On the Asymptotic Standard Errors of Residual Autocorrelations in Nonlinear Time Series Modelling
Author(s): W. K. Li
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2336859
Accessed: 08-09-2016 14:42 UTC

REFERENCES
Linked references are available on JSTOR for this article:
You may need to log in to JSTOR to access the linked references.
On the asymptotic standard errors of residual autocorrelations in nonlinear time series modelling

BY W. K. LI

Department of Statistics, University of Hong Kong, Hong Kong

SUMMARY

The asymptotic distribution of residual autocorrelations for some very general nonlinear time series models is derived. This includes the important class of threshold models. Consequently more accurate standard errors for residual autocorrelations can be obtained facilitating model diagnostic checkings in many situations. A small simulation result applying the methodology to threshold models is reported.

Some key words: Asymptotic distribution; Nonlinear models; Residual autocorrelations; Standard errors; Threshold models.

1. INTRODUCTION

Nonlinear time series analysis has in recent years attracted a lot of attention in the literature. Many classes of models have been proposed. Among these, the two most popular classes of models are the bilinear models (Granger & Andersen, 1978) and the threshold models (Tong, 1978; Tong & Lim, 1980). See the recent book by Tong (1990) for a more complete introduction to nonlinear time series modelling.

However, as is well known, for residuals obtained from a fitted autoregressive moving average model the actual standard errors could be much less than $1/\sqrt{n}$ (Box & Pierce, 1970; Ansley & Newbold, 1979). In nonlinear time series, the asymptotic distribution of the residual autocorrelations for all the major models has been unknown and consequently the value $1/\sqrt{n}$ is only a rough guide in diagnostic checking. For example, Tong (1990, p. 324) suggested the use of $1/\sqrt{n}$ but cautioned that this could be conservative especially for the first few lags. In the present paper, the distribution of the residual autocorrelations from a general nonlinear model is derived. This includes automatically the threshold autoregressive models as special cases. The asymptotic covariance matrix could be evaluated using sample averages instead of expected values. A similar approach has also been considered by Li (1991) for extensions of generalized linear models. A small simulation applied to threshold models shows that the standard errors thus obtained are much closer to the actual ones. The main result is given in § 2. Application to threshold type nonlinear models is given in § 3.

2. THE MAIN RESULT

Let $\{X_t\}$ be a stationary and ergodic time series, with $F_t$ the $\sigma$-field generated by $\{X_t, X_{t-1}, \ldots\}$. Let $\{X_t\}$ satisfy the nonlinear model

$$X_t = f(F_{t-1}; \phi) + a_t,$$

(1)

where $f$ is a known nonlinear function of past $X_t$'s and $\phi$ is a $p \times 1$ vector of parameters. The function $f$ is assumed to have continuous second order derivatives almost surely. The noise process $\{a_t\}$ is assumed to be independent, Gaussian with mean zero and variance $\sigma_a^2$. It is further assumed that (1) is invertible or equivalently $\{a_t\}$ is measurable with respect to $F_t$. 

This content downloaded from 165.91.114.21 on Thu, 08 Sep 2016 14:42:01 UTC
All use subject to http://about.jstor.org/terms
Let the length of realization be $n$. Let the lag $k$ white noise autocovariance be $C_k = \Sigma a_i a_{i-k} / n$ ($k = 1, \ldots, M$) and denote by $\hat{C}_k$ the corresponding residual autocovariances. The residuals $\{\hat{a}_i\}$ are assumed to be from a least squares fit of (1) to $\{X_t\}$. Define the lag $k$ residual autocorrelations to be $r_k = C_k / C_0$. Using a Taylor series expansion of $r_k$ it can be shown that the asymptotic distribution of $r_k$ does not depend on $C_0$ and therefore we can ignore $C_0$ in deriving the asymptotic distribution of $r_k$. The result for $r_k$ will follow from that of $\hat{C}_k$ by scaling.

