Revisiting Tests for Neglected Nonlinearity Using Artificial Neural Networks

Jin Seo Cho

jinseocho@yonsei.ac.kr

School of Economics, Yonsei University, Seoul 120-749, Korea

Isao Ishida

i-ishida@sigmath.es.osaka-u.ac.jp

CSFI, Osaka University, Osaka 560-8531, Japan

Halbert White

hwhite@weber.ucsd.edu

Department of Economics, University of California, San Diego, La Jolla 92093-0508, USA

Keywords: Artificial neural networks, linearity hypothesis, quasi-likelihood ratio test, directionally differentiable model, asymptotic null distribution, Gaussian process

Abstract

Tests for regression neglected nonlinearity based on artificial neural networks (ANNs) have so far been studied by separately analyzing the two ways in which the null of regression linearity can hold. This implies that the asymptotic behavior of general ANN-based tests for neglected nonlinearity is still an open question. Here we analyze a convenient ANN-based quasi-likelihood ratio (QLR) statistic for testing neglected

nonlinearity, paying careful attention to both components of the null. We derive the asymptotic null distribution under each component separately and analyze their interaction. Somewhat remarkably, it turns out that the previously known asymptotic null distribution for the "type 1" case still applies, but under somewhat stronger conditions than previously recognized. We present Monte Carlo experiments corroborating our theoretical results and showing that standard methods can yield misleading inference when our new, stronger regularity conditions are violated.

1 Introduction

Artificial neural networks (ANNs) have become increasingly of interest in a wide range of applied disciplines. For example, in economics, ANNs now have their own *Journal of Economic Literature* classification number, C45. This widespread interest in ANNs is due to their many useful properties. In particular, single hidden layer feedforward perceptrons permit arbitrarily accurate approximation to broad classes of functions (see, e.g., Hornik, Stinchcombe and White (1989, 1990)), supporting parametric or nonparametric estimation of conditional mean, quantile, or density functions (see, e.g., White (1990, 1992), Gallant and White (1992), Kuan and White (1994), White (1996), and Chen and White (1999)). In what follows, we focus on these specific ANNs, and it should be understood that although for convenience we refer to "ANNs," we always have in mind this particular architecture.

This universal approximation property can also be exploited to test for neglected nonlinearity in regression analysis, as in White (1989*a*) and Lee, White, and Granger (1993). When suitably constructed, such tests can be consistent against arbitrary nonlinearity. Closely related methods can be found in the work of Bierens (1987, 1990), Bierens and Hartog (1988), and Hansen (1996). As has been well recognized in this literature, using ANNs can lead to nonstandard tests, falling into the category of tests with *nuisance parameters identified only under the alternative* (Davies (1977, 1987)).

Although attempts have been made to accommodate the nonstandard nature of such tests, the previous work has not satisfactorily examined their asymptotic null distribution. This is because a correct linear ANN specification can arise in two different ways. First, the "hidden-to-output unit" coefficient, say λ^*, can be zero. Alternatively, $\boldsymbol{\delta}^*$, the "input-to-hidden unit" coefficients on the explanatory variables, which determine the hidden unit output, can be zero. We refer to these as "type 1" and "type 2" hypotheses, respectively. The previous literature has focused separately on the type 1 and type 2 hypotheses in obtaining asymptotic null distributions. For type 1, these distributions have a representation as a function of a Gaussian process indexed by the parameters unidentified under the type 1 null, as shown by Bierens (1990) and Hansen (1996), among others, using Wald- and Lagrange multiplier (LM)-type test statistics. For type 2, Luukkonen, Saikkonen, and Teräsvirta (1988), Teräsvirta, Lin, and Granger (1993), Teräsvirta (1994), and Granger and Teräsvirta (chapter 6, 1993) have proposed LM-type tests for specific hidden unit activations. These tests have convenient chi-squared asymptotic null distributions. So far, however, it is unknown how the type 1 and 2 hypotheses interact to determine the null distribution in the general case. This omission means that the properties of general ANN tests for neglected nonlinearity are still an open question.

Several possibilities arise when treating type 1 and type 2 hypotheses jointly: (i) the regularity conditions for type 1 do not suffice for type 2, or vice versa; (ii) the asymptotic distribution for type 2 differs from that for type 1; (iii) both (i) and (ii) hold; or (iv) neither (i) nor (ii) hold. It is not at all obvious a priori which of these possibilities obtains. Our goal here, therefore, is to address these issues by carefully analyzing the asymptotic null distribution of a convenient ANN-based quasi-likelihood ratio (QLR) statistic designed to test for neglected nonlinearity. We first examine the asymptotic behaviors of the QLR statistic under type 1 and type 2 nulls separately; we then examine their stochastic interrelation.

Somewhat remarkably and rather fortunately, it turns out that suitably constructed ANN tests for neglected nonlinearity fall into category (i): we require stronger regularity conditions than previously recognized, but the asymptotic null distribution that properly accounts for both type 1 and type 2 nulls and their interaction coincides with that previously obtained by neglecting type 2. That is, the previous type 1 literature obtained essentially the right answer for the general case, but without a proper foundation. We say "suitably constructed" tests, as we also find that certain choices of the hidden unit activation function (denoted Ψ) can lead to test statistics that fall into category (iii), which is much less convenient, both analytically and computationally. In fact, as our simulations show, choices that violate our conditions but are otherwise standard can lead to misleading inference using standard methods.

The plan of this paper is as follows. In Section 2, we separately derive the asymptotic distributions of the QLR statistic under type 1 and type 2 nulls. The type 1 results are essentially known; we apply results of Hansen (1996). The type 2 results turn out to require use of a fourth-order Taylor approximation. Such approximations have been studied in other contexts (Bartlett (1953*a*, 1953*b*); McCullagh (1987)), but their use in the ANN context appears to be novel. Our methods are particularly straightforward, in that we are able to avoid using the tensors employed by McCullagh (1987). Section 2 completes the analysis by deriving the stochastic interrelationship of the type 1 and type 2 weak limits. Section 3 presents some Monte Carlo experiments using a first-order autoregressive process, affirming the theoretical results of Section 2 and showing that misleading inference can result from specifications that violate our new, stronger regularity conditions. Section 4 contains a summary and concluding remarks; we collect formal mathematical proofs into the Appendix.

2 The DGP and Artificial Neural Network Model

We work with the following data generating process (DGP).

Assumption A1 (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 target and predictor variables, respectively. For convenience, $\mathbf{X}_t$ omits the constant. The mixing coefficients β_τ are

$$\beta_\tau := \sup_{s \in \mathbb{N}} E[\sup_{A \in \mathcal{F}_{s+\tau}^{\infty}} |\mathbb{P}(A \mid \mathcal{F}_{-\infty}^{s}) - \mathbb{P}(A)|]$$

where $\mathcal{F}_t^s$ is the $\sigma-$field ("information set") generated by $(Y_t, \mathbf{X}_t, ..., Y_{t+s}, \mathbf{X}_{t+s})$. The β_τ's measure the time-series dependence in the data. For more on absolutely regular (β-mixing) processes, see Doukhan, Massart, and Rio (1995, DMR hereafter).

A1 is appropriate for analyzing weakly dependent time-series data, such as non-trending data arising in economic or biological systems. It has been adopted, among others, by Hansen (1996, 2006) and Cho and White (2007). In particular, A1 helps ensure the tightness used to guarantee that our main test statistic weakly converges to a function of a Gaussian process. For this, we rely on results of DMR.

When interest focuses on forecasting Y_t using the information in $\mathbf{X}_t$, it is common to forecast using an approximation to $E[Y_t|\mathbf{X}_t]$, the conditional expectation ("regression") of Y_t given $\mathbf{X}_t$. This conditional expectation gives the mean-squared-error optimal forecast of Y_t given $\mathbf{X}_t$. Here, we approximate $E[Y_t|\mathbf{X}_t]$ using a single hidden layer feedforward network with the following structure:

Assumption A2 (Model): *Let $\Psi : \mathbb{R} \mapsto \mathbb{R}$ be such that $\Psi(\,\cdot\,)$ is a non-polynomial analytic function such that $\Psi(0) \neq 0$. Let $\mathbf{A} \subset \mathbb{R}$, $\mathbf{B} \subset \mathbb{R}^k$, $\mathbf{\Lambda} \subset \mathbb{R}$, and $\mathbf{\Delta} \subset \mathbb{R}^k$ be non-empty compact and convex sets, with $0 \in \text{int}(\mathbf{\Lambda})$ and $\mathbf{0} \in \text{int}(\mathbf{\Delta})$. Let $f(\mathbf{X}_t\,;\alpha,\boldsymbol{\beta},\lambda,\boldsymbol{\delta}) := \alpha + \mathbf{X}_t'\boldsymbol{\beta} + \lambda\Psi(\mathbf{X}_t'\boldsymbol{\delta})$, and define the model $\mathcal{M}$ as*

$$\mathcal{M} := \{f(\,\cdot\,;\alpha,\boldsymbol{\beta},\lambda,\boldsymbol{\delta}) : (\alpha,\boldsymbol{\beta},\lambda,\boldsymbol{\delta}) \in \mathbf{A} \times \mathbf{B} \times \mathbf{\Lambda} \times \mathbf{\Delta}\}.$$

Note that f is a feedforward network with direct linear input-output connections and just one hidden unit. Our interest here is in testing whether a simple linear network (no hidden units) provides an adequate approximation to $E[Y_t|\mathbf{X}_t]$ or whether there is neglected nonlinearity, so that using hidden units can improve the approximation. By choosing Ψ to be analytic (i.e., locally given by a convergent power series) and non-polynomial, we ensure that Ψ is generically comprehensively revealing (GCR; see Stinchcombe and White, 1998). This then guarantees that a single hidden unit suffices to detect arbitrary neglected nonlinearity (but also see Escanciano, 2009). Standard GCR choices for Ψ are $\Psi = \exp$ (as in Bierens, 1990), the logistic cumulative distribution function (CDF) originally used by White (1989*a*), or the ridgelets of Candès (2003). Below, we impose additional conditions on Ψ.

Note that because $\mathbf{X}_t$ omits the constant, $\Psi(\mathbf{X}_t'\boldsymbol{\delta})$ does not contain an adjustable input-to-hidden bias. Instead, we permit a fixed or "hard-wired" bias. This can be arbitrarily set without any adverse effect on the test's ability to detect arbitrary nonlinearity.[1]

Our null hypothesis is that $E[Y_t|\mathbf{X}_t]$ is linear. Formally, we test

$$\mathcal{H}_0: \text{For some } (\alpha, \boldsymbol{\beta}) \in \mathbf{A} \times \mathbf{B},\ \mathbb{P}[\, E(Y_t|\mathbf{X}_t) = \alpha + \mathbf{X}_t'\boldsymbol{\beta}\,] = 1 \quad \text{versus}$$
$$\mathcal{H}_1: \text{For all } (\alpha, \boldsymbol{\beta}) \in \mathbf{A} \times \mathbf{B},\ \mathbb{P}[\, E(Y_t|\mathbf{X}_t) = \alpha + \mathbf{X}_t'\boldsymbol{\beta}\,] < 1.$$

Under $\mathcal{H}_0$, the model is correctly specified as in White (1994).

Testing $\mathcal{H}_0$ using ANNs is not standard, as has often been noted. For example, White (1989*b*) and Bierens (1990) note that under $\mathcal{H}_0$, Davies's (1977, 1987) identification problem arises, in which nuisance parameters are not identified under the null. Davies (1977, 1987) proposes statistics whose asymptotic null distributions are functions of Gaussian processes. White (1989*a*) and Lee, White, and Granger (1993) consider statistics that avoid the need to work with functions of a Gaussian process, essentially by selecting nuisance parameters at random. Bierens (1990) and Hansen (1996) consider optimal choice of nuisance parameters, directly confronting the nuisance parameter problem. Hansen (1996) provides general regularity conditions.

Nevertheless, the literature does not take into account the *twofold* nature of the identification problem arising here. Let $(\alpha^*, \boldsymbol{\beta}^*, \lambda^*, \boldsymbol{\delta}^*)$ be parameter values satisfying $f(\mathbf{X}_t\,; \alpha^*, \boldsymbol{\beta}^*, \lambda^*, \boldsymbol{\delta}^*) = E[Y_t|\mathbf{X}_t]$ under $\mathcal{H}_0$. Then $\mathcal{H}_0$ consists of two sub-hypotheses: $\mathcal{H}_0 = \mathcal{H}_{01} \cup \mathcal{H}_{02}$, where

$$\mathcal{H}_{01}: \lambda^* = 0 \quad \text{and} \quad \mathcal{H}_{02}: \boldsymbol{\delta}^* = \mathbf{0}. \tag{1}$$

$\mathcal{H}_{01}$ and $\mathcal{H}_{02}$ are the type 1 and 2 hypotheses mentioned above. Under $\mathcal{H}_{01}$, $\boldsymbol{\delta}^*$ is not identified; that is, the representation for $E[Y_t|\mathbf{X}_t]$ has many possible values for $\boldsymbol{\delta}^*$. Under $\mathcal{H}_{02}$, only $\alpha^* + \lambda^*\Psi(0)$ is identified, and there are many combinations of α^* and λ^* such that $\alpha^* + \lambda^*\Psi(0)$ is identical to the intercept in $E[Y_t|\mathbf{X}_t]$. Thus, Davies's (1977, 1987) identification problem arises in two different ways, each of which requires its own analysis. We therefore call this the *twofold identification problem*.

There are many examples of the twofold identification problem in the statistics lit-

[1] For learning, it can be useful to permit input-to-hidden biases to adapt. Because our interest here is not learning but inference (testing), there is no loss to fixing the input-to-hidden bias. This also greatly simplifies the analysis.

erature. The first that we have been able to find is the mixture model of Neyman and Scott (1965, 1966), where they illustrate use of the locally asymptotically optimal $C(\alpha)$ statistic, also advocated by Lindsay (1995). But their test of the mixture hypothesis using $C(\alpha)$ only focuses on one of two hypotheses yielding their null. Another early example is the conditional heteroskedasticity model of Rosenberg (1973); here also, only one of the hypotheses comprising the homoskedasticity null is tested. There are many other examples of models having twofold identification problems. Nevertheless, almost all test just one component of the null.

Neglect of the twofold null is also common in the ANN context. The LM statistic in White (1989*a*) and Lee, White, and Granger (1993) is designed specifically to test $\mathcal{H}_{01}$ only. They do not consider $\mathcal{H}_{02}$. Nor does Hansen (1996) accommodate the possibility of $\mathcal{H}_{02}$. His regularity conditions may therefore not suffice under $\mathcal{H}_{02}$. The same is also true for the specification test of Bierens (1990). In the ANN context, there is, to the best of our knowledge, no analysis examining the linearity hypothesis under both $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$ simultaneously.

Accordingly, we consider a test statistic that properly takes into account both $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$ to test $\mathcal{H}_0$ versus $\mathcal{H}_1$. Specifically, the quasi-likelihood ratio (QLR) statistic can serve for this purpose. In determining its asymptotic null distribution, we explicitly accommodate the stochastic dependence between the weak limits obtained under $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$.

Our discussion follows the conventions in the literature. Specifically, Bierens (1990) estimates the model of A2 by nonlinear least squares (NLS), which maximizes the quasi-log-likelihood (QL):

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

where n is the the sample size. The QLR test is then defined as

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

where

$$\hat{\sigma}^2_{n,0} := \min_{\alpha,\boldsymbol{\beta}} n^{-1}\sum_{t=1}^{n}\{Y_t - \alpha - \mathbf{X}_t'\boldsymbol{\beta}\}^2, \text{ and}$$

$$\hat{\sigma}_{n,A}^2 := \min_{\alpha,\boldsymbol{\beta},\boldsymbol{\delta},\lambda} n^{-1}\sum_{t=1}^{n}\{Y_t - \alpha - \mathbf{X}_t'\boldsymbol{\beta} - \lambda\Psi(\mathbf{X}_t'\boldsymbol{\delta})\}^2.$$

The analysis of the QLR statistic differs between $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$. For this, it is convenient to consider three different representations for QLR:

$$QLR_n^{(1)} := \left\{n - \min_{\boldsymbol{\delta}}\min_{\lambda}\min_{\alpha,\boldsymbol{\beta}}\frac{1}{\hat{\sigma}_{n,0}^2}\sum_{t=1}^{n}\{Y_t - \alpha - \mathbf{X}_t'\boldsymbol{\beta} - \lambda\Psi(\mathbf{X}_t'\boldsymbol{\delta})\}^2\right\},$$

$$QLR_n^{(2)} := \left\{n - \min_{\lambda}\min_{\boldsymbol{\delta}}\min_{\alpha,\boldsymbol{\beta}}\frac{1}{\hat{\sigma}_{n,0}^2}\sum_{t=1}^{n}\{Y_t - \alpha - \mathbf{X}_t'\boldsymbol{\beta} - \lambda\Psi(\mathbf{X}_t'\boldsymbol{\delta})\}^2\right\}, \quad \text{and}$$

$$QLR_n^{(3)} := \left\{n - \min_{\alpha}\min_{\boldsymbol{\delta}}\min_{\lambda,\boldsymbol{\beta}}\frac{1}{\hat{\sigma}_{n,0}^2}\sum_{t=1}^{n}\{Y_t - \alpha - \mathbf{X}_t'\boldsymbol{\beta} - \lambda\Psi(\mathbf{X}_t'\boldsymbol{\delta})\}^2\right\}.$$

$QLR_n^{(1)}$ is obtained by minimizing with respect to λ before minimizing with respect $\boldsymbol{\delta}$; this makes it convenient for analysis under $\mathcal{H}_{01}$. Under $\mathcal{H}_{01}$, $\boldsymbol{\delta}^*$ is not identified, but this can be addressed by following the approach of Hansen (1996).

In $QLR_n^{(2)}$ and $QLR_n^{(3)}$, the order of minimization is reversed. This makes it convenient for analysis under $\mathcal{H}_{02}$. We first apply a Taylor expansion to QL with respect to $\boldsymbol{\delta}$; we then minimize the approximation with respect to λ and α respectively. We let $\overline{QLR}_n^{(2)}$ and $\overline{QLR}_n^{(3)}$ denote the corresponding approximations. Here, we separately consider $\overline{QLR}_n^{(2)}$ and $\overline{QLR}_n^{(3)}$ to accommodate the fact that there is a continuum of combinations of α^* and λ^* such that $\alpha^* + \lambda^*\Psi(0)$ is identical to the intercept of $E[Y_t|\mathbf{X}_t]$. We overcome this difficulty by first fixing λ^*. This enables us to identify the other parameters $(\alpha^*, \boldsymbol{\delta}^*, \boldsymbol{\beta}^*)$ and apply a Taylor approximation to QL. We then optimize with respect to λ^*. This approximation is denoted as $\overline{QLR}_n^{(2)}$. We then interchange the roles of α^* and λ^* to obtain $\overline{QLR}_n^{(3)}$. Nevertheless, as it turns out, the asymptotic null behaviors of $\overline{QLR}_n^{(2)}$ and $\overline{QLR}_n^{(3)}$ are equivalent, so that only one of them is relevant to the null asymptotic behavior of the QLR test.

We note the following simple fact that

$$QLR_n = \max[QLR_n^{(1)}, QLR_n^{(2)}, QLR_n^{(3)}] = \max[QLR_n^{(1)}, \overline{QLR}_n^{(2)}, \overline{QLR}_n^{(3)}] + o_{\mathbb{P}}(1). \tag{2}$$

The asymptotic distribution of QLR_n is thus determined by the weak limits of $QLR_n^{(1)}$,

$\overline{QLR}_n^{(2)}$, and $\overline{QLR}_n^{(3)}$. We examine these limits and their relationship in detail in the remainder of this section.

2.1 Asymptotic Null Distribution of the QLR Test under $\mathcal{H}_{01}$

The asymptotic null distribution of QLR_n under $\mathcal{H}_{01}$ is already available in the literature. We sketch its derivation to fix notation and motivate the assumptions. Concentrating QL with respect to $(\alpha, \boldsymbol{\beta}')'$ gives

$$L_n^{(1)}(\lambda; \boldsymbol{\delta}) := \max_{\alpha, \boldsymbol{\beta}} L_n(\alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta}) = -\{\mathbf{Y} - \lambda \boldsymbol{\Psi}(\boldsymbol{\delta})\}'\mathbf{M}\{\mathbf{Y} - \lambda \boldsymbol{\Psi}(\boldsymbol{\delta})\}, \tag{3}$$

where $\mathbf{M} := \mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$; $\mathbf{Z} := [\boldsymbol{\iota}, \mathbf{X}]$, with $\mathbf{X}$ the $n \times k$ regressor matrix with rows $\mathbf{X}_t'$ and $\boldsymbol{\iota}$ the $n \times 1$ vector of ones; $\boldsymbol{\Psi}(\boldsymbol{\delta}) := [\Psi(\mathbf{X}_1'\boldsymbol{\delta}), \Psi(\mathbf{X}_2'\boldsymbol{\delta}), \cdots, \Psi(\mathbf{X}_n'\boldsymbol{\delta})]'$; and $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_n)'$. For now, we assume $(\mathbf{Z}'\mathbf{Z})^{-1}$ exists**.** We ensure this below.

We define $\Psi_t(\boldsymbol{\delta}) := \Psi(\mathbf{X}_t'\boldsymbol{\delta})$, $U_t := Y_t - E[Y_t|\mathbf{X}_t]$ and $\mathbf{U} := [U_1, U_2, \cdots, U_n]'$. Since $\mathbf{MY} = \mathbf{MU}$ under $\mathcal{H}_0$, it is standard that

$$\begin{aligned}\sup_{\lambda}\{L_n^{(1)}(\lambda; \boldsymbol{\delta}) - L_n^{(1)}(0; \boldsymbol{\delta})\} &= \sup_{\lambda} 2\lambda \boldsymbol{\Psi}(\boldsymbol{\delta})'\mathbf{MU} - \lambda^2 \boldsymbol{\Psi}(\boldsymbol{\delta})'\mathbf{M}\boldsymbol{\Psi}(\boldsymbol{\delta}) \\ &= \frac{\{\boldsymbol{\Psi}(\boldsymbol{\delta})'\mathbf{MU}\}^2}{\boldsymbol{\Psi}(\boldsymbol{\delta})'\mathbf{M}\boldsymbol{\Psi}(\boldsymbol{\delta})}.\end{aligned} \tag{4}$$

Thus,

$$QLR_n^{(1)} = \sup_{\boldsymbol{\delta}} \frac{\{\boldsymbol{\Psi}(\boldsymbol{\delta})'\mathbf{MU}\}^2}{\hat{\sigma}_{n,0}^2 \boldsymbol{\Psi}(\boldsymbol{\delta})'\mathbf{M}\boldsymbol{\Psi}(\boldsymbol{\delta})}. \tag{5}$$

The asymptotic null behavior of $QLR_n^{(1)}$ is determined by that of $n^{-1/2}\boldsymbol{\Psi}(\cdot)'\mathbf{MU}$ and $n^{-1}\hat{\sigma}_{n,0}^2\boldsymbol{\Psi}(\,\cdot\,)'\,\mathbf{M}\boldsymbol{\Psi}(\,\cdot\,)$ under some regularity conditions. Theorem 1 of Hansen (1996) derives the asymptotic null distribution of an LM statistic; his regularity conditions also apply to $QLR_n^{(1)}$. For this, we impose:

Assumption A3 (Moments): *There exists a sequence of stationary ergodic random variables* $\{M_t\}$ *such that* $|U_t| \leq M_t$, $|X_{t,j}| \leq M_t$, $j = 1, 2, \cdots, k$, *and for some* $\kappa \geq 2(\rho - 1)$, $E[M_t^{4+2\kappa}] < \infty$.

Assumption A4 (Martingale Difference): (i) $E[U_t \mid \mathbf{X}_t, U_{t-1}, \mathbf{X}_{t-1}, \cdots] = 0$; (ii) $E[U_t^2 \mid \mathbf{X}_t] = \sigma_*^2$.

A3 and A4 ensure that σ_*^2, $E[U_t^4]$, and $E[X_{t,j}^4]$, $j = 1, 2, \cdots, k$, are finite.

A4 is not strictly necessary to obtain the asymptotic null distribution of $QLR_n^{(1)}$. Nor does Theorem 1 of Hansen (1996) require this. Nevertheless, the martingale difference assumption of A4(i) can often be plausibly ensured by including sufficient lags of Y_t and other variables in $\mathbf{X}_t$, and it greatly simplifies the covariance structure of the Gaussian processes relevant for our tests. The conditional homoskedasticity (constant conditional variance) assumption in A4(ii) yields further simplifications.

Next, we impose some bounds.

Assumption A5 (Bounds): (i) $\sup_{\boldsymbol{\delta}\in\boldsymbol{\Delta}} |\Psi_t(\boldsymbol{\delta})| \leq M_t$; *and* ($ii$) $\sup_{\boldsymbol{\delta}\in\boldsymbol{\Delta}} |\frac{\partial}{\partial\delta_j}\Psi_t(\boldsymbol{\delta})| \leq M_t$, $j = 1, ..., k+1$.

Assumption A5 is used to show that the numerator of (5) is tight, as a direct consequence of DMR. Assumption 2 of Hansen (1996) pertains here. By our A2, Ψ is analytic in each of its arguments, so both Ψ_t and $(\partial/\partial\delta_j)\Psi_t$ are analytic for each $\mathbf{X}_t$. This ensures that Ψ_t and $(\partial/\partial\delta_j)\Psi_t$ are also Lipschitz continuous on $\boldsymbol{\Delta}$ and therefore bounded for each $\mathbf{X}_t$. A5 places moment conditions on these bounds. In particular, A5(ii) imposes the moment condition for the Lipschitz constant as in assumption 2 of Hansen (1996). Conveniently, we can assume that Ψ and its derivatives are uniformly bounded without losing the GCR property that gives the ANN test its power.

The analysis of $QLR_n^{(1)}$ requires care, since for every n, the numerator of (5) converges to zero $a.s.(-\mathbb{P})$ as $\boldsymbol{\delta}$ tends to $\mathbf{0}$:

$$\lim_{\boldsymbol{\delta}\to\mathbf{0}} \boldsymbol{\Psi}(\boldsymbol{\delta})'\mathbf{M}\mathbf{U} = \Psi(\mathbf{0})\boldsymbol{\iota}'\mathbf{M}\mathbf{U} = 0 \quad a.s.$$

Because $\boldsymbol{\iota}$ is a column of $\mathbf{Z}$, the denominator behaves similarly:

$$\lim_{\boldsymbol{\delta}\to\mathbf{0}} \boldsymbol{\Psi}(\boldsymbol{\delta})'\mathbf{M}\boldsymbol{\Psi}(\boldsymbol{\delta}) = \Psi(\mathbf{0})^2\boldsymbol{\iota}'\mathbf{M}\boldsymbol{\iota} = 0 \quad a.s.$$

This creates difficulties in obtaining the asymptotic null distribution of $QLR_n^{(1)}$ near $\boldsymbol{\delta} = \mathbf{0}$. For now, we avoid these by restricting the parameter space. For given $\epsilon > 0$, define

$$\boldsymbol{\Delta}(\epsilon) := \left\{ \boldsymbol{\delta} \in \boldsymbol{\Delta} : \sum_{j=1}^{k} |\delta_j| \geq \epsilon \right\}.$$

We let $\epsilon \to 0$ below. To ensure asymptotic non-degeneracy, we write $\mathbf{Z}_t := (1, \mathbf{X}_t')'$ and impose

Assumption A6 (Covariance): *For each* $\epsilon > 0$ *and* $\boldsymbol{\delta} \in \boldsymbol{\Delta}(\epsilon)$, $\det \mathbf{V}_1(\boldsymbol{\delta}) > 0$ *and* $\det \mathbf{V}_2(\boldsymbol{\delta}) > 0$, *where*

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

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

Our first formal result describes the null behavior of the numerator and denominator in (5). This is a corollary of Hansen (1996, theorem 1), and it states the weak convergence in continuous function space. We write $\Psi_t^*(\boldsymbol{\delta}) := \Psi_t(\boldsymbol{\delta}) - E[\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t']\{E[\mathbf{Z}_t\mathbf{Z}_t']\}^{-1}\mathbf{Z}_t$.

