Higher-Order Approximations for Testing Neglected Nonlinearity

Halbert White

hwhite@weber.ucsd.edu

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

Jin Seo Cho

jinseocho@yonsei.ac.kr

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

Keywords: Artificial neural networks, logistic function, sixth-order approximation, quasi-likelihood ratio test, Lagrange multiplier test

Abstract

We illustrate the need to use higher-order (specifically sixth-order) expansions in order to properly determine the asymptotic distribution of a standard artificial neural network test for neglected nonlinearity. The test statistic is a Quasi-Likelihood Ratio (QLR) statistic designed to test whether the mean square prediction error improves by including an additional hidden unit with an activation function violating the "no-zero" condition in Cho, Ishida, and White (2011). This statistic is also shown to be asymptotically equivalent under the null to the Lagrange multiplier (LM) statistic of Luukkenon, Saikkonen, and Teräsvirta (1988) and Teräsvirta (1994). In addition, we compare the power properties of our QLR test to one satisfying the no-zero condition and find that

the latter is not consistent for detecting a DGP with neglected nonlinearity violating an analogous no-zero condition, whereas our QLR test is consistent.

1 Introduction

In analyzing the first-order asymptotic behavior of likelihood-based test statistics such as the Wald, Lagrange multiplier, or (quasi-) likelihood ratio (QLR) statistics, it is usually adequate to work with second-order (quadratic) approximations to the log-likelihood function. As Phillips (2011) recently notes in a related context, however, such approximations are not always adequate for first-order asymptotics, and scholars going back at least to Cramér (1946) have given careful attention to cases where higher-order approximations are required. For example, Bartlett (1953a, 1953b) analyzes models requiring higher-order approximation, and McCullagh (1984, 1987) provides a framework for this using tensor analysis. McCullagh (1986), Carrasco, Hu, and Ploberger (2004), and Cho and White (2007) also apply higher-order expansions to a variety of interesting models to obtain first-order asymptotics.

Recently, Cho, Ishida, and White (2011) showed that QLR tests for neglected nonlinearity based on artificial neural networks (ANNs) cannot be analyzed using quadratic approximation, and they provide conditions under which a quartic (fourth-order) approximation yields the desired first-order asymptotics. Nevertheless, they also discuss the fact that cases violating their assumption A7 ("no zero") require the use of even higher-order approximations to obtain the first-order asymptotics for the QLR statistic. In particular, they show how constructing the test using a hidden unit with logistic activation function – a standard choice in the ANN literature – violates A7. At present, the conditions yielding first-order asymptotics for the QLR statistic with this standard choice are unknown. Nor is it satisfactory simply to rule out such cases.

The goal of this study is to gain a deeper understanding of the asymptotic behavior of ANN-based QLR tests for neglected nonlinearity when Cho, Ishida, and White's (2011) no-zero assumption is violated. In doing so, we illustrate the use of higher-order, specifically sixth-order, expansions to obtain first-order asymptotics. Although Cho, Ishida, and White (2011) obtain the asymptotic distribution of their QLR statistic by explicitly treating the two-fold identification problem that arises in this approach to testing neglected nonlinearity, for conciseness we restrict our focus here to analyzing the QLR statistic when there is only a single source of identification failure under the null. We leave the two-fold identification problem to other work.

The plan of this paper is as follows. In Section 2, we introduce a simple ANN model employing a single hidden unit violating the "no-zero" condition, and we analyze a QLR statistic designed to test neglected nonlinearity using this model. Section 3 contains Monte Carlo simulations; these corroborate the results of Section 2 and provide insight into hidden unit selection. Section 4 contains a summary and concluding remarks

2 A QLR test for neglected nonlinearity

We begin by specifying the same data generating process (DGP) assumed by Cho, Ishida, and White (2011).

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

We note that $\mathbf{X}_t$ may contain lagged values of Y_t, as well as nonlinear transformations of these lags or of other underlying variables.

Next, we specify a model for $E[Y_t|\mathbf{X}_t]$. For this, we let $X_{t,1}$ denote the first element of X_t.

Assumption 2 (Model) *Let* $f(\mathbf{X}_t\,; \alpha, \beta, \lambda, \delta) := \alpha + \mathbf{X}_t'\boldsymbol{\beta} + \lambda\Psi(X_{t,1}\delta)$, *and define the model* $\mathcal{M}$ *as*

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

where $A \subset \mathbb{R}$, $\mathbf{B} \subset \mathbb{R}^k$, $\Lambda \subset \mathbb{R}$, *and* $\Delta \subset \mathbb{R}$ *are non-empty compact sets, with* $0 \in \Delta$, *and* $\Psi : \mathbb{R} \mapsto \mathbb{R}$ *is an analytic function with* $c_2 = 0$ *and* $c_3 \neq 0$, *where* $c_j := \frac{\partial^j}{\partial x^j}\Psi(x)|_{(x=0)}$ $j = 2, 3, \ldots$.

This model provides a natural framework in which to test for neglected nonlinearity with respect to $X_{t,1}$. We consider only a single element of $\mathbf{X}_t$ appearing inside Ψ for simplicity. It is straightforward to treat the case where $\Psi(X_{t,1}\delta)$ is replaced by $\Psi(\mathbf{X}_t'\boldsymbol{\delta})$, but the notation required to handle this case becomes extremely cumbersome. Many hidden unit functions satisfy the conditions in Assumption 2. For example, if $\Psi(\cdot)$ is the logistic function, so that $\Psi(x) := \{1 + \exp(x)\}^{-1}$, then $c_2 = 0$ but $c_3 \neq 0$. In addition, $\sin(x)$, $\arctan(x)$, $\sin[\arctan(x)]$, etc., satisfy Assumption 2.