Let $r_k = C_k / C_0$, $r = (r_1, \ldots, r_M)'$ and $\hat{r} = (\hat{r}_1, \ldots, \hat{r}_M)'$. The residual variance $\sigma^2$ is estimated by $\hat{\sigma}^2$. If $\{a_t\}$ have finite fourth order moments then it is well known that $r^2/n$ is asymptotically normally distributed with mean zero and variance $I_M$, where $I_M$ is the $M \times M$ identity matrix. Under regularity conditions as given by Klimko & Nelson (1978) the least squares estimator $\hat{\phi}$ of $\phi$ can be shown to be asymptotically normally distributed with mean $\phi$ and covariance matrix $\sigma^2 V^{-1}/n$, where

$$V = E\{n^{-1} \sum (\partial a_i / \partial \phi)(\partial a_i / \partial \phi)\}.$$

Denote $f(F_{t-1}, \phi)$ by $f_{t-1}$. Suppose that $E(\partial f_{t-1} / \partial \phi a_{t-1})$ exist for $j = 1, \ldots, M$, and that corresponding sample averages converge in probability to the respective expected values. A sufficient condition for the latter would be the covariance between $a_{t-j} \partial f_{t-1} / \partial \phi$ and $a_{t-j} \partial f_{t-1} / \partial \phi$ goes to zero as $|t-t'| \to \infty$. This seems to be a reasonable assumption in practice. The following two lemmas follow (McLeod, 1978; Li, 1991) using Taylor series expansion of $\Sigma a_t^2$ and $\hat{C}_k$.

**Lemma 1.** The asymptotic cross covariance between $\sqrt{n}(\hat{\phi} - \phi)$ and $\sqrt{n}C = (C_1, \ldots, C_M)'$ is equal to $\sigma^2 V^{-1} J$, where

$$J = E\{\sum \partial f_{t-1} / \partial \phi a_{t-1}, \ldots, \sum \partial f_{t-1} / \partial \phi a_{t-M}\} n^{-1}.$$

**Proof.** This follows from the standard result

$$(\hat{\phi} - \phi) \sim (\sum \partial f_{t-1} / \partial \phi a_{t-1})^{-1} (\sum \partial f_{t-1} / \partial \phi a_{t-M}).$$

**Lemma 2.** For large $n$, $\hat{C} \sim C - J'(\hat{\phi} - \phi)$.

**Proof.** This follows from a Taylor series expansion of $\hat{C}_k$ about $\phi$ and evaluated at $\hat{\phi}$.

From these two lemmas and the martingale central limit theorem (Billingsley, 1961) we have the following theorem.

**Theorem.** The large sample distribution of $\hat{r}^2/n$ is normal with mean zero and covariance matrix $I_M - \sigma^2 J' V^{-1} J$.

Note that, for autoregressive moving average models, $V$ and $J$ can be evaluated in terms of $\phi$. For nonlinear models closed form expressions for these quantities are usually unavailable. Our proposal here is to use observed quantities instead of the expectations. This is in some sense analogous to the use of observed rather than expected Fisher information (Efron & Hinkley, 1978). Simulation results reported in § 3 suggest that this is a viable approach.

### 3. Simulation results

A small simulation experiment was conducted to compare the asymptotic and the empirical standard errors of $\hat{r}_k$ in threshold models. The design of the experiment is as follows. We considered a simple SETAR (2; 1, 1) model $X_t = \phi_1 X_{t-1} + a_t$ if $X_{t-1} > 0$; and $X_t = \phi_2 X_{t-1} + a_t$ otherwise, where $\{a_t\}$ were normally distributed with mean 0 and variance 1. Then it can be easily shown that $V = n^{-1}(X'X)$, where $X$ is given by Tong (1983, p. 140). Similarly, elements of $J$ can be shown to be the limits in probability of the quantities $\Sigma X_{t-1} a_{k-t} I_j / n$, where $k = 1, \ldots, M$, $j = 1, 2$. Here $I_1$ indicates $X_{t-1} > 0$ and $I_2 = 1 - I_1$. For each pair $(\phi_1, \phi_2)$, 1000 independent realizations each of length 200 were generated. The values of $(\phi_1, \phi_2)$ considered were $(0.5, -0.5)$, $(0.8, -0.8)$, $(0.95, -0.95)$, $(0.8, 0.3)$, $(-0.8, -0.3)$. The series were generated and fitted using IMSL subroutines. The sample variances $V(\hat{r}_k)$ of $\hat{r}_k$ over the 1000 replications were computed for each model. Denote $\sqrt{V(\hat{r}_k)}$ by $S_{dk}$. These were taken to be the 'true' standard errors of $\hat{r}_k$. The
Table 1. Empirical results for residual autocorrelations in SETAR(2; 1, 1) models, n = 200, 1000 replications

