Revisiting Tests for Neglected Nonlinearity Using Artificial Neural Networks

JIN SEO CHO  
School of Economics  
Yonsei University  
jinseocho@yonsei.ac.kr

ISAO ISHIDA  
CSFI  
Osaka University  
i-ishida@sigmath.es.osaka-u.ac.jp

HALBERT WHITE  
Department of Economics  
University of California, San Diego  
hwhite@weber.ucsd.edu

First version: September 2006, This version: July 2010

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.

Key Words: Artificial neural networks; Testing linearity hypothesis; Quasi-likelihood ratio test; Directionally differentiable model; Asymptotic null distribution; Gaussian process.

JEL Classification: C12, C22, C45, C52.

Acknowledgements: The authors have benefited from discussions with Chang-Jin Kim, Cheolbeom Park, Sang Soo Park, Byeongseon Seo, and other participants at 2010 summer conference of Korean Econometric Society. Cho appreciates the research support by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-332-B00025).
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, they 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)).

This approximation property can also be exploited to test for neglected nonlinearity in regression analysis, as in White (1989) 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 $\lambda^*$, can be zero. Alternatively, $\delta^*$, 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 (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 $\Psi$) 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 (1953a, 1953b); 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):** \( \{(Y_t, X_t) : t = 1, 2, \ldots \} \subset \mathbb{R}^{1+k}(k \in \mathbb{N}) \) 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 \( \beta_\tau \) such that for some \( \rho > 1 \), \( \sum_{\tau=1}^{\infty} \tau^{1/(\rho-1)} \beta_\tau < \infty \).
Here, $Y_t$ and $X_t$ are target and predictor variables, respectively. For convenience, $X_t$ omits the constant. The mixing coefficients $\beta_\tau$ are

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

where $\mathcal{F}_t^s$ is the $\sigma$-field ("information set") generated by $(Y_t, X_t, ..., Y_{t+s}, X_{t+s})$. The $\beta_\tau$'s measure the time-series dependence in the data. For more on absolutely regular ($\beta$-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 $X_t$, it is common to forecast using an approximation to $E[Y_t | X_t]$, the conditional expectation ("regression") of $Y_t$ given $X_t$. This conditional expectation gives the mean-squared-error optimal forecast of $Y_t$ given $X_t$. Here, we approximate $E[Y_t | X_t]$ using a 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 $A \subset \mathbb{R}$, $B \subset \mathbb{R}^k$, $\Lambda \subset \mathbb{R}$, and $\Delta \subset \mathbb{R}^k$ be non-empty compact and convex sets, with $0 \in \text{int}(\Lambda)$ and $0 \in \text{int}(\Delta)$. Let $f(X_t; \alpha, \beta, \lambda, \delta) := \alpha + X_t \beta + \lambda \Psi(X_t \delta)$, and define the model $M$ as

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

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 $E[Y_t | X_t]$ or whether there is neglected nonlinearity, so that using hidden units can improve the approximation. By choosing $\Psi$ to be non-polynomial analytic, we ensure that $\Psi$ 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 $\Psi$ are $\Psi = \exp$ (as in Bierens, 1990), the logistic cumulative distribution function (CDF) originally used by White (1989a), or the ridgelets of Candès (1999). Below, we impose additional conditions on $\Psi$.

Note that because $X_t$ omits the constant, $\Psi(X_t \delta)$ 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.\footnote{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.}

Our null hypothesis is that $E[Y_t|X_t]$ is linear. Formally, we test

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

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 (1989b) 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 (1989a) 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 $(\alpha^*, \beta^*, \lambda^*, \delta^*)$ be parameter values satisfying $f(X_t; \alpha^*, \beta^*, \lambda^*, \delta^*) = E[Y_t|X_t] \text{ under } \mathcal{H}_0$. Then $\mathcal{H}_0$ consists of two sub-hypotheses: $\mathcal{H}_0 = \mathcal{H}_{01} \cup \mathcal{H}_{02}$, where

$$\mathcal{H}_{01} : \lambda^* = 0 \ \text{and} \ \mathcal{H}_{02} : \delta^* = 0.$$ (1)

$\mathcal{H}_{01}$ and $\mathcal{H}_{02}$ are the type 1 and 2 hypotheses mentioned above. Under $\mathcal{H}_{01}$, $\delta^*$ is not identified; that is, the representation for $E[Y_t|X_t]$ has many possible values for $\delta^*$. Under $\mathcal{H}_{02}$, only $\alpha^* + \lambda^*\Psi(0)$ is identified, and there are many combinations of $\alpha^*$ and $\lambda^*$ such that $\alpha^* + \lambda^*\Psi(0)$ is identical to the intercept in $E[Y_t|X_t]$. 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 literature. 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 (1989a)
and Lee, White, and Granger (1993) is designed specifically to test $H_{01}$ only. They do not consider $H_{02}$.
Nor does Hansen (1996) accommodate the possibility of $H_{02}$. His regularity conditions may therefore not
suffice under $H_{02}$. 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 $H_{01}$ and
$H_{02}$ simultaneously.

Accordingly, we consider a test statistic that properly takes into account both $H_{01}$ and $H_{02}$ to test $H_0$
versus $H_1$. 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 $H_{01}$ and $H_{02}$.

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):

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

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

$$QLR_n := n \left(1 - \frac{\hat{\sigma}^A_n}{\hat{\sigma}^0_n}\right),$$

where

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

and

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

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

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

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

and
\[ QLR_n^{(3)} := \left\{ n - \min_{\alpha} \min_{\delta} \min_{\lambda, \beta} \frac{1}{\sigma^2} \sum_{t=1}^{n} \{ Y_t - \alpha - X'_t \beta - \lambda \Psi(X'_t \delta) \}^2 \right\}. \]

QLR_n^{(1)} is obtained by minimizing with respect to \( \lambda \) before minimizing with respect to \( \delta \); this makes it convenient for analysis under \( \mathcal{H}_{01} \). Under \( \mathcal{H}_{01} \), \( \delta^* \) 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 \( \delta \); we then minimize the approximation with respect to \( \lambda \) and \( \alpha \) 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 \( \alpha^* \) and \( \lambda^* \) such that \( \alpha^* + \lambda^* \Psi(0) \) is identical to the intercept of \( E[Y_t | X_t] \). We overcome this difficulty by first fixing \( \lambda^* \). This enables us to identify the other parameters \( (\alpha^*, \delta^*, \beta^*) \) and apply a Taylor approximation to QL. We then optimize with respect to \( \lambda^* \). This approximation is denoted as \( \overline{QLR}_n^{(2)} \). We then interchange the roles of \( \alpha^* \) and \( \lambda^* \) 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

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

The asymptotic distribution of \( QLR_n \) is thus determined by the weak limits of \( \overline{QLR}_n^{(1)}, \overline{QLR}_n^{(2)}, \) 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 \( (\alpha, \beta')' \) gives

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

where

\[ M := I_n - Z(Z'Z)^{-1}Z' \; ; \; Z := [\iota, X] \; ; \; \Psi(\delta) := [\Psi(X'_1 \delta), \Psi(X'_2 \delta), \ldots, \Psi(X'_n \delta)]' \; ; \; \text{and} \; \mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'. \]

For now, we assume \( (Z'Z)^{-1} \) exists. We ensure this below.

We define \( \Psi_t(\delta) := \Psi(X'_t \delta), U_t := Y_t - E[Y_t | X_t] \) and \( \mathbf{U} := [U_1, U_2, \ldots, U_n]' \). Since \( \mathbf{MY} = \mathbf{MU} \)}
under $H_0$, it is standard that

$$\sup_{\lambda} \{ L_n^{(1)}(\lambda; \delta) - L_n^{(1)}(0; \delta) \} = \sup_{\lambda} 2\lambda \Psi(\delta)'MU - \lambda^2 \Psi(\delta)'M\Psi(\delta) = \frac{\{\Psi(\delta)'MU\}^2}{\Psi(\delta)'M\Psi(\delta)}.$$  (4)

Thus,

$$QLR_n^{(1)} = \sup_{\delta} \frac{\{\Psi(\delta)'MU\}^2}{\hat{\sigma}_n^2 \Psi(\delta)'M\Psi(\delta)}.$$  (5)

The asymptotic null behavior of $QLR_n^{(1)}$ is determined by that of $n^{-1/2} \Psi(\cdot)'MU$ and $n^{-1} \hat{\sigma}_n^2 \Psi(\cdot)'M\Psi(\cdot)$ 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 $|U_t| \leq M_t$, $|X_{t,j}| \leq M_t$, $j = 1, 2, \ldots, k$, and for some $\kappa \geq 2(\rho - 1)$, $E[M_t^{4+2\kappa}] < \infty$.

**Assumption A4 (Martingale Difference):** (i) $E[U_t | X_t, U_{t-1}, X_{t-1}, \ldots] = 0$; (ii) $E[U_t^2 | X_t] = \sigma^2_t$.

A3 and A4 ensure that $\sigma_t^2$, $E[U_t^4]$, and $E[X_{t,j}^4]$, $j = 1, 2, \ldots, k$, 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 $X_t$, 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) $\sup_{\delta \in \Delta} |\Psi_t(\delta)| \leq M_t$; and (ii) $\sup_{\delta \in \Delta} |(\partial/\partial \delta_j) \Psi_t(\delta)| \leq M_t$, $j = 1, \ldots, k+1$.

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, $\Psi$ is analytic in each of its arguments, so both $\Psi_t$ and $(\partial/\partial \delta_j) \Psi_t$ are analytic for each $X_t$. This ensures that $\Psi_t$ and $(\partial/\partial \delta_j) \Psi_t$ are also Lipschitz continuous on $\Delta$ and therefore bounded for each $X_t$. 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 $\Psi$ 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. $(-P)$
as $\delta$ tends to 0:

$$\lim_{\delta \to 0} \Psi'(\delta)'MU = \Psi(0)U'MU = 0 \ a.s.$$  

Because $\epsilon$ is a column of $Z$, the denominator behaves similarly:

$$\lim_{\delta \to 0} \Psi'(\delta)'M\Psi(\delta) = \Psi(0)^2U'MU = 0 \ a.s.$$  

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

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

We let $\epsilon \to 0$ below. To ensure asymptotic non-degeneracy, we write $Z_t := (1, X_t)'$ and impose

**Assumption A6 (Covariance):** For each $\epsilon > 0$ and $\delta \in \Delta(\epsilon)$, det $V_1(\delta) > 0$ and det $V_2(\delta) > 0$, where

$$V_1(\delta) := \begin{bmatrix} E[U_t^2\Psi_t(\delta)^2] & E[U_t^2\Psi_t(\delta)Z_t] \\ E[U_t^2Z_t\Psi_t(\delta)] & E[U_t^2Z_tZ_t'] \end{bmatrix} \quad \text{and} \quad V_2(\delta) := \begin{bmatrix} E[\Psi_t(\delta)^2] & E[\Psi_t(\delta)Z_t] \\ E[Z_t\Psi_t(\delta)] & E[Z_tZ_t'] \end{bmatrix}.$$  

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 $\Psi_t^* := \Psi_t(\delta) - E[\Psi_t(\delta)Z_t']\{E[Z_tZ_t']\}^{-1}Z_t$.

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

(i) $\sigma_n^0 \xrightarrow{p} \sigma_n^2 := E[U_t^2]$;

(ii) for each $\epsilon > 0$, \{n$^{-1/2}\Psi(\cdot)'MU, \sigma_n^0n^{-1}\Psi(\cdot)'M\Psi(\cdot)\} \Rightarrow \{G_0(\cdot), J(\cdot, \cdot)\}$ on $\Delta(\epsilon)$, where $G_0$ is a zero-mean continuous Gaussian process such that $E[G_0(\delta)G_0(\delta')] = T(\delta, \delta)$, where for each $\delta, \hat{\delta}$,

$$T(\delta, \hat{\delta}) := E[U_t^2\Psi_t^*(\delta)\Psi_t^*(\delta')] \quad \text{and} \quad J(\delta, \hat{\delta}) := \sigma_n^2E[\Psi_t^*(\delta)\Psi_t^*(\delta')];$$

(iii) if A4(ii) also holds, then $T(\delta, \hat{\delta}) = J(\delta, \hat{\delta}).$

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

$$QLR_n^{(1)}(\epsilon) := \sup_{\delta \in \Delta(\epsilon)} \frac{\{\Psi(\delta)'MU\}^2}{\sigma_n^0 \Psi(\delta)'M\Psi(\delta)},$$

8
To state the result, let $J(\delta) = J(\delta, \delta)$ and $G_1(\delta) := J(\delta)^{-1/2}G_0(\delta)$, so that for each $\delta$ and $\tilde{\delta}$,

$$E[G_1(\delta)G_1(\tilde{\delta})] = \rho_1(\delta, \tilde{\delta}) := \frac{T(\delta, \tilde{\delta})}{\{J(\delta, \delta)\}^{1/2}\{J(\delta, \tilde{\delta})\}^{1/2}}.$$

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

(i) $QLR_n^{(1)}(\epsilon) \Rightarrow \sup_{\delta \in \Delta(\epsilon)} G_1(\delta)^2$;

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

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

Note that for each $\delta$, $G_1(\delta)$ generally is not standard normal, although when $A4(ii)$ holds, we do have $G_1(\delta) \sim N(0, 1)$. Assuming conditional homoskedasticity ($A4(ii)$) may restrict the application of the QLR test, as many data 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 $\mathcal{H}_{02}$

#### 2.2.1 Case 1: $\lambda$ given

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

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

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

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

$$\frac{\partial}{\partial \delta_i} L_n^{(2)}(0; \lambda) = -2\lambda c_1 X_i'M[Y - \lambda c_0 t] = -2\lambda c_1 X_i'MU \equiv 0,$$

with $c_j := D^j\Psi(0), j = 0, 1, 2, \cdots$, where $D^j$ is the $j$th derivative operator with respect to the argument of $\Psi$; and $X_i := [X_{1,i}, X_{2,i}, \cdots, X_{n,i}]'$ is the $i$th column of $X$. We have $X_i'M \equiv 0$, as $M$ is an idempotent matrix constituted by $X_i$. 

9
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 $D_i := \text{diag}\{X_i\}$, $D_{ij} := D_iD_j$, $D_{ij\ell} := D_iD_jD_\ell$, and $D_{ij\ell m} := D_iD_jD_\ell D_m$, $i, j, \ell, m = 1, 2, \cdots, k$.

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

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

(ii) $\frac{\partial^2}{\partial \delta_i \partial \delta_j} L_n^{(2)}(0; \lambda) = 2\lambda c_2 t' D_{ij} \mu E$;

(iii) $\frac{\partial^3}{\partial \delta_i \partial \delta_j \partial \delta_\ell} L_n^{(2)}(0; \lambda) = 2\lambda c_3 t' D_{ij\ell} \mu E$; and

(iv) $\frac{\partial^4}{\partial \delta_i \partial \delta_j \partial \delta_\ell \partial \delta_m} L_n^{(2)}(0; \lambda) = 2\lambda c_4 t' D_{ij\ell m} \mu E - 2\lambda^2 c_2^2 t'[D_{ij} M D_{\ell m} + D_{ij\ell} M D_{jm} + D_{ij\ell m} M D_{j\ell}]$.