Lemma 1. *Given A1 to A3, A4*(i)*, A5, A6, and* $\mathcal{H}_{01}$,

(i) $\hat{\sigma}_{n,0}^2 \stackrel{\mathbb{P}}{\to} \sigma_*^2 := E[U_t^2]$;

(ii) *for each* $\epsilon > 0$, $\{n^{-1/2}\boldsymbol{\Psi}(\,\cdot\,)'\mathbf{M}\mathbf{U}, \hat{\sigma}_{n,0}^2 n^{-1}\boldsymbol{\Psi}(\,\cdot\,)'\mathbf{M}\boldsymbol{\Psi}(\,\cdot\,)\} \Rightarrow \{\mathcal{G}_0(\cdot), \mathcal{J}(\cdot,\cdot)\}$ *on* $\boldsymbol{\Delta}(\epsilon)$, *where* $\mathcal{G}_0$ *is a zero-mean continuous Gaussian process such that* $E[\mathcal{G}_0(\boldsymbol{\delta})\mathcal{G}_0(\widetilde{\boldsymbol{\delta}})] = \mathcal{T}(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}})$, *where for each* $\boldsymbol{\delta}$, $\widetilde{\boldsymbol{\delta}}$,

$$\mathcal{T}(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) := E[U_t^2 \Psi_t^*(\boldsymbol{\delta}) \Psi_t^*(\widetilde{\boldsymbol{\delta}})] \quad \textit{and} \quad \mathcal{J}(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) := \sigma_*^2\, E[\Psi_t^*(\boldsymbol{\delta}) \Psi_t^*(\widetilde{\boldsymbol{\delta}})];$$

(iii) *if A4*(ii) *also holds, then* $\mathcal{T}(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) = \mathcal{J}(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}})$.

Applying the continuous mapping theorem and Lemma 1 delivers the asymptotic null behavior of

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

To state the result, let $\mathcal{J}(\boldsymbol{\delta}) = \mathcal{J}(\boldsymbol{\delta}, \boldsymbol{\delta})$ and $\mathcal{G}_1(\boldsymbol{\delta}) := \mathcal{J}(\boldsymbol{\delta})^{-1/2}\mathcal{G}_0(\boldsymbol{\delta})$, so that for each $\boldsymbol{\delta}$ and $\widetilde{\boldsymbol{\delta}}$,

$$E[\mathcal{G}_1(\boldsymbol{\delta})\mathcal{G}_1(\widetilde{\boldsymbol{\delta}})] = \rho_1(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) := \frac{\mathcal{T}(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}})}{\{\mathcal{J}(\boldsymbol{\delta}, \boldsymbol{\delta})\}^{1/2}\{\mathcal{J}(\widetilde{\boldsymbol{\delta}}, \widetilde{\boldsymbol{\delta}})\}^{1/2}}.$$

Theorem 1. *Given A1 to A3, A4(i), A5, A6, and $\mathcal{H}_{01}$, for each $\epsilon > 0$,*

(i) $QLR_n^{(1)}(\epsilon) \Rightarrow \sup_{\boldsymbol{\delta} \in \boldsymbol{\Delta}(\epsilon)} \mathcal{G}_1(\boldsymbol{\delta})^2$;

(ii) *if A4(ii) also holds, then*

$$\rho_1(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) = \frac{\mathcal{J}(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}})}{\{\mathcal{J}(\boldsymbol{\delta}, \boldsymbol{\delta})\}^{1/2}\{\mathcal{J}(\widetilde{\boldsymbol{\delta}}, \widetilde{\boldsymbol{\delta}})\}^{1/2}}.$$

Note that for each $\boldsymbol{\delta}$, $\mathcal{G}_1(\boldsymbol{\delta})$ generally is not standard normal, although when A4(ii) holds, we do have $\mathcal{G}_1(\boldsymbol{\delta}) \sim N(0,1)$. Assuming conditional homoskedasticity (A4(ii)) may restrict the application of the QLR test, as data often exhibit conditional heteroskedasticity. Thus, we will not demand that A4(ii) holds. Nevertheless, this simplifies the analysis and yields more intuitive results, so we will record these.

2.2 Asymptotic Null Distribution of the QLR Test under $\mathcal{H}_{02}$

2.1 Case 1: λ given

We now examine the asymptotic null distribution of the QLR statistic under the type 2 hypothesis. Concentrating QL with given λ yields

$$L_n^{(2)}(\boldsymbol{\delta}; \lambda) := \max_{\alpha, \boldsymbol{\beta}} L_n(\alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta}) = -\{\mathbf{Y} - \lambda \boldsymbol{\Psi}(\boldsymbol{\delta})\}'\mathbf{M}\{\mathbf{Y} - \lambda \boldsymbol{\Psi}(\boldsymbol{\delta})\}. \tag{6}$$

Note that $L_n^{(2)}(\,\cdot\,; \lambda)$ in (6) is a function of $\boldsymbol{\delta}$, whereas $L_n^{(1)}(\,\cdot\,; \boldsymbol{\delta})$ in (3) is a function of λ.

We derive the desired type 2 asymptotic behavior of the QLR statistic using a Taylor series expansion in $\boldsymbol{\delta}$. In standard situations, a second-order Taylor expansion suffices. This fails here, because $\nabla_{\boldsymbol{\delta}} L_n^{(2)}(\mathbf{0}; \lambda) \equiv \mathbf{0}$. Specifically, when $\mathcal{H}_{02}$ holds ($\boldsymbol{\delta}^* = \mathbf{0}$),

$$\frac{\partial}{\partial \delta_i} L_n^{(2)}(\mathbf{0}; \lambda) = -2\lambda c_1 \mathbf{X}_i' \mathbf{M}[\mathbf{Y} - \lambda c_0 \boldsymbol{\iota}] = -2\lambda c_1 \mathbf{X}_i' \mathbf{M} \mathbf{U} \equiv 0,$$

with $c_j := D^j \Psi(0)$, $j = 0, 1, 2, \cdots$, where D^j is the jth derivative operator with respect to the argument of Ψ; and $\mathbf{X}_i := [X_{1,i}, X_{2,i}, \cdots, X_{n,i}]'$ is the ith column of $\mathbf{X}$. We have $\mathbf{X}_i' \mathbf{M} \equiv \mathbf{0}$, as $\mathbf{M}$ is an idempotent matrix of the form $\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, where $\mathbf{X}_i$ is a column of $\mathbf{Z}$.

As it turns out, a fourth-order Taylor approximation suffices. The next lemma collects together the relevant higher-order derivatives under $\mathcal{H}_{02}$. For this, let $\mathbf{D}_i := \text{diag}\{\mathbf{X}_i\}$, $\mathbf{D}_{ij} := \mathbf{D}_i\mathbf{D}_j$, $\mathbf{D}_{ij\ell} := \mathbf{D}_i\mathbf{D}_j\mathbf{D}_\ell$, and $\mathbf{D}_{ij\ell m} := \mathbf{D}_i\mathbf{D}_j\mathbf{D}_\ell\mathbf{D}_m$, $i, j, \ell, m = 1, 2, \cdots, k$.

Lemma 2. *Given A1 and A2, for* $i, j, \ell, m = 1, 2, \cdots, k$,

(*i*) $\frac{\partial}{\partial \delta_i} L_n^{(2)}(\mathbf{0}; \lambda) = 0$;

(*ii*) $\frac{\partial^2}{\partial \delta_i \partial \delta_j} L_n^{(2)}(\mathbf{0}; \lambda) = 2\lambda c_2 \boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{MU}$;

(*iii*) $\frac{\partial^3}{\partial \delta_i \partial \delta_j \partial \delta_\ell} L_n^{(2)}(\mathbf{0}; \lambda) = 2\lambda c_3 \boldsymbol{\iota}'\mathbf{D}_{ij\ell}\mathbf{MU}$; *and*

(*iv*) $\frac{\partial^4}{\partial \delta_i \partial \delta_j \partial \delta_\ell \partial \delta_m} L_n^{(2)}(\mathbf{0}; \lambda) = 2\lambda c_4 \boldsymbol{\iota}'\mathbf{D}_{ij\ell m}\mathbf{MU} - 2\lambda^2 c_2^2 \boldsymbol{\iota}'[\mathbf{D}_{ij}\mathbf{MD}_{\ell m} + \mathbf{D}_{i\ell}\mathbf{MD}_{jm} + \mathbf{D}_{im}\mathbf{MD}_{j\ell}]\boldsymbol{\iota}$.

As these results are easily derived, we omit the proof from the Appendix.

We can apply the law of large numbers and central limit theorem (CLT) to the second-, third-, and fourth-order derivatives above. For this, we strengthen A3 to accommodate the higher order terms of the quartic expansion.

Assumption A3* (Moments): $E|U_t|^8 < \infty$ *and* $E|X_{t,i}|^8 < \infty$; *or* $E|U_t|^4 < \infty$ *and* $E|X_{t,i}|^{16} < \infty$, $i = 1, 2, \cdots, k$.

Lemma 3. *Given A1, A2, A3*, A4(i), A6, and* $\mathcal{H}_{02}$, *for* $i, j, \ell, m = 1, 2, \cdots, k$,

(*i*) $\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{MU} = O_{\mathbb{P}}(n^{1/2})$;

(*ii*) $\boldsymbol{\iota}'\mathbf{D}_{ij\ell}\mathbf{MU} = o_{\mathbb{P}}(n^{3/4})$;

(*iii*) $\boldsymbol{\iota}'\mathbf{D}_{ij\ell m}\mathbf{MU} = o_{\mathbb{P}}(n)$; *and*

(*iv*) $\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{MD}_{\ell m}\boldsymbol{\iota} = O_{\mathbb{P}}(n)$.

Lemma 3 also implies that the other terms in Lemma 2(iv) are $O_{\mathbb{P}}(n)$, allowing us to write

$$L_n^{(2)}(\boldsymbol{\delta}; \lambda) - L_n^{(2)}(\mathbf{0}; \lambda) = \lambda c_2 \sum_{i=1}^k \sum_{j=1}^k (\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{MU})\delta_i\delta_j - \frac{1}{4}\lambda^2 c_2^2 \sum_{i=1}^k \sum_{j=1}^k \sum_{\ell=1}^k \sum_{m=1}^k (\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{MD}_{\ell m}\boldsymbol{\iota})\delta_i\delta_j\delta_\ell\delta_m + O_{\mathbb{P}}(n^{-1/4}). \quad (7)$$

We see that the second- and fourth-order terms are the main factors driving QLR asymptotically under $\mathcal{H}_{02}$, provided $c_2 \neq 0$. (Note that when $\lambda = 0$, the left hand side of (7) vanishes, so λ and c_2 play different roles here.) To avoid the complexities introduced when c_2 can be zero, we impose

Assumption A7 (No Zero): $c_2 \neq 0$.

This rules out choosing Ψ to be the logistic CDF or a ridgelet with zero input-to-hidden bias. But one can simply "bias-shift" $\mathbf{X}_t'\boldsymbol{\delta}$ to $c + \mathbf{X}_t'\boldsymbol{\delta}$, where c is chosen explicitly to ensure $c_2 \neq 0$. For this, we can replace $\Psi(\mathbf{X}_t'\boldsymbol{\delta})$ with $\Psi_c(\mathbf{X}_t'\boldsymbol{\delta}) := \Psi(c+\mathbf{X}_t'\boldsymbol{\delta})$. We leave such shifts implicit in what follows. Alternatively, $\exp(\cdot)$ or antiderivatives of CDFs are convenient admissible choices. The latter are appealing, as A3 often suffices to ensure that A5 holds for these. Further, closed form expressions exist for useful classes of antiderivatives; see, for example, Giacomini, et al. (2008), who give convenient expressions for antiderivatives of the Student $t-$distribution CDF.

Assumption A7 can be relaxed, but at a material cost. Specifically, when $c_2 = 0$, sixth- or even higher order Taylor expansions are required. This requires additional moment and other regularity conditions, and the resulting asymptotic distributions inconveniently differ from those under A7.

The quadruple sums can be simplified using matrix notation. For $i, j, \ell, m = 1, 2, \cdots, k$, let

$$\widetilde{\mathbf{M}} := [\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{U}], \quad \mathbf{W} := [\mathbf{W}_{ij}], \quad \mathbf{W}_{ij} = [\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{D}_{\ell m}\boldsymbol{\iota}].$$

Note that $\mathbf{W}$ is a $k^2 \times k^2$ matrix. Then eq. (7) becomes

$$L_n^{(2)}(\boldsymbol{\delta};\lambda) - L_n^{(2)}(\mathbf{0};\lambda) = \lambda c_2 \boldsymbol{\delta}'\widetilde{\mathbf{M}}\boldsymbol{\delta} - \frac{\lambda^2}{4}c_2^2\{\boldsymbol{\delta}'(\mathbf{I}_k \otimes \boldsymbol{\delta})'\mathbf{W}(\mathbf{I}_k \otimes \boldsymbol{\delta})\boldsymbol{\delta}\} + O_{\mathbb{P}}(n^{-1/4}). \quad (8)$$

The first two terms on the right survive asymptotically, and the third vanishes. We thus write

$$\widetilde{QLR}_n^{(2)}(\boldsymbol{\delta};\lambda) = \frac{1}{\hat{\sigma}_{n,0}^2}\left\{\lambda c_2(\boldsymbol{\delta}'\widetilde{\mathbf{M}}\boldsymbol{\delta}) - \frac{1}{4}\lambda^2 c_2^2\{\boldsymbol{\delta}'(\mathbf{I}_k \otimes \boldsymbol{\delta})'\mathbf{W}(\mathbf{I}_k \otimes \boldsymbol{\delta})\boldsymbol{\delta}\}\right\}. \quad (9)$$

As in the standard case, the asymptotic distribution of the QLR statistic under $\mathcal{H}_{02}$ obtains by maximizing (9) with respect to $\boldsymbol{\delta}$. Nevertheless, maximizing a quartic with respect to $\boldsymbol{\delta}$ is much more cumbersome than maximizing a quadratic. We simplify by decomposing $\boldsymbol{\delta} - \boldsymbol{\delta}_*$ into a direction $\boldsymbol{d}$ and a distance h:

$$\boldsymbol{\delta} = \boldsymbol{\delta}^* + h\boldsymbol{d}, \tag{10}$$

where $h \in \mathbb{R}^+$ and $\boldsymbol{d} \in \mathbb{S}^{k-1} := \{\boldsymbol{\delta} \in \mathbb{R}^k : \boldsymbol{\delta}'\boldsymbol{\delta} = 1\}$. Under $\mathcal{H}_{02}$, $\boldsymbol{\delta} = h\boldsymbol{d}$; then maximizing $\widetilde{QLR}_n^{(2)}(\,\cdot\,;\lambda)$ with respect to $\boldsymbol{\delta}$ can be written as a two-stage problem:

$$\overline{QLR}_n^{(2)}(\lambda) := \sup_{\boldsymbol{\delta}} \widetilde{QLR}_n^{(2)}(\boldsymbol{\delta};\lambda) = \sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} \sup_{h\in\mathbb{R}^+} \widetilde{QLR}_n^{(2)}(h\boldsymbol{d};\lambda). \tag{11}$$

Combining (9) and (11) gives

$$\overline{QLR}_n^{(2)}(\lambda) = \sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} \sup_{h\in\mathbb{R}^+} \frac{1}{\hat{\sigma}_{n,0}^2} \left\{ \lambda c_2(\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d})h^2 - \frac{1}{4}\lambda^2 c_2^2 \{\boldsymbol{d}'(\mathbf{I}_k \otimes \boldsymbol{d})'\mathbf{W}(\mathbf{I}_k \otimes \boldsymbol{d})\boldsymbol{d}\} h^4 \right\} \tag{12}$$

$$= \sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} \frac{\max[\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d}, 0]^2}{\hat{\sigma}_{n,0}^2 \{\boldsymbol{d}'(\mathbf{I}_k \otimes \boldsymbol{d})'\mathbf{W}(\mathbf{I}_k \otimes \boldsymbol{d})\boldsymbol{d}\}}$$

$$= \sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} \max\left[\frac{\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d}}{\sigma_*\{\boldsymbol{d}'(\mathbf{I}_k \otimes \boldsymbol{d})'\mathbf{W}(\mathbf{I}_k \otimes \boldsymbol{d})\boldsymbol{d}\}^{1/2}}, 0 \right]^2 + o_{\mathbb{P}}(1),$$

where the $\max$ operator accommodates $h \geq 0$. If $\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d} \leq 0$, then maximizing with respect to h gives $h = 0$, which implies that $\overline{QLR}_n^{(2)}(\lambda) = 0$, as when $\lambda = 0$. Otherwise,

$$h^2 = \frac{2(\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d})}{\lambda c_2\{\boldsymbol{d}'(\mathbf{I}_k \otimes \boldsymbol{d})'\mathbf{W}(\mathbf{I}_k \otimes \boldsymbol{d})\boldsymbol{d}\}}.$$

Thus, $\overline{QLR}_n^{(2)}(\lambda)$ has mass at zero under $\mathcal{H}_{02}$. Also, the factors λ and c_2 cancel in the maximization, so the $\mathcal{H}_{02}$ asymptotic distribution is nuisance parameter-free. Thus, we write $\overline{QLR}_n^{(2)} = \overline{QLR}_n^{(2)}(\lambda)$. This and eq. (7) also imply $QLR_n^{(2)} = \overline{QLR}_n^{(2)} + o_{\mathbb{P}}(1)$.

The quartic structure of eq. (12) is similar to the conventional quadratic approximation. That is, the second-order derivatives determine the asymptotic distribution, whereas the fourth-order derivatives converge to a deterministic matrix. This corresponds to the standard quadratic approximation, where the first- and second-order derivatives determine the asymptotic distribution and converge to a deterministic matrix, respectively. Further, the fourth-order derivatives are closely related to the asymptotic covariance of the second-order derivatives. This is similar to the quadratic approximation, where the second-order derivatives are closely related to the asymptotic covariance of the first-order derivatives. This can be clearly demonstrated by noting that $\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d}$ =vec$(\boldsymbol{d}\boldsymbol{d}')'vec(\widetilde{\mathbf{M}})$, $\boldsymbol{d}'(\mathbf{I}_k \otimes \boldsymbol{d})'\mathbf{W}(\mathbf{I}_k \otimes \boldsymbol{d})\boldsymbol{d} = \text{vec}(\boldsymbol{d}\boldsymbol{d}')'\mathbf{W}\text{vec}(\boldsymbol{d}\boldsymbol{d}')$, and $\text{vec}(\boldsymbol{d}\boldsymbol{d}')' \text{ vec}(\boldsymbol{d}\boldsymbol{d}') = 1$. Letting $\boldsymbol{b}$:=vec$(\boldsymbol{d}\boldsymbol{d}') \in \mathbb{S}_c^{k^2-1} := \{\boldsymbol{b} \in \mathbb{S}^{k^2-1} : \boldsymbol{b} = \text{vec}(\boldsymbol{d}\boldsymbol{d}')$, $\boldsymbol{d} \in \mathbb{S}^{k-1}\}$, we can write $\overline{QLR}_n^{(2)}$ more compactly as

$$\overline{QLR}_n^{(2)} = \sup_{\boldsymbol{b} \in \mathbb{S}_c^{k^2-1}} \max \left[\frac{\boldsymbol{b}'\text{vec}(\widetilde{\mathbf{M}})}{\{\hat{\sigma}_{n,0}^2 \boldsymbol{b}'\mathbf{W}\boldsymbol{b}\}^{1/2}}, 0 \right]^2 .$$

It is not hard to show that the variance of $n^{-1/2}\boldsymbol{b}'\text{vec } \widetilde{\mathbf{M}}$ is asymptotically equivalent to $n^{-1}\hat{\sigma}_{n,0}^2 \boldsymbol{b}'\mathbf{W}\boldsymbol{b}$ under conditional homoskedasticity, as we see below.

The limiting distribution of $\overline{QLR}_n^{(2)}$ is driven mainly by the terms in Lemma 3(i), with typical element

$$\begin{aligned} \frac{1}{\sqrt{n}}\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{U} =& \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t,i}X_{t,j}U_t \\ & - \left(\frac{1}{n}\sum_{t=1}^{n} X_{t,i}X_{t,j}\mathbf{Z}_t'\right)\left(\frac{1}{n}\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{Z}_t'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{Z}_t U_t\right), \end{aligned}$$

where $n^{-1/2}\sum_{t=1}^{n} X_{t,i}X_{t,j}U_t$ and $n^{-1/2}\sum_{t=1}^{n}\mathbf{Z}_t U_t$ are scores from the second- and first-order derivatives respectively. The joint asymptotic normality of these terms holds by the multivariate CLT. No further terms contribute, due essentially to the degeneracy of third-order derivatives under $\mathcal{H}_{02}$.

To ensure the non-degeneracy of the relevant limiting distribution, we impose

Assumption A6* (Covariance): *Let* $\mathbf{C}_t := \text{vech}(\mathbf{X}_t\mathbf{X}_t')$. *Suppose* $\det \widetilde{\mathbf{V}}_1 > 0$ *and* $\det \widetilde{\mathbf{V}}_2 > 0$, *where*

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

We use the vech operator to avoid entering the common elements of $\mathbf{X}_t\mathbf{X}_t'$ twice.

We can now obtain the limiting behavior of the components of $\overline{QLR}_n^{(2)}$. For this, we let $C_{t,ij}^* := X_{t,i}X_{t,j} - E[X_{t,i}X_{t,j}\mathbf{Z}_t']\{E[\mathbf{Z}_t\mathbf{Z}_t']\}^{-1}\mathbf{Z}_t$, and $\mathbf{C}_t^* := [C_{t,ij}^*]$, a $k \times k$ matrix with $C_{t,ij}^*$ as its i-th row and j-th column element.

Lemma 4. *Given A1, A2, A3*, A4(i), A6*, and $\mathcal{H}_{02}$, for each $\boldsymbol{b} \in \mathbb{S}_c^{k^2-1}$,*

(i) $n^{-1/2}\boldsymbol{b}'\text{vec}(\widetilde{\mathbf{M}}) \Rightarrow \boldsymbol{b}'\text{vec}(\mathcal{M})$, *where* $\mathcal{M} := [\mathcal{M}_{ij}]$ *is a* $k \times k$ *symmetric matrix of jointly normal random variables such that for* $i, j, \ell, m = 1, 2, \cdots, k$, $E(\mathcal{M}_{ij}) = 0$ *and* $E(\mathcal{M}_{ij}\mathcal{M}_{\ell m}) = E(U_t^2 C_{t,ij}^* C_{t,\ell m}^*)$;

(ii) $n^{-1}\boldsymbol{b}'\mathbf{W}\boldsymbol{b} \to \boldsymbol{b}'\mathbf{W}^*\boldsymbol{b}$ *a.s.*, *where* $\mathbf{W}^* := [\mathbf{W}_{ij}^*]$ *and* $\mathbf{W}_{ij}^* := [\tau_{ij\ell m}]$, *where* $\tau_{ij\ell m} := E(C_{t,ij}^* C_{t,\ell m}^*)$;

(iii) *if A4*(ii) *also holds, for* $i, j, \ell, m = 1, 2, \cdots, k$, $E(\mathcal{M}_{ij}\mathcal{M}_{\ell m}) = \sigma_*^2 E(C_{t,ij}^*\, C_{t,\ell m}^*)$.

We use Lemma 4 to obtain the asymptotic behavior of (11) under $\mathcal{H}_{02}$, as

$$\sup_{\boldsymbol{\delta}} \frac{1}{\hat{\sigma}_{n,0}^2}\{L_n^{(2)}(\boldsymbol{\delta};\lambda) - L_n^{(2)}(\mathbf{0};\lambda)\} = \overline{QLR}_n^{(2)} + o_{\mathbb{P}}(1)$$
$$\Rightarrow \sup_{\boldsymbol{b}\in\mathbb{S}_c^{k^2-1}} \max\left[\frac{\boldsymbol{b}'\text{vec}(\mathcal{M})}{\{\sigma_*^2\boldsymbol{b}'\mathbf{W}^*\boldsymbol{b}\}^{1/2}}, 0\right]^2.$$

Formally, we have

Theorem 2. *Given A1, A2, A3*, A4(i), A6*, A7, and $\mathcal{H}_{02}$,*

(i) $\overline{QLR}_n^{(2)} \Rightarrow \sup_{\boldsymbol{b}\in\mathbb{S}_c^{k^2-1}} \max[\mathcal{G}_2(\boldsymbol{b}), 0]^2$, *where* $\mathcal{G}_2$ *is a Gaussian process defined on* $\mathbb{S}_c^{k^2-1}$ *such that for each* $\boldsymbol{b}$ *and* $\widetilde{\boldsymbol{b}}$, $E[\mathcal{G}_2(\boldsymbol{b})] = 0$ *and*

$$E[\mathcal{G}_2(\boldsymbol{b})\mathcal{G}_2(\widetilde{\boldsymbol{b}})] = \rho_2(\boldsymbol{b}, \widetilde{\boldsymbol{b}}) := \frac{\mathcal{K}(\boldsymbol{b}, \widetilde{\boldsymbol{b}})}{\mathcal{I}(\boldsymbol{b}, \boldsymbol{b})^{1/2}\mathcal{I}(\widetilde{\boldsymbol{b}}, \widetilde{\boldsymbol{b}})^{1/2}}, \tag{13}$$

where $\mathcal{K}(\boldsymbol{b}, \widetilde{\boldsymbol{b}}) := \boldsymbol{b}'E[U_t^2\text{vec}(\mathbf{C}_t^*)\text{vec}(\mathbf{C}_t^*)']\widetilde{\boldsymbol{b}}$, *and* $\mathcal{I}(\boldsymbol{b}, \widetilde{\boldsymbol{b}}) := \sigma_*^2\boldsymbol{b}'\mathbf{W}^*\widetilde{\boldsymbol{b}}$;

(ii) *if A4*(ii) *also holds,*

$$\rho_2(\boldsymbol{b}, \widetilde{\boldsymbol{b}}) = \frac{\mathcal{I}(\boldsymbol{b}, \widetilde{\boldsymbol{b}})}{\mathcal{I}(\boldsymbol{b}, \boldsymbol{b})^{1/2}\mathcal{I}(\widetilde{\boldsymbol{b}}, \widetilde{\boldsymbol{b}})^{1/2}}.$$

Note that the Gaussian process in Theorem 2 is defined simply by

$$\mathcal{G}_2(\boldsymbol{b}) := \frac{\boldsymbol{b}'\text{vec}(\mathcal{M})}{\{\sigma_*^2\boldsymbol{b}'\mathbf{W}^*\boldsymbol{b}\}^{1/2}}.$$

We offer several remarks before proceeding. First, note that $\mathcal{G}_2$ is indexed by a direction $\boldsymbol{d}$ from the origin (through $\boldsymbol{b}$) rather than by the parameter λ unidentified under $\mathcal{H}_{02}$. QLR is thus nuisance parameter-free under $\mathcal{H}_{02}$, a remarkable fact. As we show in the next subsection, $\mathcal{G}_2$, indexed by $\boldsymbol{d}$, is a special case of $\mathcal{G}_1$, indexed by $\boldsymbol{\delta}$. This implies that under $\mathcal{H}_{02}$, the asymptotic distribution of QLR is essentially governed by $\mathcal{H}_{01}$, justifying the use of $\mathcal{G}_1$ previously employed for hypothesis testing in this context. We provide further details in Section 2.3 below. Second, the fourth-order Taylor expansion has previously been found helpful. Bartlett (1953*a*, *b*) first examines quartic approximations of statistical models. McCullagh (1987) also considers quartic approximations using tensors. Here, using h and $\boldsymbol{d}$, we can avoid the more cumbersome use of tensors, permitting us to apply the methods of Cho and White (2009) and enabling us to readily associate the asymptotic distributions under $\mathcal{H}_{02}$ and $\mathcal{H}_{01}$. Third, the approximation is especially straightforward if $k = 1$. Then $\mathbb{S}^{k-1} = \{-1, 1\}$, so for each $\boldsymbol{d} \in \mathbb{S}^{k-1}$, $\boldsymbol{b} = 1$, implying that $\mathcal{G}_2$ is free of $\boldsymbol{b}$ and $\overline{QLR}_n^{(2)} \Rightarrow \max[\mathcal{Z}, 0]^2$, where $\mathcal{Z}$ is normal with mean zero under $\mathcal{H}_{02}$. Under conditional homoskedasticity (A4(ii)), $\mathcal{Z} \sim N(0, 1)$. Further, under A4(ii), the moment conditions in A3 can be relaxed to require only $E|X_{t,i}|^6 < \infty$, without adverse consequences for Lemma 2(iii) or Theorem 2(ii). Finally, similar approaches are those of Dacunha-Castelle and Gassiat (1999) and Cho and White (2007). These authors treat a one-dimensional version of this case in obtaining the asymptotic null distribution of a likelihood ratio statistic for testing a mixture hypothesis. The current results extend this to the multi-dimensional case and resolve the associated difficulties by use of h and $\boldsymbol{d}$.

