameter space. The sample sizes we consider here are $n = 100$, 200, 400, 600, $1,000$, and $2,000$. The total number of replications is 2,000. We also obtain the asymptotic distributions by simulating $\sup_{\delta \in \Delta} \bar{\mathcal{B}}(\delta; \hat{\sigma}_n^2)^2$ 50,000 times. This environment is almost identical to those considered by CIW.

<<<<<<< Insert Figure 1 around here. >>>>>>>>

Our Monte Carlo simulation results can be summarized as follows. First, Figure 1 shows the asymptotic null distributions obtained by simulating $\sup_{\delta \in \Delta} \bar{\mathcal{B}}(\delta; \sigma_*^2)^2$. As the parameter space becomes larger, the associated critical values also increase, and the probability of obtaining a larger value of $\sup_{\delta \in \Delta} \mathcal{B}(\delta; \sigma_*^2)^2$ increases as shown in Figure 1.

<<<<<<< Insert Table 1 around here. >>>>>>>>

Second, we examine the empirical rejection rates of the QLR test. The simulation results are presented in Table 1, in which we consider two cases separately. As the first case, the critical values obtained from Figure 1 are applied, and as the second case, we assume that σ_*^2 is unknown. By generating $\sup_{\delta \in \Delta} \mathcal{B}(\delta; \widehat{\sigma}_n^2)^2$ iteratively, we obtain the critical values. As $\widehat{\sigma}_n^2$ is different for each iteration, different critical values have to be applied. Through this process, we aim to examine how estimating σ_*^2 modifies the empirical rejection rates.

<<<<<<< Insert Figure 2 around here. >>>>>>>>

We summarize the simulation results as follows. First, the simulation results are similar irrespective of whether the unknown σ_*^2 is estimated. The differences between the empirical rejection rates in Table 1 are very small for every cell in the table. Furthermore, the difference decreases as the sample size increases. Second, the empirical rejection rates converge to the nominal levels as the sample size increases. When the sample size is small, the empirical rejection rates are relatively imprecise. We can see this aspect from Figure 2, in which we draw the empirical distributions of the QLR tests and the asymptotic distribution for each parameter space. Third, more desirable results are obtained when the parameter space of δ is moderately large. When Δ_1 is selected, the empirical rejection rates are relatively imprecise, and they become worse for small sample sizes. Figure 2 also shows that the empirical distributions for Δ_1 are more or less different from the asymptotic distribution although the other empirical distributions are almost identical to their asymptotic distributions.

<<<<<<< Insert Table 2 around here. >>>>>>>>

Third, we examine the application of the weighted bootstrap to the QLR test. The covariance structure of $\mathcal{B}(\cdot;\sigma_*^2)$ is unknown if the distribution of U_t is unknown. This implies that we cannot apply the critical values obtained from Figure 1. Hansen (1996) suggests applying the weighted bootstrap for such a case, and CIW show through Monte Carlo experiments that it can be successfully exploited if the no-zero condition holds. As the weighted bootstrap is already explained by Hansen (1996), CIW, and Cho, Cheong, and White (2011), we do not explain its implementation here again. Our simulation results are presented in Table 2 and Figures 3. The estimated lines show the estimated p-values. Under the null, the estimated lines should lie along the 45-degree line to be successful.

<<<<<<< Insert Figure 3 around here. >>>>>>>>

We can summarize the simulation results as follows. First, the weighted bootstrap yields asymptotically consistent results as the sample size increases. This result differs from that obtained by CIW. Their simulation results, driven by the logistic CDF, yield slightly different results from the 45 degree line even for fairly large sample sizes. Note that the logistic CDF also requires a sixth-order expansion as does the sine function. This implies that the selection

of the activation function is very important for obtaining precise empirical rejection rates under the null. Second, the simulation results obtained by the weighted bootstrap are better than those obtained by the asymptotic critical values, although the difference is not large. In addition, the simulation results are worse when Δ_1 is applied to the data set with small sample sizes.

5 Conclusion

We revisit CIW's twofold identification problem, which arises when testing for neglected nonlinearity using an ANN framework. We replace their "no-zero" condition to the next level and apply a sixth-order expansion. Specifically, we apply a sixth-order expansion and obtain the asymptotic null distribution of the QLR test. In particular, we do not restrict the number of explanatory variables in the activation function by following the distance and direction method in Cho and White (2012).