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): $E|U_i|^8 < \infty$ and $E|X_{t,i}|^8 < \infty$; or $E|U_i|^4 < \infty$ 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) $t' D_{ij} \mu E = O_P(n^{1/2})$;

(ii) $t' D_{ij\ell} \mu E = o_P(n^{3/4})$;

(iii) $t' D_{ij\ell m} \mu E = o_P(n)$; and

(iv) $t' D_{ij} M D_{\ell m} t = O_P(n)$.

Lemma 3 also implies that the other terms in Lemma 2(iv) are $O_P(n)$, allowing us to write

\[
L_n^{(2)}(\delta; \lambda) - L_n^{(2)}(0; \lambda) = \frac{\lambda c_2}{2} \sum_{i=1}^k \sum_{j=1}^k (t' D_{ij} \mu E) \delta_i \delta_j \\
- \frac{1}{4} \lambda^2 c_2^2 \sum_{i=1}^k \sum_{j=1}^k \sum_{\ell=1}^k \sum_{m=1}^k (t' D_{ij} M D_{\ell m} t) \delta_i \delta_j \delta_\ell \delta_m + O_P(n^{-1/4}).
\]

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 $\lambda$ and $c_2$ play different roles here.) To avoid the complexities introduced when $c_2$ can be zero, we impose

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

10
This rules out choosing \( \Psi \) to be the logistic CDF or a ridgelet with zero input-to-hidden bias. But one can simply “bias-shift” \( X_t \delta \) to \( c + X_t \delta \), where \( c \) is chosen explicitly to ensure \( c^2 \neq 0 \). For this, we can replace \( \Psi(X_t \delta) \) with \( \Psi_c(X_t \delta) := \Psi(c + X_t \delta) \). 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 \( i, j, \ell, m = 1, 2, \ldots, k \), let

\[
\tilde{M} := [\ell' D_{ij} MU], \quad W := [W_{ij}], \quad W_{ij} = [\ell' D_{ij} MD_{\ell m} \ell].
\]

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

\[
L^{(2)}_n(\delta; \lambda) - L^{(2)}_n(0; \lambda) = \lambda c_2 \delta' \tilde{M} \delta - \frac{\lambda^2}{4} c_2^2 \{ \delta'(I_k \otimes \delta)' W (I_k \otimes \delta) \delta \} + O_P(n^{-1/4}).
\]

(8)

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

\[
\hat{QLR}_n^{(2)}(\delta; \lambda) = \frac{1}{\sigma_n^2} \left\{ \lambda c_2 (\delta' \tilde{M} \delta) - \frac{1}{4} \lambda^2 c_2^2 \{ \delta'(I_k \otimes \delta)' W (I_k \otimes \delta) \delta \} \right\}.
\]

(9)

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

\[
\delta = \delta^* + hd,
\]

(10)

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

\[
\hat{QLR}_n^{(2)}(\lambda) := \sup_\delta \hat{QLR}_n^{(2)}(\delta; \lambda) = \sup_{d \in S^{k-1}} \sup_{h \in \mathbb{R}^+} \hat{QLR}_n^{(2)}(hd; \lambda).
\]

(11)
Combining (9) and (11) gives

$$QLR_n^{(2)}(\lambda) = \sup_{d \in \mathbb{S}^{k-1}} \sup_{h \in \mathbb{R}^+} \frac{1}{\sigma_n^2} \left\{ \lambda c_2 (d' \tilde{M} d) h^2 - \frac{1}{4} \lambda^2 c_2^2 \{ d' (I_k \otimes d)' W (I_k \otimes d) d \} h^4 \right\}$$

$$= \sup_{d \in \mathbb{S}^{k-1}} \frac{\max [d' \tilde{M} d, 0]^2}{\sigma_n^0 \{ d' (I_k \otimes d)' W (I_k \otimes d) d \}}$$

$$= \sup_{d \in \mathbb{S}^{k-1}} \max \left[ \frac{d' \tilde{M} d}{\sigma_n^0 \{ d' (I_k \otimes d)' W (I_k \otimes d) d \}^{1/2}}, 0 \right] ^2 + o_P(1),$$

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

$$h^2 = \frac{2 (d' \tilde{M} d)}{\lambda c_2 \{ d' (I_k \otimes d)' W (I_k \otimes d) d \}}.$$

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

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 $d' \tilde{M} d = \text{vec}(dd)' \text{vec}(\tilde{M})$, $d' (I_k \otimes d)' W (I_k \otimes d) d = \text{vec}(dd)' W \text{vec}(dd')$, and $\text{vec}(dd)' \text{vec}(dd') = 1$. Letting $b := \text{vec}(dd') \in \mathbb{S}^{k^2-1} := \{ b \in \mathbb{S}^{k^2-1} : b = \text{vec}(dd'), d \in \mathbb{S}^{k-1} \}$, we can write $QLR_n^{(2)}$ more compactly as

$$QLR_n^{(2)} = \sup_{b \in \mathbb{S}^{k^2-1}} \max \left[ \frac{b' \text{vec}(\tilde{M})}{\{ \hat{\sigma}_n^0 b' W b \}^{1/2}}, 0 \right] ^2.$$

It is not hard to show that the variance of $n^{-1/2} b' \text{vec} \tilde{M}$ is asymptotically equivalent to $n^{-1} \hat{\sigma}_n^0 b' W b$ 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

$$\frac{1}{\sqrt{n}} t^i D_{ij} M U = \frac{1}{\sqrt{n}} \sum_{t=1}^n X_{t,i} X_{t,j} U_t - \left( \frac{1}{n} \sum_{t=1}^n X_{t,j} Z_t' \right) \left( \frac{1}{n} \sum_{t=1}^n Z_t Z_t' \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_t U_t \right),$$

where $n^{-1/2} \sum_{t=1}^n X_{t,i} X_{t,j} U_t$ and $n^{-1/2} \sum_{t=1}^n Z_t U_t$ 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 $C_t := \operatorname{vech}(X_t X_t')$. Suppose $\det \tilde{V}_1 > 0$ and $\det \tilde{V}_2 > 0$, where

$$\tilde{V}_1 := \begin{bmatrix} E[U_t^2 Z_t Z_t'] & E[U_t^2 Z_t C_t'] \\ E[U_t^2 C_t Z_t'] & E[U_t^2 C_t C_t'] \end{bmatrix}$$

and

$$\tilde{V}_2 := \begin{bmatrix} E[Z_t Z_t'] & E[Z_t C_t'] \\ E[C_t Z_t'] & E[C_t C_t'] \end{bmatrix}.$$

We use the vec operator to avoid entering the common elements of $X_t X_t'$ twice.

We can now obtain the limiting behavior of the components of $\overline{QLR}_n^{(2)}$. For this, we let $C_{t,ij}^* := X_{t,i} X_{t,j} - E[X_{t,i} X_{t,j} Z_t'] E[Z_t Z_t']^{-1} Z_t$, and $C_t^* := [C_{t,ij}^*]$, 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 $b \in S_c^{k^2-1}$,

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

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

(iii) if $A4(ii)$ also holds, for $i, j, \ell, m = 1, 2, \cdots, k$, $E(M_{ij} M_{\ell m}) = \sigma_2^2 E(C_{t,ij}^* C_{t,\ell m}^*)$.

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

$$\sup_{\delta} \frac{1}{\sigma_n^2} \{ L_n^{(2)}(\delta; \lambda) - L_n^{(2)}(0; \lambda) \} = \overline{QLR}_n^{(2)} + o_P(1) \Rightarrow \sup_{b \in S_c^{k^2-1}} \max_{b \in S_c^{k^2-1}} \left[ \frac{b' \operatorname{vec}(\mathcal{M})}{\sigma_n^2 b' W^* b} \right]^{1/2},$$

Formally, we have

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

(i) $\overline{QLR}_n^{(2)} \Rightarrow \sup_{b \in S_c^{k^2-1}} \max_{b \in S_c^{k^2-1}} \| \mathcal{G}_2(b) \|_2^2$, where $\mathcal{G}_2$ is a Gaussian process defined on $S_c^{k^2-1}$ such that
for each \( \mathbf{b} \) and \( \tilde{\mathbf{b}} \), \( E[G_2(\mathbf{b})] = 0 \) and

\[
E[G_2(\mathbf{b})G_2(\tilde{\mathbf{b}})] = \rho_2(\mathbf{b}, \tilde{\mathbf{b}}) := \frac{\mathcal{K}(\mathbf{b}, \tilde{\mathbf{b}})}{\mathcal{I}(\mathbf{b}, \mathbf{b})^{1/2} \mathcal{I}(\tilde{\mathbf{b}}, \tilde{\mathbf{b}})^{1/2}},
\]

where \( \mathcal{K}(\mathbf{b}, \tilde{\mathbf{b}}) := \mathbf{b}'E[U_t^2 \mathrm{vec}(C_t^*) \mathrm{vec}(C_t^*)'] \tilde{\mathbf{b}}, \) and \( \mathcal{I}(\mathbf{b}, \tilde{\mathbf{b}}) := \sigma^2 \mathbf{b}'W^* \tilde{\mathbf{b}}; \)

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

\[
\rho_2(\mathbf{b}, \tilde{\mathbf{b}}) = \frac{\mathcal{I}(\mathbf{b}, \tilde{\mathbf{b}})}{\mathcal{I}(\mathbf{b}, \mathbf{b})^{1/2} \mathcal{I}(\tilde{\mathbf{b}}, \tilde{\mathbf{b}})^{1/2}}.
\]

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

\[
G_2(\mathbf{b}) := \frac{\mathbf{b}' \mathrm{vec}(\mathcal{M})}{\{\sigma^2 \mathbf{b}'W^* \mathbf{b}\}^{1/2}}.
\]

We offer several remarks before proceeding. First, note that \( G_2 \) is indexed by a direction \( \mathbf{d} \) from the origin (through \( \mathbf{b} \)) rather than by the parameters \( \lambda \) unidentified under \( H_{02} \). QLR is thus nuisance parameter-free under \( H_{02} \), a remarkable fact. As we show in the next subsection, \( G_2 \), indexed by \( \mathbf{d} \), is a special case of \( G_1 \), indexed by \( \delta \). This implies that under \( H_{02} \), the asymptotic distribution of QLR is essentially governed by \( H_{01} \), justifying the use of \( G_1 \) 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 (1953a, b) first examines quartic approximations of statistical models. McCullagh (1987) also considers quartic approximations using tensors. Here, using \( h \) and \( \mathbf{d} \), we can avoid the more cumbersome use of tensors, permitting us to apply the methods of Cho and White (2009b) and enabling us to readily associate the asymptotic distributions under \( H_{02} \) and \( H_{01} \). Third, the approximation is especially straightforward if \( k = 1 \). Then \( S^{k-1} = \{-1, 1\} \), so for each \( \mathbf{d} \in S^{k-1}, \mathbf{b} = 1 \), implying that \( G_2 \) is free of \( \mathbf{b} \) and \( QLR^{(2)}_n \Rightarrow \max[Z, 0]^2 \), where \( Z \) is normal with mean zero under \( H_{02} \). Under conditional homoskedasticity (A4(ii)), \( Z \sim N(0, 1) \). 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 \( \mathbf{d} \).
2.2.2 Case 2: \( \alpha \) given

We continue our analysis of the QLR statistic under the type 2 hypothesis, but now we suppose that \( \alpha \) is given and concentrate QL with respect to \( \lambda \) and \( \beta \). Then

\[
L_n^{(3)}(\delta; \alpha) := \max_{\lambda, \beta} L_n(\alpha, \beta, \lambda, \delta) = -(Y - \alpha \ell)'P(\delta)(Y - \alpha \ell),
\]

where \( P(\delta) := I - Q(\delta)[Q(\delta)'Q(\delta)]^{-1}Q(\delta)' \), and \( Q(\delta) := [X, \Psi(\delta)] \). 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 \( \delta \), but \( \alpha \) is now given, unlike eq. 6, where \( \lambda \) is given.

This separate approach is needed, as fixing \( \alpha \) could yield an asymptotic null distribution different from that of Theorem 2. As we now show, however, the asymptotic null distribution does turn out to be the same.

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 \( H_{02} \) and letting \( \delta = h\delta \) as in eq. 10.

**Lemma 5.** Given A1 and A2,

1. \( \frac{\partial}{\partial \alpha} L_n^{(3)}(0; \alpha) = 0; \)
2. \( \frac{\partial^2}{\partial \alpha^2} L_n^{(3)}(0; \alpha) = 4\gamma*J_2MU + 2U'J_0H_0^{-1}J_2'MU, \)
3. \( \frac{\partial^3}{\partial \alpha^3} L_n^{(3)}(0; \alpha) = \gamma*J_0[J_0'\frac{\partial^3}{\partial h^3}P(h\delta)|_{h=0}]J_0\gamma* + o_P(n^{3/4}); \)
4. \( \frac{\partial^4}{\partial \alpha^4} L_n^{(3)}(0; \alpha) = -6\gamma*J_2MJ_2\gamma* + o_P(n). \)

Here, \( \gamma*, J_j, H_j \) are in fact functions of \( \alpha \) and/or \( d \). 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 \( H_{02} \),

1. \( \frac{\partial^2}{\partial \alpha^2} L_n^{(3)}(0; \alpha) = (\alpha* - \alpha)(c_2/c_0)d'Md + o_P(n^{1/2}); \)
2. \( \frac{\partial^3}{\partial \alpha^3} L_n^{(3)}(0; \alpha) = o_P(n^{3/4}); \) and
3. \( \frac{\partial^4}{\partial \alpha^4} L_n^{(3)}(0; \alpha) = -6(\alpha* - \alpha)^2(c_2/c_0)^2d'(I_k \otimes d)'W(I_k \otimes d)d + o_P(n). \)

By Lemma 6, the 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

\[ QLR_n^{(3)} (hd, \alpha) := \frac{1}{\sigma_n^2} \left\{ \left( \frac{c_2}{c_0} \right) (\alpha^* - \alpha) d' \tilde{M} d \right\} h^2 - \frac{1}{4} \left( \frac{c_2}{c_0} \right)^2 (\alpha^* - \alpha)^2 \left\{ d' (I_k \otimes d)' W (I_k \otimes d) d \right\} h^4 \]

and

\[ QLR_n^{(3)} (\alpha) := \sup_{d \in \mathbb{S}^{k-1}} \sup_{h \in \mathbb{R}^+} \widetilde{QLR}_n^{(3)} (hd, \alpha), \]

then it follows that

\[ QLR_n^{(3)} (\alpha) = \sup_{d \in \mathbb{S}^{k-1}} \frac{\max \{ d' \tilde{M} d, 0 \}^2}{\sigma_n^2 \left\{ d' (I_k \otimes d)' W (I_k \otimes d) d \right\} h^4}, \] (15)

and we can write \( QLR_n^{(3)} = QLR_n^{(3)} (\alpha) \), as \( \alpha \) 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 \( QLR_n^{(3)} \) coincides with that given in Theorem 2. We summarize with the following corollary.