2.2 Case 2: α given

We continue our analysis of the QLR statistic under the type 2 hypothesis, but now we suppose that α is given and concentrate QL with respect to λ and $\boldsymbol{\beta}$. Then

$$L_n^{(3)}(\boldsymbol{\delta}; \alpha) := \max_{\lambda, \boldsymbol{\beta}} L_n(\alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta}) = -(\mathbf{Y} - \alpha\boldsymbol{\iota})'\mathbf{P}(\boldsymbol{\delta})(\mathbf{Y} - \alpha\boldsymbol{\iota}), \tag{14}$$

where $\mathbf{P}(\boldsymbol{\delta}) := \mathbf{I} - \mathbf{Q}(\boldsymbol{\delta})[\mathbf{Q}(\boldsymbol{\delta})'\mathbf{Q}(\boldsymbol{\delta})]^{-1}\mathbf{Q}(\boldsymbol{\delta})'$, and $\mathbf{Q}(\boldsymbol{\delta}) := [\mathbf{X}, \boldsymbol{\Psi}(\boldsymbol{\delta})]$. Note that the concentrated QL here corresponds to $L_n^{(2)}(\,\cdot\,; \lambda)$ in eq. (6). $L_n^{(3)}(\cdot; \alpha)$ is a function of $\boldsymbol{\delta}$, but α is now given, unlike eq. (6), where λ is given.

This separate approach is needed, as fixing α could, in general, yield an asymptotic null distribution different from that of Theorem 2. For example, if the hidden unit activation is generalized to $\Psi(\mathbf{X}_t, \boldsymbol{\delta})$, so that $\boldsymbol{\delta}$ can interact arbitrarily with $\mathbf{X}_t$, we can show that a different asymptotic null distribution is obtained. This motivates us to separately consider the fixed-α case. As we now show, however, the asymptotic null distribution does turn out to be the same for our specific case. The intuition is straightforward. As one of the referees points out, we could reparameterize the model in A2 as

$$\tilde{f}(\mathbf{X}_t; \alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\delta}) := \alpha + \mathbf{X}_t'\boldsymbol{\beta} + \lambda[\Psi(\mathbf{X}_t'\boldsymbol{\delta}) - \Psi(0)]$$

and test $\mathcal{H}_{02}$. In this case, λ_* is not identified but $(\alpha_*, \boldsymbol{\beta}_*)$ is identified; further, we can obtain the same conclusion as in Theorem 2 by the invariance principle. This also implies that the fixed-α case should yield the same conclusion as in Theorem 2. We now confirm this intuition under A2.

We first examine the derivatives of $L_n^{(3)}$, as we again require a quartic expansion. We also exploit the direction and distance method of Section 2.2.1, imposing $\mathcal{H}_{02}$ and letting $\boldsymbol{\delta} = h\boldsymbol{d}$ as in eq. (10).

Lemma 5. *Given A1 and A2,*

(i) $\frac{\partial}{\partial h} L_n^{(3)}(\mathbf{0}; \alpha) = 0$;

(ii) $\frac{\partial^2}{\partial h^2} L_n^{(3)}(\mathbf{0}; \alpha) = 4\boldsymbol{\gamma}^{*\prime}\mathbf{J}_2\mathbf{M}\mathbf{U} + 2\mathbf{U}'\mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{J}_2'\mathbf{M}\mathbf{U}$, *where* $\boldsymbol{\gamma}^* := [\boldsymbol{\beta}^{*\prime *} - \alpha)/c_0]'$, $\mathbf{J}_j := \frac{\partial^j}{\partial h^j}\mathbf{Q}(h\boldsymbol{d})|_{h=0}$, $\mathbf{H}_j := \frac{\partial^j}{\partial h^j}\mathbf{Q}(h\boldsymbol{d})'\mathbf{Q}(h\boldsymbol{d})|_{h=0}$, $j = 0, 1, 2, \cdots$;

(iii) $\frac{\partial^3}{\partial h^3} L_n^{(3)}(\mathbf{0}; \alpha) = \gamma^{*\prime}\mathbf{J}_0'[\frac{\partial^3}{\partial h^3}\mathbf{P}(h\boldsymbol{d})|_{h=0}]\mathbf{J}_0\gamma^* + o_{\mathbb{P}}(n^{3/4})$;

(iv) $\frac{\partial^4}{\partial h^4} L_n^{(3)}(\mathbf{0}; \alpha) = -6\boldsymbol{\gamma}^{*\prime}\mathbf{J}_2'\mathbf{M}\mathbf{J}_2\boldsymbol{\gamma}^* + o_{\mathbb{P}}(n)$.

Here, $\boldsymbol{\gamma}^*$, $\mathbf{J}_j$, $\mathbf{H}_j$ are in fact functions of α and/or $\boldsymbol{d}$. We suppress this dependence for notational simplicity. Also, the given derivatives are well defined, as the associated parameters c_j's are well defined under our assumptions. In particular, A2 ensures that $c_0 \neq 0$, so that c_0^{-1} is also well defined.

The derivatives in Lemma 5 are not identical to those in Lemma 2, but they are asymptotically equivalent, as the following lemma shows.

Lemma 6. *Given A1, A2, A3***, A4*(i)*, A6, and* $\mathcal{H}_{02}$,

(i) $\frac{\partial^2}{\partial h^2} L_n^{(3)}(\mathbf{0}; \alpha) = 2(\alpha^* - \alpha)(c_2/c_0)\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d} + o_{\mathbb{P}}(n^{1/2})$;

(ii) $\frac{\partial^3}{\partial h^3} L_n^{(3)}(\mathbf{0};\alpha) = o_{\mathbb{P}}(n^{3/4})$; *and*

(iii) $\frac{\partial^4}{\partial h^4} L_n^{(3)}(\mathbf{0};\alpha) = -6(\alpha^* - \alpha)^2 (c_2/c_0)^2 \boldsymbol{d}'(\mathbf{I}_k \otimes \boldsymbol{d})'\mathbf{W}(\mathbf{I}_k \otimes \boldsymbol{d})\boldsymbol{d} + o_{\mathbb{P}}(n)$.

By Lemma 6, all but the second and fourth-order derivatives vanish in probability. We can therefore proceed in a manner parallel to Section 2.2.1. That is, if we let

$$\begin{aligned}\widetilde{QLR}_n^{(3)}(h\boldsymbol{d};\alpha) :=& \frac{1}{\hat{\sigma}_{n,0}^2}\left(\frac{c_2}{c_0}\right)(\alpha^* - \alpha)\, \boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d}\, h^2 \\ & - \frac{1}{4}\frac{1}{\hat{\sigma}_{n,0}^2}\left(\frac{c_2}{c_0}\right)^2 (\alpha^* - \alpha)^2 \{\boldsymbol{d}'(\mathbf{I}_k \otimes \boldsymbol{d})'\mathbf{W}(\mathbf{I}_k \otimes \boldsymbol{d})\boldsymbol{d}\} h^4\end{aligned}$$

and

$$\overline{QLR}_n^{(3)}(\alpha) := \sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} \sup_{h\in\mathbb{R}^+} \widetilde{QLR}_n^{(3)}(h\boldsymbol{d};\alpha),$$

then it follows that

$$\overline{QLR}_n^{(3)}(\alpha) = \sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} \frac{\max[\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d}, 0]^2}{\hat{\sigma}_{n,0}^2\{\boldsymbol{d}'(\mathbf{I}_k \otimes \boldsymbol{d})'\mathbf{W}(\mathbf{I}_k \otimes \boldsymbol{d})\boldsymbol{d}\}}, \tag{15}$$

and we can write $\overline{QLR}_n^{(3)} = \overline{QLR}_n^{(3)}(\alpha)$, as α does not appear on the RHS of (15). Further, the RHS of (15) is identical to that of (12), and this implies that the asymptotic distribution of $\overline{QLR}_n^{(3)}$ coincides with that given in Theorem 2. We summarize with the following corollary.

Corollary 1. *Given A1, A2, A3***, A4*(i)*, A6***, A7, and* $\mathcal{H}_{02}$,

(i) $\overline{QLR}_n^{(3)} \Rightarrow \sup_{\boldsymbol{b}\in\mathbb{S}_c^{k^2-1}} \max[\mathcal{G}_2(\boldsymbol{b}), 0]^2$; *and*

(ii) $\overline{QLR}_n^{(2)} - \overline{QLR}_n^{(3)} = o_{\mathbb{P}}(1)$.

The proof of Corollary 1(i) is identical to that of Theorem 2. Corollary 1(ii) immediately follows from the fact that the RHS of (15) is identical to that of (12). We thus do not prove Corollary 1 in the Appendix.

Since the weak limits are the same for both cases of $\mathcal{H}_{02}$, it now suffices just to relate the weak limit of Theorem 2 to that of Theorem 1 to obtain the asymptotic null distribution under $\mathcal{H}_0$.

2.3 Asymptotic Null Distribution of the QLR Statistic under $\mathcal{H}_0$

The behaviors of $QLR_n^{(1)}$ and $\overline{QLR}_n^{(2)}$ (equivalently $\overline{QLR}_n^{(3)}$) are related under $\mathcal{H}_0$. Specifically, we show that $QLR_n^{(1)}$ converges to $\overline{QLR}_n^{(2)}$ as $\boldsymbol{\delta}$ converges to $\mathbf{0}$. For this, let $\boldsymbol{\delta} = h\boldsymbol{d}$ as above, and define

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

Then we can write $\sup_\lambda\{L_n(\lambda; \boldsymbol{\delta}) - L_n(0; \boldsymbol{\delta})\}$ in eq. (4) as

$$\frac{N_n(h, \boldsymbol{d})}{D_n(h, \boldsymbol{d})} = \frac{\{\boldsymbol{\Psi}(\boldsymbol{\delta})'\mathbf{M}\mathbf{U}\}^2}{\boldsymbol{\Psi}(\boldsymbol{\delta})'\mathbf{M}\boldsymbol{\Psi}(\boldsymbol{\delta})}.$$

Our next result describes the behavior of this ratio as h converges to zero.

Lemma 7. *Given A1 and A2, for each* n *and* $\boldsymbol{d}$,

(i) *for* $\ell = 0, 1, 2, 3$, $\lim_{h\downarrow 0} N_n^{(\ell)}(h, \boldsymbol{d}) = 0$ *a.s. and* $\lim_{h\downarrow 0} D_n^{(\ell)}(h, \boldsymbol{d}) = 0$ *a.s., where* $N_n^{(\ell)}(h, \boldsymbol{d}) := (\partial^\ell/\partial h^\ell)N_n(h, \boldsymbol{d})$, *and* $D_n^{(\ell)}(h, \boldsymbol{d}) := (\partial^\ell/\partial h^\ell)D_n(h, \boldsymbol{d})$;

(ii) $\lim_{h\downarrow 0} N_n^{(4)}(h, \boldsymbol{d}) = 6c_2^2\{\sum_{i=1}^k \sum_{j=1}^k \boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{U}d_i d_j\}^2$ *a.s.; and*

(iii) $\lim_{h\downarrow 0} D_n^{(4)}(h, \boldsymbol{d}) = 6c_2^2\{\sum_{i=1}^k \sum_{j=1}^k \sum_{\ell=1}^k \sum_{m=1}^k \boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{D}_{\ell m}\boldsymbol{\iota}\, d_i d_j d_\ell d_m\}$ *a.s.*

As Lemma 7(i) trivially holds, we prove only Lemma 7(ii and iii) in the Appendix.

Given Lemma 7, L'Hôspital's rule gives

$$\lim_{h\downarrow 0} \frac{N_n(h, \boldsymbol{d})}{\hat{\sigma}_{n,0}^2 D_n(h, \boldsymbol{d})} = \frac{\lim_{h\downarrow 0} N_n^{(4)}(h, \boldsymbol{d})}{\lim_{h\downarrow 0} \hat{\sigma}_{n,0}^2 D_n^{(4)}(h, \boldsymbol{d})} = \frac{\{\boldsymbol{b}'\text{vec}(\widetilde{\mathbf{M}})\}^2}{\hat{\sigma}_{n,0}^2 \boldsymbol{b}'\mathbf{W}\boldsymbol{b}} \quad a.s. \tag{16}$$

This implies $QLR_n^{(1)} \geq \overline{QLR}_n^{(2)}$, as

$$\begin{aligned} QLR_n^{(1)} &= \sup_{h,\boldsymbol{d}} \frac{\{\boldsymbol{\Psi}(h, \boldsymbol{d})'\mathbf{M}\mathbf{U}\}^2}{\hat{\sigma}_{n,0}^2 \boldsymbol{\Psi}(h, \boldsymbol{d})'\mathbf{M}\boldsymbol{\Psi}(h, \boldsymbol{d})} \geq \sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} \lim_{h\downarrow 0} \frac{\{\boldsymbol{\Psi}(h, \boldsymbol{d})'\mathbf{M}\mathbf{U}\}^2}{\hat{\sigma}_{n,0}^2 \boldsymbol{\Psi}(h, \boldsymbol{d})'\mathbf{M}\boldsymbol{\Psi}(h, \boldsymbol{d})} \\ &= \sup_{\boldsymbol{b}\in\mathbb{S}_c^{k^2-1}} \frac{\{\boldsymbol{b}'\text{vec}(\widetilde{\mathbf{M}})\}^2}{\hat{\sigma}_{n,0}^2\, \boldsymbol{b}'\mathbf{W}\boldsymbol{b}} \geq \sup_{\boldsymbol{b}\in\mathbb{S}_c^{k^2-1}} \max\left[\frac{\boldsymbol{b}'\text{vec}(\widetilde{\mathbf{M}})}{\{\hat{\sigma}_{n,0}^2 \boldsymbol{b}'\mathbf{W}\boldsymbol{b}\}^{1/2}}, 0\right]^2 \\ &= \overline{QLR}_n^{(2)}. \end{aligned} \tag{17}$$

Because $QLR_n^{(2)} = \overline{QLR}_n^{(2)} + o_{\mathbb{P}}(1) = QLR_n^{(3)} = \overline{QLR}_n^{(3)} + o_{\mathbb{P}}(1)$, this gives

$$\begin{aligned} QLR_n &= \max[QLR_n^{(1)}, QLR_n^{(2)}, QLR_n^{(3)}] \\ &= \max[QLR_n^{(1)}, \overline{QLR}_n^{(2)}, \overline{QLR}_n^{(3)}] + o_{\mathbb{P}}(1) = QLR_n^{(1)} + o_{\mathbb{P}}(1) \end{aligned}$$

and shows that the limiting behavior of QLR_n is determined by that of $QLR_n^{(1)}$.

So far, however, we have only established the asymptotic behavior of $QLR_n^{(1)}(\epsilon)$, not that of $QLR_n^{(1)}$. Recall that $QLR_n^{(1)}(\epsilon)$ is based on eliminating $\mathbf{0}$ from $\boldsymbol{\Delta}$, whereas $\mathbf{0} \in \text{int}(\boldsymbol{\Delta})$ is explicitly assumed for $QLR_n^{(1)}$. Thus, $QLR_n^{(1)}(\epsilon)$ does not immediately provide the desired asymptotic distribution. This also implies that the asymptotic null distribution in Hansen (1996, theorem 1) cannot be literally regarded as the asymptotic null distribution of the QLR test because his regularity condition assumption 1 does not hold when $\mathbf{0} \in \text{int}(\boldsymbol{\Delta})$. This necessitates a further analysis of the QLR test treating $\mathbf{0}$ as an element of $\text{int}(\boldsymbol{\Delta})$. Interestingly, it turns out that the asymptotic null distribution we obtained in Section 2.2.1 is closely related to that of $QLR_n^{(1)}$.

We proceed by examining how the asymptotic null behavior of $QLR_n^{(1)}(\epsilon)$ varies as ϵ tends to zero. It turns out that $\overline{QLR}_n^{(2)}$ plays a key role here. To provide sufficient conditions for this, we combine the moment conditions of A3 and A3* and similarly combine the covariance conditions A6 and A6*.

Assumption A3** (Moments): *$E|U_t|^8 < \infty$ and $E|X_{t,i}|^8 < \infty$; or for some $\kappa > 2(\rho - 1)$, $E|U_t|^{4+2\kappa} < \infty$ and $E|X_{t,i}|^{16} < \infty$, $i = 1, 2, \cdots, k$.*

Assumption A6** (Covariance): *For each $\epsilon > 0$ and $\boldsymbol{\delta} \in \boldsymbol{\Delta}(\epsilon)$, $\det \bar{\mathbf{V}}_1(\boldsymbol{\delta}) > 0$ and $\det \bar{\mathbf{V}}_2(\boldsymbol{\delta}) > 0$, where*

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

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

These new conditions accommodate the fact that the score of the QL function under $\mathcal{H}_{01}$ turns out to be identical to the score obtained under $\mathcal{H}_{02}$ when direction $\boldsymbol{d}$ is given, and $\boldsymbol{\delta}$ goes to zero in the direction $\boldsymbol{d}$ by sending h to zero. Our next result involves the joint asymptotic behavior of the random functions $\mathcal{G}_1$ and $\mathcal{G}_2$ separately derived above:

Theorem 3. *Given A1, A2, A3****, A4*(i)*, A5, A6****, A7, and* $\mathcal{H}_0$,

(i) $QLR_n = QLR_n^{(1)} + o_{\mathbb{P}}(1)$; *and*

(ii) $QLR_n \Rightarrow \sup_{\boldsymbol{\delta}} \mathcal{G}(\boldsymbol{\delta})^2$, *where*

$$\mathcal{G}(\boldsymbol{\delta}) := \begin{cases} \mathcal{G}_1(\boldsymbol{\delta}), & \textit{if } h \neq 0; \\ \mathcal{G}_2(\boldsymbol{b}), & \textit{otherwise}. \end{cases} \tag{18}$$

In previous works, the asymptotic null distribution of QLR_n has been given as $\sup_{\boldsymbol{\delta}} \mathcal{G}_1(\boldsymbol{\delta})^2$, but this neglects the twofold identification problem. The true asymptotic null distribution may differ from $\sup_{\boldsymbol{\delta}} \mathcal{G}_1(\boldsymbol{\delta})^2$, because this does not properly handle the asymptotic null distribution under $\mathcal{H}_{02}$, mainly due to the regularity conditions needed for the quartic approximation. Properly accounting for $\mathcal{H}_{02}$ shows that this distribution is actually $\sup_{\boldsymbol{\delta}} \mathcal{G}(\boldsymbol{\delta})^2$, which depends on both $\mathcal{G}_1$ and $\mathcal{G}_2$. The stronger conditions A3**, A6**, and A7 are not required for $\mathcal{H}_{01}$ but are key to ensuring the validity of the quartic approximation. These conditions and the reparameterization in (10) permit Theorem 2 to extend Theorem 1 to hold on all of $\boldsymbol{\Delta}$, including $\mathbf{0}$. Theorem 3 then holds as an easy corollary, exploiting (16) and (17). We thus do not provide a proof of Theorem 3 in the Appendix.

The covariance structure of $\mathcal{G}$ necessarily accommodates the covariance of $\mathcal{G}_1$ and $\mathcal{G}_2$. Specifically, for each $\boldsymbol{\delta} = (h, \boldsymbol{d})$ and $\widetilde{\boldsymbol{\delta}} = (\widetilde{h}, \widetilde{\boldsymbol{d}})$, $E[\mathcal{G}(\boldsymbol{\delta})\mathcal{G}(\widetilde{\boldsymbol{\delta}})] = \rho(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}})$, where

$$\rho(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) := \begin{cases} \rho_1(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}), & \text{if } h \neq 0 \text{ and } \widetilde{h} \neq 0; \\ \rho_2(\boldsymbol{b}, \widetilde{\boldsymbol{b}}), & \text{if } h = 0 \text{ and } \widetilde{h} = 0; \\ \rho_3(\boldsymbol{b}, \widetilde{\boldsymbol{\delta}}), & \text{if } h = 0 \text{ and } \widetilde{h} \neq 0, \end{cases}$$

with

$$\rho_3(\boldsymbol{b}, \widetilde{\boldsymbol{\delta}}) := \frac{\mathcal{H}(\boldsymbol{b}, \widetilde{\boldsymbol{\delta}})}{\{\mathcal{I}(\boldsymbol{b}, \boldsymbol{b})\}^{1/2}\{\mathcal{J}(\widetilde{\boldsymbol{\delta}}, \widetilde{\boldsymbol{\delta}})\}^{1/2}}, \quad \text{and} \quad \mathcal{H}(\boldsymbol{b}, \widetilde{\boldsymbol{\delta}}) := E[U_t^2 \boldsymbol{b}' \text{vec}(\mathbf{C}_t^*) \Psi_t^*(\widetilde{\boldsymbol{\delta}})].$$

Thus, $\mathcal{H}(\boldsymbol{b}, \widetilde{\boldsymbol{\delta}})$ represents the covariance between the scores for $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$. If A4(ii) also holds, then $\mathcal{H}(\boldsymbol{b}, \widetilde{\boldsymbol{\delta}}) = \sigma_*^2 \boldsymbol{b}' E[\text{vec}(\mathbf{C}_t^*)\Psi_t^*(\widetilde{\boldsymbol{\delta}})]$. In this case,

$$\rho_3(\boldsymbol{b}, \widetilde{\boldsymbol{\delta}}) = \frac{\boldsymbol{b}' E[\text{vec}(\mathbf{C}_t^*)\Psi_t^*(\widetilde{\boldsymbol{\delta}})]}{\{\boldsymbol{b}' E[\text{vec}(\mathbf{C}_t^*)\text{vec}(\mathbf{C}_t^*)']\boldsymbol{b}\}^{1/2}\{E[\Psi_t^*(\widetilde{\boldsymbol{\delta}})^2]\}^{1/2}}.$$

The covariance structures $\rho_2(\boldsymbol{b}, \widetilde{\boldsymbol{b}})$ and $\rho_3(\boldsymbol{b}, \widetilde{\boldsymbol{\delta}})$ are related to $\rho_1(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}})$. Specifically, they essentially represent the limits of $\rho_1(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}})$ as h and $\widetilde{h}$ tend to zero, respectively. To show this, we define

$$\Phi_t(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) := \Psi_t(\boldsymbol{\delta})\Psi_t(\widetilde{\boldsymbol{\delta}}) \quad \textit{and} \quad \Upsilon_{t,j}(\boldsymbol{\delta}) := \Psi_t(\boldsymbol{\delta}) Z_{t,j}, \quad j = 1, 2, \cdots, k+1.$$

We ensure the applicability of the Lebesgue dominated convergence theorem by imposing the following:

Assumption A8 (Domination): (i) *For* $\ell = 0, 1, \cdots, 4$, *and each* j, $E[\sup_{\boldsymbol{\delta}} |(\partial^\ell/\partial h^\ell) \Upsilon_{t,j}(\boldsymbol{\delta})|^2] < \infty$; *and*

(ii) *for* $\ell, m = 0, 1, \cdots, 4$ *such that* $\ell + m \leq 4$, $E[\sup_{\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}} |(\partial^{\ell+m}/\partial h^\ell \partial \widetilde{h}^m)\Phi_t(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}})|^2] < \infty$.

We have the following formal result:

Lemma 8. *Given A1, A2, A3***, *A4*(i), *A5, A6***, *A7, A8, and* $\mathcal{H}_0$,

(i) $\lim_{h \downarrow 0} \rho_1(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) = \text{sgn}[c_2]\rho_3(\boldsymbol{b}, \widetilde{\boldsymbol{\delta}})$; *and*

(ii) $\lim_{h \downarrow 0} \lim_{\widetilde{h} \downarrow 0} \rho_1(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) = \rho_2(\boldsymbol{b}, \widetilde{\boldsymbol{b}})$.

Note that $\text{sgn}[c_2]$ appears in Lemma 8(i), so that we do not necessarily have $\mathcal{G}_1 \overset{d}{=} \mathcal{G}$, where $\overset{d}{=}$ denotes equality in distribution. Nevertheless, squaring these Gaussian processes makes this sign irrelevant, so that $\mathcal{G}_1^2 \overset{d}{=} \mathcal{G}^2$, and $QLR_n \Rightarrow \sup_{\boldsymbol{\delta}} \mathcal{G}_1(\boldsymbol{\delta})^2$ by Theorem 3(ii). The following states this formally.

Corollary 2. *Given A1, A2, A3***, *A4, A5, A6***, *A7, A8, and* $\mathcal{H}_0$, $QLR_n \Rightarrow \sup_{\boldsymbol{\delta}} \mathcal{G}_1(\boldsymbol{\delta})^2$.

This follows directly from Lemma 8 and our earlier discussion, so we do not prove this in the Appendix.

Under the further conditions of Corollary 2, the asymptotic null distribution $\mathcal{G}_1$ previously derived in the literature for QLR_n by neglecting the twofold identification problem is indeed correct. Nevertheless, properly accounting for $\mathcal{H}_{02}$ introduces regularity conditions stronger than previously recognized. The given conditions have the advantage of permitting a straightforward treatment of the twofold identification problem. Although it may be possible to find weaker conditions ensuring the conclusion of Corollary 2, A7 is crucial, as different null distributions may pertain when $c_2 = 0$. We demonstrate this in our Monte Carlo experiments below.

A practical implication of Corollary 2 is that the QLR test can test $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$ simultaneously under the given regularity conditions. This is an improvement, in the sense that previous statistics in the literature have been designed to test $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$ separately. But applying different tests sensitive to different alternatives can lead to size distortion when the null hypothesis is arbitrary neglected nonlinearity. On the other hand, the QLR test studied here obviates this problem.

The most crucial condition for the success of the QLR test is A7, as we emphasize above. If A7 is violated, but the researcher ignores this, the size of the test may not be properly controlled. Indeed, methods for constructing proper critical values for this case are unknown, as higher order approximations are required to derive these, and this is a topic requiring further research.

3 A Modeling Exercise and Monte Carlo Experiments

3.1 An AR(1) Example

In this section, we illustrate our theory using a Gaussian AR(1) process with DGP

$$Y_t = \theta_* + \beta_* Y_{t-1} + U_t, \quad t = 1, 2, \cdots,$$

where $|\beta_*| < 1$ and $\{U_t\} \sim$ IID $N(0, \sigma_*^2)$, so that $k = 1$ and

$$E[Y_t | Y_{t-1}, Y_{t-2}, Y_{t-3}, \cdots] = \theta_* + \beta_* Y_{t-1}.$$

To test $\mathcal{H}_0$, we take $\tilde{\Psi} = \exp$ and specify the alternative model with

$$f(Y_{t-1}\,; \alpha, \beta, \lambda, \delta) = \alpha + \beta Y_{t-1} + \lambda \exp(\delta Y_{t-1}).$$

We take $\delta \in \Delta := [\underline{\delta}, \overline{\delta}]$, where $-\infty < \underline{\delta} < 0 < \overline{\delta} < \infty$.

First, we examine the behavior of QLR under $\mathcal{H}_{01}$: $\lambda_* = 0$. Letting $\mathbf{\Psi}(\delta) := [\exp(\delta Y_0), \exp(\delta Y_1), \cdots, \exp(\delta Y_{n-1})]'$, Theorem 1 gives

$$QLR_n^{(1)}(\epsilon) = \sup_{\delta \in [\underline{\delta}, -\epsilon] \cup [\epsilon, \overline{\delta}]} \frac{\{\mathbf{U}'\mathbf{M}\mathbf{\Psi}(\delta)\}^2}{\hat{\sigma}_{n,0}^2 \mathbf{\Psi}(\delta)'\mathbf{M}\mathbf{\Psi}(\delta)} \Rightarrow \sup_{\delta \in [\underline{\delta}, -\epsilon] \cup [\epsilon, \overline{\delta}]} \mathcal{G}_1(\delta)^2,$$

with mean zero Gaussian process $\mathcal{G}_1$ such that

$$E[\mathcal{G}_1(\delta)\mathcal{G}_1(\widetilde{\delta})] = \frac{\mathcal{J}(\delta, \widetilde{\delta})}{\{\mathcal{J}(\delta, \delta)\}^{1/2}\{\mathcal{J}(\widetilde{\delta}, \widetilde{\delta})\}^{1/2}},$$

and, for each non-zero δ and $\widetilde{\delta}$,

$$\begin{aligned} \mathcal{J}(\delta, \widetilde{\delta}) :=& \sigma_*^2 E[\exp(Y_t(\delta + \widetilde{\delta}))] \\ & - \sigma_*^2 \begin{bmatrix} E[\exp(Y_t\delta)] \\ E[Y_t \exp(Y_t\delta)] \end{bmatrix}' \begin{bmatrix} 1 & E[Y_t] \\ E[Y_t] & E[Y_t^2] \end{bmatrix}^{-1} \begin{bmatrix} E[\exp(Y_t\widetilde{\delta})] \\ E[Y_t \exp(Y_t\widetilde{\delta})] \end{bmatrix}. \end{aligned}$$

Defining

$$\begin{aligned} M(\delta) :=& \exp\left\{\frac{\theta_*}{1-\beta_*}\delta + \frac{\sigma_*^2}{2(1-\beta_*^2)}\delta^2\right\} \quad \text{and} \\ J(\delta, \widetilde{\delta}) :=& \{\exp[\text{var}(Y_t)\delta\widetilde{\delta}] - 1 - \text{var}(Y_t)\delta\widetilde{\delta}\}, \end{aligned}$$

we have

$$\mathcal{J}(\delta, \widetilde{\delta}) = \sigma_*^2 M(\delta) M(\widetilde{\delta}) J(\delta, \widetilde{\delta}).$$

Note that $M(\delta) = E[\exp(\delta Y_t)]$ is the moment generating function of $Y_t \sim N[\theta_*/(1 - \beta_*), \sigma_*^2/(1-\beta_*^2)]$. It follows easily that

$$\rho_1(\delta, \widetilde{\delta}) := \frac{\mathcal{J}(\delta, \widetilde{\delta})}{\{\mathcal{J}(\delta, \delta)\}^{1/2}\{\mathcal{J}(\widetilde{\delta}, \widetilde{\delta})\}^{1/2}} = \frac{J(\delta, \widetilde{\delta})}{\{J(\delta, \delta)\}^{1/2}\{J(\widetilde{\delta}, \widetilde{\delta})\}^{1/2}}.$$

Now $\mathcal{G}_1$ is indexed by δ. We also have $\delta = hd$, with distance $h \in \mathbb{R}^+$ and direction $d = \pm 1$.