Now consider testing the linearity of conditional expectation: for some $\alpha_* \in A$ and $\boldsymbol{\beta}_* \in \mathbf{B}$, $E[Y_t|\mathbf{X}_t] = \alpha_* + \mathbf{X}_t'\beta_*$. When linearity of $E[Y_t|\mathbf{X}_t]$ holds, the pseudo-true values λ_* and δ_* satisfy $\lambda_* = 0$ or $\delta_* = 0$, implying the presence of parameters not identified under the null.

Letting $\Psi_t(\delta) = \Psi(X_{t,1}\delta)$, we define the QLR statistic for neglected nonlinearity using the quasi-log likelihood (QL):

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

As Cho, Ishida, and White (2011) show, different orders of expansion are required when testing $\lambda_* = 0$ than when testing $\delta_* = 0$. A quadratic expansion is sufficient for testing $\lambda_* = 0$ when $\delta \neq 0$ (e.g., Hansen (1996)), whereas a quartic approximation is needed for testing $\delta_* = 0$, under regularity conditions provided by Cho, Ishida, and White (2011). The most critical condition is the no-zero condition (assumption A7), which states that $c_2 \neq 0$. Without this, the quartic expansion fails. The model of Assumption 2 violates this condition, so Cho, Ishida, and White's (2011) results do not apply. Further, as their simulations show, the asymptotic distribution obtained when the no-zero condition holds does not provide a useful approximation to the required distribution when the no-zero condition fails.

We analyze the QLR statistic under $\mathbb{H}_0 : \delta_* = 0$ by adapting the approach in Cho, Ishida, and White (2011). As it turns out, a sixth-order Taylor expansion suffices. To verify this, we first concentrate the QL with respect to α and $\boldsymbol{\beta}$, obtaining

$$L_n(\delta; \lambda) := -[\mathbf{Y} - \lambda\boldsymbol{\Psi}(\delta)]'\mathbf{M}[\mathbf{Y} - \lambda\boldsymbol{\Psi}(\delta)], \tag{1}$$

where $\mathbf{Y} := [Y_1, Y_2, \ldots, Y_n]'$, $\boldsymbol{\Psi}(\delta) := [\Psi_1(\delta), \Psi_2(\delta), \ldots, \Psi_n(\delta)]'$, $\mathbf{M} := \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, $\mathbf{Z} := (\boldsymbol{\iota}, \mathbf{X})$ with $\boldsymbol{\iota}$ the $n \times 1$ vector of ones, and $\mathbf{X} := [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n]'$. The QLR statistic for testing $\delta_* = 0$ is then

$$\mathcal{QLR}_n := \sup_{\lambda \in \Lambda} QLR_n(\lambda) := \sup_{\lambda \in \Lambda} n \left\{ \frac{L_n(0;\lambda) - \sup_{\delta \in \Delta} L_n(\delta;\lambda)}{L_n(0;\lambda)} \right\}.$$

Approximating the concentrated QL, eq.(1), by a Taylor expansion around $\delta_* = 0$ requires the following partial derivatives at $\delta_* = 0$:

- $L_n^{(1)}(0;\lambda) := \frac{\partial}{\partial \delta} L_n(0;\lambda) = 0$;
- $L_n^{(2)}(0;\lambda) := \frac{\partial^2}{\partial \delta^2} L_n(0;\lambda) = 0$;
- $L_n^{(3)}(0;\lambda) := \frac{\partial^3}{\partial \delta^3} L_n(0;\lambda) = \frac{1}{4}\lambda \boldsymbol{\iota}' \mathbf{D}_3 \mathbf{M} \mathbf{U}$;
- $L_n^{(4)}(0;\lambda) := \frac{\partial^4}{\partial \delta^4} L_n(0;\lambda) = 0$;
- $L_n^{(5)}(0;\lambda) := \frac{\partial^5}{\partial \delta^5} L_n(0;\lambda) = -\frac{1}{2}\lambda \boldsymbol{\iota}' \mathbf{D}_5 \mathbf{M} \mathbf{U}$; and
- $L_n^{(6)}(0;\lambda) := \frac{\partial^6}{\partial \delta^6} L_n(0;\lambda) = -\frac{5}{16}\lambda^2 \boldsymbol{\iota}' \mathbf{D}_3 \mathbf{M} \mathbf{D}_3 \boldsymbol{\iota}$,

where $\mathbf{U} := [U_1, U_2, \ldots, U_n]'$ with $U_t := Y_t - E[Y_t|\mathbf{X}_t]$, and $\mathbf{D}_m$, the "power matrix" of order m, is $\mathbf{D}_m := \text{diag}\{X_{1,1}^m, X_{2,1}^m, ..., X_{n,1}^m\}$ for $m = 3, 5$. Here, $L_n^{(1)}(0;\lambda) = 0$ and $L_n^{(2)}(0;\lambda) = 0$, whereas Cho, Ishida, and White's (2011) no-zero condition gives $L_n^{(1)}(0;\lambda) = 0$ and $L_n^{(2)}(0;\lambda) \neq 0$. This permits them to use $L_n^{(2)}(0;\lambda)$ as the key term determining the asymptotic distribution, but this is not possible here. Instead, $L_n^{(3)}(0;\lambda)$ now plays the key role, mainly because $c_3 \neq 0$. The sixth-order Taylor expansion at $\delta_* = 0$ is then

$$L_n(\delta;\lambda) - L_n(0;\lambda) = \frac{1}{3!} L_n^{(3)}(0;\lambda)\,\delta^3 + \frac{1}{5!} L_n^{(5)}(0;\lambda)\,\delta^5 + \frac{1}{6!} L_n^{(6)}(0;\lambda)\,\delta^6 + o_{\mathbb{P}}(1).$$

Before examining the asymptotic behavior of the terms on the right, we impose the following regularity conditions:

Assumption 3 (Moments) $E|U_t^2 X_{t,1}^6| < \infty$, $E|U_t^2 X_{t,1}^{10}| < \infty$, *and for* $j > 1$, $E|U_t^2 X_{t,j}^2| < \infty$.