**Corollary 1.** Given A1, A2, A3*, A4(i), A6*, A7, and \( \mathcal{H}_{0_2} \),

(i) \( QLR_n^{(3)} \Rightarrow \sup_{b \in \mathbb{S}^{2-1}} \max \{ G_2 (b), 0 \}^2 \); and

(ii) \( QLR_n^{(2)} - QLR_n^{(3)} = o_p(1) \).

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 \( \mathcal{H}_{0_2} \), 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 \( QLR_n^{(2)} \) (equivalently \( QLR_n^{(3)} \)) are related under \( \mathcal{H}_0 \). Specifically, we show that \( QLR_n^{(1)} \) converges to \( QLR_n^{(2)} \) as \( \delta \) converges to 0. For this, let \( \delta = hd \) as above, and define

\[ N_n(h, d) = N_n(\delta) := \{ \Psi(\delta)' \mu \mu \}^2 \quad \text{and} \quad D_n(h, d) = D_n(\delta) := \Psi(\delta)' M \Psi(\delta). \]

Then we can write \( \sup_{\lambda} \{ L_n(\lambda; \delta) - L_n(0; \delta) \} \) in eq. (14) as

\[ \frac{N_n(h, d)}{D_n(h, d)} = \frac{\{ \Psi(\delta)' \mu \mu \}^2}{\Psi(\delta)' M \Psi(\delta)}. \]

Our next result describes the behavior of this ratio as \( h \) converges to zero.
Lemma 7. Given A1 and A2, for each $n$ and $d$,

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

(ii) $\lim_{h \to 0} N_n^{(4)}(h, d) = 6\epsilon_2^2 \{ \sum_{i=1}^k \sum_{j=1}^k \ell' D_{ij} M_d d_j \}^2$ a.s.; and

(iii) $\lim_{h \to 0} D_n^{(4)}(h, d) = 6\epsilon_2^2 \{ \sum_{i=1}^k \sum_{j=1}^k \sum_{m=1}^k \ell' D_{ij} M_d d_{im} d_j d_d m \}$ a.s.

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

Given Lemma 7, l’Hospital’s rule gives

$$
\lim_{h \to 0} \frac{N_n(h, d)}{\sigma_n^0 D_n(h, d)} = \lim_{h \to 0} \frac{N_n^{(4)}(h, d)}{\sigma_n^0 D_n^{(4)}(h, d)} = \frac{\{b' \vec{V}(\hat{M})\}^2}{\sigma_n^0 b'Wb} \text{ a.s.} \tag{16}
$$

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

$$QLR_n^{(1)} = \sup_{h, d} \frac{\{\Psi(h, d)'M\Psi(h, d)'\}}{\sigma_n^{(0)} \Psi(h, d)'M\Psi(h, d)'} \geq \sup_{d \in \Sigma^{k-1}} \lim_{h \to 0} \frac{\{\Psi(h, d)'M\Psi(h, d)'\}}{\sigma_n^{(0)} \Psi(h, d)'M\Psi(h, d)'}
= \sup_{b \in \Sigma_k} \left[ \frac{b' \vec{V}(\hat{M})}{\sigma_n^0 b'Wb} \right] = \frac{\{b' \vec{V}(\hat{M})\}^2}{\sigma_n^0 b'Wb} \text{ a.s.} \tag{17}
$$

Because $QLR_n^{(2)} = QLR_n^{(2)} + o_p(1) = QLR_n^{(3)} = QLR_n^{(3)} + o_p(1)$, this gives

$$QLR_n = \max\{QLR_n^{(1)}, QLR_n^{(2)}, QLR_n^{(3)}\} = \max\{QLR_n^{(1)}, QLR_n^{(2)}, QLR_n^{(3)}\} + o_p(1) = QLR_n^{(1)} + o_p(1)
$$

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 0 from $\Delta$, whereas 0 $\in$ int($\Delta$) 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 0 $\in$ int($\Delta$). This necessitates a further analysis of the QLR test treating 0 as an element of int($\Delta$). 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 $\epsilon$ tends to zero. It turns out that $QLR_n^{(2)}$ 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'.
These new conditions accommodate the fact that the score of the QL function under \( H \) is identical to the score obtained under \( \psi \) separately derived above:

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

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

\[
\bar{V}_1(\delta) := \begin{bmatrix}
E[U_t^2 \Psi_t(\delta)^2] & E[U_t^2 \Psi_t(\delta)Z_t^i] & E[U_t^2 \Psi_t(\delta)C_t^i] \\
E[U_t^2 Z_t \Psi_t(\delta)] & E[U_t^2 Z_t Z_t^i] & E[U_t^2 Z_t C_t^i] \\
E[U_t^2 C_t \Psi_t(\delta)] & E[U_t^2 C_t Z_t^i] & E[U_t^2 C_t C_t^i]
\end{bmatrix}, \quad \text{and}
\]

\[
\bar{V}_2(\delta) := \begin{bmatrix}
E[\Psi_t(\delta)^2] & E[\Psi_t(\delta)Z_t^i] & E[\Psi_t(\delta)C_t^i] \\
E[Z_t \Psi_t(\delta)] & E[Z_t Z_t^i] & E[Z_t C_t^i] \\
E[C_t \Psi_t(\delta)] & E[C_t Z_t^i] & E[C_t C_t^i]
\end{bmatrix}.
\]

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 \( d \) is given, and \( \delta \) goes to zero in the direction \( d \) by sending \( h \) to zero. Our next result involves the joint asymptotic behavior of the random functions \( G_1 \) and \( G_2 \) separately derived above:

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

(i) \( QLR_{n} = QLR_{n}^{(1)} + o_{\mathbb{P}}(1) \); and

(ii) \( QLR_{n} \Rightarrow \sup_{\delta} G(\delta)^2 \), where

\[
G(\delta) := \begin{cases}
G_1(\delta), & \text{if } h \neq 0; \\
G_2(b), & \text{otherwise}.
\end{cases}
\]

In previous works, the asymptotic null distribution of \( QLR_n \) has been given as \( \sup_{\delta} G_1(\delta)^2 \), but this neglects the twofold identification problem. The true asymptotic null distribution may differ from \( \sup_{\delta} G_1(\delta)^2 \), 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 \( \sup_{\delta} G(\delta)^2 \), which depends on both \( G_1 \) and \( G_2 \). 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 \( \Delta \), including 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 \( G \) necessarily accommodates the covariance of \( G_1 \) and \( G_2 \). Specifically, for each \( \delta = (h, \tilde{d}) \) and \( \tilde{\delta} = (\tilde{h}, \tilde{d}) \), \( E[G(\delta)G(\tilde{\delta})] = \rho(\delta, \tilde{\delta}) \), where

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

with

\[
\rho_3(b, \tilde{\delta}) := \frac{\mathcal{H}(b, \tilde{\delta})}{\{\mathcal{J}(b, \tilde{\delta})\}^{1/2}} \quad \text{and} \quad \mathcal{H}(b, \tilde{\delta}) := E[U_t^2 b' \text{vec}(C_i^*) \Psi_t^*(\tilde{\delta})].
\]

Thus, \( \mathcal{H}(b, \tilde{\delta}) \) represents the covariance between the scores for \( \mathcal{H}_01 \) and \( \mathcal{H}_02 \). If \( A4(ii) \) also holds, then \( \mathcal{H}(b, \tilde{\delta}) = \sigma^2 b'E[\text{vec}(C_i^*) \Psi_t^*(\tilde{\delta})] \). In this case,

\[
\rho_3(b, \tilde{\delta}) = \frac{b'E[\text{vec}(C_i^*) \Psi_t^*(\tilde{\delta})]}{\{b'E[\text{vec}(C_i^*) \text{vec}(C_i^*)]b\}^{1/2} \{E[\Psi_t^*(\tilde{\delta})^2]\}^{1/2}}.
\]

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

\[
\Phi_t(\delta, \tilde{\delta}) := \Psi_t(\delta) \Psi_t(\tilde{\delta}) \quad \text{and} \quad \Upsilon_{t,j}(\delta) := \Psi_t(\delta) Z_{t,j}, \quad j = 1, 2, \ldots, k + 1.
\]

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

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

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

We have the following formal result:

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

(i) \( \lim_{h \to 0} \rho_1(\delta, \tilde{\delta}) = \text{sgn}[c_2] \rho_3(b, \tilde{\delta}) \); and

(ii) \( \lim_{\tilde{h} \to 0} \lim_{h \to 0} \rho_1(\delta, \tilde{\delta}) = \rho_2(b, \tilde{b}) \).

Note that \( \text{sgn}[c_2] \) appears in Lemma 8(i), so that we do not necessarily have \( G_1 \overset{d}{=} G \), where \( \overset{d}{=} \) denotes equality in distribution. Nevertheless, squaring these Gaussian processes makes this sign irrelevant, so that \( G_1^2 \overset{d}{=} G^2 \), and \( QLR_n \Rightarrow \sup_\delta G_1(\delta)^2 \) by Theorem 3(ii). The following states this formally.

**Corollary 2.** Given \( A1, A2, A3^{**}, A4, A5, A6^{**}, A7, A8, \) and \( \mathcal{H}_0 \), \( QLR_n \Rightarrow \sup_\delta G_1(\delta)^2 \).
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 $G_1$ previously derived in the literature for $QLR_n$ by neglecting the twofold identification problem is indeed correct. Nevertheless, properly accounting for $H_{02}$ 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.

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

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

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

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

To test $H_0$, 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 $\delta \in \Delta := [\delta, \bar{\delta}]$, where $-\infty < \delta < 0 < \bar{\delta} < \infty$.

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

$$QLR_{n1}(\epsilon) = \sup_{\delta \in [\delta, \bar{\delta}] \cup [-\epsilon, \epsilon]} \frac{\{U'M\Psi(\delta)\}^2}{\sigma_n^2 \Psi(\delta)'M\Psi(\delta)} \Rightarrow \sup_{\delta \in [\delta, \bar{\delta}] \cup [-\epsilon, \epsilon]} G_1(\delta)^2,$$

with mean zero Gaussian process $G_1$ such that

$$E[G_1(\delta)G_1(\tilde{\delta})] = \frac{\mathcal{J}(\delta, \tilde{\delta})}{\{\mathcal{J}(\delta, \tilde{\delta})\}^{1/2}\{\mathcal{J}(\delta, \tilde{\delta})\}^{1/2}}.$$
and, for each non-zero $\delta$ and $\tilde{\delta}$,

$$J(\delta, \tilde{\delta}) := \sigma^2 \left\{ E[\exp(Y_t(\delta + \tilde{\delta}))] - \left[ \begin{array}{ccc} E[\exp(Y_t\delta)] & 1 & E[Y_t] \\ E[Y_t \exp(Y_t\delta)] & E[Y_t] & E[Y_t^2] \end{array} \right]^{-1} \left[ \begin{array}{c} E[\exp(Y_t\delta)] \\ E[Y_t \exp(Y_t\delta)] \end{array} \right] \right\}.$$ 

Defining

$$M(\delta) := \exp \left\{ \frac{\theta_\ast - \delta + \frac{\sigma^2}{2(1 - \beta^2_\ast)} \delta^2}{1 - \beta_\ast} \right\}$$

and

$$J(\delta, \tilde{\delta}) := \{\exp[\var(Y_t)\delta\tilde{\delta}] - 1 - \var(Y_t)\delta\tilde{\delta}\},$$

we have

$$J(\delta, \tilde{\delta}) = \sigma^2 M(\delta) M(\tilde{\delta}). J(\delta, \tilde{\delta}).$$

Note that $M(\delta) = E[\exp(\delta Y_t)]$ is the moment generating function of $Y_t \sim N[\theta_\ast/(1 - \beta_\ast), \sigma^2_\ast/(1 - \beta^2_\ast)]$.

It follows easily that

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

Now $G_1$ is indexed by $\delta$. We also have $\delta = h d$, with distance $h \in \mathbb{R}^+$ and direction $d = \pm 1$.

Next, consider the behavior of QLR under $H_{02}$: $\delta_\ast = 0$. Theorem 2 gives

$$QLR_{n}^{(2)} = \max[\ell' D_{11}MU, 0]^2 \sigma^2_n \ell' D_{11}MD_{11} \ell + o_P(1),$$

where $D_{11} := \text{diag}\{Y_0^2, Y_1^2, \ldots, Y_{n-1}^2\}$,

$$\ell' D_{11}MU = \sum_{t=1}^{n} Y_{t-1}^2 U_t - \left[ \sum_{t=1}^{n} Y_{t-1}^2 \right] \left[ \sum_{t=1}^{n} Y_{t-1} \right]^{-1} \left[ \sum_{t=1}^{n} U_t \right],$$

and

$$\ell' D_{11}MD_{11} \ell = \sum_{t=1}^{n} Y_{t-1}^4 - \left[ \sum_{t=1}^{n} Y_{t-1}^2 \right] \left[ \sum_{t=1}^{n} Y_{t-1} \right]^{-1} \left[ \sum_{t=1}^{n} Y_{t-1}^2 \right].$$

As mentioned at the end of Section 2.2, the fact that $k = 1$ makes the role of $d$ in Theorem 2 trivial, because
where \( \mathcal{G}_2 \sim N(0, 1) \). This holds because

\[
n^{-1}\hat{\sigma}_n^2 \mathbf{D}_{11} \mathbf{M} \mathbf{D}_{11} \mathbf{t} \to \mathcal{I} := \sigma^2 \left\{ \frac{1}{\text{var}(Y_t)} \{E[Y_t^4] - 2E[Y_t^2]E[Y_t^2]E[Y_t^3] + E[Y_t^3]^2\} \right\} \ a.s.
\]

Further, \( Y_t \sim N[\theta/(1 - \delta^2), \sigma^2/(1 - \beta^2)] \) implies \( \mathcal{I} = 2\sigma^2 \text{var}(Y_t)^2 \), as well as \( n^{-1/2} \mathbf{t}' \mathbf{D}_{11} \mathbf{M} \mathbf{U} \overset{\text{a.s.}}{\sim} N[0, 2\sigma^2 \text{var}(Y_t)^2] \), so \( \mathcal{G}_2 \) obtains as the weak limit of

\[
n^{-1/2} \mathbf{t}' \mathbf{D}_{11} \mathbf{M} \mathbf{U} \quad \frac{\{\sigma^2 n^{-1/2} \mathbf{t}' \mathbf{D}_{11} \mathbf{M} \mathbf{D}_{11} \mathbf{t}\}^{1/2}}{\{\sigma^2 n^{-1/2} \mathbf{t}' \mathbf{D}_{11} \mathbf{M} \mathbf{D}_{11} \mathbf{t}\}^{1/2}}.
\]