Next, consider the behavior of QLR under $\mathcal{H}_{02}$: $\delta_* = 0$. Theorem 2 gives

$$\overline{QLR}_n^{(2)} = \frac{\max[\boldsymbol{\iota}'\mathbf{D}_{11}\mathbf{MU}, 0]^2}{\hat{\sigma}_{n,0}^2 \boldsymbol{\iota}'\mathbf{D}_{11}\mathbf{MD}_{11}\boldsymbol{\iota}} + o_{\mathbb{P}}(1),$$

where $\mathbf{D}_{11} := \text{diag}\{Y_0^2, Y_1^2, \cdots, Y_{n-1}^2\}$,

$$\boldsymbol{\iota}'\mathbf{D}_{11}\mathbf{MU} = \sum_{t=1}^{n} Y_{t-1}^2 U_t - \begin{bmatrix} \sum_{t=1}^n Y_{t-1}^2 \\ \sum_{t=1}^n Y_{t-1}^3 \end{bmatrix}' \begin{bmatrix} n & \sum_{t=1}^n Y_{t-1} \\ \sum_{t=1}^n Y_{t-1} & \sum_{t=1}^n Y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^n U_t \\ \sum_{t=1}^n Y_{t-1} U_t \end{bmatrix},$$

and

$$\boldsymbol{\iota}'\mathbf{D}_{11}\mathbf{MD}_{11}\boldsymbol{\iota} = \sum_{t=1}^{n} Y_{t-1}^4 - \begin{bmatrix} \sum_{t=1}^n Y_{t-1}^2 \\ \sum_{t=1}^n Y_{t-1}^3 \end{bmatrix}' \begin{bmatrix} n & \sum_{t=1}^n Y_{t-1} \\ \sum_{t=1}^n Y_{t-1} & \sum_{t=1}^n Y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^n Y_{t-1}^2 \\ \sum_{t=1}^n Y_{t-1}^3 \end{bmatrix}.$$

As mentioned at the end of Section 2.2, the fact that $k = 1$ makes the role of d in Theorem 2 trivial, because $d \in \{-1, 1\}$. Thus, Theorem 2 implies

$$\overline{QLR}_n^{(2)} \Rightarrow \max[0, \mathcal{G}_2]^2,$$

where $\mathcal{G}_2 \sim N(0, 1)$. This holds because

$$n^{-1}\hat{\sigma}_{n,0}^2 \boldsymbol{\iota}'\mathbf{D}_{11}\mathbf{MD}_{11}\boldsymbol{\iota} \to \mathcal{I} := \sigma_*^2 \left\{ E[Y_t^4] - \frac{1}{\text{var}(Y_t)} \{E[Y_t^2]^3 - 2E[Y_t]E[Y_t^2]E[Y_t^3] + E[Y_t^3]^2\} \right\} \quad a.s.$$

Further, $Y_t \sim N[\theta_*/(1-\beta_*), \sigma_*^2/(1-\beta_*^2)]$ implies $\mathcal{I} = 2\sigma_*^2 \,\text{var}(Y_t)^2$, as well as $n^{-1/2}\boldsymbol{\iota}'\mathbf{D}_{11}\mathbf{MU} \overset{\text{A}}{\sim} N[0, 2\sigma_*^2 \,\text{var}(Y_t)^2]$, so $\mathcal{G}_2$ obtains as the weak limit of

$$\frac{n^{-1/2}\boldsymbol{\iota}'\mathbf{D}_{11}\mathbf{MU}}{\{\sigma_*^2 n^{-1}\boldsymbol{\iota}'\mathbf{D}_{11}\mathbf{MD}_{11}\boldsymbol{\iota}\}^{1/2}}.$$

Finally, we combine these separate results. By Theorem 3, we have $QLR_n \Rightarrow$

$\sup_{\delta\in\Delta} \mathcal{G}(\delta)^2$, where

$$\mathcal{G}(\delta) = \begin{cases} \mathcal{G}_1(\delta), & \text{if } \delta \neq 0; \\ \mathcal{G}_2, & \text{otherwise,} \end{cases}$$

with

$$E[\mathcal{G}(\delta)\mathcal{G}(\widetilde{\delta})] = \begin{cases} \rho_1(\delta, \widetilde{\delta}), & \text{if } \delta \neq 0 \text{ and } \widetilde{\delta} \neq 0; \\ 1, & \text{if } \delta = 0 \text{ and } \widetilde{\delta} = 0; \\ \rho_3(\delta), & \text{if } \widetilde{\delta} = 0 \text{ and } \delta \neq 0, \end{cases}$$

where

$$\rho_3(\delta) = \frac{\mathcal{H}(\delta)}{\{\mathcal{J}(\delta,\delta)\}^{1/2}\{\mathcal{I}\}^{1/2}} = \frac{\mathcal{H}(\delta)}{\{\sigma_*^2 M(\delta)M(\delta)J(\delta,\delta)\}^{1/2}\{\mathcal{I}\}^{1/2}},$$

with

$$\mathcal{H}(\delta) := \sigma_*^2\, E[Y_t^2 \exp(Y_t\delta)] - \frac{\sigma_*^2}{\text{var}(Y_t)} \left\{(E[Y_t^2]^2 - E[Y_t]E[Y_t^3])E[\exp(\delta Y_t)]\right\} \\ - \frac{\sigma_*^2}{\text{var}(Y_t)} \left\{(E[Y_t^3] - E[Y_t]E[Y_t^2])E[Y_t \exp(\delta Y_t)]\right\}.$$

Using the normality of Y_t and its moment generating function $M(\delta)$, it is straightforward to show that $\mathcal{H}(\delta) = \sigma_*^2\, \text{var}(Y_t)^2 M(\delta)\delta^2$. Using the definition of $\mathcal{J}(\delta, \widetilde{\delta})$ and the fact that $\mathcal{I} = 2\sigma_*^2\, \text{var}(Y_t)^2$, we have

$$\rho_3(\delta) = \frac{\text{var}(Y_t)\delta^2}{\{2[\exp[\text{var}(Y_t)\delta^2] - 1 - \text{var}(Y_t)\delta^2]\}^{1/2}}.$$

Finally, we find that the covariance kernel of $\mathcal{G}$ is just $\rho_1(\delta, \widetilde{\delta})$. This follows because

$$\lim_{\delta\to 0}\lim_{\widetilde{\delta}\to 0} \rho_1(\delta, \widetilde{\delta}) = \lim_{\widetilde{\delta}\to 0}\lim_{\delta\to 0} \rho_1(\delta, \widetilde{\delta}) = 1 \quad \text{and} \quad \lim_{\delta\to 0} \rho_1(\delta, \widetilde{\delta}) = \rho_3(\widetilde{\delta}).$$

These represent $E[\mathcal{G}(\delta)\mathcal{G}(0)]$ and $E[\mathcal{G}(0)^2]$ respectively.

Thus, $\mathcal{G} \overset{d}{=} \mathcal{G}_1$, so that $QLR_n \Rightarrow \sup_{\delta\in\Delta} \mathcal{G}_1(\delta)^2$ under $\mathcal{H}_0$.

3.2 Monte Carlo Experiments

In this subsection, we present the results of Monte Carlo experiments designed to investigate how well our asymptotic results approximate the finite-sample null behavior of our QLR statistic. We continue to study the AR(1) example. For conciseness, we restrict attention to behavior under the null, as the power of such tests under this distribution has already been well studied, both theoretically in the context of contiguity (e.g., Le Cam (1960), Hájek and Šidák (1967), and van der Vaart (1998)), and via Monte Carlo experiments (e.g., Bierens (1990) and Hansen (1996)). We first discuss a method for obtaining the asymptotic null distribution alternative to that of Hansen (1996), and we show how this embodies the features of the QLR statistic developed in Section 2. We then examine the performance of Hansen's weighted bootstrap, paying particular attention to what happens when A7 holds or is violated.

3.1 Simulating the Asymptotic Null Distribution

Hansen (1996) proposes a bootstrap procedure for constructing critical values for tests of the sort considered here; we discuss this in the next section. Because this procedure is computationally very intensive, we first discuss a less demanding procedure, available in particular cases. This method directly constructs and simulates a Gaussian process equivalent to that obeyed asymptotically by the statistic of interest. This approach has been taken by Phillips (1998), Andrews (2001), and Cho and White (2007, 2009, 2010), among others. This method is feasible for our AR(1) example, due to the assumed normality and conditional homoskedasticity. For other distributions or with conditional heteroskedasticity, this approach may not be possible; Hansen's method is especially useful in such cases.

Specifically, a process identical in distribution to $\mathcal{G} \overset{d}{=} \mathcal{G}_1$ of Section 3.1 is

$$\widetilde{\mathcal{G}}(\delta) := \frac{\sum_{k=2}^{\infty}\{\text{var}(Y_t)\}^{k/2}\delta^k Z_k/\sqrt{k!}}{\{\exp(\text{var}(Y_t)\delta^2) - 1 - \text{var}(Y_t)\delta^2\}^{1/2}},$$

with $\{Z_k\} \sim$ IID $N(0,1)$. To show this, we note that for any δ and $\widetilde{\delta}$ in Δ, $E[\mathcal{G}(\delta)] = E[\widetilde{\mathcal{G}}(\delta)]$ and $E[\mathcal{G}(\delta)\mathcal{G}(\widetilde{\delta})] = E[\widetilde{\mathcal{G}}(\delta)\widetilde{\mathcal{G}}(\widetilde{\delta})]$. Thus, $\mathcal{G} \overset{d}{=} \mathcal{G}_1 \overset{d}{=} \widetilde{\mathcal{G}}$ by a well known property of Gaussian processes. Further, $\lim_{\delta\to 0}\widetilde{\mathcal{G}}(\delta) = Z_2$ $a.s.$ and

$$E[Z_2\widetilde{\mathcal{G}}(\delta)] = \frac{\text{var}(Y_t)\delta^2}{\{2[\exp[\text{var}(Y_t)\delta^2] - 1 - \text{var}(Y_t)\delta^2]\}^{1/2}} = \rho_3(\delta).$$

This verifies that Z_2 has the same stochastic properties as $\mathcal{G}_2$.

For any given δ,

$$\lim_{k\to\infty} \text{var}\left[\frac{1}{\sqrt{k!}}\{\text{var}(Y_t)\delta^2\}^{k/2} Z_k\right] = \lim_{k\to\infty} \frac{1}{k!}\{\text{var}(Y_t)\}^k \delta^{2k} = 0.$$

Thus, if K is sufficiently large, simulating $\widetilde{\mathcal{G}}(\delta; K)$ can yield a useful approximation to $\mathcal{G} \stackrel{d}{=} \mathcal{G}_1 \stackrel{d}{=} \widetilde{\mathcal{G}}$, where

$$\widetilde{\mathcal{G}}(\delta; K) := \frac{\sum_{k=2}^{K}\{\text{var}(Y_t)\}^{k/2}\delta^k Z_k/\sqrt{k!}}{\{\exp(\text{var}(Y_t)\delta^2) - 1 - \text{var}(Y_t)\delta^2\}^{1/2}}.$$

Because $\text{var}(Y_t)$ is unknown, we replace it with a sample estimator, say,

$$\widehat{\text{var}}_n(Y_t) := \frac{1}{n}\sum_{t=1}^{n} Y_t^2 - \left\{\frac{1}{n}\sum_{t=1}^{n} Y_t\right\}^2,$$

and obtain critical values for QLR by simulating $\sup_{\delta\in\Delta} \widehat{\mathcal{G}}_n(\delta; K)^2$, where

$$\widehat{\mathcal{G}}_n(\delta; K) := \frac{\sum_{k=2}^{K}\{\widehat{\text{var}}_n(Y_t)\}^{k/2}\delta^k Z_k/\sqrt{k!}}{\{\exp(\widehat{\text{var}}_n(Y_t)\delta^2) - 1 - \widehat{\text{var}}_n(Y_t)\delta^2\}^{1/2}}.$$

Because $\sup_{\delta\in\Delta} \mathcal{G}_n(\delta; K)^2$ and $\sup_{\delta\in\Delta} \widehat{\mathcal{G}}_n(\delta; K)^2$ depend on K and Δ, we emphasize this by writing

$$\widehat{\mathcal{QLR}}_n(\Delta; K) := \sup_{\delta\in\Delta} \widehat{\mathcal{G}}_n(\delta; K)^2, \qquad \text{and}$$

$$\mathcal{QLR}_n(\Delta; K) := \sup_{\delta\in\Delta} \widetilde{\mathcal{G}}(\delta; K)^2.$$

We examine the properties of the QLR statistic under the null for a variety of relevant cases. We let $K = 150$ and consider four choices for Δ: $\Delta_{0.5} := [-0.5, 0.5]$, $\Delta_{1.0} := [-1.0, 1.0]$, $\Delta_{1.5} := [-1.5, 1.5]$, and $\Delta_{2.0} := [-2.0, 2.0]$. We also let $(\theta_*, \beta_*, \sigma_*^2) = (0, 0.5, 1)$, so that $\text{var}(Y_t) = 4/3$. The distribution of $\widehat{\mathcal{QLR}}_n(\Delta; K)$ is obtained by grid search for the maximum over Δ. The grid distances for $\Delta_{0.5}$, $\Delta_{1.0}$, $\Delta_{1.5}$, and $\Delta_{2.0}$ are $1/101$, $2/201$, $3/301$, and $4/401$, respectively. This avoids the zero grid point, where $\widehat{\mathcal{QLR}}_n(\Delta; K) = \mathcal{QLR}_n(\Delta; K) = 0$.

INSERT Table 1 AROUND HERE.

Table 1 presents the asymptotic critical values obtained for $\mathcal{QLR}_n(\Delta; K)$. We see immediately that these depend on Δ. As Δ gets larger, the asymptotic critical values increase, as the definition of $\mathcal{QLR}_n(\Delta; K)$ implies.

INSERT Table 2 AROUND HERE.

Table 2 presents the finite-sample properties of the QLR statistic. As Corollary 2 implies, for every Δ, the finite-sample distribution of the QLR statistic approaches the asymptotic distribution of $\mathcal{QLR}_n(\Delta; K)$ as n increases. We also see that the empirical rejection rates for nominal levels 1%, 5%, and 10% approach these levels from below. Figures 1 and 2 respectively show the empirical distribution and estimated density function of the QLR statistic for each Δ. The density functions are obtained by kernel density estimation method using the standard normal density function as kernel. As can be seen from Figures 1 and 2, the empirical distributions uniformly approach the asymptotic null distribution as the sample size increases. We also see that the QLR statistics have better finite sample properties when the associated parameter space Δ is smaller. The nominal rejection rates are closest to the asymptotic distribution for $\mathcal{QLR}_n(\Delta; K)$ when $\Delta = \Delta_{0.5}$. If $\Delta = \Delta_{2.0}$, the finite sample distribution for QLR is still quite far from that of $\mathcal{QLR}_n(\Delta_{2.0}; K)$, even when the sample size is 5,000.

INSERT Figures 1 and 2 AROUND HERE.

Table 3 presents simulation results for the case in which $\text{var}(Y_t)$ is estimated. Simulating $\widehat{\mathcal{QLR}}_n(\Delta; K)$ for every realized estimate of $\text{var}(Y_t)$ requires an immense amount of computation time. Consequently, we obtain critical values by interpolating values obtained from Table 2. Specifically, in Table 2, we analyze seven sample sizes (50, 100, $\cdots$, and 5,000), each of which is replicated 10,000 times for each choice of Δ. We collect the minimum and maximum values of the estimates of $\text{var}(Y_t)$ from the replications for each sample size, giving 14 estimated values that we denote $\widehat{\text{var}}(Y_t)$. Using these, we put $K = 150$ as before and generate null distributions of $\widehat{\mathcal{QLR}}_n(\Delta; K)$, simulating 50,000 times to obtain precise critical values. Denote the critical values for a nominal level $\boldsymbol{\alpha}$ and choice of Δ obtained in this way by $cv(\widehat{\text{var}}(Y_t), \Delta, \boldsymbol{\alpha})$. From this simulation, we observe that, for each $\boldsymbol{\alpha}$ and Δ, $cv(\widehat{\text{var}}(Y_t), \Delta, \boldsymbol{\alpha})$ monotonically increases as $\widehat{\text{var}}(Y_t)$ increases. Thus, if the sample QLR statistic is less than $cv(\widehat{\text{var}}(Y_t), \Delta, \boldsymbol{\alpha})$

and its associated variance estimate is less than $\widehat{\text{var}}(Y_t)$, then the null shouldn't rejected. On the other hand, we reject the null if the QLR statistic is strictly greater than $cv(\widehat{\text{var}}(Y_t), \Delta, \boldsymbol{\alpha})$ and the estimated variance is less than $\widehat{\text{var}}(Y_t)$. We find that for each Δ and $\boldsymbol{\alpha}$, better than 99% of the 10,000 replications of Table 2 can be handled by this rule. For those replications that cannot be handled in this way, we obtain the critical values by interpolating the 14 combinations of $(\widehat{\text{var}}(Y_t), cv(\widehat{\text{var}}(Y_t), \Delta, \boldsymbol{\alpha}))$ and apply the standard decision rule.

INSERT Table 3 AROUND HERE.

Table 3 presents the empirical rejection rates obtained in this way; the results are almost identical to those of Table 2. As the sample size increases, the nominal levels are more closely matched. Also as before, the levels are better when the associated parameter space is smaller. Thus, the findings of Table 2 are preserved, even when $\text{var}(Y_t)$ is estimated.

3.2 Hansen's Weighted Bootstrap

In this section, we apply Hansen's (1996) weighted bootstrap to estimate the asymptotic null distribution and to examine how the weighted bootstrap behaves when our regularity conditions are or are not met. For this, we continue to study the AR(1) DGP of Section 3.2.1 and the choice $\Psi = \exp$ ("Model 1"). We also consider the choice $\Psi =$ logistic CDF, so that $\Psi(\delta) = 1/\{1+\exp(\delta Y_{t-1})\}$ ("Model 2"). Further, we let $\delta \in \Delta_{0.5}$ for Models 1 and 2. Note that for Model 2 we have $c_2 = 0$, violating A7.

The specific procedure for applying Hansen's weighted bootstrap is as follows: First, for each grid point $\boldsymbol{\delta} \in \boldsymbol{\Delta}$, we compute the scores $\hat{S}_{nt}(\boldsymbol{\delta}) := \{\hat{D}_{nt}(\boldsymbol{\delta})\}^{-\frac{1}{2}}\hat{W}_{nt}(\boldsymbol{\delta})$, where

$$\begin{aligned}\hat{D}_{nt}(\boldsymbol{\delta}) :=& \frac{1}{n}\sum_{t=1}^{n}[\hat{U}_{nt}\Psi_t(\boldsymbol{\delta})]^2 \\ &- \left[\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{nt}^2\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t'\right]\left[\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{nt}^2\mathbf{Z}_t\mathbf{Z}_t'\right]^{-1}\left[\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{nt}^2\mathbf{Z}_t\Psi_t(\boldsymbol{\delta})\right],\end{aligned}$$

$$\hat{W}_{nt}(\boldsymbol{\delta}) := \Psi_t(\boldsymbol{\delta})\hat{U}_{nt} - \left[\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{nt}^2\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t'\right]\left[\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{nt}^2\mathbf{Z}_t\mathbf{Z}_t'\right]^{-1}\mathbf{Z}_t\hat{U}_{nt},$$

$\hat{U}_{nt} := Y_t - \mathbf{Z}_t'(\hat{\alpha}_n, \hat{\beta}_n)'$, and $(\hat{\alpha}_n, \hat{\beta}_n)$ is the least squares estimator obtained using the null model. Grid points with grid distance 1/101 are selected from Δ as before. Thus, there are 102 grid points for Models 1 and 2.

Second, for $j = 1, \cdots, J$, we generate $Z_{jt} \sim$ IID $N(0, 1)$, $t = 1, 2, \cdots, n$, and simulate the asymptotic distribution of the QLR statistic by computing the empirical distribution of

$$\mathcal{QLR}_{jn} := \sup_{\boldsymbol{\delta} \in \boldsymbol{\Delta}} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \hat{S}_{nt}(\boldsymbol{\delta}) Z_{jt} \right)^2 .$$

We let $J = 500$, as the computational burden is immense. Although J is not large enough for highly precise estimates, Hansen (1996) suggests that simulation results with this choice should generate solid results.

Third, we compute the proportion of simulated outcomes exceeding the QLR statistic. That is, we compute the empirical level $\hat{p}_n \equiv J^{-1} \sum_{j=1}^{J} I[QLR_n < \mathcal{QLR}_{jn}]$, where $I[\,\cdot\,]$ denotes the indicator function.

Finally, we repeat the entire exercise $4,000$ times, generating $\hat{p}_n^{(i)}, i = 1, \cdots, 4,000$, and we compute the proportion of outcomes whose $\hat{p}_n^{(i)}$ is less than the specified nominal level (e.g., $\boldsymbol{\alpha} = 5\%$). That is, we compute $\frac{1}{4000} \sum_{i=1}^{4000} I[\hat{p}_n^{(i)} < \boldsymbol{\alpha}]$. Under the null, this converges to the significance level corresponding to the specified nominal level, $\boldsymbol{\alpha}$. Plotted as a function of $\boldsymbol{\alpha}$, this should converge to a 45-degree line on the unit interval, if the weighted bootstrap is successful.

INSERT Table 4 AROUND HERE.

We present these estimates in Table 4 and Figure 3. The first panel of Table 4 indicates the obtained empirical levels for $\boldsymbol{\alpha} = 1\%$, 5%, 10%, 30%, 50%, 80%, 90%, and 95%, when the exponential function is used for the activation function. This model satisfies all of our regularity conditions. We see that the empirical rejection rates converge to the specified nominal levels as the sample size increases. This shows that even if Hansen's (1996) regularity condition A1 is not met (i.e., $\lim_{\boldsymbol{\delta} \to \mathbf{0}} \mathcal{T}(\boldsymbol{\delta}, \boldsymbol{\delta}) = 0$), his weighted bootstrap still consistently delivers the specified nominal levels. This is mainly because the numerator and denominator of the QLR statistic converge to zero at the same rate, so that applying L'Hôspital's rule delivers the asymptotic distribution of the QLR statistic as $\boldsymbol{\delta}$ converges to zero. The same result is especially obvious from Figure 3. The first panel of Figure 3 shows the estimated value of $\frac{1}{4000} \sum_{i=1}^{4000} I[\hat{p}_n^{(i)} < \boldsymbol{\alpha}]$

for each $\boldsymbol{\alpha} \in [0,1]$. As the sample size increases, the estimated relation uniformly approaches the 45-degree line, affirming that Hansen's (1996) weighted bootstrap is successful when our regularity conditions hold.

INSERT Figure 3 AROUND HERE.

The second panel of Table 4 shows what happens when the logistic CDF is used. As mentioned above, this violates A7. Indeed, we observe that the weighted bootstrap does not work for this case and that large sample sizes are required to achieve asymptotic behavior. The empirical rejection rates differ substantially from the nominal levels when $n = 6,000$, and this difference does not vanish with increasing sample size. In particular, although the right-hand tail probability is relatively well approximated by the weighted bootstrap, the difference is persistent for the left-hand tail probability: when $n = 40,000$, the empirical rejection rate for the left-hand tail probability does not converge to the nominal size. Essentially, the quartic approximation is insufficient. The second panel of Figure 3 also shows the value of $\frac{1}{4000}\sum_{i=1}^{4000} I[\hat{p}_n^{(i)} < \boldsymbol{\alpha}]$ for each $\boldsymbol{\alpha} \in [0,1]$. This exhibits the same behavior. The relation does not converge to a 45-degree line as the sample size increases, even for $n = 40,000$. Use of the weighted bootstrap does not deliver reliable inference when A7 is violated.

4 Conclusion

This study revisits testing for neglected nonlinearity in regression using ANNs, motivated by the fact that the literature so far has not accommodated the twofold identification problem: using ANNs, the linear null can be generated in two different ways. Previously, the possibilities under the null have only been analyzed separately, and this is not enough to obtain the desired asymptotic null distribution of ANN-based nonlinearity tests.

This asymptotic behavior is therefore still an open question. Here we analyze a convenient ANN-based quasi-likelihood ratio (QLR) statistic for testing neglected nonlinearity, paying careful attention to both components of the null. We derive the asymptotic null distribution under each component separately and analyze their interaction. Somewhat remarkably, we find that the previously known asymptotic null distribution for the type 1 case still applies, but under somewhat stronger conditions than previously recognized. We present Monte Carlo experiments corroborating our theoretical results, and showing that standard methods can yield misleading inference when our new, stronger regularity conditions are violated.

Acknowledgements

The authors thank the Editor and two anonymous referees for helpful comments and suggestions. The authors have also benefited from discussions with Chang-Jin Kim, Cheolbeom Park, Sang Soo Park, Byeongseon Seo, and other participants at the 2010 summer conference of the Korean Econometric Society and the Sixteenth Joint Seminar of Yonsei University and Hokkaido University. Cho appreciates research support by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-332-B00025).

Appendix: Proofs

Proof of Lemma 1: (i) Given Conditions A1, A3, and A4, we have

$$\begin{aligned}\hat{\sigma}^2_{n,0} =& n^{-1}\sum_{t=1}^{n} U_t^2 - (n^{-1}\sum_{t=1}^{n} U_t\mathbf{Z}_t')(n^{-1}\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{Z}_t')^{-1}(n^{-1}\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{U})\\ &\to \sigma_*^2 - \mathbf{0}'(E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}\mathbf{0} = \sigma_*^2\end{aligned}$$

in probability by the ergodic theorem.

(ii) We separate our proof into two parts. First in (a), we show the weak convergence of $n^{-1/2}\mathbf{\Psi}(\,\cdot\,)'\mathbf{M}\mathbf{U}$. In (b), we show that $n^{-1}\hat{\sigma}^2_{n,0}\mathbf{\Psi}(\,\cdot\,)'\mathbf{M}\mathbf{\Psi}(\,\cdot\,) \to \mathcal{J}(\,\cdot\,)$ uniformly on $\mathbf{\Delta}(\epsilon)$ in probability. Given these, the desired result follows from the converging-together lemma (Billingsley (1999, p. 151)).

