Higher-Order Approximations for Testing Neglected Nonlinearity

HALBERT WHITE
Department of Economics
University of California, San Diego
9500 Gilman Dr. 0508
La Jolla 92093-0508, U.S.A.
hwhite@ucsd.edu

JIN SEO CHO
School of Economics
Yonsei University
50 Yonsei-ro, Seodaemun-gu
Seoul 120-749, Korea
jinseocho@yonsei.ac.kr

First version: October 2009, This version: April 2011

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 a logistic activation function. This statistic is also shown to be asymptotically equivalent under the null to the Lagrange multiplier (LM) statistic of Luukkonen, Saikkonen, and Teräsvirta (1988), and Teräsvirta (1994).

Key Words: Artificial neural networks; logistic function; sixth-order approximation; quasi-likelihood ratio test; Lagrange multiplier test

Subject Class: C12, C22, C45, C52.

Acknowledgements: The authors 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).
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 with a logistic activation function, 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. Section 4 contains a summary and concluding remarks.
2 A QLR test for neglected nonlinearity

We begin by specifying the same data generating process as that 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 \(\{(Y_t, X'_t) : t = 1, 2, \ldots\} \in \mathbb{R}^{1+k}\) with mixing coefficients \(\beta_r\) such that for some \(\rho > 1\),

\[
\sum_{t=1}^{\infty} \tau^1/(\rho-1) \beta_t < \infty.
\]

Further, \(E(Y_t) < \infty\).

We note that \(X_t\) 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 \(E[Y_t|X_t]\).

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

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

where \(\Psi : \mathbb{R} \mapsto \mathbb{R}\) is the logistic function, so that \(\Psi(x) := \{1 + \exp(x)\}^{-1}\), and \(A \subset \mathbb{R}, B \subset \mathbb{R}^k, \Lambda \subset \mathbb{R},\) and \(\Delta \subset \mathbb{R}\) are non-empty compact sets, with \(0 \in \Delta\). Here, \(X_{t,1}\) is the first element of \(X_t\).

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 \(X_t\) appearing inside \(\Psi\) for simplicity. It is straightforward to treat the case where \(\Psi(X_{t,1}\delta)\) is replaced by \(\Psi(X'_t\delta)\), but the notation required to handle this case becomes extremely cumbersome.

Now consider testing the linearity of conditional expectation: for some \(\alpha_* \in A\) and \(\beta_* \in B, \) \(E[Y_t|X_t] = \alpha_* + X'_t \beta_*\). When linearity of \(E[Y_t|X_t]\) holds, the pseudo-true values \(\lambda_*\) and \(\delta_*\) 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):

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

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 \(c_2 \neq 0\), where

\[
c_2 := \frac{\partial^2}{\partial x^2} \Psi(x) \bigg|_{(x=0)}.
\]
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 $\alpha$ and $\beta$, obtaining

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

where $\mathbf{Y} := [Y_1, Y_2, \ldots, Y_n]'$, $\Psi(\delta) := [\Psi_1(\delta), \Psi_2(\delta), \ldots, \Psi_n(\delta)]'$, $\mathbf{M} := \mathbf{I} - \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}'$, $\mathbf{Z} := (\mathbf{t}, \mathbf{X})$ with $\mathbf{t}$ 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

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

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

- $L_n^{(1)}(0; \lambda) := \frac{\partial}{\partial \lambda} L_n(0; \lambda) = 0$;
- $L_n^{(2)}(0; \lambda) := \frac{\partial^2}{\partial \lambda^2} L_n(0; \lambda) = 0$;
- $L_n^{(3)}(0; \lambda) := \frac{\partial^3}{\partial \lambda^3} L_n(0; \lambda) = \frac{1}{2} \lambda \mathbf{t}' \mathbf{D}_3 \mathbf{MU}$;
- $L_n^{(4)}(0; \lambda) := \frac{\partial^4}{\partial \lambda^4} L_n(0; \lambda) = 0$;
- $L_n^{(5)}(0; \lambda) := \frac{\partial^5}{\partial \lambda^5} L_n(0; \lambda) = -\frac{1}{2} \lambda \mathbf{t}' \mathbf{D}_3 \mathbf{MU}$; and
- $L_n^{(6)}(0; \lambda) := \frac{\partial^6}{\partial \lambda^6} L_n(0; \lambda) = -\frac{5}{16} \lambda^2 \mathbf{t}' \mathbf{D}_3 \mathbf{MD}_3 \mathbf{t}$,