We show that the QLR test can be used to combine the null distributions obtained separately by type I and II methodologies as defined in this paper. This implies that a further higher-order expansion can be applied to capture the asymptotic null distribution even if the no-zero condition does not hold. Furthermore, Hansen's (1996) weighted bootstrap can be usefully exploited for the QLR test statistic.

References

ANDREWS, D. (2001): "Testing When a Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

BAEK, Y. AND CHO, J. (2012): "Testing the Linearity Hypothesis Using Power Transformations," Discussion Paper, Yonsei University.

BIERENS, H. (1987): "ARMAX Model Specification Testing with an Application to Unemployment in the Netherlands," *Journal of Econometrics*, 35, 161–190.

BIERENS, H. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.

BIERENS, H. AND HARTDOG, J. (1988): "Nonlinear Regression with Discrete Explanatory Variables with an Application to the Earnings Function," *Journal of Econometrics*, 38, 269–299.

CHO, J. (2012): "Quasi-Maximum Likelihood Estimation Revisited Using the Distance and Direction Method," *Journal of Economic Theory and Econometrics*, 23, 89–112.

CHO, J. AND WHITE, H. (2007): "Testing for Regime Switching," *Econometrica*, 75, 1671–1720.

CHO, J. AND WHITE, H. (2010): "Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Econometrics*, 157, 458–480.

CHO, J. AND ISHIDA, I. (2012): "Testing for the Effects of Omitted Power Transformations," *Economics Letters*, 117, 287–290.

CHO, J. AND WHITE, H. (2012): "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., CHEONG, T., AND WHITE, H. (2011): "Experience with the Weighted Bootstrap in Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Economic Theory and Econometrics*, 22:2, 60–91

DAVIES, R. (1977): "Hypothesis Testing When a Nuisance Parameter is Present only under the Alternative," *Biometrika*, 64, 247–254.

DAVIES, R. (1987): "Hypothesis Testing When a Nuisance Parameter is Present only under the Alternative," *Biometrika*, 74, 33–43.

DOUKHAN, P. (1994): *Mixing: Properties and Examples*, Lecture Notes in Statistics. Springer.

DOUKHAN, P., P. MASSART, AND E. RIO. (1995): "Invariance Principles for Absolutely Regular Empirical Processes," *Annales de l'Institut Henri Poincaré, Probabilites et Statistiques*, 31, 393–427.

HANSEN, B. (1996): "Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–440.

KING, M. AND T. SHIVELY. (1993): "Locally Optimal Testing When a Nuisance Parameter is Present Only Under the Alternative," *The Review of Economics and Statistics*, 75, 1–7.

LEE, T.-H., WHITE, H. AND GRANGER, C. (1993): "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests," *Journal of Econometrics*, 56, 269–290.

LUUKKONEN, R., SAIKKONEN, P., AND TERÄSVIRTA, T. (1988): "Testing Linearity against Smooth Transition Autoregressive Models," *Biometrika*, 75, 491–499.

RAMSEY, J.B. (1969): "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," *Journal of Royal Statistical Society B.*, 31, 350–371.

ROSENBERG, B. (1973): "The Analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regression," *Annals of Economic and Social Measurement*, 2, 399–428.

STINCHCOMBE, M. AND WHITE, H. (1998): "Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative," *Econometric Theory*, 14, 295–324.

TERÄSVIRTA, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89, 208–218.

TERÄSVIRTA, T., C.-F. LIN, AND C.W.J. GRANGER (1993): "Power of the Neural Network Linearity Test," *Journal of Time Series Analysis*, 14, 309–323.

WHITE, H. (1992): "Nonparametric Estimation of Conditional Quantiles Using Neural Networks," in *Proceedings of the Symposium on the Interface*. New York: Springer-Verlag, 190–199.

WHITE, H. (1994): *Estimation, Inference, and Specification Analysis*. New York: Cambridge University Press.

WHITE, H. (2001): *Asymptotic Theory for Econometricians*. Orlando: Academic Press.

WHITE, H. AND CHO, J. (2012): "Higher-Order Approximations for Testing Neglected Nonlinearity," *Neural Computation*, 24, 273–287.