Assumption 4 (MDS) $E[U_t|\mathbf{X}_t, U_{t-1}, \mathbf{X}_{t-1}, ...] = 0$.

Assumption 5 (Covariance) $\det[E[U_t^2 \widetilde{\mathbf{Z}}_t \widetilde{\mathbf{Z}}_t']] > 0$ *and* $\det[E[\widetilde{\mathbf{Z}}_t \widetilde{\mathbf{Z}}_t']] > 0$, *where* $\widetilde{\mathbf{Z}}_t := (\mathbf{Z}_t', X_{t,1}^3)'$ *and* $\mathbf{Z}_t := (1, \mathbf{X}_t')'$.

These conditions ensure the regular asymptotic behavior of each derivative term. In particular, Assumption 3 holds by the Cauchy-Schwarz inequality if $E|U_t|^4 < \infty$, $E|X_{t,1}|^{20} < \infty$, and for $j > 1$, $E|X_{t,j}|^4 < \infty$. Also, the ergodic theorem and the central limit theorem for martingale difference sequences ensure that

- $n^{-1}L_n(0;\lambda) \overset{\text{a.s.}}{\to} -\sigma_*^2$;
- $Z_{n,3} := n^{-1/2}\boldsymbol{\iota}'\mathbf{D}_3\mathbf{M}\mathbf{U} \Rightarrow \mathcal{Z}_3 \sim N(0,\tau_3^*)$;
- $Z_{n,5} := n^{-1/2}\boldsymbol{\iota}'\mathbf{D}_5\mathbf{M}\mathbf{U} \Rightarrow \mathcal{Z}_5 \sim N(0,\tau_5^*)$; and
- $W_{n,6} := n^{-1}\boldsymbol{\iota}'\mathbf{D}_3\mathbf{M}\mathbf{D}_3\boldsymbol{\iota} \overset{\text{a.s.}}{\to} \xi_3^*$,

where $\sigma_*^2 := E[U_t^2]$; for $m = 3$ and 5,

$$\tau_m^* := E[U_t^2 X_{t,1}^{2m}] - 2E[U_t^2 X_{t,1}^m \mathbf{Z}_t']E[\mathbf{Z}_t\mathbf{Z}_t']^{-1}E[X_{t,1}^m\mathbf{Z}_t] \\ + E[X_{t,1}^m\mathbf{Z}_t']E[\mathbf{Z}_t\mathbf{Z}_t']^{-1}E[U_t^2\mathbf{Z}_t\mathbf{Z}_t']E[\mathbf{Z}_t\mathbf{Z}_t']^{-1}E[X_{t,1}^m\mathbf{Z}_t];$$

and $\xi_m^* := E[X_{t,1}^{2m}] - E[X_{t,1}^m\mathbf{Z}_t']E[\mathbf{Z}_t\mathbf{Z}_t']^{-1}E[X_{t,1}^m\mathbf{Z}_t]$. As these results are elementary, we do not prove them.

Substituting appropriately gives

$$L_n(\delta;\lambda) - L_n(0;\lambda) \\ = \frac{\lambda}{3!4}Z_{n,3}\{n^{1/6}\delta\}^3 - \frac{\lambda}{5!2}\cdot\frac{Z_{n,5}}{n^{1/3}}\{n^{1/6}\delta\}^5 - \frac{5\lambda^2}{6!16}W_{n,6}\{n^{1/6}\delta\}^6 + o_{\mathbb{P}}(1).$$

The numerator of the QLR statistic is the opposite of $\sup_{\delta\in\Delta}\{L_n(\delta;\lambda) - L_n(0;\lambda)\}$. Letting $\mathcal{D}_n := n^{1/6}\delta$ and maximizing the non-vanishing expression on the right above gives first order conditions

$$3\frac{\lambda}{3!4}Z_{n,3}\,\mathcal{D}_n^2 - 5\frac{\lambda}{5!2}\cdot\frac{Z_{n,5}}{n^{1/3}}\,\mathcal{D}_n^4 - 6\frac{5\lambda^2}{6!16}W_{n,6}\,\mathcal{D}_n^5 = 0.$$

The solution $\mathcal{D}_n = 0$ gives the minimum, so we have $\mathcal{D}_n^2 > 0$, and we can divide both sides by $\mathcal{D}_n^2$ to obtain

$$3\frac{\lambda}{3!4}Z_{n,3} - 5\frac{\lambda}{5!2}\cdot\frac{Z_{n,5}}{n^{1/3}}\,\mathcal{D}_n^2 - 6\frac{5\lambda^2}{6!16}W_{n,6}\,\mathcal{D}_n^3 = 0.$$

This is a cubic equation in $\mathcal{D}_n$.

Inspecting the cubic discriminant and noting that $n^{-1/3}Z_{n,5} = o_{\mathbb{P}}(1)$, we find that with probability approaching one, this equation has one real root. This root, say $\hat{\mathcal{D}}_n$, is a continuous function of $Z_{n,3}$ and $W_{n,6}$, both of which converge in distribution and are thus $O_{\mathbb{P}}(1)$. It follows that $\hat{\mathcal{D}}_n \Rightarrow \mathcal{D}_*$, say, and that $\hat{\mathcal{D}}_n = O_{\mathbb{P}}(1)$. We therefore have

$$\sup_{\delta\in\Delta}\{L_n(\delta;\lambda) - L_n(0;\lambda)\} \Rightarrow \frac{\lambda}{3!4}\mathcal{Z}_3\,\mathcal{D}_*^3 - \frac{5\lambda^2}{6!16}\xi_3^*\,\mathcal{D}_*^6 = \sup_{\mathcal{D}}\frac{\lambda}{3!4}\mathcal{Z}_3\,\mathcal{D}^3 - \frac{5\lambda^2}{6!16}\xi_3^*\,\mathcal{D}^6.$$

From the final equality, it is straightforward to verify that

$$\mathcal{D}_*^3 := \left(\frac{48}{\xi_3^*\lambda}\right)\mathcal{Z}_3 \sim N\left[0, \tau_3^*\left(\frac{48}{\xi_3^*\lambda}\right)^2\right].$$

Thus, it follows that

$$\sup_{\delta\in\Delta}\{L_n(\delta;\lambda) - L_n(0;\lambda)\} = \frac{Z_{n,3}^2}{W_{n,6}} + o_{\mathbb{P}}(1) \Rightarrow \frac{\lambda}{3!4}\mathcal{Z}_3\mathcal{D}_*^3 - \frac{5\lambda^2}{6!16}\xi_3^*\mathcal{D}_*^6 = \frac{\mathcal{Z}_3^2}{\xi_3^*}.$$

Observe that the unidentified parameter λ cancels out, so the asymptotic null distribution is free of λ, implying that Davies's (1977, 1987) identification problem does not arise.