Finally, we combine these separate results. By Theorem 3, we have \( QLR_n \Rightarrow \sup_{\delta \in \Delta} \mathcal{G}(\delta)^2 \), where

\[
\mathcal{G}(\delta) = \left\{ \begin{array}{ll} G_1(\delta), & \text{if } \delta \neq 0; \\ \mathcal{G}_2, & \text{otherwise}, \end{array} \right.
\]

with

\[
E[\mathcal{G}(\delta) \mathcal{G}(\bar{\delta})] = \left\{ \begin{array}{ll} \rho_1(\delta, \bar{\delta}), & \text{if } \delta \neq 0 \text{ and } \bar{\delta} \neq 0; \\ 1, & \text{if } \delta = 0 \text{ and } \bar{\delta} = 0; \\ \rho_3(\delta), & \text{if } \bar{\delta} = 0 \text{ and } \delta \neq 0, \end{array} \right.
\]

where

\[
\rho_3(\delta) = \frac{\mathcal{H}(\delta)}{\{\mathcal{J}(\delta, \bar{\delta})\}^{1/2} \{\mathcal{I}\}^{1/2}} = \frac{\mathcal{H}(\delta)}{\{\sigma^2 \text{var}(Y_t)M(\delta)M(\bar{\delta})J(\delta, \bar{\delta})\}^{1/2} \{\mathcal{I}\}^{1/2}},
\]

with

\[
\mathcal{H}(\delta) := \frac{\sigma^2}{\text{var}(Y_t)} E[Y_t^2 \exp(Y_t \delta)] - \frac{\sigma^2}{\text{var}(Y_t)} \{ (E[Y_t^2] - E[Y_t]E[Y_t^2])E[\exp(Y_t \delta)] \}
\]

\[
- \frac{\sigma^2}{\text{var}(Y_t)} \{ (E[Y_t^3] - E[Y_t]E[Y_t^2])E[Y_t \exp(Y_t \delta)] \}.
\]

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 \( \mathcal{J}(\delta, \bar{\delta}) \) and the fact that \( \mathcal{I} = 2\sigma^2 \text{var}(Y_t)^2 \), we have

\[
\rho_3(\delta) = \frac{\text{var}(Y_t) \delta^2}{\{2[\text{var}(Y_t) \delta^2] - 1 - \text{var}(Y_t) \delta^2\}^{1/2}}.
\]
Finally, we find that the covariance kernel of $G$ is just $\rho_1(\delta, \bar{\delta})$. This follows because

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

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

Thus, $G \overset{d}{=} G_1$, so that $QLR_n \Rightarrow \sup_{\delta \in \Delta} \G_1(\delta)^2$ 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.

#### 3.2.1 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, 2009a, 2009b), 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 $G \overset{d}{=} G_1$ of Section 3.1 is

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

with $\{Z_k\} \sim \text{IID } N(0, 1)$. To show this, we note that for any $\delta$ and $\bar{\delta}$ in $\Delta$, $E[G(\delta)] = E[\bar{G}(\delta)]$ and
Thus, if $K$ is sufficiently large, simulating $\tilde{G}(\delta; K)$ can yield a useful approximation to $G \overset{d}{=} G_1 \overset{d}{=} \tilde{G}$, where

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

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

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

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

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

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

$$\overline{QLR}_n(\Delta; K) := \sup_{\delta \in \Delta} \tilde{G}_n(\delta; K)^2,$$

and

$$\overline{QLR}_n(\Delta; K) := \sup_{\delta \in \Delta} G_n(\delta; K)^2.$$ 

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 $\Delta$: $\Delta_{0.5} := [-0.5, 0.5]$, $\Delta_{1.0} := [-1.0, 1.0]$, $\Delta_{1.5} := [-1.5, 1.5]$, and $\Delta_{2.0} := [-2.0, 2.0]$. We also let $(\theta_*, \beta_*, \sigma^2_{\epsilon}) = (0, 0.5, 1)$, so that $\text{var}(Y_t) = 4/3$. The distribution of $\overline{QLR}_n(\Delta; K)$ is obtained by grid search for the maximum over $\Delta$. 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 $\overline{QLR}_n(\Delta; K) = QLR_n(\Delta; K) = 0.$
Table 1 presents the asymptotic critical values obtained for $QLR_n(\Delta; K)$. We see immediately that these depend on $\Delta$. As $\Delta$ gets larger, the asymptotic critical values increase, as the definition of $QLR_n(\Delta; K)$ implies.

Table 2 presents the finite-sample properties of the QLR statistic. As Corollary 2 implies, for every $\Delta$, the finite-sample distribution of the QLR statistic approaches the asymptotic distribution of $QLR_n(\Delta; K)$ 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 $\Delta$. 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 $\Delta$ is smaller. The nominal rejection rates are closest to the asymptotic distribution for $QLR_n(\Delta; K)$ when $\Delta = \Delta_{0.5}$. If $\Delta = \Delta_{2.0}$, the finite sample distribution for QLR is still quite far from that of $QLR_n(\Delta_{2.0}; K)$, even when the sample size is 5,000.

Table 3 presents simulation results for the case in which $\text{var}(Y_t)$ is estimated. Simulating $\hat{QLR}_n(\Delta; K)$ 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, \ldots, and 5,000), each of which is replicated 10,000 times for each choice of $\Delta$. 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 $\hat{\text{var}}(Y_t)$. Using these, we put $K = 150$ as before and generate null distributions of $\hat{QLR}_n(\Delta; K)$, simulating 50,000 times to obtain precise critical values. Denote the critical values for a nominal level $\alpha$ and choice of $\Delta$ obtained in this way by $cv(\hat{\text{var}}(Y_t), \Delta, \alpha)$.

From this simulation, we observe that, for each $\alpha$ and $\Delta$, $cv(\hat{\text{var}}(Y_t), \Delta, \alpha)$ monotonically increases as $\hat{\text{var}}(Y_t)$ increases. Thus, if the sample QLR statistic is less than $cv(\hat{\text{var}}(Y_t), \Delta, \alpha)$ and its associated variance estimate is less than $\hat{\text{var}}(Y_t)$, then the null shouldn’t rejected. On the other hand, we reject the null if the QLR statistic is strictly greater than $cv(\hat{\text{var}}(Y_t), \Delta, \alpha)$ and the estimated variance is less than $\hat{\text{var}}(Y_t)$. We find that for each $\Delta$ and $\alpha$, 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 $(\hat{\text{var}}(Y_t), cv(\hat{\text{var}}(Y_t), \Delta, \alpha))$ and apply the standard decision rule.

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.

### 3.2.2 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 \( \Psi = \text{logistic CDF} \), 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 \( \delta \in \Delta \), we compute the scores \( \hat{S}_{nt}(\delta) := \left\{ \hat{D}_{nt}(\delta) \right\}^{-\frac{1}{2}} \hat{W}_{nt}(\delta) \), where

\[
\hat{D}_{nt}(\delta) := \frac{1}{n} \sum_{t=1}^{n} [\hat{U}_{nt} \Psi_t(\delta)]^2 - \left[ \frac{1}{n} \sum_{t=1}^{n} \hat{U}_{nt}^2 \Psi_t(\delta) \right] \left[ \frac{1}{n} \sum_{t=1}^{n} \hat{U}_{nt} Z_t Z_t' \right]^{-1} \left[ \frac{1}{n} \sum_{t=1}^{n} \hat{U}_{nt}^2 \Psi_t(\delta) \right],
\]

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

\( \hat{U}_{nt} := Y_t - Z_t' (\hat{\alpha}_n, \hat{\beta}_n)' \), 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 \( \Delta \) as before. Thus, there are 102 and grid points for Models 1 and 2.

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

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

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 \( \hat{\rho}_n \equiv J^{-1} \sum_{j=1}^{J} I[\text{QLR}_n < \text{QLR}_{jn}] \), where \( I[\cdot] \) denotes the indicator function.

Finally, we repeat the entire exercise 4,000 times, generating \( \hat{\rho}_n^{(i)}, i = 1, \cdots, 4,000 \), and we compute the proportion of outcomes whose \( \hat{\rho}_n^{(i)} \) is less than the specified nominal level (e.g., \( \alpha = 5\% \)). That is, we compute \( \frac{1}{4000} \sum_{i=1}^{4000} I[\hat{\rho}_n^{(i)} < \alpha] \). Under the null, this converges to the significance level corresponding to the specified nominal level, \( \alpha \). Plotted as a function of \( \alpha \), this should converge to a 45-degree line on the
We present these estimates in Table 4 and Figure 3. The first panel of Table 4 indicates the obtained empirical levels for $\alpha = 1\%, 5\%,$ and $10\%,$ 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., $\lim_{\delta \to 0} T(\delta, \delta) = 0$), 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’Hospital’s rule delivers the asymptotic distribution of the QLR statistic as $\delta$ converges to zero. The same result is especially obvious from Figure 3. The first panel of Figure 3 shows the estimated value of $\frac{1}{4000} \sum_{i=1}^{4000} I[\hat{p}_n^{(i)} < \alpha]$ for each $\alpha \in [0, 1]$. 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.

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. The empirical rejection rates differ from the nominal levels even when $n = 6,000,$ and this difference does not vanish with increasing sample size. Essentially, the quartic approximation is insufficient. The second panel of Figure 3 also shows the value of $\frac{1}{4000} \sum_{i=1}^{4000} I[\hat{p}_n^{(i)} < \alpha]$ for each $\alpha \in [0, 1]$. This exhibits the same behavior. The relation does not converge to a 45-degree line as the sample size increases even for $n = 6,000.$ 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. Only one of the possibilities under the null has previously been analyzed, 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 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.

5 Appendix: Proofs

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

\[ \hat{\sigma}_n^0 = n^{-1} \sum_{t=1}^{n} U_t^2 - (n^{-1} \sum_{t=1}^{n} U_t Z_t^\prime) (n^{-1} \sum_{t=1}^{n} Z_t Z_t^\prime)^{-1} (n^{-1} \sum_{t=1}^{n} Z_t U_t) \rightarrow \sigma^2 - \mathbf{0}^\prime (E[Z_t Z_t^\prime])^{-1} \mathbf{0} = \sigma^2 \]

in probability by the ergodic theorem.

(ii) We separate our proof into two parts. First in (a), we show the weak convergence of \( n^{-1/2} \Psi(\cdot)^\prime MU \).

In (b), we show that \( n^{-1} \hat{\sigma}_n^0 \Psi(\cdot)^\prime \Psi(\cdot) \rightarrow J(\cdot) \) uniformly on \( \Delta(\epsilon) \) in probability. Given these, the desired result follows from the converging-together lemma (Billingsley (1999, p. 151)).

(a) To show the weak convergence of \( n^{-1/2} \Psi(\cdot)^\prime MU \), we note that

\[ n^{-1/2} \Psi(\delta)^\prime MU = n^{-1/2} \sum_{t=1}^{n} \Psi_t(\delta) U_t - (\sum_{t=1}^{n} \Psi_t(\delta) Z_t^\prime) (\sum_{t=1}^{n} Z_t Z_t^\prime)^{-1} n^{-1/2} \sum_{t=1}^{n} Z_t U_t. \]

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

\[ \hat{\Psi}_{n,t}(\delta) := \Psi_t(\delta) U_t - (\sum_{t=1}^{n} \Psi_t(\delta) Z_t^\prime) (\sum_{t=1}^{n} Z_t Z_t^\prime)^{-1} Z_t U_t, \]

and let \( \hat{\Psi}_t(\delta) \) be defined as

\[ \hat{\Psi}_t(\delta) := \Psi_t(\delta) U_t - E(\Psi_t(\delta) Z_t^\prime) (E[Z_t Z_t^\prime])^{-1} Z_t U_t. \]

We show that

\[ \sup_{\delta \in \Delta(\epsilon)} \left| n^{-1/2} \sum_{t=1}^{n} [\hat{\Psi}_{n,t}(\delta) - \hat{\Psi}_t(\delta)] \right| = o_p(1) \quad (19) \]
and then show the weak convergence of \( \{ n^{-1/2} \sum_{t=1}^{n} \tilde{\Psi}_t(\cdot) \} \) on \( \Delta(\epsilon) \). First, we note that

\[
\sup_{\delta \in \Delta(\epsilon)} \left| n^{-1/2} \sum_{t=1}^{n} [\tilde{\Psi}_{n,t}(\delta) - \tilde{\Psi}_t(\delta)] \right| 
\leq \sup_{\delta \in \Delta(\epsilon)} \left| (n^{-1} \sum_{t=1}^{n} \Psi_t(\delta)Z_t') \left\{ (n^{-1} \sum_{t=1}^{n} Z_tZ_t')^{-1} - (E[Z_tZ_t'])^{-1} \right\} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_tU_t 
+ \sup_{\delta \in \Delta(\epsilon)} \left\{ (n^{-1} \sum_{t=1}^{n} \Psi_t(\delta)Z_t') - E(\Psi_t(\delta)Z_t') \right\} (E[Z_tZ_t'])^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_tU_t \right| .
\]

We note that \( \{Z_tU_t, \mathcal{F}_t\} \) is a martingale difference sequence (MDS) so that \( E[Z_tU_t] = 0 \), with \( E[|X_{t,i}|^4] = E[U_t^4]^{1/2} E[|X_{t,i}|^4]^{1/2} \leq E[M_t^4]^{1/2} E[X_{t,i}^4]^{1/2} < \infty \); also, \( E[U_t^2Z_t] \) is positive definite by A6, where \( \mathcal{F}_{t-1} := \sigma(X_t, U_{t-1}, \cdots) \) and then show the weak convergence of \( \{ \tilde{\Psi}_t(\cdot) \} \) on \( \Delta(\epsilon) \). Further, \( \sum_{t=1}^{n} Z_tU_t = O_P(n^{1/2}) \).

\[
\sup_{\delta \in \Delta} \left\| n^{-1} \sum_{t=1}^{n} \Psi_t(\delta)Z_t - E[\Psi_t(\delta)Z_t] \right\|_1 = o_P(1),
\]
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,

\[
\sup_{\delta \in \Delta(\epsilon)} \left\{ (n^{-1} \sum_{t=1}^{n} \Psi_t(\delta)Z_t') - E(\Psi_t(\delta)Z_t') \right\} (E[Z_tZ_t'])^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_tU_t = o_P(1).
\]

Also, \( n^{-1} \sum_{t=1}^{n} Z_tZ_t' - E[Z_tZ_t'] = o_P(1) \) and for each \( i = 1, 2, \cdots, k \), \( \sup_{\delta \in \Delta} |\sum_{t=1}^{n} \Psi_t(\delta)X_{t,i}| \leq \sum_{t=1}^{n} M_t |X_{t,i}| = O_P(n) \) by applying A5 and the ergodic theorem. This implies that

\[
\sup_{\delta \in \Delta(\epsilon)} \left| n^{-1} \sum_{t=1}^{n} \Psi_t(\delta)Z_t' \left\{ (n^{-1} \sum_{t=1}^{n} Z_tZ_t')^{-1} - (E[Z_tZ_t'])^{-1} \right\} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_tU_t \right| = o_P(1).
\]

Thus, \((19)\) follows.