(a) To show the weak convergence of $n^{-1/2}\mathbf{\Psi}(\,\cdot\,)'\mathbf{M}\mathbf{U}$, we note that

$$n^{-1/2}\mathbf{\Psi}(\boldsymbol{\delta})'\mathbf{M}\mathbf{U} = n^{-1/2}\sum_{t=1}^{n}\Psi_t(\boldsymbol{\delta})U_t - (\sum_{t=1}^{n}\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t')(\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{Z}_t')^{-1}n^{-1/2}\sum_{t=1}^{n}\mathbf{Z}_tU_t.$$

For each $\boldsymbol{\delta} \in \boldsymbol{\Delta}(\epsilon)$, let $\hat{\Psi}_{n,t}(\boldsymbol{\delta})$ be defined as

$$\hat{\Psi}_{n,t}(\boldsymbol{\delta}) := \Psi_t(\boldsymbol{\delta})U_t - (\sum_{t=1}^{n} \Psi_t(\boldsymbol{\delta})\mathbf{Z}_t')(\sum_{t=1}^{n} \mathbf{Z}_t\mathbf{Z}_t')^{-1}\mathbf{Z}_t U_t,$$

and let $\tilde{\Psi}_t(\boldsymbol{\delta})$ be defined as

$$\tilde{\Psi}_t(\boldsymbol{\delta}) := \Psi_t(\boldsymbol{\delta})U_t - E(\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t')(E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}\mathbf{Z}_t U_t.$$

We show that

$$\sup_{\boldsymbol{\delta}\in\boldsymbol{\Delta}(\epsilon)} \left| n^{-1/2}\sum_{t=1}^{n}[\hat{\Psi}_{n,t}(\boldsymbol{\delta}) - \tilde{\Psi}_t(\boldsymbol{\delta})] \right| = o_{\mathbb{P}}(1) \tag{19}$$

and then show the weak convergence of $\{n^{-1/2}\sum_{t=1}^{n} \tilde{\Psi}_t(\cdot)\}$ on $\boldsymbol{\Delta}(\epsilon)$. First, we note that

$$\begin{aligned}
&\sup_{\boldsymbol{\delta}\in\boldsymbol{\Delta}(\epsilon)} \left| n^{-1/2}\sum_{t=1}^{n}[\hat{\Psi}_{n,t}(\boldsymbol{\delta}) - \tilde{\Psi}_t(\boldsymbol{\delta})] \right| \\
&\qquad \le \sup_{\boldsymbol{\delta}\in\boldsymbol{\Delta}(\epsilon)} \left| (n^{-1}\sum_{t=1}^{n} \Psi_t(\boldsymbol{\delta})\mathbf{Z}_t') \left\{ (n^{-1}\sum_{t=1}^{n} \mathbf{Z}_t\mathbf{Z}_t')^{-1} - (E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1} \right\} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{Z}_t U_t \right| \\
&\qquad\quad + \sup_{\boldsymbol{\delta}\in\boldsymbol{\Delta}(\epsilon)} \left| \left\{ (n^{-1}\sum_{t=1}^{n} \Psi_t(\boldsymbol{\delta})\mathbf{Z}_t') - E(\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t') \right\} (E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{Z}_t U_t \right|.
\end{aligned}$$

We note that $\{\mathbf{Z}_t U_t, \mathcal{F}_t\}$ is a martingale difference sequence (MDS) so that $E[\mathbf{Z}_t U_t] = \mathbf{0}$, with $E[|X_{t,i}U_t|^2] = E[U_t^4]^{1/2}E[|X_{t,i}|^4]^{1/2} \le E[M_t^4]^{1/2}E[X_{t,i}^4]^{1/2} < \infty$; also, $E[U_t^2 \mathbf{Z}_t\mathbf{Z}_t']$ is positive definite by A6, where $\mathcal{F}_{t-1} := \sigma(\mathbf{X}_t, U_{t-1}, \mathbf{X}_{t-1}, U_{t-2}, \cdots)$. This implies that $n^{-1/2}\sum \mathbf{Z}_t U_t$ is asymptotically normal by, e.g., theorem 5.25 of White (2001). Therefore, $\sum_{t=1}^{n} \mathbf{Z}_t U_t = O_{\mathbb{P}}(n^{1/2})$. Further,

$$\sup_{\boldsymbol{\delta}\in\boldsymbol{\Delta}} \left\| n^{-1}\sum_{t=1}^{n} \Psi_t(\boldsymbol{\delta})\mathbf{Z}_t - E[\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t] \right\|_1 = o_{\mathbb{P}}(1),$$

as shown in (b), by applying Ranga Rao's uniform law of large numbers (ULLN), where $\|[a_{ij}]\|_1 = (\sum_j \sum_i |a_{ij}|)$. Therefore,

$$\sup_{\boldsymbol{\delta}\in\boldsymbol{\Delta}(\epsilon)} \left| \left\{ (n^{-1}\sum_{t=1}^{n} \Psi_t(\boldsymbol{\delta})\mathbf{Z}_t') - E(\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t') \right\} (E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{Z}_t U_t \right| = o_{\mathbb{P}}(1).$$

Also, $|n^{-1}\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{Z}_t' - E[\mathbf{Z}_t\mathbf{Z}_t']| = o_{\mathbb{P}}(1)$ and for each $i = 1, 2, \cdots, k$, $\sup_{\boldsymbol{\delta}\in\boldsymbol{\Delta}} |\sum_{t=1}^{n} \Psi_t(\boldsymbol{\delta})X_{t,i}| \leq \sum_{t=1}^{n} M_t|X_{t,i}| = O_{\mathbb{P}}(n)$ by applying A5 and the ergodic theorem. This implies that

$$\sup_{\boldsymbol{\delta}\in\boldsymbol{\Delta}(\epsilon)}\left|(n^{-1}\sum_{t=1}^{n}\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t')\left\{(n^{-1}\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{Z}_t')^{-1} - (E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}\right\}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbf{Z}_tU_t\right| = o_{\mathbb{P}}(1).$$

Thus, (19) follows.

Next, we verify that the two terms on the RHS of $\tilde{\Psi}_t(\boldsymbol{\delta})$ satisfy the sufficiency conditions for weak convergence. First, we note that $|\Psi_t(\boldsymbol{\delta})U_t - \Psi_t(\widetilde{\boldsymbol{\delta}})U_t| \leq M_t \cdot |U_t| \cdot \|\boldsymbol{\delta} - \widetilde{\boldsymbol{\delta}}\|$ from the differentiability and bound conditions in A2 and A5 respectively, so that for $\kappa > 0$ as in A3, it follows that

$$\sup_{\|\boldsymbol{\delta}-\widetilde{\boldsymbol{\delta}}\|<\eta} |\Psi_t(\boldsymbol{\delta})U_t - \Psi_t(\widetilde{\boldsymbol{\delta}})U_t|^{2+\kappa} \leq M_t^{2+\kappa}|U_t|^{2+\kappa}\eta^{2+\kappa} \leq M_t^{4+2\kappa}\eta^{2+\kappa},$$

implying that

$$E\left[\sup_{\|\boldsymbol{\delta}-\widetilde{\boldsymbol{\delta}}\|<\eta} |\Psi_t(\boldsymbol{\delta})U_t - \Psi_t(\widetilde{\boldsymbol{\delta}})U_t|^{2+\kappa}\right]^{\frac{1}{2+\kappa}} \leq E[M_t^{4+2\kappa}]^{\frac{1}{2+\kappa}}\eta. \tag{20}$$

Likewise, by the moment condition in A3, there is some C such that

$$\begin{aligned}
&|E(\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t')(E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}\,\mathbf{Z}_tU_t - E(\Psi_t(\widetilde{\boldsymbol{\delta}})\mathbf{Z}_t')(E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}\,\mathbf{Z}_tU_t| \\
&\qquad = |E[\{\Psi_t(\boldsymbol{\delta}) - \Psi_t(\widetilde{\boldsymbol{\delta}})\}\mathbf{Z}_t']\,(E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}\,\mathbf{Z}_tU_t| \\
&\qquad \leq CM_t^2\|(E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}\|_1 \cdot \|E[\{\Psi_t(\boldsymbol{\delta}) - \Psi_t(\widetilde{\boldsymbol{\delta}})\}\mathbf{Z}_t]\|_1.
\end{aligned}$$

Then

$$\begin{aligned}
&|E(\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t')(E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}\,\mathbf{Z}_tU_t - E(\Psi_t(\widetilde{\boldsymbol{\delta}})\mathbf{Z}_t')(E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}\,\mathbf{Z}_tU_t| \\
&\leq C \cdot M_t^2 \cdot \|\{E[\mathbf{Z}_t\mathbf{Z}_t']\}^{-1}\|_1 \cdot \|E[\mathbf{Z}_t\Psi_t(\boldsymbol{\delta})] - E[\mathbf{Z}_t\Psi_t(\widetilde{\boldsymbol{\delta}})]\|_1 \\
&\leq k \cdot C \cdot M_t^2 \cdot \|\{E[\mathbf{Z}_t\mathbf{Z}_t']\}^{-1}\|_1 \cdot E[M_t^2] \cdot \|\boldsymbol{\delta} - \widetilde{\boldsymbol{\delta}}\|,
\end{aligned}$$

where the last inequality follows from the Lipschitz continuity condition implied by

A5. That is, for $j = 1, 2, \cdots, k+1$, $|Z_{j,t}[\Psi_t(\boldsymbol{\delta}) - \Psi_t(\widetilde{\boldsymbol{\delta}})]| \leq M_t^2 \|\boldsymbol{\delta} - \widetilde{\boldsymbol{\delta}}\|$. Hence,

$$E\left[\sup_{\|\boldsymbol{\delta}-\widetilde{\boldsymbol{\delta}}\|<\eta} |E[(\Psi_t(\boldsymbol{\delta}) - \Psi_t(\widetilde{\boldsymbol{\delta}}))\mathbf{Z}_t')](E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}\,\mathbf{Z}_tU_t|^{2+\kappa}\right]^{\frac{1}{2+\kappa}}$$
$$\leq C \cdot E[M_t^{4+2\kappa}]^{\frac{1}{2+\kappa}} \cdot \|\{E[\mathbf{Z}_t\mathbf{Z}_t']\}^{-1}\|_1 \cdot E[M_t^2] \cdot \eta. \qquad (21)$$

Given (20) and (21), it follows that that for some $B < \infty$, $E[\sup_{\|\boldsymbol{\delta}-\boldsymbol{\delta}\|\leq\eta} |\tilde{\Psi}_t(\boldsymbol{\delta}) - \tilde{\Psi}_t(\widetilde{\boldsymbol{\delta}})$ $|^{2+\kappa}] \leq B\eta$, implying that Ossiander's $L^{2+\kappa}$ entropy is finite. Thus, $\{n^{-1/2}\sum_{t=1}^n \tilde{\Psi}_t(\,\cdot\,)\}$ is tight by Theorem 1 of DMR (1995). This and (19) imply that $\{n^{-1/2}\sum_{t=1}^n \hat{\Psi}_{n,t}(\,\cdot\,)\}$ is tight, and the multivariate CLT gives the finite-dimensional weak convergence, which we do not prove, as this is straightforward. These two facts ensure the weak convergence of $\{n^{-1/2}\sum_{t=1}^n \hat{\Psi}_{n,t}(\,\cdot\,)\}$ on $\boldsymbol{\Delta}(\epsilon)$.

Finally, the given covariance structure follows by the finite moment conditions.

(b) Next, for each $\boldsymbol{\delta}$,

$$n^{-1}\boldsymbol{\Psi}(\boldsymbol{\delta})'\mathbf{M}\boldsymbol{\Psi}(\boldsymbol{\delta}) = n^{-1}\sum_{t=1}^n \Psi_t(\boldsymbol{\delta})^2 - \{n^{-1}\sum_{t=1}^n \Psi_t(\boldsymbol{\delta})\mathbf{Z}_t'\}\{n^{-1}\sum_{t=1}^n \mathbf{Z}_t\mathbf{Z}_t'\}^{-1}\{n^{-1}\sum_{t=1}^n \Psi_t(\boldsymbol{\delta})\mathbf{Z}_t\}.$$

It easily follows that $\sup_{\boldsymbol{\delta}} |n^{-1}\sum_{t=1}^n \boldsymbol{\Psi}_t(\boldsymbol{\delta})^2 - E[\boldsymbol{\Psi}_t(\boldsymbol{\delta})^2]| \to 0$ in probability and $\sup_{\boldsymbol{\delta}} \|n^{-1}\sum_{t=1}^n \Psi_t(\boldsymbol{\delta})\,\mathbf{Z}_t - E[\Psi_t(\boldsymbol{\delta})\mathbf{Z}_t]\|_1 \to 0$ in probability given A1, A2, A3, and A5 by Ranga Rao's (1962) ULLN. Therefore, from this and the weak law of large numbers (WLLN) applied to $n^{-1}\sum_{t=1}^n \mathbf{Z}_t\mathbf{Z}_t'$,

$$\sup_{\boldsymbol{\delta}} |n^{-1}\boldsymbol{\Psi}(\boldsymbol{\delta})'\mathbf{M}\boldsymbol{\Psi}(\boldsymbol{\delta}) - \{E[\boldsymbol{\Psi}_t(\boldsymbol{\delta})^2] - E[\boldsymbol{\Psi}_t(\boldsymbol{\delta})\mathbf{Z}_t']\{E[\mathbf{Z}_t\mathbf{Z}_t']\}^{-1}E[\boldsymbol{\Psi}_t(\boldsymbol{\delta})\mathbf{Z}_t]\}| \to 0$$

in probability. Therefore, Lemma 1(i) proves the desired result.

(iii) A4(ii) immediately implies

$$\mathcal{T}(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}}) := E[U_t^2\Psi_t^*(\boldsymbol{\delta})\Psi_t^*(\widetilde{\boldsymbol{\delta}})] = E[E[U_t^2|\mathbf{X}_t]\Psi_t^*(\boldsymbol{\delta})\Psi_t^*(\widetilde{\boldsymbol{\delta}})] = \sigma_*^2 E[\Psi_t^*(\boldsymbol{\delta})\Psi_t^*(\widetilde{\boldsymbol{\delta}})].$$

■

Proof of Theorem 1: (i) Given Lemma 1(i and ii), the continuous mapping theorem completes the proof.

(ii) We use Lemma 1(iii) and the definition of $\rho_1(\boldsymbol{\delta}, \widetilde{\boldsymbol{\delta}})$ to obtain the desired result.

■

Proof of Lemma 3: (i) By the definitions of $\mathbf{D}_{ij}$ and $\mathbf{M}$,

$$\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{U} = \sum_{t=1}^{n} X_{t,i}X_{t,j}U_t - \sum_{t=1}^{n} X_{t,i}X_{t,j}\mathbf{Z}_t'(\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{Z}_t')^{-1}\sum_{t=1}^{n}\mathbf{Z}_tU_t.$$

We note that $n^{-1}\sum X_{t,i}X_{t,j}\mathbf{Z}_t'$ and $n^{-1}\sum \mathbf{Z}_t\mathbf{Z}_t'$ obey the WLLN by the ergodic theorem and A3* and further show that $n^{-1/2}\sum X_{t,i}X_{t,j}U_t$ satisfies the CLT. We already showed that $n^{-1/2}\sum \mathbf{Z}_tU_t$ obeys the asymptotic normality in the proof of Lemma 1(i), which also holds under the conditions of Lemma 3. For $n^{-1/2}\sum X_{t,i}X_{t,j}U_t$, we verify the conditions of theorem 5.25 of White (2001). First, we note that $\{X_{t,i}X_{t,j}U_t, \mathcal{F}_t\}$ is an MDS under the null, so that $E[X_{t,i}X_{t,j}U_t|\mathcal{F}_{t-1}] = 0$. Second, $E[|X_{t,i}X_{t,j}U_t|^2] < \infty$ by A3*. This follows from the fact that

$$\begin{aligned} E[|X_{t,i}X_{t,j}U_t|^2] &\leq E[U_t^4]^{1/2}E[|X_{t,i}X_{t,j}|^4]^{1/2} \\ &\leq E[M_t^4]^{1/2}E[|X_{t,i}|^8]^{1/4}E[|X_{t,j}|^8]^{1/4} < \infty, \end{aligned}$$

where the first two inequalities and the last inequality follow from Cauchy-Schwarz inequality and A3*, respectively. This implies that $n^{-1/2}\sum X_{t,i}X_{t,j}U_t$ is asymptotically normal by theorem 5.25 of White (2001). Thus, $\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{U} = O_{\mathbb{P}}(n^{1/2})$.

(ii) The proof is almost identical to (i). We note that

$$\boldsymbol{\iota}'\mathbf{D}_{ij\ell}\mathbf{M}\mathbf{U} = \sum_{t=1}^{n} X_{t,i}X_{t,j}X_{t,\ell}U_t - \sum_{t=1}^{n} X_{t,i}X_{t,j}X_{t,\ell}\mathbf{Z}_t'(\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{Z}_t')^{-1}\sum_{t=1}^{n}\mathbf{Z}_tU_t,$$

and that $n^{-1}\sum_{t=1}^{n} X_{t,i}X_{t,j}X_{t,\ell}\mathbf{Z}_t$ obeys the WLLN. Also, $\{X_{t,i}X_{t,j}X_{t,\ell}U_t, \mathcal{F}_t\}$ is a MDS, implying that $E[X_{t,i}X_{t,j}X_{t,\ell}U_t|\mathcal{F}_{t-1}] = 0$. Cauchy-Schwarz inequality yields

$$\begin{aligned} E[|X_{t,i}X_{t,j}X_{t,\ell}U_t|^2] &\leq E[|X_{t,i}X_{t,j}|^4]^{1/2}E[|X_{t,\ell}U_t|^4]^{1/2} \\ &\leq E[|X_{t,i}|^8]^{1/4}E[|X_{t,j}|^8]^{1/4}E[|X_{t,\ell}|^8]^{1/4}E[|U_t|^8]^{1/4} < \infty; \end{aligned}$$

alternatively, we have

$$\begin{aligned}E[|X_{t,i}X_{t,j}X_{t,\ell}U_t|^2] &\leq E[|X_{t,i}X_{t,j}X_{t,\ell}|^4]^{1/2}E[|U_t|^4]^{1/2}\\ &\leq E[|X_{t,i}X_{t,j}|^8]^{1/4}E[|X_{t,\ell}|^8]^{1/4}E[|U_t|^4]^{1/2}\\ &\leq E[|X_{t,i}|^{16}]^{1/8}E[|X_{t,j}|^{16}]^{1/8}E[|X_{t,\ell}|^8]^{1/4}E[|U_t|^4]^{1/2} < \infty\end{aligned}$$

by A3*. Finally, $E[X_{t,i}^2X_{t,j}^2X_{t,\ell}^2U_t^2] > 0$. Thus, $n^{-1/2}\sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell}U_t$ is asymptotically normal by theorem 5.25 of White (2001). Hence, $\boldsymbol{\iota}'\mathbf{D}_{ij\ell}\mathbf{M}\mathbf{U} = O_{\mathbb{P}}(n^{1/2})$, so that $\boldsymbol{\iota}'\mathbf{D}_{ij\ell}\mathbf{M}\mathbf{U} = o_{\mathbb{P}}(n^{3/4})$.

(iii) We have

$$\boldsymbol{\iota}'\mathbf{D}_{ij\ell m}\mathbf{M}\mathbf{U} = \sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}U_t - \sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}\mathbf{Z}_t'(\sum_{t=1}^n \mathbf{Z}_t\mathbf{Z}_t')^{-1}\sum_{t=1}^n \mathbf{Z}_tU_t.$$

By the ergodic theorem and A3*, $n^{-1}\sum X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}\mathbf{Z}_t$ and $n^{-1}\sum \mathbf{Z}_tU_t$ respectively converge to $E[X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}\mathbf{Z}_t]$ and $\mathbf{0}$ in probability. Also, $E[X_{t,i}X_{t,j}X_{t,\ell}X_{t,m} U_t|\mathbf{Z}_t] = 0$, so that the desired result follows by the ergodic theorem if $E[|X_{t,i}X_{t,j}X_{t,\ell} X_{t,m}U_t|] < \infty$. For this, we note that

$$\begin{aligned}&E[|X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}U_t|] \leq E[|X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}|^2]^{1/2}E[|U_t|^2]^{1/2}\\ &\quad \leq E[|X_{t,i}|^8]^{1/8}E[|X_{t,j}|^8]^{1/8}E[|X_{t,\ell}|^8]^{1/8}E|X_{t,m}|^8]^{1/8}E[|U_t|^2]^{1/2} < \infty,\end{aligned}$$

by A3*.

(iv) By our definitions, $\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{D}_{\ell m}\boldsymbol{\iota}$ is identical to

$$\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{D}_{\ell m}\boldsymbol{\iota} = \sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell}X_{t,m} - \sum_{t=1}^n X_{t,i}X_{t,j}\mathbf{Z}_t'(\sum_{t=1}^n \mathbf{Z}_t\mathbf{Z}_t')^{-1}\sum_{t=1}^n \mathbf{Z}_tX_{t,\ell}X_{t,m}.$$

We can apply the ergodic theorem to each element, so that $n^{-1}\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{D}_{\ell m}\boldsymbol{\iota} \to \tau_{ij\ell m}$ $a.s.$, where

$$\begin{aligned}\tau_{ij\ell m} &:= E[C_{t,ij}^*C_{t,\ell m}^*]\\ &= E[X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}] - E[X_{t,i}X_{t,j}\mathbf{Z}_t]'\{E[\mathbf{Z}_t\mathbf{Z}_t']\}^{-1}E[\mathbf{Z}_tX_{t,\ell}X_{t,m}],\end{aligned}$$

which is finite by Conditions A3* and A6. This yields the desired result. ■

Proof of Lemma 4: (i) First, we note that

$$\text{vech}(n^{-1/2}\widetilde{\mathbf{M}}) = n^{-1/2}\sum_{t=1}^{n}\mathbf{C}_t U_t - (n^{-1}\sum_{t=1}^{n}\mathbf{C}_t\mathbf{Z}_t')(n^{-1}\sum_{t}^{n}\mathbf{Z}_t\mathbf{Z}_t')^{-1}(n^{-1/2}\sum_{t=1}^{n}\mathbf{Z}_t U_t).$$

Second, the MDS CLT ensures that under our conditions, and, in particular, A6*,

$$n^{-1/2}\sum_{t=1}^{n}[\mathbf{Z}_t', \mathbf{C}_t']' U_t \overset{\text{A}}{\sim} N(\mathbf{0}, \widetilde{\mathbf{V}}_1).$$

Third, $n^{-1}\sum \mathbf{C}_t\mathbf{Z}_t'$ and $n^{-1}\sum \mathbf{Z}_t\mathbf{Z}_t'$ respectively converge to $E[\mathbf{C}_t\mathbf{Z}_t']$ and $E[\mathbf{Z}_t\mathbf{Z}_t]$ in probability by the WLLN. Therefore,

$$\text{vech}(n^{-1/2}\widetilde{\mathbf{M}}) \overset{\text{A}}{\sim} N(\mathbf{0}, E[U_t^2[\mathbf{C}_t - E[\mathbf{C}_t\mathbf{Z}_t']E[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbf{Z}_t][\mathbf{C}_t - E[\mathbf{C}_t\mathbf{Z}_t']E[\mathbf{Z}_t\mathbf{Z}_t']^{-1}\mathbf{Z}_t]']).$$

Note that the typical element of the covariance matrix is $E[U_t^2 C_{t,ij}^* C_{t,\ell m}^*]$. Thus, it follows that $n^{-1/2}\widetilde{\mathbf{M}} \Rightarrow \mathcal{M}$ from the symmetry of $\widetilde{\mathbf{M}}$ and the fact that $n^{-1/2}\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d} \Rightarrow \boldsymbol{d}'\mathcal{M}\boldsymbol{d} = \text{vec}(\boldsymbol{d}\boldsymbol{d}')'\text{vec}(\mathcal{M})$ for any $\boldsymbol{d} \in \mathbb{S}^{k-1}$ by the continuous mapping theorem. Therefore, $n^{-1/2}\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d} \Rightarrow \boldsymbol{b}'\text{vec}(\mathcal{M})$ by the definition of $\boldsymbol{b}$.

(ii) The desired result trivially follows from the fact that $n^{-1}\boldsymbol{b}'\mathbf{W}\boldsymbol{b} = \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{\ell=1}^{k}\sum_{m=1}^{k} n^{-1}(\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{D}_{\ell m}\boldsymbol{\iota})d_i d_j d_\ell d_m$ and that $n^{-1}\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{D}_{\ell m}\boldsymbol{\iota} \to \tau_{ij\ell m}$ $a.s.$, as shown in the proof of Lemma 3(iv). Thus, $n^{-1}\boldsymbol{b}'\mathbf{W}\boldsymbol{b} \to \boldsymbol{b}'\mathbf{W}^*\boldsymbol{b}$ a.s.

(iii) For each $i, j, \ell, m = 1, 2, \cdots, k$, $E[\mathcal{M}_{ij}\mathcal{M}_{\ell m}] = E[U_t^2 C_{t,ij}^* C_{t,\ell m}^*] = E[E[U_t^2 |\mathbf{X}_t] C_{t,ij}^* C_{t,\ell m}^*] = \sigma_*^2 E[C_{t,ij}^* C_{t,\ell m}^*]$ under A4(ii). This completes the proof. ■

Proof of Theorem 2: (i) We separate the proof into three parts: (a), (b), and (c). In (a), we prove the weak convergence of $n^{-1/2}\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d}$ as a function of $\boldsymbol{d}$. In (b), we show that $\boldsymbol{b}'\mathbf{W}\boldsymbol{b}$ obeys the ULLN as a function of $\boldsymbol{b}$. Finally, the covariance structure of $\mathcal{G}_2$ is derived in (c).

(a) Given that $n^{-1/2}\widetilde{\mathbf{M}} \Rightarrow \mathcal{M}$, the tightness of $\{n^{-1/2}\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d}\}$ as a function of $\boldsymbol{d}$ follows easily. For this, we first show that

$$\sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} n^{-1/2}\boldsymbol{d}'(\widetilde{\mathbf{M}} - \mathbf{M}^*)\boldsymbol{d} = o_{\mathbb{P}}(1),$$

where $\mathbf{M}^* := [\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{U} - \mathbf{P}_{ij}\mathbf{Z}'\mathbf{U}]$ and $\mathbf{P}_{ij} := (E[X_{t,i}X_{t,j}\mathbf{Z}_t'])(E[\mathbf{Z}_t\mathbf{Z}_t'])^{-1}$; and we then show that $\{n^{-1/2}\boldsymbol{d}'\mathbf{M}^*\boldsymbol{d}\}$ is tight as a function of $\boldsymbol{d}$.

First, we note that $\widetilde{\mathbf{M}} = [\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{U}]$ and

$$\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{U} - \sum_{t=1}^{n} C_{t,ij}^* U_t = \left[(\sum_{t=1}^{n} X_{t,i}X_{t,j}\mathbf{Z}_t')(\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{Z}_t')^{-1} - \mathbf{P}_{ij} \right] \sum_{t=1}^{n} \mathbf{Z}_t U_t,$$

which is $o_{\mathbb{P}}(n^{1/2})$, because $\sum_{t=1}^{n}\mathbf{Z}_t U_t = O_{\mathbb{P}}(n^{1/2})$, and $n^{-1}\sum_{t=1}^{n} X_{t,i}X_{t,j}\mathbf{Z}_t'$ and n^{-1} $\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{Z}_t'$ obey the WLLN, as we saw in the proof of Lemma 3(i). Thus, $n^{-1/2}(\widetilde{\mathbf{M}} - \mathbf{M}^*) = o_{\mathbb{P}}(1)$. This implies that $\sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} n^{-1/2}\boldsymbol{d}'(\widetilde{\mathbf{M}} - \mathbf{M}^*)\boldsymbol{d} = o_{\mathbb{P}}(1)$, because $\mathbb{S}^{k-1}$ is bounded. Next, for any $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\left[\sup_{\|\boldsymbol{d}-\widetilde{\boldsymbol{d}}\|<\delta} \frac{1}{\sqrt{n}}|\boldsymbol{d}'\mathbf{M}^*\boldsymbol{d} - \widetilde{\boldsymbol{d}}'\mathbf{M}^*\widetilde{\boldsymbol{d}}| > \varepsilon\right] \leq & \mathbb{P}\left[\sup_{\|\boldsymbol{d}-\widetilde{\boldsymbol{d}}\|<\delta} \frac{1}{\sqrt{n}}|\boldsymbol{d}'\mathbf{M}^*\boldsymbol{d} - \boldsymbol{d}'\mathbf{M}^*\widetilde{\boldsymbol{d}}| > \frac{\varepsilon}{2}\right] \\ & + \mathbb{P}\left[\sup_{\|\boldsymbol{d}-\widetilde{\boldsymbol{d}}\|<\delta} \frac{1}{\sqrt{n}}|\boldsymbol{d}'\mathbf{M}^*\widetilde{\boldsymbol{d}} - \widetilde{\boldsymbol{d}}'\mathbf{M}^*\widetilde{\boldsymbol{d}}| > \frac{\varepsilon}{2}\right].\end{aligned} \tag{22}$$

We have

$$\begin{aligned}\sup_{\|\boldsymbol{d}-\widetilde{\boldsymbol{d}}\|<\delta} |\boldsymbol{d}'\mathbf{M}^*\boldsymbol{d} - \boldsymbol{d}'\mathbf{M}^*\widetilde{\boldsymbol{d}}| &\leq \sup_{\|\boldsymbol{d}-\widetilde{\boldsymbol{d}}\|<\delta} \sum_{i=1}^{k}\sum_{j=1}^{k} |\{\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{U} - \mathbf{P}_{ij}\mathbf{Z}'\mathbf{U}\}d_i(d_j - \widetilde{d}_j)| \\ &\leq \delta \sum_{i=1}^{k}\sum_{j=1}^{k} |\{\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{U} - \mathbf{P}_{ij}\mathbf{Z}'\mathbf{U}\}|.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{P}\left[\sup_{\|\boldsymbol{d}-\widetilde{\boldsymbol{d}}\|<\delta} \frac{1}{\sqrt{n}}|\boldsymbol{d}'\mathbf{M}^*\boldsymbol{d} - \boldsymbol{d}'\mathbf{M}^*\widetilde{\boldsymbol{d}}| > \frac{\varepsilon}{2}\right] &\leq \mathbb{P}\left[\delta\sum_{i=1}^{k}\sum_{j=1}^{k} \frac{1}{\sqrt{n}}|\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{U} - \mathbf{P}_{ij}\mathbf{Z}'\mathbf{U}| > \frac{\varepsilon}{2}\right] \\ &\leq \sum_{i=1}^{k}\sum_{j=1}^{k} \mathbb{P}\left[\frac{1}{\sqrt{n}}|\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{U} - \mathbf{P}_{ij}\mathbf{Z}'\mathbf{U}| > \frac{\varepsilon}{2\delta k^2}\right] \\ &\leq \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{4\delta^2 k^4}{\varepsilon^2}\sigma_{ijij}\end{aligned}$$

using Markov's inequality, where $i, j, \ell, m = 1, 2, \cdots, k$, $\sigma_{i,j,\ell,m} := E[U_t^2 C_{t,ij}^* C_{t,\ell m}^*]$.