We offer the following remarks. First, by the definition of the QLR statistic, its asymptotic null behavior is given by

$$\mathcal{QLR}_n := \sup_{\lambda\in\Lambda} QLR_n(\lambda) \Rightarrow \left(\frac{\mathcal{Z}_3^2}{\sigma_*^2\xi_3^*}\right).$$

In contrast, under their no-zero condition, Cho, Ishida, and White (2011) obtain the square of the half-normal distribution as the limiting distribution, implying that the QLR statistic has a probability mass at zero. Our result is different from theirs, because $\mathcal{D}_*$ captures the asymptotic behavior of $n^{1/6}\hat{\delta}_n$ under the null, where $\hat{\delta}_n$ is the nonlinear least squares (NLS) estimator. This enables the QLR test to have a continuous distribution under the null. When the no-zero condition of Cho, Ishida, and White (2011) holds, the NLS estimator is squared, leading to the square of the half-normal distribution. Second, we see that under the null, the convergence speed of the NLS estimator is quite slow; specifically, it is $n^{1/6}$. This does not necessarily imply that testing for neglected nonlinearity using the model considered here is inferior to the $c_2 \neq 0$ case. Although we have a slower convergence rate than the $n^{1/4}$ rate in Cho, Ishida, and White (2011), if the neglected nonlinearity has $c_2 = 0$ and $c_3 \neq 0$, the QLR test constructed using an activation function with $c_2 \neq 0$ may not have power, because such tests neglect the higher order terms needed to detect $c_3 \neq 0$ alternatives. On the other hand, if a QLR test constructed using an activation function with $c_2 = 0$ and $c_3 \neq 0$ is applied to a $c_2 \neq 0$ nonlinearity, we expect its power to be less than a $c_2 \neq 0$ QLR test, although it may still have some power. We examine these features in our Monte Carlo simulations in the next section. Third, our analysis here relies on $c_2 = 0$ and $c_3 \neq 0$. If the activation function has both $c_2 = 0$ and $c_3 = 0$, then our analysis no longer applies. Instead, further higher-order approximations are required. For example, if $c_4 \neq 0$, an eighth-order expansion may be useful. Fourth, a Lagrange multiplier (LM) statistic can be equivalently defined:

$$\mathcal{LM}_n := \frac{Z_{n,3}^2}{\hat{\sigma}_n^0 W_{n,3}} = \frac{(\boldsymbol{\iota}'\mathbf{D}_3\hat{\mathbf{U}}_0)^2}{\hat{\sigma}_n^0(\boldsymbol{\iota}'\mathbf{D}_3\mathbf{M}\mathbf{D}_3\boldsymbol{\iota})},$$

where $\hat{\mathbf{U}}_0 := \mathbf{Y} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ and $\hat{\sigma}_n^0 := -n^{-1}L_n(0;\lambda)$. In particular, $\hat{\mathbf{U}}_0 := \mathbf{U} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{U}$ under the null, and it easily follows that under the null,

$$\mathcal{LM}_n \Rightarrow \left(\frac{\mathcal{Z}_3^2}{\sigma_*^2\xi_3^*}\right).$$

This LM test differs from the standard LM test statistic, as it is defined using the third-order derivative. Luukkonen, Saikkonen, and Teräsvirta's (1988) LM test for the linear autoregressive model versus a smooth transition alternative is derived similarly. Finally, the availability of the LM test is useful in corroborating our theory, as our Monte Carlo experiments of the next section show.

3 Simulations

We divide this section into two subsections. First, we examine the relationship between the QLR and LM tests. This serves to corroborate the theory developed in Section 2. Second, we examine the level and power of the QLR and LM tests in different environments and compare their performances.

3.1 Comparison of the QLR and LM Tests

We consider the following simulation environment for the first goal stated above:

- (X_t, U_t) is identically and independently distributed;
- $Y_t \equiv X_t + U_t$; and
- $(X_t, U_t)' \sim N(\mathbf{0}, \mathbf{I}_2)$.

Using this DGP, we examine the asymptotic null behavior of the QLR test based on a logistic activation function. We denote the QLR and LM tests as $\mathcal{QLR}_n^{(\mathcal{L})}$ and $\mathcal{LM}_n^{(\mathcal{L})}$, respectively, and let the parameter spaces be $A = [-2.0, 2.0]$, $B = [-2.0, 2.0]$, $\Lambda = [0.5, 1.5]$, and $\Delta = [-2.0, 2.0]$. The alternative model is denoted as "$\mathcal{L}$" (for logistic) to distinguish it from the models considered below. That is,

- $\mathcal{L} := \{f(\cdot; \alpha, \beta, \lambda, \delta) : f(x; \alpha, \beta, \lambda, \delta) = \alpha + \beta x + \lambda\{1 + \exp(\delta x)\}^{-1} : \alpha, \beta, \delta \in [-2.0, 2.0], \lambda \in [0.5, 1.5]\}$.