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

\[
\sup_{\|\delta - \tilde{\delta}\| < \eta} |\Psi_t(\delta)U_t - \Psi_t(\delta)U_t|^{2+\kappa} \leq M_t^{2+\kappa} |U_t|^{2+\kappa} \eta^{2+\kappa} \leq M_t^{4+2\kappa} \eta^{2+\kappa},
\]

and then show the weak convergence of \( \{ n^{-1/2} \sum_{t=1}^{n} \tilde{\Psi}_t(\cdot) \} \) on \( \Delta(\epsilon) \). First, we note that
implying that
\[ E \left[ \sup_{\|\delta - \tilde{\delta}\| < \eta} |\Psi_t(\delta) U_t - \Psi_t(\tilde{\delta}) U_t|^{2+\kappa} \right]^{\frac{1}{2+\kappa}} \leq E[M_t^{4+2\kappa}]^{\frac{1}{2+\kappa}} \eta. \]  

(20)

Likewise, by the moment condition in A3, there is some \( C \) such that
\[ |E(\Psi_t(\delta)Z_t')(E[Z_tZ_t'])^{-1} Z_t U_t - E(\Psi_t(\tilde{\delta})Z_t')(E[Z_tZ_t'])^{-1} Z_t U_t| \]
\[ = |E[(\{\Psi_t(\delta) - \Psi_t(\tilde{\delta})\}Z_t')(E[Z_tZ_t'])^{-1} Z_t U_t] | \]
\[ \leq CM_t^2 \|\{E[Z_tZ_t']\}^{-1}\|_1 \cdot \|E[(\{\Psi_t(\delta) - \Psi_t(\tilde{\delta})\}Z_t)]\|_1. \]

Then
\[ |E(\Psi_t(\delta)Z_t')(E[Z_tZ_t'])^{-1} Z_t U_t - E(\Psi_t(\tilde{\delta})Z_t')(E[Z_tZ_t'])^{-1} Z_t U_t| \]
\[ \leq C \cdot M_t^2 \cdot \|\{E[Z_tZ_t']\}^{-1}\|_1 \cdot \|E[Z_t \Psi_t(\delta)] - E[Z_t \Psi_t(\tilde{\delta})]\|_1 \]
\[ \leq k \cdot M_t^2 \cdot \|\{E[Z_tZ_t']\}^{-1}\|_1 \cdot E[M_t^2] \cdot \|\delta - \tilde{\delta}\|, \]

where the last inequality follows from the Lipschitz continuity condition implied by A5. That is, for \( j = 1, 2, \cdots, k + 1, |Z_{j,t}[\Psi_t(\delta) - \Psi_t(\tilde{\delta})]| \leq M_t^2 \|\delta - \tilde{\delta}\|. \) Hence,
\[ E \left[ \sup_{\|\delta - \tilde{\delta}\| < \eta} |E[(\{\Psi_t(\delta) - \Psi_t(\tilde{\delta})\}Z_t')(E[Z_tZ_t'])^{-1} Z_t U_t^{2+\kappa}] \right]^{\frac{1}{2+\kappa}} \]
\[ \leq C \cdot E[M_t^{4+2\kappa}]^{\frac{1}{2+\kappa}} \cdot \|\{E[Z_tZ_t']\}^{-1}\|_1 \cdot E[M_t^2] \cdot \eta. \]  

(21)

Given (20) and (21), it follows that that for some \( B < \infty, E \left[ \sup_{\|\delta - \tilde{\delta}\| \leq \eta} |\Psi_t(\delta) - \Psi_t(\tilde{\delta})|^{2+\kappa} \right] \leq B \eta, \) implying that Ossiander’s \( L^{2+\kappa} \) entropy is finite. Thus, \( \{n^{-1/2} \sum_{t=1}^n \hat{\Psi}_{n,t}(\cdot)\} \) 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 \( \Delta(e) \).

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

(b) Next, for each \( \delta, \)
\[ n^{-1} \Psi(\delta)' M \Psi(\delta) = n^{-1} \sum_{t=1}^n \Psi_t(\delta)^2 - \{n^{-1} \sum_{t=1}^n \Psi_t(\delta)Z_t\} \{n^{-1} \sum_{t=1}^n Z_tZ_t'\}^{-1} \{n^{-1} \sum_{t=1}^n \Psi_t(\delta)Z_t\}. \]

It easily follows that \( \sup_{\delta} |n^{-1} \sum_{t=1}^n \Psi_t(\delta)^2 - E[\Psi_t(\delta)^2]| \to 0 \) in probability and \( \sup_{\delta} \|n^{-1} \sum_{t=1}^n \Psi_t(\delta) \to 0 \).
\( \mathbf{Z}_t - E[\Psi_t(\delta)\mathbf{Z}_t] \|_1 \to 0 \) 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 \( n^{-1} \sum_{t=1}^n \mathbf{Z}_t \mathbf{Z}_t' \),

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

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

(iii) A4(ii) immediately implies

\[
T(\delta, \tilde{\delta}) := E[U_t^2 \Psi_t^*(\delta) \Psi_t^*(\tilde{\delta})] = E[U_t^2 X_t \Psi_t^*(\delta) \Psi_t^*(\tilde{\delta})] = \sigma_t^2 E[\Psi_t^*(\delta) \Psi_t^*(\tilde{\delta})].
\]

\[\blacksquare\]

\textbf{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 \( \rho_t(\delta, \tilde{\delta}) \) to obtain the desired result. \[\blacksquare\]

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

\[
i' \mathbf{D}_{ij} \mathbf{M} \mathbf{U} = \sum_{t=1}^n X_{t,i} X_{t,j} U_t - \sum_{t=1}^n X_{t,i} X_{t,j} \mathbf{Z}_t' \{ \sum_{t=1}^n \mathbf{Z}_t \mathbf{Z}_t' \}^{-1} \sum_{t=1}^n \mathbf{Z}_t U_t.
\]

We note that \( n^{-1} \sum_{t=1}^n X_{t,i} X_{t,j} \mathbf{Z}_t' \) and \( n^{-1} \sum \mathbf{Z}_t \mathbf{Z}_t' \) obey the WLLN by the ergodic theorem and A3* and further show that \( n^{-1/2} \sum_{t=1}^n X_{t,i} X_{t,j} U_t \) satisfies the CLT. We already showed that \( n^{-1/2} \sum \mathbf{Z}_t U_t \) obeys the asymptotic normality in the proof of Lemma 1(i), which also holds under the conditions of Lemma 3. For \( n^{-1/2} \sum_{t=1}^n X_{t,i} X_{t,j} U_t \), 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

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

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, \( i' \mathbf{D}_{ij} \mathbf{M} \mathbf{U} = O_p(n^{1/2}) \).

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

\[
i' \mathbf{D}_{ij} \mathbf{M} \mathbf{U} = \sum_{t=1}^n X_{t,i} X_{t,j} X_{t,i} U_t - \sum_{t=1}^n X_{t,i} X_{t,j} X_{t,i} \mathbf{Z}_t' \{ \sum_{t=1}^n \mathbf{Z}_t \mathbf{Z}_t' \}^{-1} \sum_{t=1}^n \mathbf{Z}_t U_t,
\]
and that $n^{-1} \sum_{t=1}^{n} X_{t,i}X_{t,j}X_{t,\ell}Z_t$ 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

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

alternatively, we have

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

by A3*. Finally, $E[X_{t,i}^2X_{t,j}X_{t,\ell}^2U_t^2] > 0$. 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, $\nu D_{ij\ell}MU = O_p(n^{1/2})$, so that $\nu D_{ij\ell}MU = o_p(n^{3/4})$.

(iii) We have

$$\nu D_{ijm}MU = \sum_{t=1}^{n} X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}U_t - \sum_{t=1}^{n} X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}Z_t(\sum_{t=1}^{n} Z_t Z_t')^{-1} \sum_{t=1}^{n} Z_t U_t.$$ 

By the ergodic theorem and A3*, $n^{-1} \sum X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}Z_t$ and $n^{-1} \sum Z_t U_t$ respectively converge to $E[X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}Z_t]$ and 0 in probability. Also, $E[X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}U_t|Z_t] = 0$, 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

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

by A3*.

(iv) By our definitions, $\nu D_{ijm}MD_{lm}$ is identical to

$$\nu D_{ijm}MD_{lm} = \sum_{t=1}^{n} X_{t,i}X_{t,j}X_{t,\ell}X_{t,m} - \sum_{t=1}^{n} X_{t,i}X_{t,j}Z_t(\sum_{t=1}^{n} Z_t Z_t')^{-1} \sum_{t=1}^{n} Z_t X_{t,\ell}X_{t,m}.$$ 

We can apply the ergodic theorem to each element, so that $n^{-1} \nu D_{ijm}MD_{lm} \rightarrow \tau_{ijm} a.s.,$ where

$$\tau_{ijm} := E[C_{t,ij}^{*} C_{t,lm}^{*}] = E[X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}] - E[X_{t,i}X_{t,j}Z_t] E[Z_t Z_t']^{-1} E[Z_t X_{t,\ell}X_{t,m}],$$

32
which is finite by Conditions A3* and A6. This yields the desired result.

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

$$\text{vech}(n^{-1/2}\tilde{M}) = n^{-1/2} \sum_{t=1}^{n} C_t U_t - (n^{-1} \sum_{t=1}^{n} C_t Z_t')(n^{-1} \sum_{t=1}^{n} Z_t Z_t')^{-1}(n^{-1/2} \sum_{t=1}^{n} Z_t U_t).$$

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

$$n^{-1/2} \sum_{t=1}^{n} [Z_t', C_t']'^t U_t \overset{\psi}{\to} N(0, \tilde{V}_1).$$

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

$$\text{vech}(n^{-1/2}\tilde{M}) \overset{\psi}{\to} N(0, \{E[U_t^2(C_t - E[C_t Z_t'])E[Z_t Z_t']^{-1} Z_t]\{C_t - E[C_t Z_t']E[Z_t Z_t']^{-1} Z_t\}').$$

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

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

(iii) For each $i, j, \ell, m = 1, 2, \cdots, k,$ $E[\mathcal{M}_{i j} \mathcal{M}_{\ell m}] = E[U_t^2 C_{t, i j}^* C_{t, t lm}^*] = E[E[U_t^2|X_t]C_{t, i j}^* C_{t, t lm}^*] = \sigma_{i j}^2 E[C_{t, i j}^* C_{t, t lm}^*]$ 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 $n^{-1/2}d'\tilde{M}d$ as a function of $d.$ In (b), we show that $b'\mathcal{W}b$ obeys the ULLN as a function of $b.$ Finally, the covariance structure of $\mathcal{G}_2$ is derived in (c).

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

$$\sup_{d \in \mathbb{S}^{k-1}} n^{-1/2}d'(\tilde{M} - \mathcal{M}^*)d = o_p(1),$$

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

33
First, we note that $\tilde{M} = [\ell^* D_{ij} \ell \ell']$ and

$$
\ell^* D_{ij} \ell \ell' \mu - \sum_{t=1}^{n} C_{t, i, j} U_t = \left[ \left( \sum_{t=1}^{n} X_{t,i} X_{t,j} Z_t' \right) \left( \sum_{t=1}^{n} Z_t Z_t' \right)^{-1} - P_{ij} \right] \sum_{t=1}^{n} Z_t U_t,
$$

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

$$
P \left( \sup_{\|d-d'\|< \delta} \frac{1}{\sqrt{n}} |d'M^* d - \tilde{d}' M^* \tilde{d}| > \varepsilon \right) \leq P \left( \sup_{\|d-d'\|< \delta} \frac{1}{\sqrt{n}} |d'M^* d - d'M^* \tilde{d}| > \frac{\varepsilon}{2} \right)
$$

$$
+ P \left( \sup_{\|d-d'\|< \delta} \frac{1}{\sqrt{n}} |d'M^* \tilde{d} - \tilde{d}' M^* \tilde{d}| > \frac{\varepsilon}{2} \right). \tag{22}
$$

We have

$$
\sup_{\|d-d'\|< \delta} |d'M^* d - d'M^* \tilde{d}| \leq \sup_{\|d-d'\|< \delta} \sum_{i=1}^{k} \sum_{j=1}^{k} |\{\ell^* D_{ij} \ell - P_{ij} Z' U\} \ell_i (d_j - \tilde{d}_j)|
$$

$$
\leq \delta \sum_{i=1}^{k} \sum_{j=1}^{k} |\{\ell^* D_{ij} \ell - P_{ij} Z' U\}|.
$$

Therefore,

$$
P \left( \sup_{\|d-d'\|< \delta} \frac{1}{\sqrt{n}} |d'M^* d - d'M^* \tilde{d}| > \frac{\varepsilon}{2} \right) \leq P \left( \delta \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{1}{\sqrt{n}} |\ell^* D_{ij} \ell - P_{ij} Z' U| > \frac{\varepsilon}{2} \right)
$$

$$
\leq \sum_{i=1}^{k} \sum_{j=1}^{k} P \left( \frac{1}{\sqrt{n}} |\ell^* D_{ij} \ell - P_{ij} Z' U| > \frac{\varepsilon}{2 \delta k^2} \right)
$$

$$
\leq \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{4\delta^2 k^4}{\varepsilon^2} \sigma_{i,j}
$$

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

$$
\limsup_{n \to \infty} P \left( \sup_{\|d-d'\|< \delta} \frac{1}{\sqrt{n}} |d'M^* d - \tilde{d}' M^* \tilde{d}| > \varepsilon \right) \leq \frac{8\delta^2 k^4}{\varepsilon^2} \sum_{i=1}^{k} \sum_{j=1}^{k} \sigma_{i,j} \tag{23}
$$

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 $\delta$ be small. It follows that $\{n^{-1/2} d'M^* d\}$ is tight. Its finite-dimensional dis-
tribution also follows from the multivariate CLT, which ensures \( n^{-1/2}\tilde{M} \Rightarrow \mathcal{M} \). As this is straightforward, we leave the details aside here. Therefore, \( n^{-1/2}d'\tilde{M}d \) weakly converges to \( d'\mathcal{M}d \) as a function of \( d \).

(b) Since \( b'Wb = \text{vec}(dd')'W\text{vec}(dd') \) and

\[
\sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{\ell=1}^{k} \sum_{m=1}^{k} \tau_{ij\ell m}d_idjd\ell d_m = \text{vec}(dd')'W^*\text{vec}(dd'),
\]

it follows that

\[
\sup_{d \in \mathbb{R}^{k-1}} \left| n^{-1}\text{vec}(dd')'W\text{vec}(dd') - \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{\ell=1}^{k} \sum_{m=1}^{k} \tau_{ij\ell m}d_idjd\ell d_m \right| \\
\leq \sup_{d \in \mathbb{R}^{k-1}} \left| n^{-1}(d'D_{ij}\mathcal{M}D_{\ell m} - \tau_{ij\ell m}) \times |d_idjd\ell d_m| \right|
\]

We already saw that \( n^{-1}D_{ij}\mathcal{M}D_{\ell m} \rightarrow \tau_{ij\ell m} \) a.s. in the proof of Lemma 3(ii). Therefore,

\[
\sup_{d \in \mathbb{R}^{k-1}} \left| n^{-1}\text{vec}(dd')'W\text{vec}(dd') - \text{vec}(dd')'W^*\text{vec}(dd') \right| \rightarrow 0 \ a.s.
\]

This implies that \( n^{-1}b'Wb = n^{-1}\text{vec}(dd')'W\text{vec}(dd') \) obeys the ULLN as a function of \( d \).