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 $\mathbf{D}_m := \text{diag}\{X_{1,1}^m, X_{2,1}^m, \ldots, X_{n,1}^m\}$ 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. The sixth-order Taylor expansion at $\delta_* = 0$ is then

$$L_n(\delta; \lambda) - L_n(0; \lambda) = \frac{1}{3!} L_n^{(3)}(0; \lambda) \delta^3 + \frac{1}{5!} L_n^{(5)}(0; \lambda) \delta^5 + \frac{1}{6!} L_n^{(6)}(0; \lambda) \delta^6 + o_p(1).$$
Before examining the asymptotic behavior of the terms on the right, we impose the following regularity conditions:

**Assumption 3 (Moments)** \( E|U_t|^4 < \infty, E|X_{t,1}|^{10} < \infty \), and for \( j > 1 \), \( E|X_{t,j}|^4 < \infty \).

**Assumption 4 (MDS)** \( E[U_t|X_t, U_{t-1}, X_{t-1}, \ldots] = 0 \).

**Assumption 5 (Covariance)** \( \det[E[U_t^2 \tilde{Z}_t \tilde{Z}_t^\prime]] > 0 \) and \( \det[E[Z_t Z_t^\prime]] > 0 \), where \( \tilde{Z}_t := (Z_t, X_{t,1}^3)' \) and \( Z_t := (1, X_t)' \).

These conditions ensure the regular asymptotic behavior of each derivative term. Specifically, the ergodic theorem and the central limit theorem for martingale difference sequences ensure that

- \( n^{-1} L_n(0; \lambda) \xrightarrow{a.s.} -\sigma_3^2 \); 
- \( Z_{n,3} := n^{-1/2} \tau^3 \mathbf{D}_3 \mu \Rightarrow Z_3 \sim N(0, \tau_3^*) \); 
- \( Z_{n,5} := n^{-1/2} \tau^5 \mathbf{D}_5 \mu \Rightarrow Z_5 \sim N(0, \tau_5^*) \); and 
- \( W_{n,6} := n^{-1} \tau^6 \mathbf{D}_6 \mu \xrightarrow{a.s.} \xi_3^* \),

where \( \sigma_3^2 := E[U_t^2] \); for \( m = 3 \) and \( 5 \),

\[
\tau_m^* := E[U_t^2 X_{t,1}^{2m}] - 2E[U_t^2 X_{t,1}^m Z_t | Z_t, Z_t^\prime] E[Z_t Z_t^\prime]^{-1} E[X_{t,1}^m Z_t] \\
+ E[X_{t,1}^m Z_t] E[Z_t Z_t^\prime]^{-1} E[U_t^2 Z_t Z_t^\prime] E[Z_t Z_t^\prime]^{-1} E[X_{t,1}^m Z_t];
\]

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

Substituting appropriately gives

\[
L_n(\delta; \lambda) - L_n(0; \lambda) = \frac{\lambda}{3!4} Z_{n,3} \{n^{1/6} \delta\}^3 - \frac{\lambda}{5!2} Z_{n,5} \frac{n^{1/3}}{n^{1/6}} \{n^{1/6} \delta\}^5 - \frac{5\lambda^2}{6!16} W_{n,6} \{n^{1/6} \delta\}^6 + o_P(1).
\]

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

\[
3 \frac{\lambda}{3!4} Z_{n,3} D_n^2 - 5 \frac{\lambda}{5!2} \frac{Z_{n,5}}{n^{1/3}} D_n^4 - \frac{5\lambda^2}{6!16} W_{n,6} D_n^6 = 0.
\]

The solution \( D_n = 0 \) gives the minimum, so we have \( D_n^2 > 0 \), and we can divide both sides by \( D_n^2 \) to obtain

\[
3 \frac{\lambda}{3!4} Z_{n,3} - 5 \frac{\lambda}{5!2} \frac{Z_{n,5}}{n^{1/3}} D_n^2 - \frac{5\lambda^2}{6!16} W_{n,6} D_n^4 = 0.
\]
This is a cubic equation in $D_n$.