Here, Λ does not contain 0, so the QLR statistic is not affected by the two-fold identification problem arising when $\lambda_* = 0$. The results of Section 2 and the independence of X_t and U_t ensure that $\mathcal{QLR}_n^{(\mathcal{L})} \overset{\text{A}}{\sim} \mathcal{X}_1^2$ and $\mathcal{LM}_n^{(\mathcal{L})} \overset{\text{A}}{\sim} \mathcal{X}_1^2$ under the null. We also have $\mathcal{QLR}_n^{(\mathcal{L})} = \mathcal{LM}_n^{(\mathcal{L})} + o_{\mathbb{P}}(1)$ under the null, so we expect that their correlation should converge to one as the sample size increases.

INSERT Figure 1 AROUND HERE.

We proceed with our simulations as follows. First, we obtain the empirical distributions of the QLR and LM tests for a sample size of $50,000$. This rather large sample size is chosen to accommodate the slow convergence of $\hat{\delta}_n$. Figure 1 shows these empirical distributions. There are four lines in Figure 1, obtained by repeating the experiments 5,000 times. The solid and dashed lines are reference lines, respectively the distribution functions of the chi-square random variable with one degree of freedom and the squared half standard normal random variable, $\max[0, Z]^2$, where $Z \sim N(0, 1)$. The other two lines (dotted and small-dashed lines) are the empirical distributions of the QLR and LM tests, respectively. We see that they closely match the chi-square distribution with one degree of freedom. They do not match the squared half-normal distribution applicable when the no-zero condition holds.

INSERT Figure 2 AROUND HERE.

To examine the role of the no-zero condition further, suppose that we modify the logistic hidden unit activation function to $\widetilde{\Psi}(x) := \{1 + \exp(1 + x)\}^{-1}$ so that $c_2 \neq 0$. The other assumptions are the same as before. Our modified QLR test, which we denote as $\mathcal{QLR}_n^{(\mathcal{M})}$, is expected to weakly converge to $\max[0, Z]^2$ by theorem 2 of Cho, Ishida, and White (2011). This is affirmed by Figure 2. That is, the empirical distribution of the QLR test (dotted line) is essentially identical to that of $\max[0, Z]^2$, and we can conclude from this that the order of expansion required when $c_2 = 0$ is different from that required when $c_2 \neq 0$. On the other hand, the LM statistic still has the $\mathcal{X}_1^2$ distribution. This illustrates the fact that when the no-zero condition holds, the QLR and LM statistics are no longer asymptotically equivalent under the null.

INSERT Table 1 AROUND HERE.

Finally, we examine the relation between the QLR and LM statistics when $c_2 = 0$. According to our theory, these are asymptotically equivalent under the null. To check this empirically, we tabulate the correlation coefficients between the QLR and LM tests for various sample sizes, n, using 1,000 replications for each n. Table 1 presents these results. As expected, the correlation coefficient approaches one as n increases, corroborating our theoretical results and confirming that a sixth-order expansion is indeed necessary to analyze the QLR statistic.

3.2 Level and Power of the QLR and LM Tests

Next, we compare the QLR and LM tests for a variety of cases. The main goal of this comparison is to investigate circumstances under which the performances of the tests may be poor.

We define two different QLR tests by specifying two different models, "$\mathcal{S}$" for "sine" and "$\mathcal{C}$" for "cosine":

- $\mathcal{S} := \{f(\cdot; \alpha, \beta, \lambda, \delta) : f(x; \alpha, \beta, \lambda, \delta) = \alpha + \beta x + \lambda \sin(\delta x) : \alpha, \beta, \delta \in [-5.0, 5.0], \lambda \in [1.0, 5.0]\}$;
- $\mathcal{C} := \{f(\cdot; \alpha, \beta, \lambda, \delta) : f(x; \alpha, \beta, \lambda, \delta) = \alpha + \beta x + \lambda \cos(\delta x) : \alpha, \beta, \delta \in [-5.0, 5.0], \lambda \in [1.0, 5.0]\}$.

The only difference between models $\mathcal{S}$ and $\mathcal{C}$ is that the activation functions differ. Nevertheless, their properties are quite different. Model $\mathcal{S}$ satisfies Assumption 2, but model $\mathcal{C}$ doesn't. On the other hand, model $\mathcal{C}$ satisfies the no-zero condition in Cho, Ishida, and White (2011), but model $\mathcal{S}$ doesn't. This implies that the null behaviors of the QLR tests based on models $\mathcal{S}$ and $\mathcal{C}$ need to be approximated by sixth - and fourth-order expansions, respectively. We denote the QLR tests constructed from models $\mathcal{S}$ and $\mathcal{C}$ as $\mathcal{QLR}_n^{(\mathcal{S})}$ and $\mathcal{QLR}_n^{(\mathcal{C})}$, respectively.

In addition to the QLR tests, we also consider LM test statistics. The first LM statistic is constructed as in Section 2 and denoted $\mathcal{LM}_n^{(\mathcal{S})}$. This notation recognizes its correspondence to $\mathcal{QLR}_n^{(\mathcal{S})}$ in the sense that a sixth-order expansion is exploited, as in Section 2. We also consider an LM statistic defined as

$$\mathcal{LM}_n^{(\mathcal{C})} := \frac{\max[0, \boldsymbol{\iota}'\mathbf{D}_2\hat{\mathbf{U}}_0]^2}{\hat{\sigma}_n^0\boldsymbol{\iota}'\mathbf{D}_2\mathbf{M}\mathbf{D}_2\boldsymbol{\iota}}.$$

Note that this $\mathcal{LM}_n^{(\mathcal{C})}$ is defined using the score function used for $\mathcal{QLR}_n^{(\mathcal{C})}$, which is based upon the quartic expansion as in theorem 2 of Cho, Ishida, and White (2011). Thus, both $\mathcal{LM}_n^{(\mathcal{C})}$ and $\mathcal{QLR}_n^{(\mathcal{C})}$ are asymptotically equivalent under the null, and their asymptotic null distribution is $\max[0, Z]^2$.