When σ_*^2 is known						
	$\mathcal{QLR}(\Delta_1)$					
Nominal Level \ Sample Size	100	200	400	600	1,000	2,000
1.00 %	0.10	0.65	1.10	0.75	0.80	1.20
5.00 %	2.60	2.65	4.40	4.50	4.55	4.85
10.0 %	7.00	6.50	9.10	9.20	9.25	10.10
15.0 %	10.55	11.40	13.30	14.40	14.15	15.85
	$\mathcal{QLR}(\Delta_2)$					
Nominal Level \ Sample Size	100	200	400	600	1,000	2,000
1.00 %	0.75	0.90	1.00	0.45	1.10	0.75
5.00 %	3.70	4.05	4.15	4.25	4.95	4.90
10.0 %	8.20	9.30	8.50	8.80	9.85	10.05
15.0 %	12.25	13.70	14.10	14.75	14.90	14.15
	$\mathcal{QLR}(\Delta_3)$					
Nominal Level \ Sample Size	100	200	400	600	1,000	2,000
1.00 %	0.70	0.85	0.50	0.70	0.95	1.00
5.00 %	3.65	4.30	4.75	4.05	4.70	4.65
10.0 %	7.80	9.05	9.90	9.20	10.30	9.70
15.0 %	11.30	13.95	15.35	14.40	15.55	15.45
	$\mathcal{QLR}(\Delta_4)$					
Nominal Level \ Sample Size	100	200	400	600	1,000	2,000
1.00 %	0.80	1.00	1.10	1.25	0.95	1.20
5.00 %	4.55	5.40	6.20	6.55	5.30	5.85
10.0 %	8.80	10.70	10.85	11.75	11.20	11.60
15.0 %	13.80	15.20	16.80	17.50	16.70	17.05
When σ_*^2 is estimated						
	$\mathcal{QLR}(\Delta_1)$					
Nominal Level \ Sample Size	100	200	400	600	1,000	2,000
1.00 %	0.15	0.50	1.05	0.80	0.75	1.35
5.00 %	2.75	3.00	4.55	4.40	4.60	5.00
10.0 %	6.80	6.90	9.25	9.70	9.25	10.10
15.0 %	10.20	11.20	13.35	14.85	14.25	16.00
	$\mathcal{QLR}(\Delta_2)$					
Nominal Level \ Sample Size	100	200	400	600	1,000	2,000
1.00 %	0.75	0.95	0.80	0.45	1.20	0.65
5.00 %	3.85	4.10	4.30	4.20	5.20	4.90
10.0 %	8.25	8.90	8.75	9.35	10.25	9.80
15.0 %	12.70	14.10	14.35	14.25	15.15	14.25
	$\mathcal{QLR}(\Delta_3)$					
Nominal Level \ Sample Size	100	200	400	600	1,000	2,000
1.00 %	0.75	1.10	0.40	0.75	0.90	1.15
5.00 %	3.80	4.45	4.85	4.20	5.00	4.65
10.0 %	7.70	9.30	10.20	9.70	10.60	9.90
15.0 %	12.10	14.00	15.50	15.35	15.65	15.80
	$\mathcal{QLR}(\Delta_4)$					
Nominal Level \ Sample Size	100	200	400	600	1,000	2,000
1.00 %	0.95	1.05	1.00	1.15	0.95	1.20
5.00 %	3.90	5.10	5.60	6.40	5.10	5.45
10.0 %	9.05	10.50	11.05	11.35	11.00	11.15
15.0 %	13.75	15.00	16.30	17.55	16.25	16.80

Table 1: EMPIRICAL REJECTION RATES USING THE ASYMPTOTIC DISTRIBUTIONS (IN PERCENT). Number of Replications: 2,000, DGP: $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim$ IID $N(0,1)$, Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda \sin(\delta Y_{t-1}) + U_t$, $\Delta_1 = [-0.5, 0.5]$, $\Delta_2 = [-1.0, 1.0]$, $\Delta_3 = [-1.5, 1.5]$, $\Delta_4 = [-2.0, 2.0]$, and $K = 150$