Thus,

$$\limsup_{n\to\infty} \mathbb{P}\left[\sup_{\|\boldsymbol{d}-\widetilde{\boldsymbol{d}}\|<\delta} \frac{1}{\sqrt{n}} |\boldsymbol{d}'\mathrm{M}^*\boldsymbol{d} - \widetilde{\boldsymbol{d}}'\mathrm{M}^*\widetilde{\boldsymbol{d}}| > \varepsilon\right] \le \frac{8\delta^2 k^4}{\varepsilon^2} \sum_{i=1}^{k}\sum_{j=1}^{k} \sigma_{ijij} \tag{23}$$

using the same method applied to the second component in the RHS of (22). We can make the RHS of (23) as small as we wish by letting δ be small. It follows that $\{n^{-1/2}\boldsymbol{d}'\mathrm{M}^*\boldsymbol{d}\}$ is tight. Its finite-dimensional distribution also follows from the multivariate CLT, which ensures $n^{-1/2}\widetilde{\mathrm{M}} \Rightarrow \mathcal{M}$. As this is straightforward, we leave the details aside here. Therefore, $n^{-1/2}\boldsymbol{d}'\widetilde{\mathrm{M}}\boldsymbol{d}$ weakly converges to $\boldsymbol{d}'\mathcal{M}\boldsymbol{d}$ as a function of $\boldsymbol{d}$.

(b) Since $\boldsymbol{b}'\mathbf{W}\boldsymbol{b} = \mathrm{vec}(\boldsymbol{d}\boldsymbol{d}')'\mathbf{W}\mathrm{vec}(\boldsymbol{d}\boldsymbol{d}')$ and

$$\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{\ell=1}^{k}\sum_{m=1}^{k} \tau_{ij\ell m} d_i d_j d_\ell d_m = \mathrm{vec}(\boldsymbol{d}\boldsymbol{d}')'\mathbf{W}^*\mathrm{vec}(\boldsymbol{d}\boldsymbol{d}'),$$

it follows that

$$\begin{aligned}\sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} &\left| n^{-1}\mathrm{vec}(\boldsymbol{d}\boldsymbol{d}')'\mathbf{W}\mathrm{vec}(\boldsymbol{d}\boldsymbol{d}') - \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{\ell=1}^{k}\sum_{m=1}^{k} \tau_{ij\ell m} d_i d_j d_\ell d_m \right| \\ &\le \sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{\ell=1}^{k}\sum_{m=1}^{k} \left| n^{-1}(\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{D}_{\ell m}\boldsymbol{\iota}) - \tau_{ij\ell m}\right| \times |d_i d_j d_\ell d_m| .\end{aligned}$$

We already saw that $n^{-1}\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{M}\mathbf{D}_{\ell m}\boldsymbol{\iota} \to \tau_{ij\ell m}\ a.s.$ in the proof of Lemma 3(ii). Therefore,

$$\sup_{\boldsymbol{d}\in\mathbb{S}^{k-1}} \left| n^{-1}\mathrm{vec}(\boldsymbol{d}\boldsymbol{d}')'\mathbf{W}\mathrm{vec}(\boldsymbol{d}\boldsymbol{d}') - \mathrm{vec}(\boldsymbol{d}\boldsymbol{d}')'\mathbf{W}^*\mathrm{vec}(\boldsymbol{d}\boldsymbol{d}')\right| \to 0 \quad a.s.$$

This implies that $n^{-1}\boldsymbol{b}'\mathbf{W}\boldsymbol{b} = n^{-1}\mathrm{vec}(\boldsymbol{d}\boldsymbol{d}')'\mathbf{W}\mathrm{vec}(\boldsymbol{d}\boldsymbol{d}')$ obeys the ULLN as a function of $\boldsymbol{d}$.

(c) Further, $\{n^{-1/2}\boldsymbol{d}'\widetilde{\mathrm{M}}\boldsymbol{d}, n^{-1}\mathrm{vec}(\boldsymbol{d}\boldsymbol{d}')'\mathbf{W}\mathrm{vec}(\boldsymbol{d}\boldsymbol{d}'), \hat{\sigma}^2_{n,0}\} \Rightarrow \{\boldsymbol{d}'\mathcal{M}\boldsymbol{d}, \mathrm{vec}(\boldsymbol{d}\boldsymbol{d}')'\mathbf{W}^*\mathrm{vec}(\boldsymbol{d}\boldsymbol{d}'), \sigma^2_*\}$ as functions of $\boldsymbol{d}$ by (a), (b), and the fact that $\hat{\sigma}^2_{n,0} \to \sigma^2_*\ a.s.$ Therefore, as a function of $\boldsymbol{d}$, $\{\hat{\sigma}^2_{n,0}\boldsymbol{b}'\mathbf{W}\boldsymbol{b}\}^{-1/2}\boldsymbol{d}'\widetilde{\mathrm{M}}\boldsymbol{d} = \{\hat{\sigma}^2_{n,0}\boldsymbol{b}'\mathbf{W}\boldsymbol{b}\}^{-1/2}\boldsymbol{b}'\mathrm{vec}(\widetilde{\mathrm{M}}) \Rightarrow \mathcal{G}_2(\boldsymbol{b}) := \{\sigma^2_*\boldsymbol{b}'\mathbf{W}^*\boldsymbol{b}\}^{-1/2}\boldsymbol{b}'\mathrm{vec}(\mathcal{M}) = \{\mathcal{I}(\boldsymbol{b},\boldsymbol{b})\}^{-1/2}\boldsymbol{b}'\mathrm{vec}(\mathcal{M})$ by the converging together lemma (theorem 3.9 of Billingsley (1999, p. 37)) and the definition of $\mathcal{I}(\boldsymbol{b},\boldsymbol{b})$.

As the final step, we derive the covariance structure of $\mathcal{G}_2$ from the asymptotic covariance between $n^{-1/2}\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d}$ and $n^{-1/2}\widetilde{\boldsymbol{d}}'\widetilde{\mathbf{M}}\widetilde{\boldsymbol{d}}$. We already proved in (a) that $n^{-1/2}(\widetilde{\mathbf{M}}-\mathbf{M}^*) = o_{\mathbb{P}}(1)$. Further, we note that

$$
\begin{aligned}
n^{-1}E[\boldsymbol{d}'\mathbf{M}^*\boldsymbol{d}\,\widetilde{\boldsymbol{d}}'\mathbf{M}^*\widetilde{\boldsymbol{d}}] &= \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{\ell=1}^{k}\sum_{m=1}^{k} E[U_t^2 C_{t,ij}^* C_{t,\ell m}^*] d_i d_j \widetilde{d}_\ell \widetilde{d}_m \\
&= \boldsymbol{b}'E[\text{vec}(\mathcal{M})\text{vec}(\mathcal{M})']\widetilde{\boldsymbol{b}}.
\end{aligned}
$$

This shows that $n^{-1}E[\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d}\,\widetilde{\boldsymbol{d}}'\widetilde{\mathbf{M}}\widetilde{\boldsymbol{d}}] \to \boldsymbol{b}'E[\text{vec}(\mathcal{M})\text{vec}(\mathcal{M})']\widetilde{\boldsymbol{b}}$ in probability, which is $\mathcal{K}(\boldsymbol{b},\widetilde{\boldsymbol{b}})$ from the fact that $E[U_t^2 C_{t,ij}^* C_{t,\ell m}^*] = E[\mathcal{M}_{ij}\mathcal{M}_{\ell m}]$. Therefore, it follows that $E[\mathcal{G}_2(\boldsymbol{b})\mathcal{G}_2(\widetilde{\boldsymbol{b}})] = \rho_2(\boldsymbol{b},\widetilde{\boldsymbol{b}})$. This is the desired covariance structure.

The desired result now follows from (a), (b), and (c).

(ii) By Lemma 4(iii), $\mathcal{K}(\boldsymbol{b}, \widetilde{\boldsymbol{b}}) = \mathcal{I}(\boldsymbol{b},\widetilde{\boldsymbol{b}})$, so that $\rho_2(\boldsymbol{b},\widetilde{\boldsymbol{b}}) = \mathcal{I}(\boldsymbol{b},\boldsymbol{b})^{-1/2}\mathcal{I}(\boldsymbol{b},\widetilde{\boldsymbol{b}})$ $\mathcal{I}(\widetilde{\boldsymbol{b}},\widetilde{\boldsymbol{b}})^{-1/2}$. This completes the proof. ■

Before proving Lemmas 5 and 6, we provide some supplementary lemmas.

Lemma A1: *Given A1, A2, A3**, *A4*(i), *A6**, *and* $\mathcal{H}_{02}$,

(i) *for* $j = 0, 1, 2, 3,$ *and* 4, $\mathbf{U}'\mathbf{J}_j = O_{\mathbb{P}}(n^{1/2})$;

(ii) *for* $j = 0, 1, 2, 3,$ *and* 4, $\mathbf{J}_0'\mathbf{J}_j = O_{\mathbb{P}}(n)$;

(iii) *for* $j = 1, 2,$ *and* 3, $\mathbf{J}_1'\mathbf{J}_j = O_{\mathbb{P}}(n)$; *and*

(iv) $\mathbf{J}_2'\mathbf{J}_2 = O_{\mathbb{P}}(n)$.

Proof of Lemma A1: (i) First, we note that $\mathbf{J}_0 := [\mathbf{X}, c_0\boldsymbol{\iota}]$ and that $\mathbf{U}'\mathbf{Z} = O_{\mathbb{P}}(n^{1/2})$ as shown in the proof of Lemma 4(i). This shows that $\mathbf{U}'\mathbf{J}_0 = O_{\mathbb{P}}(n^{1/2})$. Second, we note that $\mathbf{J}_1 = c_1[\mathbf{0}_{n\times k}, \mathbf{G}\boldsymbol{\iota}]$, where $\mathbf{G} := \text{diag}\{\mathbf{X}\boldsymbol{d}\}$, so that $\mathbf{U}'\mathbf{J}_1 = c_1[\mathbf{0}_{n\times k}, \mathbf{U}'\mathbf{X}\boldsymbol{d}]$. We further note that $\boldsymbol{d}$ is bounded and $\mathbf{U}'\mathbf{X} = O_{\mathbb{P}}(n^{1/2})$, implying that $\mathbf{U}'\mathbf{J}_1 = O_{\mathbb{P}}(n^{1/2})$. Third, $\mathbf{U}'\mathbf{J}_2 = c_2[\mathbf{0}_{n\times k}, \mathbf{U}'\mathbf{G}^2\boldsymbol{\iota}]$ from the fact that $\mathbf{J}_2 = c_2[\mathbf{0}_{n\times k}, \mathbf{G}^2\boldsymbol{\iota}]$ and that $\mathbf{U}'\mathbf{G}^2\boldsymbol{\iota} = \sum_i d_i \sum_j d_j \sum_{t=1}^n (X_{t,i}X_{t,j}U_t)$. We now note that the proof of Lemma 3(i) shows that for each i and j, $\sum X_{t,i}X_{t,j}U_t = O_{\mathbb{P}}(n^{1/2})$. Therefore, $\mathbf{U}'\mathbf{J}_2 = O_{\mathbb{P}}(n^{1/2})$. Fourth, $\mathbf{U}'\mathbf{J}_3 = c_3[\mathbf{0}_{n\times k}, \mathbf{U}'\mathbf{G}^3\boldsymbol{\iota}]$ because $\mathbf{J}_3 = c_3[\mathbf{0}_{n\times k}, \mathbf{G}^3\boldsymbol{\iota}]$. The proof of Lemma 3$(ii)$ now shows that for each i, j and ℓ, $\sum X_{t,i}X_{t,j}X_{t,\ell}U_t = O_{\mathbb{P}}(n^{1/2})$, and $\mathbf{U}'\mathbf{G}^3\boldsymbol{\iota} = \sum_i d_i \sum_j d_j \sum_\ell d_\ell \sum_{t=1}^n (X_{t,i}X_{t,j}X_{t,\ell}U_t)$, so that $\mathbf{U}'\mathbf{J}_3 = O_{\mathbb{P}}(n^{1/2})$. Finally, $\mathbf{U}'\mathbf{J}_4 = c_4[\mathbf{0}_{n\times k}, \mathbf{U}'\mathbf{G}^4\boldsymbol{\iota}]$ using the fact that $\mathbf{J}_4 = c_4[\mathbf{0}_{n\times k}, \mathbf{G}^4\boldsymbol{\iota}]$. The proof of

Lemma 3(iii) proves that for each i, j, ℓ, and m, $\sum X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}U_t = O_{\mathbb{P}}(n^{1/2})$ and $\mathbf{U}'\mathbf{G}^4\boldsymbol{\iota} = \sum_i d_i \sum_j d_j \sum_\ell d_\ell \sum_m d_m \sum_{t=1}^n (X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}U_t)$, so that $\mathbf{U}'\mathbf{J}_4 = O_{\mathbb{P}}(n^{1/2})$.

(ii) First, we note that $\mathbf{J}_0'\mathbf{J}_0 = \mathbf{H}_0 = O_{\mathbb{P}}(n)$ by the fact that $\mathbf{Z}'\mathbf{Z} = O_{\mathbb{P}}(n)$ as given in the proof of Lemma 4(i). Second, $\mathbf{J}_0'\mathbf{J}_1 = c_1[\mathbf{0}_{(k+1)\times k}, \mathbf{Z}'\mathbf{G}\boldsymbol{\iota}] = c_1[\mathbf{0}_{(k+1)\times k}, \mathbf{Z}'\mathbf{X}\boldsymbol{d}]$. Given that $\boldsymbol{d}$ is bounded and that for each i, $\sum X_{t,i}\mathbf{Z}_t = O_{\mathbb{P}}(n)$, we note that $\mathbf{Z}'\mathbf{X}\boldsymbol{d} = \sum_i d_i \sum_{t=1}^n X_{t,i}\mathbf{Z}_t$, so that $\mathbf{J}_0'\mathbf{J}_1 = O_{\mathbb{P}}(n)$. Third, $\mathbf{J}_0'\mathbf{J}_2 = c_2[\mathbf{0}_{(k+1)\times k}, \mathbf{Z}'\mathbf{G}^2\boldsymbol{\iota}]$. Now, for each i and j, $\sum X_{t,i}X_{t,j}\mathbf{Z}_t = O_{\mathbb{P}}(n)$ as given in the proof of Lemma 3(i), and $\mathbf{Z}'\mathbf{G}^2\boldsymbol{\iota} = \sum_i d_i \sum_j d_j \sum_{t=1}^n X_{t,i}X_{t,j}\mathbf{Z}_t$. Thus, $\mathbf{J}_0'\mathbf{J}_2 = O_{\mathbb{P}}(n)$. Fourth, we note that $\mathbf{J}_0'\mathbf{J}_3 = c_3[\mathbf{0}_{(k+1)\times k}, \mathbf{Z}'\mathbf{G}^3\boldsymbol{\iota}]$. Now, for each i, j, and ℓ, $\sum X_{t,i}X_{t,j}X_{t,\ell}\mathbf{Z}_t = O_{\mathbb{P}}(n)$, as we saw in the proof of Lemma 3(ii), and $\mathbf{Z}'\mathbf{G}^3\boldsymbol{\iota} = \sum_i d_i \sum_j d_j \sum_\ell d_\ell \sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell}$ $\mathbf{Z}_t$. Thus, $\mathbf{J}_0'\mathbf{J}_3 = O_{\mathbb{P}}(n)$. Finally, $\mathbf{J}_0'\mathbf{J}_4 = c_4[\mathbf{0}_{(k+1)\times k}, \mathbf{Z}'\mathbf{G}^4\boldsymbol{\iota}]$. Now, for each i, j, m, and ℓ, $\sum X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}\mathbf{Z}_t = O_{\mathbb{P}}(n)$, as we saw in the proof of Lemma 3(iii), and $\mathbf{Z}'\mathbf{G}^4\boldsymbol{\iota} = \sum_i d_i \sum_j d_j \sum_\ell d_\ell \sum_m d_m \sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}\mathbf{Z}_t$. Thus, $\mathbf{J}_0'\mathbf{J}_4 = O_{\mathbb{P}}(n)$.

(iii) First, $\mathbf{J}_1'\mathbf{J}_1 = c_1^2\boldsymbol{\iota}'\mathbf{G}^2\boldsymbol{\iota}\boldsymbol{\xi}\boldsymbol{\xi}'$, where we let $\boldsymbol{\xi} := [\mathbf{0}_{(k\times 1)}', 1]'$. Given the definition of $\mathbf{G} := \text{diag}\{\mathbf{X}\boldsymbol{d}\}$ and the fact that $\boldsymbol{d}$ is bounded, $\boldsymbol{\iota}'\mathbf{G}^2\boldsymbol{\iota} = O_{\mathbb{P}}(n)$ because for each i and j, $\sum X_{t,i}X_{t,j} = O_{\mathbb{P}}(n)$ and $\boldsymbol{\iota}'\mathbf{G}^2\boldsymbol{\iota} = \sum_i d_i \sum_j d_j \sum_{t=1}^n X_{t,i}X_{t,j}$. Thus, $\mathbf{J}_1'\mathbf{J}_1 = O_{\mathbb{P}}(n)$. Second, we also note that $\mathbf{J}_1'\mathbf{J}_2 = c_1c_2\boldsymbol{\iota}'\mathbf{G}^3\boldsymbol{\iota}\boldsymbol{\xi}\boldsymbol{\xi}'$, so that $\boldsymbol{\iota}'\mathbf{G}^3\boldsymbol{\iota} = O_{\mathbb{P}}(n)$ because for each i, j, and ℓ, $\sum X_{t,i}X_{t,j}X_{t,\ell} = O_{\mathbb{P}}(n)$ and $\boldsymbol{\iota}'\mathbf{G}^3\boldsymbol{\iota} = \sum_i d_i \sum_j d_j \sum_\ell d_\ell$ $\sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell}$, implying that $\mathbf{J}_1'\mathbf{J}_2 = O_{\mathbb{P}}(n)$. Finally, $\mathbf{J}_1'\mathbf{J}_3 = c_1c_3\boldsymbol{\iota}'\mathbf{G}^4\boldsymbol{\iota}\boldsymbol{\xi}\boldsymbol{\xi}'$, and for each i, j, ℓ, and m, $\sum X_{t,i}X_{t,j}X_{t,\ell}X_{t,m} = O_{\mathbb{P}}(n)$. This implies that $\mathbf{J}_1'\mathbf{J}_3 = O_{\mathbb{P}}(n)$ because $\boldsymbol{\iota}'\mathbf{G}^4\boldsymbol{\iota} = \sum_i d_i \sum_j d_j \sum_\ell d_\ell \sum_m d_m \sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}$.

(iv) We note that $\mathbf{J}_2'\mathbf{J}_2 = c_2^2\boldsymbol{\iota}'\mathbf{G}^4\boldsymbol{\iota}\boldsymbol{\xi}\boldsymbol{\xi}'$, and we have already seen that $\boldsymbol{\iota}'\mathbf{G}^4\boldsymbol{\iota} = O_{\mathbb{P}}(n)$ in the proof of Lemma A1(iii), so that $\mathbf{J}_2'\mathbf{J}_2 = O_{\mathbb{P}}(n)$. This completes the proof. ■

Lemma A2: *Given A1 and A2*, *when* $\mathbf{V}(h;\boldsymbol{d})$ is defined as $[\mathbf{Q}(h\boldsymbol{d})'\mathbf{Q}(h\boldsymbol{d})]^{-1}$,

(i) $\frac{\partial}{\partial h}\mathbf{V}(0;\boldsymbol{d}) = -\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}$;

(ii) $\frac{\partial^2}{\partial h^2}\mathbf{V}(0;\boldsymbol{d}) = 2\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1} - \mathbf{H}_0^{-1}\mathbf{H}_2\mathbf{H}_0^{-1}$;

(iii) $\frac{\partial^3}{\partial h^3}\mathbf{V}(0;\boldsymbol{d}) = \mathbf{H}_0^{-1}\{-6\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_1 + 3\mathbf{H}_2\mathbf{H}_0^{-1}\mathbf{H}_1 + 3\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_2 - \mathbf{H}_3\}\mathbf{H}_0^{-1}$; *and*

(iv)

$$\frac{\partial^4}{\partial h^4}\mathbf{V}(0;\boldsymbol{d}) = 24\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}$$
$$+ 6\mathbf{H}_0^{-1}\{\mathbf{H}_2\mathbf{H}_0^{-1}\mathbf{H}_2 - 2\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_2\mathbf{H}_0^{-1}\mathbf{H}_1 - 2\mathbf{H}_2\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_1\}\mathbf{H}_0^{-1}$$
$$+ 4\mathbf{H}_0^{-1}\{-3\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_2 + \mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_3 + \mathbf{H}_3\mathbf{H}_0^{-1}\mathbf{H}_1\}\mathbf{H}_0^{-1} - \mathbf{H}_0^{-1}\mathbf{H}_4\mathbf{H}_0^{-1}.$$

As proving Lemma A2 is straightforward but tedious, we omit this.

Lemma A3: *Given A1, A2, A3***, A4*(i)*, A6***, and* $\mathcal{H}_{02}$,

(i) *for* $j = 0, 1, 2, 3,$ *and* 4, $\frac{\partial^j}{\partial h^j}\mathbf{V}(h;\boldsymbol{d}) = O_{\mathbb{P}}(n^{-1})$;

(ii) $\mathbf{J}_0'(\mathbf{K}+\mathbf{K}')\mathbf{U} = o_{\mathbb{P}}(n^{3/4})$ *and* $\mathbf{U}'(\mathbf{K}+\mathbf{K}')\mathbf{U} = o_{\mathbb{P}}(n^{3/4})$, *where*

$$\mathbf{K} := \mathbf{J}_3\mathbf{H}_0^{-1}\mathbf{J}_0' - 3\mathbf{J}_2\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{J}_0' + 6\mathbf{J}_1\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{J}_0' - 3\mathbf{J}_1\mathbf{H}_0^{-1}\mathbf{H}_2\mathbf{H}_0^{-1}\mathbf{J}_1'$$
$$3\mathbf{J}_2\mathbf{H}_0^{-1}\mathbf{J}_1' - 3\mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{J}_0' + 3\mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_2\mathbf{H}_0^{-1}\mathbf{J}_0';\ \textit{and}$$

(iii) $\mathbf{J}_0'(\mathbf{L}+\mathbf{L}')\mathbf{U} = o_{\mathbb{P}}(n)$ *and* $\mathbf{U}'(\mathbf{L}+\mathbf{L}')\mathbf{U} = o_{\mathbb{P}}(n)$, *where*

$$\mathbf{L} := \left\{\mathbf{J}_4\mathbf{V}(0;\boldsymbol{d}) + 4\mathbf{J}_3\frac{\partial}{\partial h}\mathbf{V}(0;\boldsymbol{d}) + 6\mathbf{J}_2\frac{\partial^2}{\partial h^2}\mathbf{V}(0;\boldsymbol{d}) + 4\mathbf{J}_1\frac{\partial^3}{\partial h^3}\mathbf{V}(0;\boldsymbol{d})\right\}\mathbf{J}_0'$$
$$+ \left\{4\mathbf{J}_3\mathbf{V}(0;\boldsymbol{d}) + 12\mathbf{J}_2\frac{\partial}{\partial h}\mathbf{V}(0;\boldsymbol{d}) + 6\mathbf{J}_1\frac{\partial^2}{\partial h^2}\mathbf{V}(0;\boldsymbol{d})\right\}\mathbf{J}_1' + 3\mathbf{J}_2\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_2'\ \ \textit{and;}$$

(iv) $\mathbf{J}_0'\mathbf{M} = \mathbf{0}_{(k+1)\times n}$ *and* $\mathbf{M}\mathbf{J}_0 = \mathbf{0}_{n\times(k+1)}$; *and*

(v) $\mathbf{J}_1'\mathbf{M} = \mathbf{0}_{(k+1)\times n}$ *and* $\mathbf{M}\mathbf{J}_1 = \mathbf{0}_{n\times(k+1)}$.

Proof of Lemma A3: (i) First, we note that $\mathbf{H}_0 = \mathbf{J}_0'\mathbf{J}_0$ and $\mathbf{Z}'\mathbf{Z} = O_{\mathbb{P}}(n)$ as proved in the proof of Lemma 4(i), implying that $\mathbf{V}(0,\boldsymbol{d}) = \mathbf{H}_0^{-1} = O_{\mathbb{P}}(n^{-1})$. Second, we note that $\mathbf{H}_1 = \mathbf{J}_0'\mathbf{J}_1 + \mathbf{J}_1'\mathbf{J}_0$. Lemma A1$(ii)$ shows that $\mathbf{J}_0'\mathbf{J}_1 = O_{\mathbb{P}}(n)$, so that $\mathbf{H}_1 = O_{\mathbb{P}}(n)$. Given that $\mathbf{H}_0^{-1} = O_{\mathbb{P}}(n^{-1})$, $\frac{\partial}{\partial h}\mathbf{V}(0,\boldsymbol{d}) = -\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1} = O_{\mathbb{P}}(n^{-1})$. Third, note that $\mathbf{H}_2 = \mathbf{J}_2'\mathbf{J}_0 + \mathbf{J}_1'\mathbf{J}_1 + \mathbf{J}_0'\mathbf{J}_2$, so that Lemmas A1$(ii$ and $iii)$ imply that $\mathbf{H}_2 = O_{\mathbb{P}}(n)$. Now, Lemma A2$(ii)$ shows that $\frac{\partial^2}{\partial h^2}\mathbf{V}(0,\boldsymbol{d}) = O_{\mathbb{P}}(n^{-1})$. Fourth, we note that $\mathbf{H}_3 = \mathbf{J}_3'\mathbf{J}_0 + 3\mathbf{J}_2'\mathbf{J}_1 + 3\mathbf{J}_1'\mathbf{J}_2 + \mathbf{J}_0'\mathbf{J}_3$, and Lemmas A1$(ii$ and $iii)$ imply that $\mathbf{H}_3 = O_{\mathbb{P}}(n)$. Therefore, Lemma A2$(iii)$ shows that $\frac{\partial^3}{\partial h^3}\mathbf{V}(0,\boldsymbol{d}) = O_{\mathbb{P}}(n^{-1})$. Finally, $\mathbf{H}_4 = \mathbf{J}_4'\mathbf{J}_0 + 4\mathbf{J}_3'\mathbf{J}_1 + 6\mathbf{J}_2'\mathbf{J}_2 + 4\mathbf{J}_1'\mathbf{J}_3 + \mathbf{J}_0'\mathbf{J}_4$, and Lemmas A1$(ii$, iii, and $iv)$ imply that $\mathbf{H}_4 = O_{\mathbb{P}}(n)$, so that $\frac{\partial^4}{\partial h^4}\mathbf{V}(0,\boldsymbol{d}) = O_{\mathbb{P}}(n^{-1})$ by applying Lemma A2(iv).