INSERT Table 2 AROUND HERE.

We compare the level and power of these test statistics. First, we apply the four test statistics to the same data sets as in the previous subsection and examine their empirical levels. Simulation results are presented in Table 2. We compare these four test statistics at three different levels of significance: 1%, 5%, and 10%. The overall performances of the four test statistics are very satisfactory. Even when the sample size is as small as 50, the empirical levels are well recovered by the asymptotic critical values. Further, when the sample size is greater than or equal to 100, the differences between the nominal levels and empirical rejection rates are less than 1 percentage point for every case. This suggests that we can trust the QLR and LM test statistics in terms of levels even when the sample size is not so large. On the other hand, if only the $n = 50$ case is considered, we see that the QLR and LM tests indexed by $\mathcal{C}$ are better than those indexed by $\mathcal{S}$ overall. This finite sample property is not surprising, given that $\mathcal{QLR}_n^{(\mathcal{C})}$ and $\mathcal{LM}_n^{(\mathcal{C})}$ have a faster convergence rate than $\mathcal{QLR}_n^{(\mathcal{S})}$ and $\mathcal{LM}_n^{(\mathcal{S})}$. Nevertheless, this is a minor difference.

Second, we apply the four test statistics to data sets generated as follows:

- (X_t, U_t) is identically and independently distributed;
- $Y_t \equiv X_t + \exp(X_t) + U_t$; and
- $(X_t, U_t)' \sim N(\mathbf{0}, \mathbf{I}_2)$.

This DGP satisfies the alternative; it enables us to compare the power properties of the QLR and LM test statistics. In particular, the neglected nonlinearity is exponential, which satisfies the no-zero condition. Although both models $\mathcal{S}$ and $\mathcal{C}$ are misspecified for this DGP, we expect that the QLR test statistic will perform better if it is constructed using an activation function satisfying the no-zero condition. Thus, $\mathcal{QLR}_n^{(\mathcal{C})}$ is expected to perform better than $\mathcal{QLR}_n^{(\mathcal{S})}$. Also, we should expect similar patterns from the LM tests, as they are constructed using the scores used for $\mathcal{QLR}_n^{(\mathcal{C})}$ and $\mathcal{QLR}_n^{(\mathcal{S})}$, respectively.

INSERT Table 3 AROUND HERE.

This expectation is affirmed by our simulation results, presented in Table 3. Note that the powers of $\mathcal{QLR}_n^{(\mathcal{C})}$ and $\mathcal{LM}_n^{(\mathcal{C})}$ are much higher than $\mathcal{QLR}_n^{(\mathcal{S})}$ and $\mathcal{LM}_n^{(\mathcal{S})}$, respectively. Even when the sample size is as small as 50, more than 98% of the replications reject the null hypothesis at the 5% level, whereas the empirical rejection rates are only around 50% for $\mathcal{QLR}_n^{(\mathcal{S})}$ and $\mathcal{LM}_n^{(\mathcal{S})}$. This demonstrates the importance of properly choosing the activation function for the test. Note, however, that $\mathcal{QLR}_n^{(\mathcal{S})}$ and $\mathcal{LM}_n^{(\mathcal{S})}$ still appear to be consistent. Also, we observe that the LM tests perform better than the QLR tests.

Finally, we apply the four test statistics to the following alternative DGP:

- (X_t, U_t) is identically and independently distributed;
- $Y_t \equiv X_t + \{1 + \exp(X_t)\}^{-1} + U_t$; and
- $(X_t, U_t)' \sim N(\mathbf{0}, \mathbf{I}_2)$.

Note that this DGP has neglected nonlinearity driven by the logistic function, which violates $c_2 \neq 0$, but instead satisfies $c_2 = 0$ and $c_3 \neq 0$. Again, models $\mathcal{S}$ and $\mathcal{C}$ are misspecified. Using this DGP, we expect that $\mathcal{QLR}_N^{(\mathcal{S})}$ will perform better than $\mathcal{QLR}_n^{(\mathcal{C})}$, for the reasons discussed earlier. In fact, given that $\mathcal{QLR}_n^{(\mathcal{C})}$ is designed to test $c_2 \neq 0-$ type nonlinearities, but here we have a $c_2 = 0$ nonlinearity, and because $\mathcal{QLR}_n^{(\mathcal{C})}$ treats higher-order terms associated with c_3, c_4, ..., as negligible in probability, we expect that $\mathcal{QLR}_n^{(\mathcal{C})}$ may have very little power.

INSERT Table 4 AROUND HERE.

Table 4 presents the simulation results. As we can see, $\mathcal{QLR}_n^{(\mathcal{S})}$ has power almost identical to $\mathcal{LM}_n^{(\mathcal{S})}$, and both are consistent. As the sample size increases, the empirical rejection rates converge to one, as expected. Although their convergence speed is not as fast as seen in Table 3, they are still consistent for detecting the neglected nonlinearity. This slow convergence is due to the use of a sixth-order expansion. On the other hand, $\mathcal{QLR}_n^{(\mathcal{C})}$ and $\mathcal{LM}_n^{(\mathcal{C})}$ have no power for any sample size. Indeed, power is always close to the nominal levels, and the empirical rejection rates don't improve even for $n = 30,000$.

In addition to the DGPs we report here, we conducted other experiments using $\mathcal{QLR}_n^{(\mathcal{C})}$, $\mathcal{LM}_n^{(\mathcal{C})}$, and alternative DGPs also exhibiting $c_2 = 0-$ type nonlinearities. Specifically, we considered the $\arctan$ and $\sin[\arctan]$ functions as sources of neglected nonlinearity. Our findings are substantially the same. Although the empirical distributions of $\mathcal{QLR}_n^{(\mathcal{C})}$ and $\mathcal{LM}_n^{(\mathcal{C})}$ are not identical to their asymptotic null distributions, the differences are slight, and the empirical distributions remain stable as the sample size increases. From this, we again see that $\mathcal{QLR}_n^{(\mathcal{C})}$ and $\mathcal{LM}_n^{(\mathcal{C})}$ are not consistent against $c_2 = 0-$ type neglected nonlinearities.