Asymptotic Distribution	$\mathcal{QLR}(\Delta_1)$					
Nominal Level \ Sample Size	100	200	400	600	1,000	2,000
1.00 %	0.50	0.90	0.50	0.80	0.90	1.30
5.00 %	3.35	4.45	4.35	4.25	4.60	4.95
10.0 %	6.65	8.00	8.85	9.30	8.95	10.45
15.0 %	11.15	12.60	13.85	14.45	13.50	14.65
Asymptotic Distribution	$\mathcal{QLR}(\Delta_2)$					
Nominal Level \ Sample Size	100	200	400	600	1,000	2,000
1.00 %	0.80	0.70	0.95	0.95	0.95	0.50
5.00 %	4.75	4.65	4.85	4.45	5.45	4.10
10.0 %	10.15	9.65	10.55	8.95	9.80	9.05
15.0 %	14.95	14.50	15.15	14.05	15.00	14.20
Asymptotic Distribution	$\mathcal{QLR}(\Delta_3)$					
Nominal Level \ Sample Size	100	200	400	600	1,000	2,000
1.00 %	0.75	1.05	0.85	1.00	0.75	1.05
5.00 %	4.60	4.90	4.05	4.95	3.70	4.55
10.0 %	8.90	10.00	8.85	10.25	8.35	9.80
15.0 %	14.70	15.45	14.00	14.55	13.55	14.25
Asymptotic Distribution	$\mathcal{QLR}(\Delta_4)$					
Nominal Level \ Sample Size	100	200	400	600	1,000	2,000
1.00 %	0.85	0.65	1.50	0.90	0.60	0.80
5.00 %	4.55	4.50	5.35	4.25	4.45	4.35
10.0 %	9.70	9.45	10.50	7.95	9.30	9.05
15.0 %	14.55	14.45	15.20	12.35	14.85	13.40

Table 2: EMPIRICAL REJECTION RATES USING THE WEIGHTED BOOTSTRAP (IN PERCENT). Number of Replications: 2,000, DGP: $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim$ IID $N(0,1)$, Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda \sin(\delta Y_{t-1}) + U_t$, $\Delta_1 = [-0.5, 0.5]$, $\Delta_2 = [-1.0, 1.0]$, $\Delta_3 = [-1.5, 1.5]$, $\Delta_4 = [-2.0, 2.0]$, and $K = 150$