(c) Further, \( \{n^{-1/2}d'\tilde{M}d, n^{-1}\text{vec}(dd')'W\text{vec}(dd'), \tilde{\sigma}_n^0 \} \Rightarrow \{d'\mathcal{M}d, \text{vec}(dd')'W^*\text{vec}(dd'), \sigma^2 \} \) as functions of \( d \) by (a), (b), and the fact that \( \tilde{\sigma}_n^0 \rightarrow \sigma^2 \) a.s. Therefore, as a function of \( d \), \( \{\tilde{\sigma}_n^0 b'Wb\}^{-1/2}d'\tilde{M}d = \{\tilde{\sigma}_n^0 b'Wb\}^{-1/2}b'\text{vec}(\tilde{M}) \Rightarrow G_2(b) := \{\sigma^2 b'W^*b\}^{-1/2}b'\text{vec}(\mathcal{M}) = \{I(b, b)\}^{-1/2}b'\text{vec}(\mathcal{M}) \) by the converging together lemma (theorem 3.9 of Billingsley (1999, p. 37)) and the definition of \( G_2(b, b) \).

As the final step, we derive the covariance structure of \( G_2 \) from the asymptotic covariance between \( n^{-1/2}d'\tilde{M}d \) and \( n^{-1/2}d'\tilde{M}d \). We already proved in (a) that \( n^{-1/2}(\tilde{M} - \mathcal{M}) = o_p(1) \). Further, we note that

\[
n^{-1}E[d'M^*d'\tilde{M}^*d'] = \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{\ell=1}^{k} \sum_{m=1}^{k} E[U_i^2C_{i,j}^*C_{\ell,m}^*]d_idjd\ell d_m = b'E[\text{vec}(\mathcal{M})\text{vec}(\mathcal{M})']b.
\]

This shows that \( n^{-1}E[d'\tilde{M}d d'\tilde{M}d] \rightarrow b'E[\text{vec}(\mathcal{M})\text{vec}(\mathcal{M})']b \) in probability, which is \( K(b, \bar{b}) \) from the fact that \( E[U_i^2C_{i,j}^*C_{\ell,m}^*] = E[M_{ij}, \mathcal{M}_{\ell m}] \). Therefore, it follows that \( E[G_2(b)G_2(\bar{b})] = \rho_2(b, \bar{b}) \). This is the desired covariance structure.

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

(ii) By Lemma 4(iii), \( K(b, \bar{b}) = I(b, \bar{b}) \), so that \( \rho_2(b, \bar{b}) = I(b, b)^{-1/2}I(b, \bar{b})I(\bar{b}, \bar{b})^{-1/2} \). 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 $H_{02}$,

(i) for $j = 0, 1, 2, 3,$ and $4$, $U'J_j = O_P(n^{1/2})$;

(ii) for $j = 0, 1, 2, 3,$ and $4$, $J_0'J_j = O_P(n)$;

(iii) for $j = 1, 2,$ and $3$, $J_1'J_j = O_P(n)$; and

(iv) $J_2'J_2 = O_P(n)$.

**Proof of Lemma A1:** (i) First, we note that $J_0 := [X, c_0^t]$ and that $U'Z = O_P(n^{1/2})$ as shown in the proof of Lemma 4(i). This shows that $U'J_0 = O_P(n^{1/2})$. Second, we note that $J_1 = c_1[0_{n \times k}, G_t]$, where $G := \text{diag}(Xd)$, so that $U'J_1 = c_1[0_{n \times k}, U'Xd]$. We further note that $d$ is bounded and $U'X = O_P(n^{1/2})$, implying that $U'J_1 = O_P(n^{1/2})$. Third, $U'J_2 = c_2[0_{n \times k}, U'G^2t]$ from the fact that $J_2 = c_2[0_{n \times k}, G^2t]$ and that $U'G^2t = \sum_i d_i \sum_j d_j \sum_{\ell=1}^n (X_{t,i}X_{t,j}U_t)$. We now note that the proof of Lemma 3(i) shows that for each $i$ and $j$, $\sum X_{t,i}X_{t,j}U_t = O_P(n^{1/2})$. Therefore, $U'J_2 = O_P(n^{1/2})$. Fourth, $U'J_3 = c_3[0_{n \times k}, U'G^3t]$ because $J_3 = c_3[0_{n \times k}, G^3t]$. The proof of Lemma 3(ii) now shows that for each $i$, $j$, and $\ell$, $\sum X_{t,i}X_{t,j}X_{t,\ell}U_t = O_P(n^{1/2})$, and $U'G^3t = \sum_i d_i \sum_j d_j \sum_{\ell} d_\ell \sum_{t=1}^n (X_{t,i}X_{t,j}X_{t,\ell}U_t)$, so that $U'J_3 = O_P(n^{1/2})$. Finally, $U'J_4 = c_4[0_{n \times k}, U'G^4t]$ using the fact that $J_4 = c_4[0_{n \times k}, G^4t]$. The proof of Lemma 3(iii) proves that for each $i$, $j$, and $\ell$, $\sum X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}U_t = O_P(n^{1/2})$ and $U'G^4t = \sum_i d_i \sum_j d_j \sum_{\ell} d_\ell \sum_m d_m \sum_{t=1}^n (X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}U_t)$, so that $U'J_4 = O_P(n^{1/2})$.

(iii) First, we note that $J_0'J_0 = H_0 = O_P(n)$ by the fact that $Z'Z = O_P(n)$ as given in the proof of Lemma 4(i). Second, $J_0'J_1 = c_1[0_{(k+1) \times k}, Z'Gt] = c_1[0_{(k+1) \times k}, Z'Xd]$. Given that $d$ is bounded and that for each $i$, $\sum X_{t,i}Z_t = O_P(n)$, we note that $Z'Xd = \sum_i d_i \sum_{t=1}^n X_{t,i}Z_t$, so that $J_0'J_1 = O_P(n)$. Third, $J_0'J_2 = c_2[0_{(k+1) \times k}, Z'G^2t]$. Now, for each $i$ and $j$, $\sum X_{t,i}X_{t,j}Z_t = O_P(n)$ as given in the proof of Lemma 3(i), and $Z'G^2t = \sum_i d_i \sum_j d_j \sum_{t=1}^n X_{t,i}X_{t,j}Z_t$. Thus, $J_0'J_2 = O_P(n)$. Fourth, we note that $J_0'J_3 = c_3[0_{(k+1) \times k}, Z'G^3t]$. Now, for each $i$, $j$, and $\ell$, $\sum X_{t,i}X_{t,j}X_{t,\ell}Z_t = O_P(n)$, as we saw in the proof of Lemma 3(ii), and $Z'G^3t = \sum_i d_i \sum_j d_j \sum_{\ell} d_\ell \sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell}Z_t$. Thus, $J_0'J_3 = O_P(n)$. Finally, $J_0'J_4 = c_4[0_{(k+1) \times k}, Z'G^4t]$. Now, for each $i$, $j$, and $m$, and $\ell$, $\sum X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}Z_t = O_P(n)$, as we saw in the proof of Lemma 3(iii), and $Z'G^4t = \sum_i d_i \sum_j d_j \sum_{\ell} d_\ell \sum_m d_m \sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}Z_t$. Thus, $J_0'J_4 = O_P(n)$.

(iii) First, $J_1'J_1 = c_1^2 t'G^2t \xi \xi'$, where we let $\xi := [0'_{(k \times 1)}, 1]'$. Given the definition of $G := \text{diag}(Xd)$ and the fact that $d$ is bounded, $t'G^2t = O_P(n)$ because for each $i$ and $j$, $\sum X_{t,i}X_{t,j} = O_P(n)$ and $t'G^2t = \sum_i d_i \sum_j d_j \sum_{\ell} d_\ell \sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell}X_{t,m}Z_t$. Thus, $J_0'J_4 = O_P(n)$. 

36
\[ \sum_i d_i \sum_j d_j \sum_{t=1}^n X_{t,i}X_{t,j}. \] Thus, \( J'_1 J_1 = O_F(n) \). Second, we also note that \( J'_1 J_2 = c_1 c_2 \nu' G^3 \nu \xi \xi' \), so that \( \nu' G^3 \nu = O_F(n) \) because for each \( i, j, \) and \( \ell, \sum X_{t,i}X_{t,j}X_{t,\ell} = O_F(n) \) and \( \nu' G^3 \nu = \sum_i d_i \sum_j d_j \sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell} \), implying that \( J'_1 J_2 = O_F(n) \). Finally, \( J'_1 J_3 = c_1 c_3 \nu' G^4 \nu \xi \xi' \), and for each \( i, j, \) and \( m, \sum X_{t,i}X_{t,j}X_{t,\ell}X_{t,m} = O_F(n) \). This implies that \( J'_1 J_3 = O_F(n) \) because \( \nu' G^4 \nu = \sum_i d_i \sum_j d_j \sum_{t=1}^n X_{t,i}X_{t,j}X_{t,\ell}X_{t,m} \). (iv) We note that \( J'_2 J_2 = c_2^2 \nu' G^4 \nu \xi \xi' \), and we have already seen that \( \nu' G^4 \nu = O_F(n) \) in the proof of Lemma A1(iii), so that \( J'_2 J_2 = O_F(n) \). This completes the proof.

\[ \textbf{Lemma A2: Given A1 and A2, when } V(h; d) \text{ is defined as } [Q(hd)'Q(hd)]^{-1}, \]

(i) \( \frac{\partial}{\partial h} V(0; d) = -H_0^{-1}H_1 H_0^{-1}; \)

(ii) \( \frac{\partial^2}{\partial h^2} V(0; d) = 2H_0^{-1}H_1 H_0^{-1}H_1 H_0^{-1} - H_0^{-1}H_2 H_0^{-1}; \)

(iii) \( \frac{\partial^3}{\partial h^3} V(0; d) = H_0^{-1}\{-6H_1 H_0^{-1}H_1 H_0^{-1}H_1 + 3H_2 H_0^{-1}H_1 + 3H_1 H_0^{-1}H_2 - H_3\}H_0^{-1}; \) and

(iv)

\[
\frac{\partial^4}{\partial h^4} V(0; d) = 24H_0^{-1}H_1 H_0^{-1}H_1 H_0^{-1}H_1 H_0^{-1}H_1 H_0^{-1} + 6H_0^{-1}\{H_2 H_0^{-1}H_2 - 2H_1 H_0^{-1}H_2 H_0^{-1}H_1 - 2H_2 H_0^{-1}H_1 H_0^{-1}H_1 - 2H_1 H_0^{-1}H_1 H_0^{-1}H_2\}H_0^{-1} + 4H_0^{-1}\{H_1 H_0^{-1}H_3 + H_3 H_0^{-1}H_1\}H_0^{-1} - H_0^{-1}H_4 H_0^{-1}. \]

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

\[ \textbf{Lemma A3: Given A1, A2, A3*, A4(i), A6*, and } H_{02}, \]

(i) for \( j = 0, 1, 2, 3, \) and \( 4, \) \( \frac{\partial}{\partial h} V(h; d) = O_F(n^{-1}); \)

(ii) \( J'_0(K + K')U = o_F(n^{-3/4}) \) and \( U'(K + K')U = o_F(n^{-3/4}), \) where

\[
K := 3J_1 H_0^{-1}J'_1 - 3J_2 H_0^{-1}H_1 H_0^{-1}J'_1 + 6J_1 H_0^{-1}H_1 H_0^{-1}H_1 H_0^{-1}J'_0 - 3J_1 H_0^{-1}H_2 H_0^{-1}J'_1 \]

\[
3J_2 H_0^{-1}J'_1 - 3J_0 H_0^{-1}H_1 H_0^{-1}H_1 H_0^{-1}J'_0 + 3J_0 H_0^{-1}H_1 H_0^{-1}H_2 H_0^{-1}J'_0; \] and

(iii) \( J'_0(L + L')U = o_F(n) \) and \( U'(L + L')U = o_F(n), \) where

\[
L := \left\{ J_1 V(0; d) + 4J_3 \frac{\partial}{\partial h} V(0; d) + 6J_2 \frac{\partial^2}{\partial h^2} V(0; d) + 4J_1 \frac{\partial^3}{\partial h^3} V(0; d) \right\} J'_0 \]

\[
+ \left\{ 4J_3 V(0; d) + 12J_2 \frac{\partial}{\partial h} V(0; d) + 6J_1 \frac{\partial^2}{\partial h^2} V(0; d) \right\} J'_1 + 3J_2 V(0; d) J'_2 \] and

(iv) \( J'_0 M = 0_{(k+1) \times n} \) and \( M J_0 = 0_{n \times (k+1)}; \) and
(v) $J'_1 M = 0_{(k+1) \times n}$ and $M J_1 = 0_{n \times (k+1)}$.

Proof of Lemma A3: (i) First, we note that $H_0 = J'_0 J_0$ and $Z'Z = O_\varphi(n)$ as proved in the proof of Lemma 4(i), implying that $V(0, d) = H_0^{-1} = O_\varphi(n^{-1})$. Second, we note that $H_1 = J'_0 J_1 + J'_1 J_0$. Lemma A1(ii) shows that $J'_0 J_1 = O_\varphi(n)$, so that $H_1 = O_\varphi(n)$. Given that $H_0^{-1} = O_\varphi(n^{-1})$, $\frac{\partial^2}{\partial h} V(0, d) = -H_0^{-1} H_1 H_0^{-1} = O_\varphi(n^{-1})$. Third, note that $H_2 = J'_2 J_0 + J'_1 J_1 + J'_0 J_2$, so that Lemmas A1(ii and iii) imply that $H_2 = O_\varphi(n)$. Now, Lemma A2(ii) shows that $\frac{\partial^2}{\partial h^2} V(0, d) = O_\varphi(n^{-1})$. Fourth, we note that $H_3 = J'_3 J_0 + 3J'_2 J_1 + 3J'_1 J_2 + J'_0 J_3$, and Lemmas A1(ii and iii) imply that $H_3 = O_\varphi(n)$. Therefore, Lemma A2(iii) shows that $\frac{\partial^3}{\partial h^3} V(0, d) = O_\varphi(n^{-1})$. Finally, $H_4 = J'_4 J_0 + 4J'_3 J_1 + 6J'_2 J_2 + 4J'_1 J_3 + J'_0 J_4$, and Lemmas A1(iii, iii, and iv) imply that $H_4 = O_\varphi(n)$, so that $\frac{\partial^4}{\partial h^4} V(0, d) = O_\varphi(n^{-1})$ by applying Lemma A2(iv).