(ii) First, we note that $\mathbf{K}+\mathbf{K}'=\mathbf{K}_1+\mathbf{K}_2+\mathbf{K}_2'$, where $\mathbf{K}_1:=\mathbf{J}_0\frac{\partial^3}{\partial h^3}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0'$ and

$$\mathbf{K}_2:=\left\{\mathbf{J}_3\mathbf{V}(0;\boldsymbol{d})+3\mathbf{J}_2\frac{\partial}{\partial h}\mathbf{V}(0;\boldsymbol{d})+3\mathbf{J}_1\frac{\partial^2}{\partial h^2}\mathbf{V}(0;\boldsymbol{d})\right\}\mathbf{J}_0'+3\mathbf{J}_2\frac{\partial}{\partial h}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_1'.$$

Second, we prove that $\mathbf{J}_0'(\mathbf{K}+\mathbf{K}')\mathbf{U}=O_{\mathbb{P}}(n^{1/2})$. For this, note that $\mathbf{J}_0'\mathbf{K}_1\mathbf{U}=\mathbf{J}_0'\mathbf{J}_0\frac{\partial^3}{\partial h^3}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0'\mathbf{U}$, $\mathbf{J}_0'\mathbf{J}_0=O_{\mathbb{P}}(n)$, $\frac{\partial^3}{\partial h^3}\mathbf{V}(0;\boldsymbol{d})=O_{\mathbb{P}}(n^{-1})$, and $\mathbf{J}_0'\mathbf{U}=O_{\mathbb{P}}(n^{1/2})$, implying that $\mathbf{J}_0'\mathbf{K}_1\mathbf{U}=O_{\mathbb{P}}(n^{1/2})$. Also, $\mathbf{U}'\mathbf{K}_1\mathbf{U}=\mathbf{U}'\mathbf{J}_0\frac{\partial^3}{\partial h^3}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0'\mathbf{U}=O_{\mathbb{P}}(1)$. Further, the main components constituting $\mathbf{J}_0'\mathbf{K}_2\mathbf{U}$ are $\mathbf{J}_0'\mathbf{J}_{3-j}\frac{\partial^j}{\partial h^j}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0'\mathbf{U}$ $(j=0,1,2)$ and $\mathbf{J}_0'\mathbf{J}_2\frac{\partial}{\partial h}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_1'\mathbf{U}$. By Lemmas A1($i\sim iii$), it follows that $\mathbf{J}_0'\mathbf{J}_{3-j}=O_{\mathbb{P}}(n)$, $\frac{\partial^j}{\partial h^j}\mathbf{V}(0;\boldsymbol{d})=O_{\mathbb{P}}(n^{-1})$, and both $\mathbf{J}_0'\mathbf{U}$ and $\mathbf{J}_1'\mathbf{U}$ are $O_{\mathbb{P}}(n^{1/2})$. These imply that $\mathbf{J}_0'\mathbf{K}_2\mathbf{U}=O_{\mathbb{P}}(n^{1/2})$. We can apply the same arguments for $\mathbf{J}_0'\mathbf{K}_2'\mathbf{U}$ to obtain that $\mathbf{J}_0'\mathbf{K}_2'\mathbf{U}=O_{\mathbb{P}}(n^{1/2})$. This implies that $\mathbf{J}_0'(\mathbf{K}_1+\mathbf{K}_2+\mathbf{K}_2')\mathbf{U}=\mathbf{J}_0'(\mathbf{K}+\mathbf{K}')\mathbf{U}=O_{\mathbb{P}}(n^{1/2})=o_{\mathbb{P}}(n^{3/4})$ as desired.

Finally, we prove that $\mathbf{U}'(\mathbf{K}+\mathbf{K}')\mathbf{U}=O_{\mathbb{P}}(1)$. For this, note that $\mathbf{U}'\mathbf{K}_1\mathbf{U}=\mathbf{U}'\mathbf{J}_0\frac{\partial^3}{\partial h^3}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0'\mathbf{U}$, $\frac{\partial^3}{\partial h^3}\mathbf{V}(0;\boldsymbol{d})=O_{\mathbb{P}}(n^{-1})$, and $\mathbf{J}_0'\mathbf{U}=O_{\mathbb{P}}(n^{1/2})$, implying that $\mathbf{U}'\mathbf{K}_1\mathbf{U}=O_{\mathbb{P}}(n^{1/2})$. Also, $\mathbf{U}'\mathbf{K}_1\mathbf{U}=\mathbf{U}'\mathbf{J}_0\frac{\partial^3}{\partial h^3}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0'\mathbf{U}=O_{\mathbb{P}}(1)$. Further, we note that the main components constituting $\mathbf{U}'\mathbf{K}_2\mathbf{U}$ are $\mathbf{U}'\mathbf{J}_{3-j}\frac{\partial^j}{\partial h^j}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0'\mathbf{U}$ $(j=0,1,2)$ and $\mathbf{U}'\mathbf{J}_2\frac{\partial}{\partial h}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_1'\mathbf{U}$. By Lemmas A1($i$ and iii), $\mathbf{U}'\mathbf{J}_{3-j}=O_{\mathbb{P}}(n^{1/2})$, $\frac{\partial^j}{\partial h^j}\mathbf{V}(0;\boldsymbol{d})=O_{\mathbb{P}}(n^{-1})$. These imply that $\mathbf{U}'\mathbf{K}_2\mathbf{U}=O_{\mathbb{P}}(1)$. We can apply the same arguments to $\mathbf{U}'\mathbf{K}_2'\mathbf{U}$ to obtain that $\mathbf{U}'\mathbf{K}_2'\mathbf{U}=O_{\mathbb{P}}(1)$, so that $\mathbf{U}'(\mathbf{K}_1+\mathbf{K}_2+\mathbf{K}_2')\mathbf{U}=\mathbf{U}'(\mathbf{K}+\mathbf{K}')\mathbf{U}=O_{\mathbb{P}}(1)=o_{\mathbb{P}}(n^{3/4})$.

(iii) First, the main components constituting $\mathbf{J}_0'\mathbf{L}\mathbf{U}$ are $\mathbf{J}_0'\mathbf{J}_{4-j}\frac{\partial^j}{\partial h^j}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0'\mathbf{U}$ $(j=0,1,2,3)$, $\mathbf{J}_0'\mathbf{J}_{3-j}\frac{\partial^j}{\partial h^j}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_1'\mathbf{U}$ $(j=0,1,2)$, and $\mathbf{J}_0'\mathbf{J}_2\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_2'\mathbf{U}$. Lemma A1($i$) shows that $\mathbf{J}_j'\mathbf{U}=O_{\mathbb{P}}(n^{1/2})$; Lemma A1($ii$) shows that $\mathbf{J}_0'\mathbf{J}_j=O_{\mathbb{P}}(n)$; and Lemma A3($i$) shows that $\frac{\partial^j}{\partial h^j}\mathbf{V}(0;\boldsymbol{d})=O_{\mathbb{P}}(n^{-1})$. These facts imply that $\mathbf{J}_0'\mathbf{L}\mathbf{U}=O_{\mathbb{P}}(n^{1/2})=o_{\mathbb{P}}(n)$, and it can be easily proved that $\mathbf{J}_0'\mathbf{L}'\mathbf{U}=o_{\mathbb{P}}(n)$ in a similar way, so that $\mathbf{J}_0'(\mathbf{L}+\mathbf{L}')\mathbf{U}=o_{\mathbb{P}}(n)$.

Next, we note that the main components constituting $\mathbf{U}'\mathbf{L}\mathbf{U}$ are $\mathbf{U}'\mathbf{J}_{4-j}\frac{\partial^j}{\partial h^j}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0'\mathbf{U}$ $(j=0,1,2,3)$, $\mathbf{U}'\mathbf{J}_{3-j}\frac{\partial^j}{\partial h^j}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_1'\mathbf{U}$ $(j=0,1,2)$, and $\mathbf{U}'\mathbf{J}_2\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_2'\mathbf{U}$. Lemma A1($i$) shows that $\mathbf{J}_j'\mathbf{U}=O_{\mathbb{P}}(n^{1/2})$; and Lemma A3($i$) shows that $\frac{\partial^j}{\partial h^j}\mathbf{V}(0;\boldsymbol{d})=O_{\mathbb{P}}(n^{-1})$. These facts imply that $\mathbf{U}'\mathbf{L}\mathbf{U}=O_{\mathbb{P}}(1)=o_{\mathbb{P}}(n)$. We can also prove in the same way that $\mathbf{U}'\mathbf{L}'\mathbf{U} = O_{\mathbb{P}}(1) = o_{\mathbb{P}}(n)$, implying that $\mathbf{U}'(\mathbf{L}+\mathbf{L}')\mathbf{U} = O_{\mathbb{P}}(1) = o_{\mathbb{P}}(n)$.

(iv) We note that $\mathbf{M} := \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' = \mathbf{I} - \mathbf{J}_0(\mathbf{J}_0'\mathbf{J}_0)^{-1}\mathbf{J}_0'$, implying that $\mathbf{M}\mathbf{J}_0 = \mathbf{J}_0 - \mathbf{J}_0 = \mathbf{0}_{n\times(k+1)}$. This also implies that $\mathbf{J}_0'\mathbf{M} = \mathbf{0}_{(k+1)\times n}$.

(v) We note that $\mathbf{J}_1 = [\mathbf{0}, \mathbf{G}\boldsymbol{\iota}]$, so that $\mathbf{M}\mathbf{J}_1 = [\mathbf{0}, \mathbf{M}\mathbf{G}\boldsymbol{\iota}] = [\mathbf{0}, \mathbf{M}\mathbf{X}\boldsymbol{d}] = \mathbf{0}_{n\times(k+1)}$. This also implies that $\mathbf{J}_1'\mathbf{M} = \mathbf{0}_{(k+1)\times n}$ and completes the proof. ■

Proof of Lemma 5: (i) When $\mathbf{P}(h;\boldsymbol{d}) := \mathbf{Q}(h\boldsymbol{d})\mathbf{V}(h;\boldsymbol{d})\mathbf{Q}(h\boldsymbol{d})'$, we note that $\frac{\partial}{\partial h}\mathbf{P}(0;\boldsymbol{d}) = \mathbf{J}_1\mathbf{H}_0^{-1}\mathbf{J}_0' - \mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{J}_0' + \mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{J}_1'$ by Lemma A2(i). Given that $\mathbf{H}_1 = \mathbf{J}_0'\mathbf{J}_1 + \mathbf{J}_1'\mathbf{J}_0$, if we plug this into $\frac{\partial}{\partial h}\mathbf{P}(0;\boldsymbol{d})$, then it follows that $\frac{\partial}{\partial h}\mathbf{P}(0;\boldsymbol{d}) = \mathbf{0}$. Finally, the consequence follows from the fact that $\frac{\partial}{\partial h}L_n(0;\boldsymbol{d}) = (\mathbf{Y} - \alpha\boldsymbol{\iota})'\frac{\partial}{\partial h}\mathbf{P}(0;\boldsymbol{d})(\mathbf{Y} - \alpha\boldsymbol{\iota}) = (\boldsymbol{\gamma}^{*\prime}\mathbf{C}' + \mathbf{U}')\frac{\partial}{\partial h}\mathbf{P}(0;\boldsymbol{d})(\mathbf{C}\boldsymbol{\gamma}^* + \mathbf{U}) = \mathbf{0}$.

(ii) Tedious algebra using Lemma A2(ii) shows that $\frac{\partial^2}{\partial h^2}\mathbf{P}(0;\boldsymbol{d}) = \mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{J}_2'\mathbf{M} + \mathbf{M}\mathbf{J}_2\mathbf{H}_0^{-1}\mathbf{J}_0'$. Thus, $\frac{\partial^2}{\partial h^2}L_n(0;\boldsymbol{d}) = (\boldsymbol{\gamma}^{*\prime}\mathbf{J}_0' + \mathbf{U}')(\mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{J}_2'\mathbf{M} + \mathbf{M}\mathbf{J}_2\mathbf{H}_0^{-1}\mathbf{J}_0')(\mathbf{J}_0\boldsymbol{\gamma}^* + \mathbf{U})$. We now note that (a) $\mathbf{J}_0'\frac{\partial^2}{\partial h^2}\mathbf{P}(0;\boldsymbol{d})\mathbf{J}_0 = \mathbf{0}$; (b) $\mathbf{J}_0'\frac{\partial^2}{\partial h^2}\mathbf{P}(0;\boldsymbol{d})\mathbf{U} = 4\mathbf{J}_2'\mathbf{M}\mathbf{U}$; and (c) $\mathbf{U}'\frac{\partial^2}{\partial h^2}\mathbf{P}(0;\boldsymbol{d})\mathbf{U} = 2\mathbf{U}'\mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{J}_2\mathbf{M}\mathbf{U}$ after using Lemma A3(v) and the facts that $\mathbf{H}_1 = \mathbf{J}_1'\mathbf{J}_0 + \mathbf{J}_0'\mathbf{J}_1$ and $\mathbf{H}_2 = \mathbf{J}_2'\mathbf{J}_0 + 2\mathbf{J}_1'\mathbf{J}_1 + \mathbf{J}_0'\mathbf{J}_2$. The desired result now follows from (a), (b), and (c).

(iii) Tedious algebra using Lemma A2(iii) shows that $\frac{\partial^3}{\partial h^3}\mathbf{P}(0;\boldsymbol{d}) = \mathbf{K} + \mathbf{K}'$, and we obtain that $\frac{\partial^3}{\partial h^3}L_n(0;\boldsymbol{d}) = (\boldsymbol{\gamma}^{*\prime}\mathbf{J}_0' + \mathbf{U}')(\mathbf{K} + \mathbf{K}')(\mathbf{J}_0\boldsymbol{\gamma}^* + \mathbf{U}) = \boldsymbol{\gamma}^{*\prime}\mathbf{J}_0'(\mathbf{K} + \mathbf{K}')\mathbf{J}_0\boldsymbol{\gamma}^* + 2\boldsymbol{\gamma}^{*\prime}\mathbf{J}_0'(\mathbf{K} + \mathbf{K}')\mathbf{U} + 2\mathbf{U}'\mathbf{K}\mathbf{U}$. Given this, Lemma A3$(ii)$ implies that $2\boldsymbol{\gamma}^{*\prime}\mathbf{J}_0'(\mathbf{K} + \mathbf{K}')\mathbf{U} + 2\mathbf{U}'\mathbf{K}\mathbf{U} = o_{\mathbb{P}}(n^{3/4})$. Thus, $\frac{\partial^3}{\partial h^3}L_n(0;\boldsymbol{d}) = \boldsymbol{\gamma}^{*\prime}\mathbf{J}_0'(\mathbf{K} + \mathbf{K}')\mathbf{J}_0\boldsymbol{\gamma}^* + o_{\mathbb{P}}(n^{3/4})$.

(iv) First, by some algebra, it follows that $\frac{\partial^4}{\partial h^4}\mathbf{P}(0;\boldsymbol{d}) = \mathbf{L} + \mathbf{L}' + \mathbf{J}_0\frac{\partial^4}{\partial h^4}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0'$. Second, we note that $\frac{\partial^4}{\partial h^4}L_n(0;\boldsymbol{d}) = (\boldsymbol{\gamma}^{*\prime}\mathbf{J}_0' + \mathbf{U}')\frac{\partial^4}{\partial h^4}\mathbf{P}(0;\boldsymbol{d})(\mathbf{J}_0\boldsymbol{\gamma}^* + \mathbf{U}) = \boldsymbol{\gamma}^{*\prime}\mathbf{J}_0'\frac{\partial^4}{\partial h^4}\mathbf{P}(0;\boldsymbol{d})\mathbf{J}_0\boldsymbol{\gamma}^* + 2\mathbf{U}'\frac{\partial^4}{\partial h^4}\mathbf{P}(0;\boldsymbol{d})\mathbf{J}_0\boldsymbol{\gamma}^* + \mathbf{U}'\frac{\partial^4}{\partial h^4}\mathbf{P}(0;\boldsymbol{d})\mathbf{U}$. Third, further tedious algebra shows that $\mathbf{J}_0'\frac{\partial^4}{\partial h^4}\mathbf{P}(0;\boldsymbol{d})\mathbf{J}_0 = -6\mathbf{J}_2'\mathbf{M}\mathbf{J}_2$ using the fact that $\frac{\partial^4}{\partial h^4}\mathbf{P}(0;\boldsymbol{d}) = \mathbf{L} + \mathbf{L}' + \mathbf{J}_0\frac{\partial^4}{\partial h^4}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0'$ and Lemma A2(iv). Finally, using Lemmas A3(iii), A1(i), A3(i), and A1(ii) shows that $\mathbf{U}'[\mathbf{L} + \mathbf{L}' + \mathbf{J}_0\frac{\partial^4}{\partial h^4}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0']\mathbf{J}_0 = O_{\mathbb{P}}(n^{1/2}) = o_{\mathbb{P}}(n)$; and using Lemmas A3$(iii)$, A1$(i)$, and A3$(i)$ shows that $\mathbf{U}'[\mathbf{L} + \mathbf{L}' + \mathbf{J}_0\frac{\partial^4}{\partial h^4}\mathbf{V}(0;\boldsymbol{d})\mathbf{J}_0']\mathbf{U} = O_{\mathbb{P}}(1) = o_{\mathbb{P}}(n)$. Therefore, $\frac{\partial^4}{\partial h^4}L_n(0;\boldsymbol{d}) = -6\boldsymbol{\gamma}^{*\prime}\mathbf{J}_2'\mathbf{M}\mathbf{J}_2\boldsymbol{\gamma}^* + o_{\mathbb{P}}(n)$. This completes the proof. ■

Proof of Lemma 6: (i) To show the given result, we examine each component in Lemma 5(ii) separately. First, if we let $\mathbf{G}_j := \text{diag}\{\mathbf{X}_j\}$, then $\sum_{j=1}^k d_j\mathbf{G}_j = \sum_{j=1}^k \text{diag}\{\mathbf{X}_j d_j\}$

$= \text{diag}\{\mathbf{X}\boldsymbol{d}\} = \mathbf{G}$, so that

$$\begin{aligned}\boldsymbol{\iota}'\mathbf{G}^2\mathbf{MU} &= \boldsymbol{\iota}'\sum_{i=1}^{k}\text{diag}\{\mathbf{X}_i d_i\}\sum_{j=1}^{k}\text{diag}\{\mathbf{X}_j d_j\}\mathbf{MU}\\ &= \boldsymbol{\iota}'\sum_{i=1}^{k}\sum_{j=1}^{k} d_i d_j \text{diag}\{\mathbf{X}_i\}\text{diag}\{\mathbf{X}_j\}\mathbf{MU} = \sum_{i=1}^{k}\sum_{j=1}^{k} d_i d_j \boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{MU} = \boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d}\end{aligned}$$

by noting that $\widetilde{\mathbf{M}} := [\boldsymbol{\iota}'\mathbf{D}_{ij}\mathbf{MU}]$. Therefore, $\boldsymbol{\iota}'\mathbf{G}^2\mathbf{MU} = \boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d}$. Next, we note that Lemmas A1($i \sim iii$) imply that $\mathbf{U}'\mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{J}_2'\mathbf{MU} = \mathbf{U}'\mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{J}_2'\mathbf{U} - \mathbf{U}'\mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{J}_2'\mathbf{J}_0 \mathbf{H}_0^{-1}\mathbf{J}_0'\mathbf{U} = O_{\mathbb{P}}(1)$, because $\mathbf{U}'\mathbf{J}_0 = O_{\mathbb{P}}(n^{1/2})$, $\mathbf{J}_2'\mathbf{U} = O_{\mathbb{P}}(n^{1/2})$ by Lemma A1(i); $\mathbf{H}_0^{-1} = \mathbf{V}(0;\boldsymbol{d}) = O_{\mathbb{P}}(n^{-1})$ by Lemma A1(iii); and $\mathbf{J}_2'\mathbf{J}_0 = O_{\mathbb{P}}(1)$ by Lemma A1(ii). From this, it trivially follows that $\mathbf{U}'\mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{J}_2'\mathbf{MU} = o_{\mathbb{P}}(n^{1/2})$. These two facts now show that $\frac{\partial^2}{\partial h^2}L_n^{(3)}(\mathbf{0};\alpha) = 2(\alpha^* - \alpha)(c_2/c_0)\boldsymbol{d}'\widetilde{\mathbf{M}}\boldsymbol{d} + o_{\mathbb{P}}(n^{1/2})$.

(ii) We note that $\mathbf{H}_0^{-1}\mathbf{J}_0'\mathbf{J}_0 = \mathbf{J}_0'\mathbf{J}_0\mathbf{H}_0^{-1} = \mathbf{I}_k$ and $\mathbf{H}_1 = \mathbf{J}_1'\mathbf{J}_0 + \mathbf{J}_0'\mathbf{J}_1$. Using these facts, we have that

$$\begin{aligned}\mathbf{J}_0'\mathbf{K}\mathbf{J}_0 = -3[&\mathbf{J}_0'\mathbf{H}_2\mathbf{H}_0^{-1}\mathbf{J}_0'\mathbf{J}_1 - \mathbf{J}_0'\mathbf{J}_1\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{J}_0'\mathbf{J}_1\\ &+ \mathbf{J}_1'\mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{H}_1\mathbf{H}_0^{-1}\mathbf{H}_1 - \mathbf{J}_1'\mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{H}_2 + \mathbf{J}_1'\mathbf{J}_2],\end{aligned}$$

so that exploiting the fact that $\mathbf{H}_1 = \mathbf{J}_1'\mathbf{J}_0 + \mathbf{J}_0'\mathbf{J}_1$ and $\mathbf{H}_2 = \mathbf{J}_2'\mathbf{J}_0 + \mathbf{J}_1'\mathbf{J}_1 + \mathbf{J}_0'\mathbf{J}_2$ yields that

$$\mathbf{J}_0'(\mathbf{K}+\mathbf{K}')\mathbf{J}_0 = 6\mathbf{J}_1'\mathbf{J}_0\mathbf{H}_0^{-1}\mathbf{J}_1'\mathbf{M}\mathbf{J}_1 + \mathbf{J}_1'\mathbf{M}\mathbf{J}_1\mathbf{H}_0^{-1}\mathbf{J}_0'\mathbf{J}_1 - 3\mathbf{J}_1'\mathbf{M}\mathbf{J}_2 - 3\mathbf{J}_2'\mathbf{M}\mathbf{J}_1.$$

Finally, we now note that $\mathbf{J}_1'\mathbf{M} = \mathbf{M}\mathbf{J}_1 = \mathbf{0}$ by Lemma A3(iv), implying that $\mathbf{J}_0'(\mathbf{K} + \mathbf{K}')\mathbf{J}_0 = \mathbf{0}$. The desired result follows from this.

(iii) We now note that $\boldsymbol{\gamma}^{*\prime}\mathbf{J}_2'\mathbf{M}\mathbf{J}_2\boldsymbol{\gamma}^* = (\alpha^* - \alpha)^2(c_2/c_0)^2\boldsymbol{\iota}'\mathbf{G}^2\mathbf{M}\mathbf{G}^2\boldsymbol{\iota}$. Thus, if $\boldsymbol{\iota}'\mathbf{G}^2\mathbf{M}\mathbf{G}^2\boldsymbol{\iota} = \boldsymbol{d}'(\mathbf{I}_k\otimes\boldsymbol{d})'\mathbf{W}(\mathbf{I}_k\otimes\boldsymbol{d})\boldsymbol{d}$, then the desired result follows. We derive this. First, we note that

$$\boldsymbol{d}'(\mathbf{I}_k\otimes\boldsymbol{d})'\mathbf{W}(\mathbf{I}_k\otimes\boldsymbol{d})\boldsymbol{d} = \sum_{i=1}^{k}\sum_{j=1}^{k} d_i(\boldsymbol{d}'\mathbf{W}_{ij}\boldsymbol{d})d_j,$$

and that $\boldsymbol{d}'\mathbf{W}_{ij}\boldsymbol{d} = \sum_{\ell=1}^{k}\sum_{m=1}^{k}\boldsymbol{\iota}'\mathbf{G}_i\mathbf{G}_j\mathbf{M}d_\ell d_m\mathbf{G}_\ell\mathbf{G}_m\boldsymbol{\iota} = \boldsymbol{\iota}'\mathbf{G}_i\mathbf{G}_j\mathbf{M}\sum_{\ell=1}^{k} d_\ell\mathbf{G}_\ell\sum_{m=1}^{k}$

$d_m \mathbf{G}_m \boldsymbol{\iota} = \boldsymbol{\iota}' \mathbf{G}_i \mathbf{G}_j \mathbf{M} \mathbf{G}^2 \boldsymbol{\iota}$. Next, we also note that

$$\sum_{i=1}^{k} \sum_{j=1}^{k} d_i (\boldsymbol{d}' \mathbf{W}_{ij} \boldsymbol{d}) d_j = \sum_{i=1}^{k} \sum_{j=1}^{k} d_i \boldsymbol{\iota}' \mathbf{G}_i \mathbf{G}_j \mathbf{M} \mathbf{G}^2 \boldsymbol{\iota} d_j$$
$$= \boldsymbol{\iota}' \sum_{i=1}^{k} d_i \mathbf{G}_i \sum_{j=1}^{k} d_j \mathbf{G}_j \mathbf{M} \mathbf{G}^2 \boldsymbol{\iota} = \boldsymbol{\iota}' \mathbf{G}^2 \mathbf{M} \mathbf{G}^2 \boldsymbol{\iota}.$$

That is, $\boldsymbol{d}'(\mathbf{I}_k \otimes \boldsymbol{d})' \mathbf{W} (\mathbf{I}_k \otimes \boldsymbol{d}) \boldsymbol{d} = \boldsymbol{\iota}' \mathbf{G}^2 \mathbf{M} \mathbf{G}^2 \boldsymbol{\iota}$. This completes the proof. ■

Proof of Lemma 7: (ii) For given n, we obtain

$$N_n^{(4)}(h, \boldsymbol{d}) = 8 \mathbf{U}' \mathbf{M} \{(\partial/\partial h) \boldsymbol{\Psi}(h, \boldsymbol{d})\} \mathbf{U}' \mathbf{M} \{(\partial^3/\partial h^3) \boldsymbol{\Psi}(h, \boldsymbol{d})\}$$
$$+ 6 \left\{ \mathbf{U}' \mathbf{M} (\partial^2/\partial h^2) \boldsymbol{\Psi}(h, \boldsymbol{d}) \right\}^2 + 2 \mathbf{U}' \mathbf{M} \boldsymbol{\Psi}(h, \boldsymbol{d}) \mathbf{U}' \mathbf{M} \left\{ (\partial^4/\partial h^4) \boldsymbol{\Psi}(h, \boldsymbol{d}) \right\}.$$

We also note that $\lim_{h \downarrow 0} \mathbf{U}' \mathbf{M} \boldsymbol{\Psi}(h, \boldsymbol{d}) = c_0 \mathbf{U}' \mathbf{M} \boldsymbol{\iota} = 0 \ a.s.$, and $\lim_{h \downarrow 0} \mathbf{U}' \mathbf{M} (\partial/\partial h)$ $\boldsymbol{\Psi}(h, \boldsymbol{d}) = c_1 \mathbf{U}' \mathbf{M} \mathbf{X} \boldsymbol{d} = 0 \ a.s.$, so that

$$\lim_{h \downarrow 0} N_n^{(4)}(h, \boldsymbol{d}) = \lim_{h \downarrow 0} 6 \left\{ \mathbf{U}' \mathbf{M} (\partial^2/\partial h^2) \boldsymbol{\Psi}(h, \boldsymbol{d}) \right\}^2$$
$$= 6 c_2^2 \left\{ \sum_{i=1}^{k} \sum_{j=1}^{k} \boldsymbol{\iota}' \mathbf{D}_{ij} \mathbf{M} \mathbf{U} d_i d_j \right\}^2 \quad a.s.$$

(iii) Similarly, for given $\boldsymbol{d}$ and n, we obtain that

$$D_n^{(4)}(h, \boldsymbol{d}) = 6 \left\{ (\partial^2/\partial h^2) \boldsymbol{\Psi}(h, \boldsymbol{d}) \right\}' \mathbf{M} \left\{ (\partial^2/\partial h^2) \boldsymbol{\Psi}(h, \boldsymbol{d}) \right\}$$
$$+ 8 \left\{ (\partial/\partial h) \boldsymbol{\Psi}(h, \boldsymbol{d}) \right\}' \mathbf{M} \left\{ (\partial^3/\partial h^3) \boldsymbol{\Psi}(h, \boldsymbol{d}) \right\}$$
$$+ 2 \boldsymbol{\Psi}(h, \boldsymbol{d})' \mathbf{M} \left\{ (\partial^4/\partial h^4) \boldsymbol{\Psi}(h, \boldsymbol{d}) \right\}.$$

Also, we note that $\lim_{h \downarrow 0} \boldsymbol{\Psi}(h, \boldsymbol{d})' \mathbf{M} = c_0 \boldsymbol{\iota}' \mathbf{M} = \mathbf{0} \ a.s.$, and $\lim_{h \downarrow 0} \{(\partial/\partial h) \boldsymbol{\Psi}(h, \boldsymbol{d})\}'$ $\mathbf{M} = c_1 \boldsymbol{d}' \mathbf{X}' \mathbf{M} = \mathbf{0} \ a.s.$, as we saw in the proof of Lemma 7(ii). Therefore,

$$\lim_{h \downarrow 0} D_n^{(4)}(h, \boldsymbol{d}) = \lim_{h \downarrow 0} 6 \left\{ (\partial^2/\partial h^2) \boldsymbol{\Psi}(h, \boldsymbol{d}) \right\}' \mathbf{M} \left\{ (\partial^2/\partial h^2) \boldsymbol{\Psi}(h, \boldsymbol{d}) \right\}$$
$$= 6 c_2^2 \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{\ell=1}^{k} \sum_{m=1}^{k} \boldsymbol{\iota}' \mathbf{D}_{ij} \mathbf{M} \mathbf{D}_{\ell m} \boldsymbol{\iota} \, d_i d_j d_\ell d_m \quad a.s.$$

This completes the proof. ■

Before proving Lemma 8, we simplify our notation by suppressing all function arguments but h and $\widetilde{h}$. That is, we let $\mathcal{J}(h)$ and $\mathcal{T}(h, \widetilde{h})$ denote $\mathcal{J}(h, \boldsymbol{d}, h, \boldsymbol{d})$ and $\mathcal{T}(h, \boldsymbol{d}, \widetilde{h}, \widetilde{\boldsymbol{d}})$ respectively. We first provide the following supplementary lemmas. As these results hold by the Lebesgue dominated convergence theorem and tedious algebra, we omit the proofs for brevity.