4 Conclusion

We illustrate the need to use higher-order expansions in order to properly determine the asymptotic distribution of a standard artificial neural network statistic designed to test for neglected nonlinearity. The test statistic is a Quasi-Likelihood Ratio (QLR) statistic for an ANN model that uses a hidden unit with a logistic activation function. This model violates Cho, Ishida, and White's (2011) no-zero condition, for which a fourth order expansion suffices. Instead, a sixth-order expansion delivers the desired first-order asymptotics. We also show that when the no-zero condition fails, the QLR statistic is asymptotically equivalent under the null to the Lagrange multiplier (LM) statistic of Luukkonen, Saikkonen, and Teräsvirta (1988), and Teräsvirta (1994). Finally, we compare the level and power of QLR tests satisfying and violating the no-zero condition in Cho, Ishida, and White (2011). This shows that when the neglected nonlinearity has $c_2 = 0$ and $c_3 \neq 0$, the QLR test constructed by a hidden layer with $c_2 \neq 0$ does not have power, whereas the QLR test with $c_2 = 0$ and $c_3 \neq 0$ does.

Acknowledgements

The authors are most grateful to the editor and two anonymous referees for their helpful comments. We also acknowledge helpful discussions with Bibo Jiang and other participants at the 5th Economics Symposium of Five Leading East Asian Universities held at Fudan University, Shanghai. Cho appreciates research support from the National Research Foundation of Korea Grant, funded by the Korean Government (NRF-2010-332-B00025).

References

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

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

Carrasco, M., Hu, L., and Ploberger, W. (2004). Optimal Test for Markov Switching. Mimeo, Economics Department, Rochester University.

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

Cho, J.S., Ishida, I., and White, H. (2011). Revisiting Tests for Neglected Nonlinearity Using Artificial Neural Networks. *Neural Computation*, *23*, 1133 – 1186.

Cramér, H. (1946). *Mathematical Methods of Statistics*. NJ: Princeton University Press.

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

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

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

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

McCullagh, P. (1984). Tensor Notation and Cumulants of Polynomials. *Biometrika*, *71*, 461–476.

McCullagh, P. (1986). The Conditional Distribution of Goodness-of-Fit Statistics for Discrete Data. *Journal of the American Statistical Association*, *81*, 104–107.

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

Phillips, P. (2011). Folklore Theorems, Implicit Maps and New Unit Root Limit Theory. Cowles Foundation Discussion Paper No. 1781, Cowles Foundation for Research in Economics, Yale University.

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

Table 1: CORRELATION COEFFICIENT BETWEEN QLR AND LM STATISTICS
Number of Replications: 1,000
DGP: $Y_t = X_t + U_t$, $(X_t, U_t) \sim$ IID $N(\mathbf{0}, \mathbf{I}_2)$
MODEL $\mathcal{L}$: $\alpha + \beta X_t + \lambda\{1 + \exp(\delta X_t)\}^{-1}$, $\alpha, \beta, \delta \in [-2.0, 2.0]$, AND $\lambda \in [0.5, 1.5]$

Sample Size	Correlation Coefficient
50	0.7929
100	0.8513
500	0.8725
$1,000$	0.8890
$5,000$	0.9349
$10,000$	0.9585
$50,000$	0.9829
$100,000$	0.9884
$200,000$	0.9932
$300,000$	0.9964

Table 2: LEVELS OF THE QLR AND LM TEST STATISTICS
NUMBER OF REPLICATIONS: 5,000
DGP: $Y_t = X_t + U_t$, $(X_t, U_t) \sim$ IID $N(\mathbf{0}, \mathbf{I}_2)$
MODEL $\mathcal{S}$: $\alpha + \beta X_t + \lambda \sin(\delta X_t)$, $\alpha, \beta, \delta \in [-5.0, 5.0]$, AND $\lambda \in [1.0, 5.0]$
MODEL $\mathcal{C}$: $\alpha + \beta X_t + \lambda \cos(\delta X_t)$, $\alpha, \beta, \delta \in [-5.0, 5.0]$, AND $\lambda \in [1.0, 5.0]$

Statistics	Levels $\backslash\ n$	50	100	150	200	250	300
	1%	1.00	0.92	1.30	1.08	0.86	0.98
$\mathcal{QLR}_n^{(\mathcal{S})}$	5%	4.28	4.82	5.28	4.92	4.48	4.84
	10%	8.70	9.60	9.56	9.12	8.50	9.46
	1%	1.16	0.94	1.28	1.06	0.94	0.90
$\mathcal{LM}_n^{(\mathcal{S})}$	5%	4.94	5.02	5.58	5.32	4.86	5.16
	10%	10.12	10.36	10.62	10.08	9.32	10.70
	1%	1.14	1.04	0.82	1.20	1.02	1.12
$\mathcal{QLR}_n^{(\mathcal{C})}$	5%	5.26	4.96	4.54	5.48	4.80	5.74
	10%	10.36	10.24	9.98	10.36	10.02	10.92
	1%	0.98	0.94	0.78	1.18	1.04	1.10
$\mathcal{LM}_n^{(\mathcal{C})}$	5%	5.08	4.80	4.36	5.32	4.72	5.52
	10%	9.94	9.92	9.68	10.12	9.96	10.72

Table 3: POWERS OF THE QLR AND LM TEST STATISTICS
NUMBER OF REPLICATIONS: 5,000
DGP: $Y_t = X_t + \exp(X_t) + U_t$, $(X_t, U_t) \sim$ IID $N(\mathbf{0}, \mathbf{I}_2)$
MODEL $\mathcal{S}$: $\alpha + \beta X_t + \lambda \sin(\delta X_t)$, $\alpha, \beta, \delta \in [-5.0, 5.0]$, AND $\lambda \in [1.0, 5.0]$
MODEL $\mathcal{C}$: $\alpha + \beta X_t + \lambda \cos(\delta X_t)$, $\alpha, \beta, \delta \in [-5.0, 5.0]$, AND $\lambda \in [1.0, 5.0]$