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

$$
\sup_{\delta \in \Delta} \{ L_n(\delta; \lambda) - L_n(0; \lambda) \} \Rightarrow \frac{\lambda}{3!4} Z_3 \hat{D}_*^3 - \frac{5\lambda^2}{6!16} \xi_3^* \hat{D}_*^6 = \sup \frac{\lambda}{3!4} Z_3 D^3 - \frac{5\lambda^2}{6!16} \xi_3^* D^6.
$$

From the final equality, it is straightforward to verify that

$$
D_*^3 := \left( \frac{48}{\xi^* \lambda} \right) Z_3 \sim N \left[ 0, \tau_*^3 \left( \frac{48}{\xi^* \lambda} \right)^2 \right].
$$

Thus, it follows that

$$
\sup_{\delta \in \Delta} \{ L_n(\delta; \lambda) - L_n(0; \lambda) \} = \frac{Z_{n,3}^2}{W_{n,6}} + o_P(1) \Rightarrow \frac{\lambda}{3!4} Z_3 D_*^3 - \frac{5\lambda^2}{6!16} \xi_3^* D_*^6 = \frac{Z_3^2}{\xi_3^*}.
$$

Observe that the unidentified parameter $\lambda$ cancels out, so the asymptotic null distribution is free of $\lambda$, 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

$$
QLR_n := \sup_{\lambda \in \Lambda} QLR_n(\lambda) \Rightarrow \left( \frac{Z_3^2}{\sigma_3^2 \xi_3^*} \right).
$$

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 $D_*$ captures the asymptotic behavior of $n^{1/6} \delta_n$ under the null, where $\hat{D}_n$ 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 $n^{1/6}$. Third, a Lagrange multiplier (LM) statistic can be equivalently defined:

$$
LM_n := \frac{Z_{n,3}^2}{\sigma_n^2 W_{n,3}} = \frac{(\nu' D_3 \hat{U}_0)^2}{\sigma_n^2 (\nu' D_3 \hat{M} \hat{D}_3 \nu)},
$$

5
where $\hat{U}_0 := Y - Z(Z'Z)^{-1}Z'Y$ and $\hat{\sigma}^2_n := -n^{-1}L_n(0; \lambda)$. It easily follows that under the null,

$$\mathcal{L}\mathcal{M}_n \Rightarrow \left(\frac{Z_2^2}{\sigma^2_3\xi^2_3}\right).$$

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 consider the following simple simulation environment:

- $(X_t, U_t)$ is identically and independently distributed;
- $Y_t \equiv X_t + U_t$; and
- $(X_t, U_t)' \sim N(0, I_2)$.

We let the parameter spaces be $A = [-2.0, 2.0]$, $B = [-2.0, 2.0]$, and $\Lambda = [0.1, 1.5]$. Here, $\Lambda$ 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 $Q\mathcal{L}\mathcal{R}_n \overset{\Lambda}{\sim} \chi^2_1$ and $\mathcal{L}\mathcal{M}_n \overset{\Lambda}{\sim} \chi^2_1$ under the null. We also have $Q\mathcal{L}\mathcal{R}_n = \mathcal{L}\mathcal{M}_n + o_P(1)$ under the null, so we expect that their correlation should converge to one as the sample size increases.

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.

To examine the role of the no-zero condition further, suppose that we modify the logistic hidden unit activation function to $\tilde{\Psi}(x) := \{1 + \exp(1 + x)\}^{-1}$ so that $c_2 \neq 0$. The other assumptions are the same as before. Our QLR test 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)
Table 1: Correlation Coefficient between QLR and LM Statistics

<table>
<thead>
<tr>
<th>Number of Replications: 1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>500</td>
</tr>
<tr>
<td>1,000</td>
</tr>
<tr>
<td>5,000</td>
</tr>
<tr>
<td>10,000</td>
</tr>
<tr>
<td>50,000</td>
</tr>
<tr>
<td>100,000</td>
</tr>
<tr>
<td>200,000</td>
</tr>
<tr>
<td>300,000</td>
</tr>
</tbody>
</table>

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 \( \chi^2_1 \) 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.

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.

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).
References


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