Lemma B1: *Given A1, A2, A3**, A4, A5, A6**, A7, A8, and* $\mathcal{H}_0$,

(i) *for* $\ell = 0$, 1, *and each* $\widetilde{h} \geq 0$, $\lim_{h\downarrow 0} \mathcal{T}^{(\ell,0)}(h, \widetilde{h}) = 0$, *where* $\mathcal{T}^{(\ell,m)}(h, \widetilde{h}) := (\partial^m \partial^\ell / \partial \widetilde{h}^m \partial h^\ell)\, \mathcal{T}(h, \widetilde{h})$;

(ii) *for* $\ell = 0, 1, 2,$ *and* 3, $\lim_{h\downarrow 0} \mathcal{J}^{(\ell)}(h) = 0$, *where* $\mathcal{J}^{(\ell)}(h) := (\partial^\ell / \partial h^\ell)\mathcal{J}(h)$;

(iii) $\mathcal{T}^{(2,0)}(0, \widetilde{h}) = c_2 \mathcal{H}(\boldsymbol{b}, \widetilde{h}, \widetilde{\boldsymbol{b}})$; *and*

(iv) $\mathcal{J}^{(4)}(0) = 6c_2^2 \mathcal{I}(\boldsymbol{b}, \boldsymbol{b})$.

Lemma B2: *Given A1, A2, A3**, A4, A5, A6**, A7, A8, and* $\mathcal{H}_0$,

(i) $\lim_{\widetilde{h}\downarrow 0} \lim_{h\downarrow 0} \mathcal{T}(h, \widetilde{h}) = 0$;

(ii) *for* $\ell = 0$ *and* 1, $\lim_{\widetilde{h}\downarrow 0} \lim_{h\downarrow 0} \mathcal{T}^{(1,\ell)}(h, \widetilde{h}) = 0$;

(iii) *for* $\ell = 0$ *and* 1, $\lim_{\widetilde{h}\downarrow 0} \lim_{h\downarrow 0} \mathcal{T}^{(2,\ell)}(h, \widetilde{h}) = 0$;

(iv) *for* $\ell = 0$ *and* 1, $\lim_{\widetilde{h}\downarrow 0} \lim_{h\downarrow 0} \mathcal{T}^{(3,\ell)}(h, \widetilde{h}) = 0$;

(v) $\lim_{\widetilde{h}\downarrow 0} \lim_{h\downarrow 0} \mathcal{T}^{(4,0)}(h, \widetilde{h}) = 0$; *and*

(vi) $\lim_{\widetilde{h}\downarrow 0} \lim_{h\downarrow 0} \mathcal{T}^{(2,2)}(h, \widetilde{h}) = c_2^2 \mathcal{K}(\boldsymbol{b}, \widetilde{\boldsymbol{b}})$.

Proof of Lemma 8: (i) Given the definition of ρ_1,

$$\rho_1(h, \widetilde{h}) := \frac{\mathcal{T}(h, \widetilde{h})}{\{\mathcal{J}(h)\}^{1/2} \{\mathcal{J}(\widetilde{h})\}^{1/2}},$$

Lemma B1(i and ii) implies that $\lim_{h\downarrow 0} \mathcal{T}(h, \widetilde{h}) = 0$ and $\lim_{h\downarrow 0} \mathcal{J}(h) = 0$. Therefore, we apply L'Hôspital's rule. We note that

$$\mathcal{T}(h, \widetilde{h}) = \mathcal{T}(0, \widetilde{h}) + \mathcal{T}^{(1,0)}(0, \widetilde{h})h + \frac{1}{2}\mathcal{T}^{(2,0)}(0, \widetilde{h})2h^2 + o(h^2) = \frac{1}{2} c_2 \mathcal{H}(\boldsymbol{b}, \widetilde{h}, \widetilde{\boldsymbol{b}})h^2 + o(h^2) \tag{24}$$

by Lemma B1(i and iii) and also that

$$\mathcal{J}(h) = \mathcal{J}(0) + \mathcal{J}^{(1)}(0)h + \frac{1}{2}\mathcal{J}^{(2)}(0)h^2 + \frac{1}{3!}\mathcal{J}^{(3)}(0)h^3 + \frac{1}{4!}\mathcal{J}^{(4)}(0)h^4 + o(h^4)$$
$$= \frac{1}{24}\mathcal{J}^{(4)}(0)h^4 + o(h^4) = \frac{1}{4}c_2^2\mathcal{I}(\boldsymbol{b},\boldsymbol{b})h^4 + o(h^4) \quad (25)$$

by Lemma B1(ii and iv). Hence,

$$\lim_{h\downarrow 0} \frac{\mathcal{T}(h,\widetilde{h})}{\{\mathcal{J}(h)\}^{1/2}\{\mathcal{J}(\widetilde{h})\}^{1/2}} = \frac{c_2\mathcal{H}(\boldsymbol{b},\widetilde{h},\widetilde{\boldsymbol{b}},)}{\{c_2^2\mathcal{I}(\boldsymbol{b},\boldsymbol{b})\}^{1/2}\{\mathcal{J}(\widetilde{h},\widetilde{\boldsymbol{b}})\}^{1/2}} = \frac{\text{sgn}[c_2]\mathcal{H}(\boldsymbol{b},\widetilde{h},\widetilde{\boldsymbol{d}})}{\{\mathcal{I}(\boldsymbol{b},\boldsymbol{b})\}^{1/2}\{\mathcal{J}(\widetilde{h},\widetilde{\boldsymbol{b}})\}^{1/2}}.$$

(ii) We apply Taylor's expansion to $\mathcal{T}(h,\widetilde{h})$. Lemma B2 then implies that

$$\mathcal{T}(h,\widetilde{h}) = \mathcal{T}(0,0) + \sum_{i=1}^{4}\sum_{j=0}^{i}\frac{1}{i!}\binom{i}{j}\mathcal{T}^{(i-j,j)}(0,0)h^{i-j}\widetilde{h}^j + o((h^2+\widetilde{h}^2)^2)$$
$$= \frac{1}{4}c_2^2\mathcal{K}(\boldsymbol{b},\widetilde{\boldsymbol{b}})h^2\widetilde{h}^2 + o((h^2+\widetilde{h}^2)^2),$$

where the first and second equalities hold by Lemmas B2(i) and B2($ii{\sim}v$), respectively. Also,

$$\mathcal{J}(h) = \frac{1}{24}\mathcal{J}^{(4)}(0)h^4 + o(h^4) = \frac{1}{4}c_2^2\mathcal{I}(\boldsymbol{b},\boldsymbol{b})h^4 + o(h^4)$$

by (25). Thus,

$$\lim_{\widetilde{h}\downarrow 0}\lim_{h\downarrow 0} \frac{\mathcal{T}(h,\widetilde{h})}{\{\mathcal{J}(h)\}^{1/2}\{\mathcal{J}(\widetilde{h})\}^{1/2}} = \frac{c_2^2\mathcal{K}(\boldsymbol{b},\widetilde{\boldsymbol{b}})}{\{c_2^2\mathcal{I}(\boldsymbol{b},\boldsymbol{b})\}^{1/2}\{c_2^2\mathcal{I}(\widetilde{\boldsymbol{b}},\widetilde{\boldsymbol{b}})\}^{1/2}}$$
$$= \frac{\mathcal{K}(\boldsymbol{b},\widetilde{\boldsymbol{b}})}{\{\mathcal{I}(\boldsymbol{b},\boldsymbol{b})\}^{1/2}\{\mathcal{I}(\widetilde{\boldsymbol{b}},\widetilde{\boldsymbol{b}})\}^{1/2}}.$$

Note that this is $\rho_2(\boldsymbol{b},\widetilde{\boldsymbol{b}})$, and this completes the proof. ■

References

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

Bartlett, M. (1953*a*). Approximate Confidence Intervals. *Biometrika*, *40*, 12 – 19.

Bartlett, M. (1953*b*). Approximate Confidence Intervals II. More Than One Unknown Parameter. *Biometrika*, *40*, 306 – 317.

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

Bierens, H. and Hartog, J. (1988). Nonlinear Regression with Discrete Explanatory Variables with an Application to the Earnings Function. *Journal of Econometrics*, *38*, 269 – 299.

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

Billingsley, P. (1999). *Convergence of Probability Measures*. New York: Wiley.

Candès, E. (2003). "Ridgelets: Estimating with Ridge Functions," *Annals of Statistics*, *31*, 1561 – 1599.

Chen, X. and White, H. (1999). Improved Rates and Asymptotic Normality for Nonparametric Neural Network Estimators. *IEEE Transactions on Information Theory*, *45*, 682 – 691.

Cho, J. and White, H. (2007). Testing for Regime Switching. *Econometrica*, *75*, 1671 – 1720.

Cho, J. and White, H. (2009). Directionally Differentiable Econometric Models. Yonsei University Discussion Paper.

Cho, J. and White, H. (2010). Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models. *Journal of Econometrics*, *157*, 458 – 480.

Dacunha–Castelle, D. and Gassiat, E. (1999). Testing the Order of a Model Using Locally Conic Parametrization: Population Mixtures and Stationary ARMA Processes. *Annals of Statistics*, *27*, 1178 – 1209.

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., Massart, P., and Rio, E. (1995). Invariance Principles for Absolutely Regular Empirical Processes. *Annales de l'Institut Henri Poincaré, Probabilites et Statistiques*, *31*, 393 – 427.

Escanciano, J.C. (2009). On the Lack of Power of Omnibus Specification Tests. *Econometric Theory*, *25*, 162 – 194.

Gallant, R. and White, H. (1992). On Learning the Derivatives of an Unknown Mapping with Multilayer Feedforward Networks. *Neural Networks*, *5*, 129 – 138.

Giacomini, R., Gottschling, A., Haefke, C., and White, H. (2008). Mixtures of t-Distributions for Finance and Forecasting. *Journal of Econometrics*, *144*, 175 – 192.

Granger, C. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. New York: Oxford University Press.

Hansen, B. (1996). Inference When a Nuisance Parameter is Not Identified under the Null Hypothesis. *Econometrica*, *64*, 413 – 430.

Hansen, B. (2006). Interval Forecasts and Parameter Uncertainty. *Journal of Econometrics*, *135*, 377 – 398.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, *2*, 359 – 366.

Hornik, K., Stinchcombe, M., and White, H. (1990). Universal Approximation of an Unknown Mapping and Its Derivatives Using Multi-layer Feedforward Networks. *Neural Networks*, *3*, 551 – 560.

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*. NY: Academic Press.

Kuan, C.-M. and White, H. (1994). Artificial Neural Networks: An Econometric Perspective. *Econometric Reviews*, *13*, 1 – 92.

Le Cam, L. (1960). Local Asymptotically Normal Families of Distributions. *University of California Publications in Statistics*, *3*, 37 – 98.

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.

Lindsay, B. (1995). *Mixture Models*: *Theory, Geometry, and Applications*. Hayward: Institute for Mathematical Statistics.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988). Testing Linearity against Smooth Transition Autoregressive Models. *Biometrika*, *75*, 491-499.

McCullagh, P. (1987). *Tensor Methods in Statistics*. London: Chapman and Hall.

Neyman, J. and Scott, E. (1965). Asymptotically Optimal Tests of Composite Hypotheses for Randomized Experiments with Noncontrolled Predictor Variables. *Journal of the American Statistical Association*, *60*, 699 – 721.

Neyman, J. and Scott, E. (1966). On the Use of $C(\alpha)$ Optimal Tests of Composite Hypothesis. *Bulletin of the International Statistics Institute*, *41*, 477 – 497.

Phillips, P. (1998). New Tools for Understanding Spurious Regressions. *Econometrica*, *66*, 1299 – 1326.

Ranga Rao, R. (1962). Relations Between Weak and Uniform Convergence of Measures with Applications. *Annals of Mathematical Statistics*, *33*, 659 – 680.

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., C.-F. Lin, and C.W.J. Granger (1993). Power of the Neural Network Linearity Test. *Journal of Time Series Analysis*, *14*, 309 – 323.

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

van der Vaart, A. (1998). *Asymptotic Statistics*. New York: Cambridge University Press.

White, H. (1989*a*). An Additional Hidden Unit Test for Neglected Nonlinearity in Multilayer Feedforward Networks. *Proceedings of the International Joint Conference on Neural Networks*, II, New York, NY: IEEE Press, 451 – 455.

White, H. (1989*b*). Some Asymptotic Results for Learning in Single Hidden Layer Feedforward Network Models. *Journal of the American Statistical Association*, *84*, 1003 – 1013.

White, H. (1990). Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings. *Neural Networks*, *3*, 535 – 549.

White, H. (1992). Nonparametric Estimation of Conditional Quantiles Using Neural Networks. *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. (1996). Parametric Statistical Estimation Using Artificial Neural Networks. P. Smolensky, M.C. Mozer, and D.E. Rumelhart, eds., *Mathematical Perspectives on Neural Networks*. HillDale, NJ: L. Erlbaum Associates, 719 – 775.

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

Table 1: Critical Values
Number of Replications: 50,000
DGP: $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim$ IID $N(0, 1)$
Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda \exp(\delta Y_{t-1}) + U_t$
$\Delta_{0.5} = [-0.5, 0.5]$, $\Delta_{1.0} = [-1, 1]$, $\Delta_{1.5} = [-1.5, 1.5]$, $\Delta_{2.0} = [-2, 2]$, and $K = 150$

Nominal Level $\backslash$ Δ	$\Delta_{0.5}$	$\Delta_{1.0}$	$\Delta_{1.5}$	$\Delta_{2.0}$
1.00 %	7.7974	8.4051	9.1206	9.7248
5.00 %	4.7399	5.4245	6.0594	6.6222
10.0 %	3.4747	4.1282	4.6833	5.2558

Table 2: Empirical Rejection Rates (in Percent)
Number of Replications: 10,000
DGP: $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim$ IID $N(0, 1)$
Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda \exp(\delta Y_{t-1}) + U_t$
$\Delta_{0.5} = [-0.5, 0.5]$, $\Delta_{1.0} = [-1, 1]$, $\Delta_{1.5} = [-1.5, 1.5]$, $\Delta_{2.0} = [-2, 2]$, and $K = 150$

Asymptotic Distribution	$\mathcal{QLR}(\Delta_{0.5}; K)$						
Nominal Level $\backslash$ Sample Size	50	100	200	500	1,000	2,000	5,000
1.00 %	0.26	0.36	0.60	0.64	0.83	0.90	0.97
5.00 %	2.38	3.00	3.53	4.57	4.80	4.70	5.05
10.0 %	6.00	6.87	8.20	9.13	9.49	9.25	9.79
Asymptotic Distribution	$\mathcal{QLR}(\Delta_{1.0}; K)$						
Nominal Level $\backslash$ Sample Size	50	100	200	500	1,000	2,000	5,000
1.00 %	0.37	0.44	0.61	0.72	0.84	0.79	0.92
5.00 %	2.45	2.82	3.31	3.66	4.12	4.01	4.49
10.0 %	5.68	6.11	7.12	8.06	8.32	8.41	8.98
Asymptotic Distribution	$\mathcal{QLR}(\Delta_{1.5}; K)$						
Nominal Level $\backslash$ Sample Size	50	100	200	500	1,000	2,000	5,000
1.00 %	0.19	0.28	0.37	0.49	0.80	0.65	0.82
5.00 %	1.64	2.19	2.59	3.08	3.82	3.78	4.09
10.0 %	4.23	5.12	5.74	6.81	7.73	8.06	8.53
Asymptotic Distribution	$\mathcal{QLR}(\Delta_{2.0}; K)$						
Nominal Level $\backslash$ Sample Size	50	100	200	500	1,000	2,000	5,000
1.00 %	0.02	0.36	0.47	0.50	0.49	0.50	0.71
5.00 %	1.48	1.64	2.21	2.67	2.50	2.79	3.56
10.0 %	3.35	3.78	4.50	5.47	5.58	5.98	7.27

Table 3: Empirical Rejection Rates (in Percent)
Number of Replications: 10,000
DGP: $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim$ IID $N(0, 1)$
Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda \exp(\delta Y_{t-1}) + U_t$
$\Delta_{0.5} = [-0.5, 0.5]$, $\Delta_{1.0} = [-1, 1]$, $\Delta_{1.5} = [-1.5, 1.5]$, $\Delta_{2.0} = [-2, 2]$, and $K = 150$

Asymptotic Distribution	$\widehat{\mathcal{QLR}}_n(\Delta_{0.5}; K)$						
Nominal Level $\backslash$ Sample Size	50	100	200	500	1,000	2,000	5,000
1.00 %	0.71	0.39	0.60	0.71	0.89	0.93	1.01
5.00 %	2.43	2.94	3.48	4.39	4.64	4.54	4.88
10.0 %	6.06	6.90	8.17	9.04	9.33	8.93	9.49
Asymptotic Distribution	$\widehat{\mathcal{QLR}}_n(\Delta_{1.0}; K)$						
Nominal Level $\backslash$ Sample Size	50	100	200	500	1,000	2,000	5,000
1.00 %	0.30	0.50	0.74	0.71	0.67	0.80	0.99
5.00 %	2.44	2.56	3.20	3.91	4.02	4.15	5.13
10.0 %	5.39	6.00	6.71	7.97	8.34	8.88	10.30
Asymptotic Distribution	$\widehat{\mathcal{QLR}}_n(\Delta_{1.5}; K)$						
Nominal Level $\backslash$ Sample Size	50	100	200	500	1,000	2,000	5,000
1.00 %	0.12	0.18	0.23	0.37	0.54	0.48	0.59
5.00 %	1.33	1.65	1.87	2.24	2.96	2.92	3.15
10.0 %	3.13	3.89	4.11	5.05	5.96	6.11	6.43
Asymptotic Distribution	$\widehat{\mathcal{QLR}}_n(\Delta_{2.0}; K)$						
Nominal Level $\backslash$ Sample Size	50	100	200	500	1,000	2,000	5,000
1.00 %	0.15	0.24	0.40	0.39	0.39	0.37	0.54
5.00 %	1.26	1.34	1.84	2.22	2.03	2.21	2.92
10.0 %	2.73	2.95	3.47	4.31	4.47	4.74	5.68

Table 4: Empirical Rejection Rates (in Percent)
Number of Replications: 4,000
DGP: $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim$ IID $N(0,1)$

Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda \exp(\delta Y_{t-1}) + U_t$, $\Delta_{0.5} = [-0.5, 0.5]$						
Nominal Level $\setminus$ Sample Size	50	100	200	500	1,000	2,000
1.00 %	0.35	0.45	0.75	0.85	1.15	0.90
5.00 %	2.55	3.62	4.10	4.67	5.20	5.27
10.0 %	6.57	8.30	8.57	8.97	9.50	10.25
30.0 %	25.77	26.67	28.05	28.85	30.20	30.75
50.0 %	46.77	46.95	47.47	49.90	48.90	50.82
80.0 %	77.55	78.35	78.77	79.22	79.67	79.85
90.0 %	86.75	88.22	89.10	89.35	88.80	90.02
95.0 %	91.60	93.62	94.32	94.92	93.95	94.95
Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda\{1 + \exp(\delta Y_{t-1})\}^{-1} + U_t$, $\Delta_{0.5} = [-0.5, 0.5]$						
Nominal Level $\setminus$ Sample Size	6,000	8,000	10,000	20,000	30,000	40,000
1.00 %	1.60	1.85	1.50	1.32	1.85	1.67
5.00 %	3.10	3.80	3.32	5.30	5.87	5.77
10.0 %	5.70	7.27	7.25	10.50	10.47	10.45
30.0 %	26.52	27.27	27.80	29.05	27.87	28.55
50.0 %	48.42	47.37	47.50	49.02	46.72	48.00
80.0 %	76.35	76.37	77.72	77.92	76.60	76.82
90.0 %	87.47	86.37	86.97	87.55	87.37	87.17
95.0 %	92.02	92.40	92.30	92.67	92.75	93.20

Figure 1: Empirical Null Distributions of the QLR Statistics
Number of Replications: 10,000
DGP: $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim$ IID $N(0,1)$
Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda \exp(\delta Y_{t-1}) + U_t$
$\Delta_{0.5} = [-0.5, 0.5]$, $\Delta_{1.0} = [-1, 1]$, $\Delta_{1.5} = [-1.5, 1.5]$, $\Delta_{2.0} = [-2, 2]$, and $K = 150$

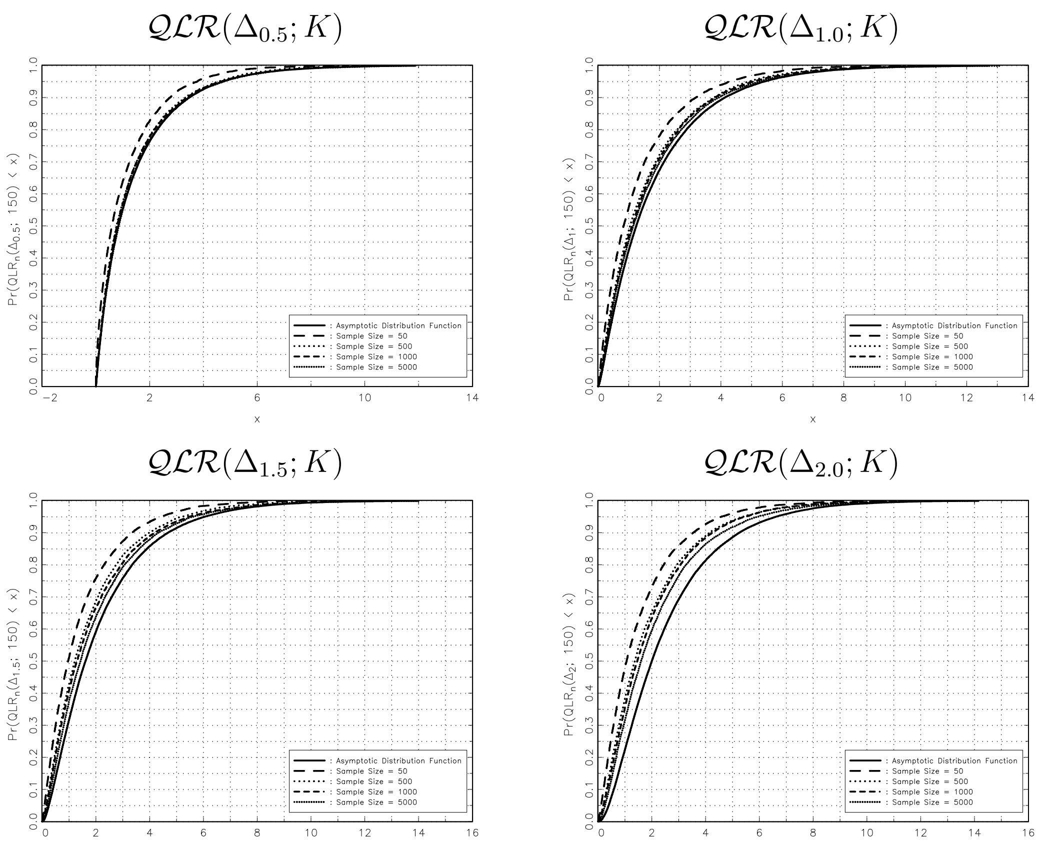


Figure 2: Empirical Density Functions of the QLR Statistics
Number of Replications: 10,000
DGP: $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim$ IID $N(0,1)$
Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda \exp(\delta Y_{t-1}) + U_t$
$\Delta_{0.5} = [-0.5, 0.5]$, $\Delta_{1.0} = [-1, 1]$, $\Delta_{1.5} = [-1.5, 1.5]$, $\Delta_{2.0} = [-2, 2]$, and $K = 150$

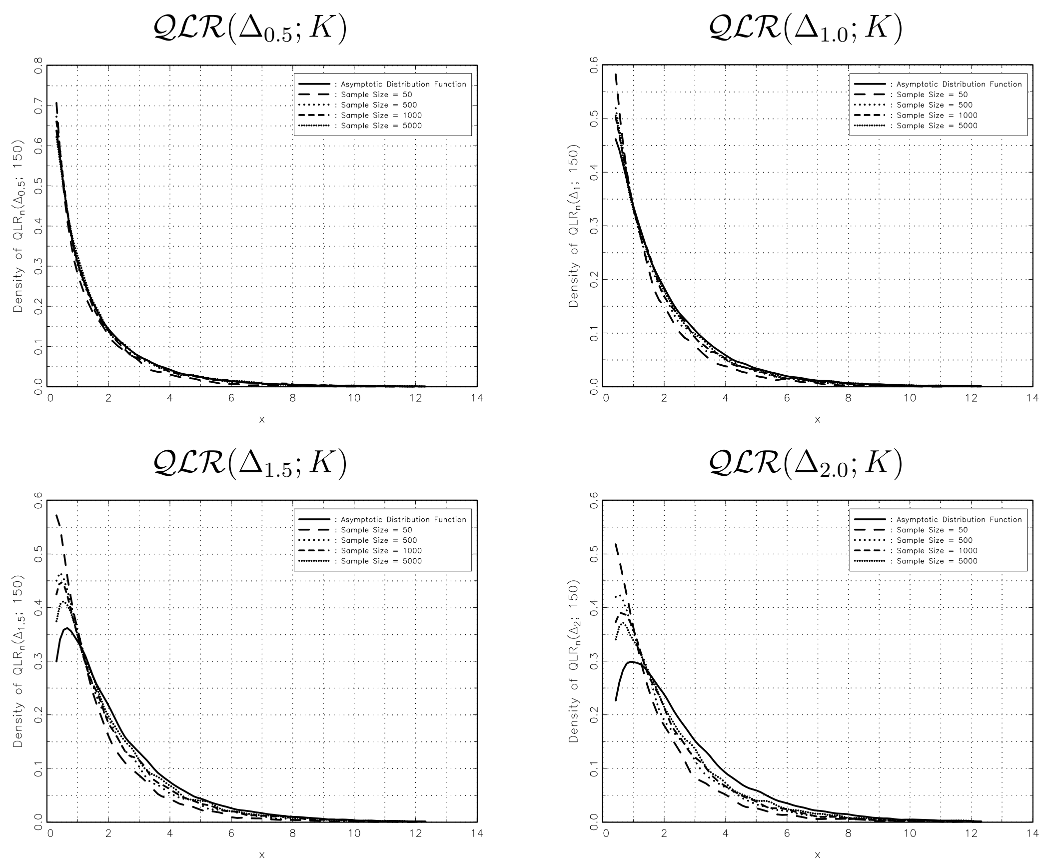
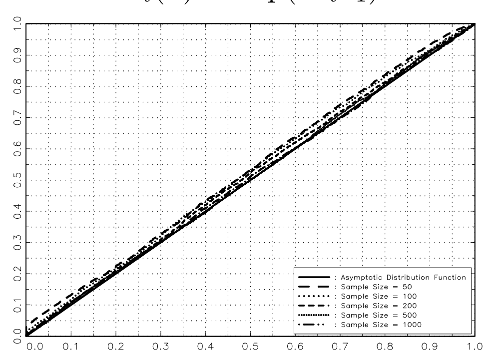


Figure 3: Empirical Distribution of Bootstrapped QLR Statistics
Number of Replications: 4,000
DGP: $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim$ IID $N(0,1)$
Model: $Y_t = \alpha + \beta Y_{t-1} + \lambda \Psi_t(\boldsymbol{\delta}) + U_t$ and $\Delta_{0.5} = [-0.5, 0.5]$

$\Psi_t(\boldsymbol{\delta}) = \exp(\delta Y_{t-1})$



$\Psi_t(\boldsymbol{\delta}) = 1/\{1 + \exp(\delta Y_{t-1})\}$