(ii) First, we note that $K + K' = K_1 + K_2 + K_2'$, where $K_1 := J'_0 \frac{\partial^3}{\partial h^3} V(0, d) J'_0$ and

$$K_2 := \left\{ J_3 V(0, d) + 3J_2 \frac{\partial}{\partial h} V(0, d) + 3J_1 \frac{\partial^2}{\partial h^2} V(0, d) \right\} J'_0 + 3J_2 \frac{\partial}{\partial h} V(0, d) J'_1.$$  

Second, we prove that $J'_0 (K + K') = U_0 (n^{1/2})$. For this, note that $J'_0 K_1 U = J'_0 \frac{\partial^3}{\partial h^3} V(0, d) J'_0$, $J'_0 J_0 = O_\varphi(n)$, $\frac{\partial^3}{\partial h^3} V(0, d) = O_\varphi(n^{-1})$, and $J'_0 U = O_\varphi(n^{1/2})$, implying that $J'_0 K_1 U = O_\varphi(n^{1/2})$. Also, $U' K_1 U = U' J'_0 \frac{\partial^3}{\partial h^3} V(0, d) J'_0 U = O_\varphi(1)$. Further, the main components constituting $J'_0 K_2 U$ are $J'_0 J_3 V(0, d) J'_0 U (j = 0, 1, 2)$ and $J'_0 J_2 \frac{\partial}{\partial h} V(0, d) J'_1 U$. By Lemmas A1(i $\sim$ iii), it follows that $J'_0 J_3 V(0, d) = O_\varphi(n^{-1})$, and both $J'_0 U$ and $J'_1 U$ are $O_\varphi(n^{1/2})$. These imply that $J'_0 K_2 U = O_\varphi(n^{1/2})$. We can apply the same arguments for $J'_0 K_2' U$ to obtain that $J'_0 K_2' U = O_\varphi(n^{1/2})$. This implies that $J'_0 (K_1 + K_2 + K_2') U = J'_0 (K + K') U = O_\varphi(n^{1/2}) = o_\varphi(n^{3/4})$ as desired.

Finally, we prove that $U'(K + K') U = O_\varphi(1)$. For this, note that $U' K_1 U = U' J'_0 \frac{\partial^3}{\partial h^3} V(0, d) J'_0 U$, $\frac{\partial^3}{\partial h^3} V(0, d) = O_\varphi(n^{-1})$, and $J'_0 U = O_\varphi(n^{1/2})$, implying that $U' K_1 U = O_\varphi(n^{1/2})$. Also, $U' K_1 U = U' J'_0 \frac{\partial^3}{\partial h^3} V(0, d) J'_0 U = O_\varphi(1)$. Further, we note that the main components constituting $U' K_2 U$ are $U' J_3 V(0, d) J'_0 U (j = 0, 1, 2)$ and $U' J_2 \frac{\partial}{\partial h} V(0, d) J'_1 U$. By Lemmas A1(i and iii), $U' J_3 V(0, d) = O_\varphi(n^{-1})$, and both $U' K_2 U$ and $U' (K + K') U = O_\varphi(1) = o_\varphi(n^{3/4})$.

(iii) First, the main components constituting $J'_0 L U$ are $J'_0 J_1 J_3 V(0, d) J'_0 U (j = 0, 1, 2, 3)$, $J'_0 J_3 J_3 V(0, d) J'_0 U (j = 0, 1, 2)$, and $J'_0 J_2 V(0, d) J'_2 U$. Lemma A1(i) shows that $J'_j U = O_\varphi(n^{1/2})$; Lemma A1(ii) shows that $J'_j J_0 = O_\varphi(n)$; and Lemma A3(i) shows that $\frac{\partial}{\partial h^3} V(0, d) = O_\varphi(n^{-1})$. These facts imply that $J'_0 L U = O_\varphi(n^{1/2}) = o_\varphi(n)$, and it can be easily proved that $J'_0 L' U = o_\varphi(n)$ in a similar way, so that
\[ J'_0(L + L')U = o_P(n). \]

Next, we note that the main components constituting \( U'L \) are \( U'J_{4-j}\frac{\partial}{\partial h} V(0; d)J'_0U \) \((j = 0, 1, 2, 3)\), \( U'J_{3-j}\frac{\partial}{\partial h} V(0; d)J'_0U \) \((j = 0, 1, 2)\), and \( U'J_2V(0; d)J'_2U \). Lemma A1(i) shows that \( J'_jU = O_P(n^{1/2}) \); and Lemma A3(i) shows that \( \frac{\partial}{\partial h} V(0; d) = O_P(n^{-1}) \). These facts imply that \( U'L = O_P(1) = o_P(n) \). We can also prove in the same way that \( U'L'U = O_P(1) = o_P(n) \), implying that \( U'(L + L')U = O_P(1) = o_P(n) \).

(iii) We note that \( M := I - Z(Z'Z)^{-1}Z' = I - J_0'(J_0'J_0)^{-1}J'_0 \), implying that \( MJ_1 = J_0 - J_0 = 0_{n \times (k+1)} \). This also implies that \( J'_0M = 0_{(k+1) \times n} \).

(ii) We note that \( J_1 = [0, Gt] \), so that \( MJ_1 = [0, MGt] = [0, MXd] = 0_{n \times (k+1)} \). This also implies that \( J'_1M = 0_{(k+1) \times n} \) and completes the proof.

Proof of Lemma 5: (i) When \( P(h; d) := Q(hd)V(h; d)Q(hd)' \), we note that \( \frac{\partial}{\partial h} P(0; d) = J_1H_0^{-1}J'_0 - J_0H_0^{-1}H_1H_0^{-1}J'_0 + J_0H_0^{-1}J'_1 \) by Lemma A2(ii). Given that \( H_1 = J'_0J_1 + J'_1J_0 \), if we plug this into \( \frac{\partial}{\partial h} L_n(0; d) \), then it follows that \( \frac{\partial}{\partial h} P(0; d) = 0 \). Finally, the consequence follows from the fact that \( \frac{\partial}{\partial h} L_n(0; d) = (Y - \alpha)(\gamma^*C' + U')\frac{\partial}{\partial h} P(0; d) (C\gamma^* + U) = 0 \).

(ii) Tediuous algebra using Lemma A2(ii) shows that \( \frac{\partial^2}{\partial h^2} P(0; d) = J_0H_0^{-1}J'_2M + MJ2H_0^{-1}J'_0 \). Thus, \( \frac{\partial^2}{\partial h^2} L_n(0; d) = (\gamma^*J'_0 + U')(J_0H_0^{-1}J'_2M + MJ2H_0^{-1}J'_0)(J_0\gamma^* + U) \). We now note that \( (a)J'_0\frac{\partial^2}{\partial h^2} P(0; d)J_0 = 0 \); (b) \( J'_0\frac{\partial^2}{\partial h^2} P(0; d)U = 4J'_2M \); and (c) \( U'\frac{\partial^2}{\partial h^2} P(0; d)U = 2U'J_0H_0^{-1}J'_2M \) after using Lemma A3(v) and the facts that \( H_1 = J'_0J_0 + J'_1J_1 \) and \( H_2 = J'_2J_0 + 2J'_1J_1 + J'_0J'_2 \). The desired result now follows from (a), (b), and (c).

(iii) Tediuous algebra using Lemma A2(iii) shows that \( \frac{\partial^3}{\partial h^3} P(0; d) = K + K' \), and we obtain that \( \frac{\partial^3}{\partial h^3} L_n(0; d) = (\gamma^*J'_0 + U')(K + K')(J_0\gamma^* + U) = \gamma^*J'_0(K + K')\gamma^* + 2\gamma^*J'_0(K + K')U + 2U'KU \). Given this, Lemma A3(ii) implies that \( 2\gamma^*J'_0(K + K')U + 2U'KU = o_P(n^{3/4}) \). Thus, \( \frac{\partial^3}{\partial h^3} L_n(0; d) = \gamma^*J'_0(K + K')\gamma^* + o_P(n^{3/4}) \).

(iv) First, by some algebra, it follows that \( \frac{\partial^4}{\partial h^4} P(0; d) = L + L' + J_0\frac{\partial^3}{\partial h^3} V(0; d)J'_0 \). Second, we note that \( \frac{\partial^4}{\partial h^4} L_n(0; d) = (\gamma^*J'_0 + U')\frac{\partial^3}{\partial h^3} P(0; d)(J_0\gamma^* + U) = \gamma^*J'_0\frac{\partial^3}{\partial h^3} P(0; d)J_0\gamma^* + 2U'\frac{\partial^3}{\partial h^3} P(0; d)J_0\gamma^* + U'\frac{\partial^3}{\partial h^3} P(0; d)U \). Third, further tedious algebra shows that \( J'_0\frac{\partial^3}{\partial h^3} P(0; d)J_0 = -6J'_2MJ_2 \) using the fact that \( \frac{\partial^4}{\partial h^4} P(0; d) = L + L' + J_0\frac{\partial^3}{\partial h^3} V(0; d)J'_0 \) and Lemma A2(iv). Finally, using Lemmas A3(iii), A1(i), A3(i), and A1(ii) shows that \( U'[L + L' + J_0\frac{\partial}{\partial h} V(0; d)J'_0]J_0 = O_P(n^{1/2}) = o_P(n) \); and using Lemmas A3(iii), A1(i), and A3(i) shows that \( U'[L + L' + J_0\frac{\partial}{\partial h} V(0; d)J'_0]U = O_P(1) = o_P(n) \). Therefore, \( \frac{\partial^4}{\partial h^4} L_n(0; d) = 6\gamma^*J'_2MJ_2\gamma^* + o_P(n) \). 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 $G_j := \text{diag} \{X_j\}$, then $\sum_{j=1}^k d_j G_j = \sum_{j=1}^k \text{diag} \{X_j d_j\} = \text{diag} \{X d\} = G$, so that

$$
\iota' G^2 MU = \iota' \sum_{i=1}^k \text{diag} \{X_i d_i\} \sum_{j=1}^k \text{diag} \{X_j d_j\} MU
$$

$$
= \iota' \sum_{i=1}^k \sum_{j=1}^k d_i d_j \text{diag} \{X_i\} \text{diag} \{X_j\} MU = \sum_{i=1}^k \sum_{j=1}^k d_i d_j \iota' D_{ij} MU = d' \tilde{M} d
$$

by noting that $\tilde{M} := [\iota' D_{ij} MU]$. Therefore, $\iota' G^2 MU = d' \tilde{M} d$. Next, we note that Lemmas A1(i $\sim$ iii) imply that $U' J_0 H_0^{-1} J_2' MU = U' J_0 H_0^{-1} J_2' U - U' J_0 H_0^{-1} J_2 J_0 H_0^{-1} J_0' U = O_P(1)$, because $U' J_0 = O_P(n^{1/2})$, $J_2' U = O_P(n^{1/2})$ by Lemma A1(i); $H_0^{-1} = V(0; d) = O_P(n^{-1})$ by Lemma A1(iii); and $J_2' J_0 = O_P(1)$ by Lemma A1(ii). From this, it trivially follows that $U' J_0 H_0^{-1} J_2' MU = o_P(n^{1/2})$. These two facts now show that $\frac{\bar{d}^2}{\bar{d}' h^2} L_n^3 (0; \alpha) = 2(\alpha^* - \alpha) (c_2/c_0) d' \tilde{M} d + o_P(n^{1/2})$.

(ii) We note that $H_0^{-1} J_0' J_0 = J_0' J_0 H_0^{-1} = I_k$ and $H_1 = J_1' J_0 + J_0' J_1$. Using these facts, we have that

$$
J_0' K J_0 = -3 [J_0' H_2 H_0^{-1} J_0' J_1 - J_0' J_1 H_0^{-1} H_1^{-1} J_0' J_1 + J_0' J_0 H_0^{-1} H_1^{-1} H_0^{-1} J_1 - J_0' J_0 H_0^{-1} H_2 + J_1' J_2],
$$

so that exploiting the fact that $H_1 = J_1' J_0 + J_0' J_1$ and $H_2 = J_2' J_0 + J_1' J_0 + J_0' J_2$ yields that

$$
J_0' (K + K') J_0 = 6 J_1' J_0 H_0^{-1} J_1' MJ_1 + J_1' M J_1 H_0^{-1} J_0' J_1 - 3 J_1' MJ_2 - 3 J_2' MJ_1.
$$

Finally, we now note that $J_0' M = MJ_1 = 0$ by Lemma A3(iv), implying that $J_0' (K + K') J_0 = 0$. The desired result follows from this.

(iii) We now note that $\gamma^* J_2' MJ_2 \gamma^* = (\alpha^* - \alpha)^2 (c_2/c_0)^2 \iota' G^2 MG^2 \iota$. Thus, if $\iota' G^2 MG^2 \iota = d' (I_k \otimes d)' W (I_k \otimes d) d$, then the desired result follows. We derive this. First, we note that

$$
d' (I_k \otimes d)' W (I_k \otimes d) d = \sum_{i=1}^k \sum_{j=1}^k d_i (d' W_{ij} d) d_j,
$$

and that $d' W_{ij} d = \sum_{\ell=1}^k \sum_{m=1}^k \iota' G_\ell G_j M d_\ell d_m G_\ell G_m t = \iota' G_1 G_j M \sum_{\ell=1}^k d_\ell G_\ell \sum_{m=1}^k d_m G_m t = \iota' G_1 G_j MG^2 t$. Next, we also note that

$$
\sum_{i=1}^k \sum_{j=1}^k d_i (d' W_{ij} d) d_j = \sum_{i=1}^k \sum_{j=1}^k d_i \iota' G_1 G_j MG^2 i d_j = \iota' \sum_{i=1}^k d_i G_i \sum_{j=1}^k d_j G_j MG^2 t = \iota' G^2 MG^2 t.
$$

That is, $d' (I_k \otimes d)' W (I_k \otimes d) d = \iota' G^2 MG^2 t$. This completes the proof.
Proof of Lemma 7: (ii) For given \( n \), we obtain

\[
N_{n}^{(4)}(h, d) = 8U'\mathbf{M}\{(\partial / \partial h)\Psi(h, d)\}U'\mathbf{M}\{\left(\partial^{2} / \partial h^{2}\right)\Psi(h, d)\} \\
+ 6\left\{ U'\mathbf{M}\left(\partial^{2} / \partial h^{2}\right)\Psi(h, d)\right\}^{2} + 2U'\mathbf{M}\Psi(h, d)U'\mathbf{M}\left\{ \left(\partial^{4} / \partial h^{4}\right)\Psi(h, d)\right\}.
\]

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

\[
\lim_{\bar{h} \downarrow 0} N_{n}^{(4)}(h, d) = \lim_{\bar{h} \downarrow 0} 6\left\{ U'\mathbf{M}\left(\partial^{2} / \partial h^{2}\right)\Psi(h, d)\right\}^{2} = 6c_2^2 \left\{ \sum_{i=1}^{k} \sum_{j=1}^{k} \mathbf{t}'D_{ij}\mathbf{M}d_{i}d_{j} \right\}^{2} \quad \text{a.s.}
\]

(iii) Similarly, for given \( d \) and \( n \), we obtain that

\[
D_{n}^{(4)}(h, d) = 6\left\{ \left(\partial^{2} / \partial h^{2}\right)\Psi(h, d)\right\}^{\prime} \mathbf{M}\left\{ \left(\partial^{2} / \partial h^{2}\right)\Psi(h, d)\right\} \\
+ 8\left\{ (\partial / \partial h)\Psi(h, d)\right\}^{\prime} \mathbf{M}\left\{ \left(\partial^{3} / \partial h^{3}\right)\Psi(h, d)\right\} \\
+ 2\Psi(h, d)^{\prime} \mathbf{M}\left\{ \left(\partial^{4} / \partial h^{4}\right)\Psi(h, d)\right\}.
\]

Also, we note that \( \lim_{\bar{h} \downarrow 0} \Psi(h, d)^{\prime} \mathbf{M} = c_0 \mathbf{t}' \mathbf{M} = 0 \) a.s., and \( \lim_{\bar{h} \downarrow 0} \{ (\partial / \partial h)\Psi(h, d)\}^{\prime} \mathbf{M} = c_1 \mathbf{d}' \mathbf{X}' \mathbf{M} = 0 \) a.s., as we saw in the proof of Lemma 7(ii). Therefore,

\[
\lim_{\bar{h} \downarrow 0} D_{n}^{(4)}(h, d) = \lim_{\bar{h} \downarrow 0} 6\left\{ \left(\partial^{2} / \partial h^{2}\right)\Psi(h, d)\right\}^{\prime} \mathbf{M}\left\{ \left(\partial^{2} / \partial h^{2}\right)\Psi(h, d)\right\} \\
= 6c_2^2 \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{\ell=1}^{k} \sum_{m=1}^{k} \mathbf{t}'D_{ij}\mathbf{M}\ell_{\ell_{m}}d_{i}d_{j}d_{\ell}d_{m} \quad \text{a.s.}
\]

This completes the proof.