Statistics	Levels $\backslash\ n$	50	100	150	200	250	300
$\mathcal{QLR}_n^{(\mathcal{S})}$	1%	34.06	45.48	52.56	56.04	59.86	62.16
	5%	42.10	51.74	56.94	59.34	62.34	64.08
	10%	46.52	55.02	59.20	61.00	63.46	65.04
$\mathcal{LM}_n^{(\mathcal{S})}$	1%	42.94	55.36	64.10	69.56	75.66	80.02
	5%	53.44	64.78	72.00	76.44	81.26	84.98
	10%	60.02	69.82	76.26	80.18	83.78	87.24
$\mathcal{QLR}_n^{(\mathcal{C})}$	1%	96.90	99.50	99.68	99.84	99.84	99.88
	5%	98.86	99.70	99.78	99.84	99.90	99.90
	10%	99.38	99.80	99.80	99.86	99.92	99.92
$\mathcal{LM}_n^{(\mathcal{C})}$	1%	96.80	99.96	100.0	100.0	100.0	100.0
	5%	99.02	99.98	100.0	100.0	100.0	100.0
	10%	99.58	100.0	100.0	100.0	100.0	100.0

Table 4: POWERS OF THE QLR AND LM TEST STATISTICS
NUMBER OF REPLICATIONS: 5,000
DGP: $Y_t = X_t + \{1 + \exp(X_t)\}^{-1} + U_t$, $(X_t, U_t) \sim$ IID $N(\mathbf{0}, \mathbf{I}_2)$
MODEL $\mathcal{S}$: $\alpha + \beta X_t + \lambda \sin(\delta X_t)$, $\alpha, \beta, \delta \in [-5.0, 5.0]$, AND $\lambda \in [1.0, 5.0]$
MODEL $\mathcal{C}$: $\alpha + \beta X_t + \lambda \cos(\delta X_t)$, $\alpha, \beta, \delta \in [-5.0, 5.0]$, AND $\lambda \in [1.0, 5.0]$

Statistics	Levels $\backslash\ n$	100	200	500	1,000	2,000	5,000	10,000	20,000	30,000
$\mathcal{QLR}_n^{(\mathcal{S})}$	1%	1.08	1.72	2.44	4.28	7.96	21.66	49.34	84.88	96.74
	5%	4.86	6.50	9.30	13.10	20.82	43.54	73.14	95.00	99.36
	10%	9.94	11.48	15.72	20.74	31.04	56.64	82.38	97.44	99.76
$\mathcal{LM}_n^{(\mathcal{S})}$	1%	1.02	1.70	2.32	4.06	7.56	20.90	48.48	84.56	96.52
	5%	5.10	6.70	9.40	12.80	20.40	42.80	72.66	94.86	99.34
	10%	10.86	12.34	16.10	20.76	30.56	56.34	82.18	97.34	99.76
$\mathcal{QLR}_n^{(\mathcal{C})}$	1%	0.96	1.18	0.90	1.08	0.98	1.04	1.20	1.10	0.98
	5%	5.42	4.88	5.10	5.88	5.28	4.64	5.14	4.58	5.00
	10%	10.44	10.40	10.46	11.12	9.60	10.24	10.34	10.00	10.20
$\mathcal{LM}_n^{(\mathcal{C})}$	1%	0.94	1.14	0.82	1.04	0.96	1.04	1.18	1.08	0.98
	5%	5.32	4.74	4.92	5.82	5.20	4.68	5.16	4.56	5.00
	10%	10.20	10.30	10.36	11.04	9.50	10.18	10.00	10.18	10.40

Figure 1: Empirical Distributions of the QLR and LM Statistics: $c_2 = 0$
Number of Replications: 5,000
Sample Size: 50,000

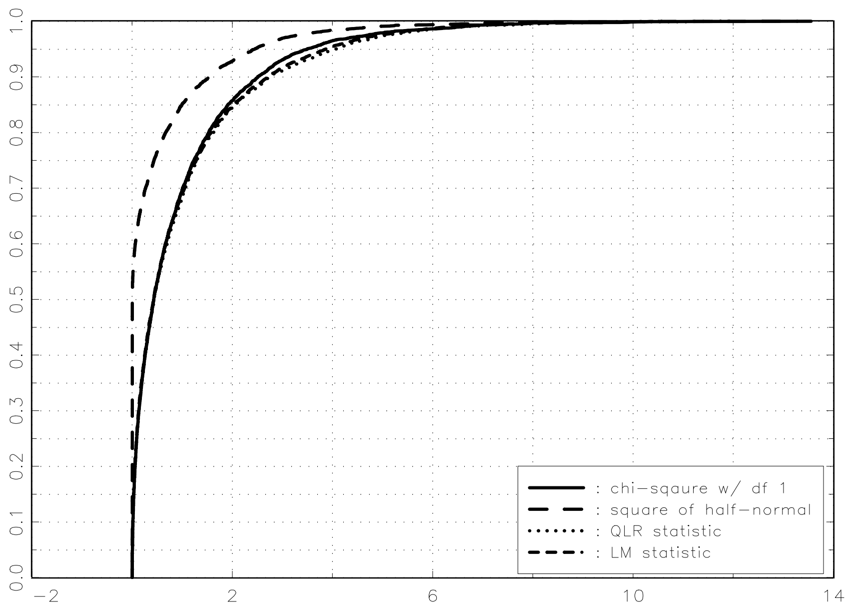

Figure 2: Empirical Distributions of the QLR and LM Statistics: $c_2 \neq 0$
Number of Replications: 5,000
Sample Size: 50,000