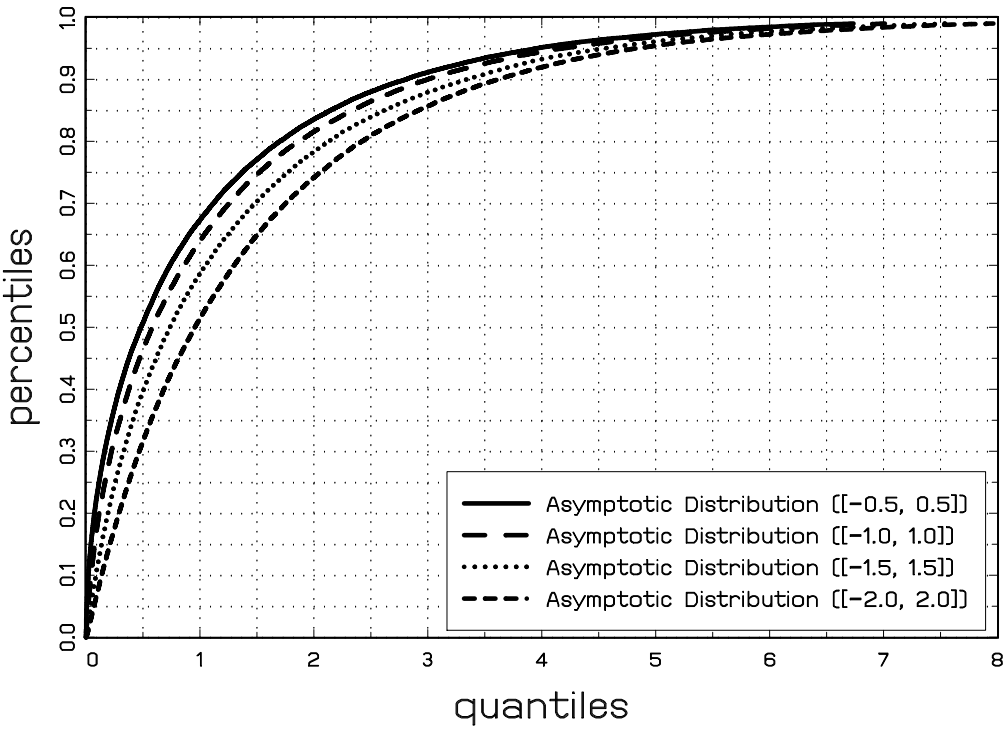


Figure 1: ASYMPTOTIC NULL DISTRIBUTIONS OF THE QLR STATISTICS. Number of Replications: 50,000, Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda \sin(\delta Y_{t-1}) + U_t$, $\Delta_1 = [-0.5, 0.5]$, $\Delta_2 = [-1.0, 1.0]$, $\Delta_3 = [-1.5, 1.5]$, $\Delta_4 = [-2.0, 2.0]$, and $K = 150$

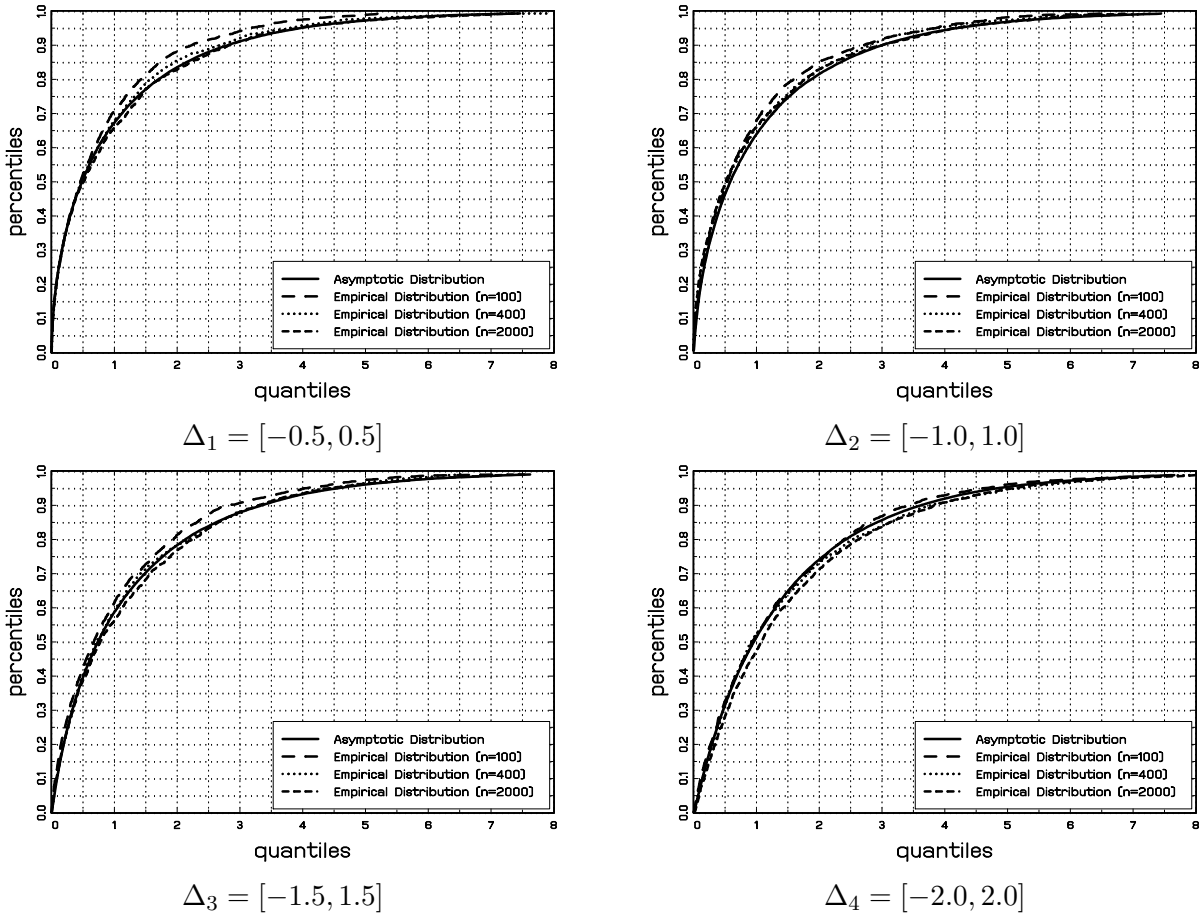
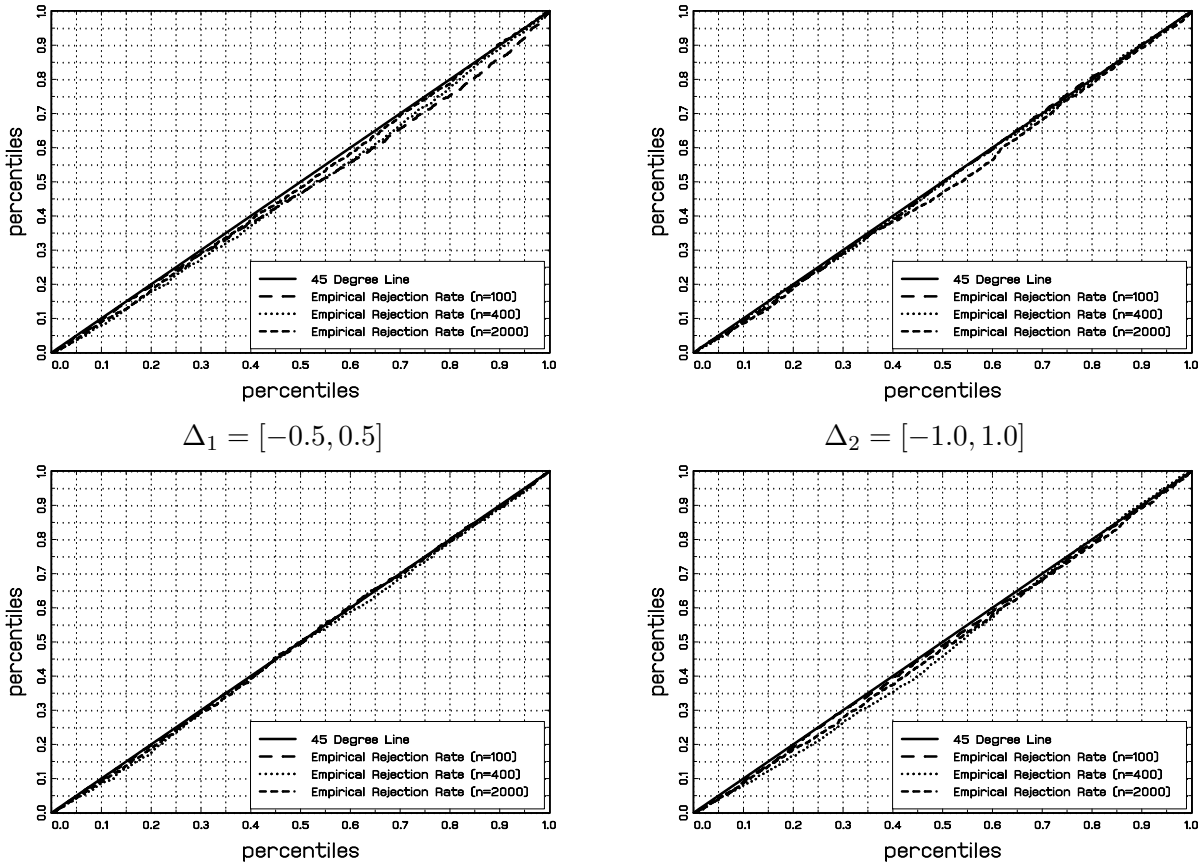


Figure 2: ASYMPTOTIC AND EMPIRICAL NULL DISTRIBUTIONS OF THE QLR STATISTICS. Number of Replications: 2,000, DGP: $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim$ IID $N(0,1)$, Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda \sin(\delta Y_{t-1}) + U_t$, and $K = 150$



$\Delta_1 = [-0.5, 0.5]$ $\Delta_2 = [-1.0, 1.0]$

$\Delta_3 = [-1.5, 1.5]$ $\Delta_4 = [-2.0, 2.0]$

Figure 3: ESTIMATED p-VALUES OF THE QLR STATISTICS BY THE WEIGHTED BOOTSTRAP. Number of Replications: 2,000, DGP: $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim$ IID $N(0,1)$, Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda \sin(\delta Y_{t-1}) + U_t$, and $K = 150$