Before proving Lemma 8, we simplify our notation by suppressing all function arguments but \( h \) and \( \bar{h} \). That is, we let \( \mathcal{J}(h) \) and \( \mathcal{T}(h, \bar{h}) \) denote \( \mathcal{J}(h, d, h, d) \) and \( \mathcal{T}(h, d, \bar{h}, \bar{d}) \) 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 \( \bar{h} \geq 0, \) \( \lim_{\bar{h} \downarrow 0} \mathcal{T}(\ell, 0)(h, \bar{h}) = 0 \), where \( \mathcal{T}(\ell, m)(h, \bar{h}) := (\partial^{m} \partial^{\ell} / \partial h^{m} \partial \bar{h}^{\ell}) \)

\( \mathcal{T}(h, \bar{h}) \);

(ii) for \( \ell = 0, 1, 2, \) and \( 3, \lim_{\bar{h} \downarrow 0} \mathcal{J}(\ell)(h) = 0 \), where \( \mathcal{J}(\ell)(h) := (\partial^{\ell} / \partial h^{\ell})\mathcal{J}(h) \);
(iii) $T^{(2,0)}(0, \tilde{h}) = c_2 \mathcal{H}(b, \tilde{h}, \tilde{b})$; and
(iv) $\mathcal{J}^{(4)}(0) = 6c_2^2 \mathcal{I}(b, b)$.

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

(i) $\lim_{h \to 0} \lim_{h \to 0} T(h, \tilde{h}) = 0$;
(ii) for $\ell = 0$ and 1, $\lim_{h \to 0} \lim_{h \to 0} T^{(1, \ell)}(h, \tilde{h}) = 0$;
(iii) for $\ell = 0$ and 1, $\lim_{h \to 0} \lim_{h \to 0} T^{(2, \ell)}(h, \tilde{h}) = 0$;
(iv) for $\ell = 0$ and 1, $\lim_{h \to 0} \lim_{h \to 0} T^{(3, \ell)}(h, \tilde{h}) = 0$;
(v) $\lim_{h \to 0} \lim_{h \to 0} T^{(4,0)}(h, \tilde{h}) = 0$; and
(vi) $\lim_{h \to 0} \lim_{h \to 0} T^{(2,2)}(h, \tilde{h}) = c_2^2 \mathcal{K}(b, \tilde{b})$.

Proof of Lemma 8: (i) Given the definition of $\rho_1$,

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

Lemma B1(i and ii) implies that $\lim_{h \to 0} T(h, \tilde{h}) = 0$ and $\lim_{h \to 0} \mathcal{J}(h) = 0$. Therefore, we apply L’Hospital’s rule. We note that

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

(24)

by Lemma B1(i and iii) and also that

$$
\mathcal{J}(h) = \mathcal{J}(0) + \mathcal{J}^{(1)}(0)h + \frac{1}{2} \mathcal{J}^{(2)}(0)h^2 + \frac{1}{3!} \mathcal{J}^{(3)}(0)h^3 + \frac{1}{4!} \mathcal{J}^{(4)}(0)h^4 + o(h^4)
$$

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

(25)

by Lemma B1(ii and iv). Hence,

$$
\lim_{h \to 0} \frac{T(h, \tilde{h})}{\{\mathcal{J}(h)\}^{1/2} \{\mathcal{J}(\tilde{h})\}^{1/2}} = \frac{c_2 \mathcal{H}(b, \tilde{h}, \tilde{b})}{\{c_2^2 \mathcal{I}(b, b)\}^{1/2} \{\mathcal{J}(h, \tilde{b})\}^{1/2}} = \frac{\text{sgn}(c_2) \mathcal{H}(b, \tilde{h}, \tilde{d})}{\{\mathcal{I}(b, b)\}^{1/2} \{\mathcal{J}(h, \tilde{b})\}^{1/2}}.
$$

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

$$
T(h, \tilde{h}) = T(0, 0) + \sum_{i=1}^{4} \sum_{j=0}^{i} \frac{1}{i!} \binom{i}{j} T^{(i-j,j)}(0, 0) h^{i-j} \tilde{h}^j + o((h^2 + \tilde{h}^2)^2)
$$

$$
= \frac{1}{4} c_2^2 \mathcal{K}(b, \tilde{b})h^2 \tilde{h}^2 + o((h^2 + \tilde{h}^2)^2),
$$

42
where the first and second equalities hold by Lemmas B2(i) and B2(ii∼v), respectively. Also,

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

by (25). Thus,

\[ \lim_{h \downarrow 0} \lim_{h \downarrow 0} \frac{T(h, \tilde{h})}{\{J(h)\}^{1/2}\{J(\hat{h})\}^{1/2}} = \frac{c_2^2 K(b, \tilde{b})}{\{c_2^2 I(b, b)\}^{1/2}\{c_2^2 I(\hat{b}, \tilde{b})\}^{1/2}} = \frac{K(b, \tilde{b})}{\{I(b, b)\}^{1/2}\{I(\hat{b}, \tilde{b})\}^{1/2}}. \]

Note that this is \( \rho_2(b, \tilde{b}) \), and this completes the proof.

References

[1]


Econometrica, 69, 683–734.


BARTLETT, M. (1953b): “Approximate Confidence Intervals II. More Than One Unknown Parameter,”


BIERENS, H. (1990): “A Consistent Conditional Moment Test of Functional Form,” Econometrica, 58,

1443–1458.


### Table 1: Critical Values

**Number of Replications:** 50,000

DGP: \( Y_t = 0.5Y_{t-1} + U_t \) and \( U_t \sim \text{IID } N(0, 1) \)

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 \( K = 150 \)

<table>
<thead>
<tr>
<th>Nominal Level \ \Delta</th>
<th>( \Delta_{0.5} )</th>
<th>( \Delta_{1.0} )</th>
<th>( \Delta_{1.5} )</th>
<th>( \Delta_{2.0} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 %</td>
<td>7.7974</td>
<td>8.4051</td>
<td>9.1206</td>
<td>9.7248</td>
</tr>
<tr>
<td>5.00 %</td>
<td>4.7399</td>
<td>5.4245</td>
<td>6.0594</td>
<td>6.6222</td>
</tr>
<tr>
<td>10.0 %</td>
<td>3.4747</td>
<td>4.1282</td>
<td>4.6833</td>
<td>5.2558</td>
</tr>
</tbody>
</table>

### Table 2: Empirical Rejection Rates (in Percent)

**Number of Replications:** 10,000

DGP: \( Y_t = 0.5Y_{t-1} + U_t \) and \( U_t \sim \text{IID } N(0, 1) \)

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 \( K = 150 \)

<table>
<thead>
<tr>
<th>Asymptotic Distribution</th>
<th>( \text{QLR}(\Delta_{0.5}; K) )</th>
<th>( \text{QLR}(\Delta_{1.0}; K) )</th>
<th>( \text{QLR}(\Delta_{1.5}; K) )</th>
<th>( \text{QLR}(\Delta_{2.0}; K) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal Level \ \Sample Size</td>
<td>50 100 200 500 1,000 2,000 5,000</td>
<td>50 100 200 500 1,000 2,000 5,000</td>
<td>50 100 200 500 1,000 2,000 5,000</td>
<td>50 100 200 500 1,000 2,000 5,000</td>
</tr>
<tr>
<td>1.00 % \ \Sample Size</td>
<td>0.26 0.36 0.60 0.64 0.83 0.90 0.97</td>
<td>0.37 0.44 0.61 0.72 0.84 0.79 0.92</td>
<td>0.19 0.28 0.37 0.49 0.80 0.65 0.82</td>
<td>0.02 0.36 0.47 0.50 0.49 0.50 0.71</td>
</tr>
<tr>
<td>5.00 % \ \Sample Size</td>
<td>2.38 3.00 3.53 4.57 4.80 4.70 5.05</td>
<td>2.45 2.82 3.31 3.66 4.12 4.01 4.49</td>
<td>1.64 2.19 2.59 3.08 3.82 3.78 4.09</td>
<td>1.48 1.64 2.21 2.67 2.50 2.79 3.56</td>
</tr>
<tr>
<td>10.0 % \ \Sample Size</td>
<td>6.00 6.87 8.20 9.13 9.49 9.25 9.79</td>
<td>5.68 6.11 7.12 8.06 8.32 8.41 8.98</td>
<td>4.23 5.12 5.74 6.81 7.73 8.06 8.53</td>
<td>3.35 3.78 4.50 5.47 5.58 5.98 7.27</td>
</tr>
</tbody>
</table>
Table 3: **Empirical Rejection Rates (in Percent)**

**Number of Replications:** 10,000

**DGP:** $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim \text{IID } N(0, 1)$

**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 $K = 150$

<table>
<thead>
<tr>
<th>Asymptotic Distribution</th>
<th>$QLR_n(\Delta_{0.5}; K)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nominal Level \ Sample Size</strong></td>
<td>50</td>
</tr>
<tr>
<td>1.00 %</td>
<td>0.71</td>
</tr>
<tr>
<td>5.00 %</td>
<td>2.43</td>
</tr>
<tr>
<td>10.0 %</td>
<td>6.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Asymptotic Distribution</th>
<th>$QLR_n(\Delta_{1.0}; K)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nominal Level \ Sample Size</strong></td>
<td>50</td>
</tr>
<tr>
<td>1.00 %</td>
<td>0.30</td>
</tr>
<tr>
<td>5.00 %</td>
<td>2.44</td>
</tr>
<tr>
<td>10.0 %</td>
<td>5.39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Asymptotic Distribution</th>
<th>$QLR_n(\Delta_{1.5}; K)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nominal Level \ Sample Size</strong></td>
<td>50</td>
</tr>
<tr>
<td>1.00 %</td>
<td>0.12</td>
</tr>
<tr>
<td>5.00 %</td>
<td>1.33</td>
</tr>
<tr>
<td>10.0 %</td>
<td>3.13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Asymptotic Distribution</th>
<th>$QLR_n(\Delta_{2.0}; K)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nominal Level \ Sample Size</strong></td>
<td>50</td>
</tr>
<tr>
<td>1.00 %</td>
<td>0.15</td>
</tr>
<tr>
<td>5.00 %</td>
<td>1.26</td>
</tr>
<tr>
<td>10.0 %</td>
<td>2.73</td>
</tr>
</tbody>
</table>

Table 4: **Empirical Rejection Rates (in Percent)**

**Number of Replications:** 4,000

**DGP:** $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim \text{IID } N(0, 1)$

**Model:** $Y_t = \alpha + \beta Y_{t-1} + \lambda \exp(\delta Y_{t-1}) + U_t, \Delta_{0.5} = [-0.5, 0.5]$

<table>
<thead>
<tr>
<th>Nominal Level \ Sample Size</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>1,000</th>
<th>2,000</th>
<th>4,000</th>
<th>5,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 %</td>
<td>0.35</td>
<td>0.45</td>
<td>0.75</td>
<td>0.85</td>
<td>1.15</td>
<td>0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.00 %</td>
<td>2.55</td>
<td>3.62</td>
<td>4.10</td>
<td>4.67</td>
<td>5.20</td>
<td>5.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.0 %</td>
<td>6.57</td>
<td>8.30</td>
<td>8.57</td>
<td>8.97</td>
<td>9.50</td>
<td>10.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Model:** $Y_t = \alpha + \beta Y_{t-1} + \lambda(1 + \exp(\delta Y_{t-1})) + U_t, \Delta_{0.5} = [-0.5, 0.5]$

<table>
<thead>
<tr>
<th>Nominal Level \ Sample Size</th>
<th>200</th>
<th>500</th>
<th>1,000</th>
<th>2,000</th>
<th>4,000</th>
<th>6,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00 %</td>
<td>3.52</td>
<td>3.52</td>
<td>3.22</td>
<td>3.02</td>
<td>2.12</td>
<td>1.60</td>
</tr>
<tr>
<td>5.00 %</td>
<td>5.77</td>
<td>6.37</td>
<td>5.25</td>
<td>4.55</td>
<td>3.60</td>
<td>3.10</td>
</tr>
<tr>
<td>10.0 %</td>
<td>7.62</td>
<td>7.65</td>
<td>6.42</td>
<td>5.62</td>
<td>5.20</td>
<td>5.70</td>
</tr>
</tbody>
</table>

48
Figure 1: **Empirical Null Distributions of the QLR Statistics**

**Number of Replications:** 10,000

**DGP:** \( Y_t = 0.5Y_{t-1} + U_t \) and \( U_t \sim \text{iid } \mathcal{N}(0,1) \)

**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 \( K = 150 \)

QLR(\( \Delta_{0.5}; K \))

QLR(\( \Delta_{1.0}; K \))

QLR(\( \Delta_{1.5}; K \))

QLR(\( \Delta_{2.0}; K \))

\( 49 \)
Figure 2: **Empirical Density Functions of the QLR Statistics**

**Number of Replications:** 10,000

**DGP:** $Y_t = 0.5Y_{t-1} + U_t$ and $U_t \sim \text{IID } N(0, 1)$

**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 $K = 150$

QLR($\Delta_{0.5}; K$)  
QLR($\Delta_{1.0}; K$)  
QLR($\Delta_{1.5}; K$)  
QLR($\Delta_{2.0}; K$)
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: $Y_t = \alpha + \beta Y_{t-1} + \lambda \Psi_t(\delta) + U_t$ and $\Delta_{0.5} = [-0.5, 0.5]$

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

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