<table>
<thead>
<tr>
<th>$\phi_1, \phi_2$</th>
<th>$k = 1$</th>
<th>$k = 2$</th>
<th>$k = 3$</th>
<th>$k = 4$</th>
<th>$k = 5$</th>
<th>$k = 6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.5, -0.5) Sdk</td>
<td>0.0282</td>
<td>0.0703</td>
<td>0.0674</td>
<td>0.0719</td>
<td>0.0709</td>
<td>0.0706</td>
</tr>
<tr>
<td>$\sqrt{C_k}$</td>
<td>0.0277</td>
<td>0.0698</td>
<td>0.0704</td>
<td>0.0704</td>
<td>0.0704</td>
<td>0.0704</td>
</tr>
<tr>
<td>(-0.8, 0.8) Sdk</td>
<td>0.0489</td>
<td>0.0688</td>
<td>0.0663</td>
<td>0.0719</td>
<td>0.0709</td>
<td>0.0706</td>
</tr>
<tr>
<td>$\sqrt{C_k}$</td>
<td>0.0477</td>
<td>0.0675</td>
<td>0.0695</td>
<td>0.0701</td>
<td>0.0703</td>
<td>0.0704</td>
</tr>
<tr>
<td>(0.95, -0.95) Sdk</td>
<td>0.0626</td>
<td>0.0672</td>
<td>0.0660</td>
<td>0.0714</td>
<td>0.0704</td>
<td>0.0702</td>
</tr>
<tr>
<td>$\sqrt{C_k}$</td>
<td>0.0601</td>
<td>0.0679</td>
<td>0.0689</td>
<td>0.0694</td>
<td>0.0697</td>
<td>0.0699</td>
</tr>
<tr>
<td>(0.8, 0.3) Sdk</td>
<td>0.0475</td>
<td>0.0630</td>
<td>0.0653</td>
<td>0.0711</td>
<td>0.0704</td>
<td>0.0701</td>
</tr>
<tr>
<td>$\sqrt{C_k}$</td>
<td>0.0459</td>
<td>0.0636</td>
<td>0.0678</td>
<td>0.0693</td>
<td>0.0699</td>
<td>0.0702</td>
</tr>
<tr>
<td>(-0.8, -0.3) Sdk</td>
<td>0.0385</td>
<td>0.0637</td>
<td>0.0659</td>
<td>0.0719</td>
<td>0.0706</td>
<td>0.0705</td>
</tr>
<tr>
<td>$\sqrt{C_k}$</td>
<td>0.0376</td>
<td>0.0632</td>
<td>0.0689</td>
<td>0.0700</td>
<td>0.0704</td>
<td>0.0704</td>
</tr>
</tbody>
</table>

Asymptotic variances $C(\hat{r}_k)$ were also estimated for each realization using the Theorem in § 2. The sample averages of $C(\hat{r}_k)$ were obtained and were denoted as $\hat{C}_k$. The results for $\sqrt{\hat{C}_k}$ and $Sdk$ ($k = 1, \ldots, 6$) are reported in Table 1.

As in the linear autoregressive situation the results in Table 1 showed that the 'true' standard errors for $\hat{r}_k$ could be smaller than the value $1/\sqrt{200} = 0.0707$. This discrepancy is more prominent if the values of $k$ are small. Consequently, using $1.96/\sqrt{n}$ as critical value would give a very conservative confidence limit for the first few residual autocorrelations. Note also the much closer match between $\sqrt{\hat{C}_k}$ and $Sdk$. This suggests that the result in § 2 could be usefully applied to give more accurate standard errors in practice. Our result will give a more stringent criterion in diagnostic checking for threshold models. This is consistent with the comments by Tong (1990, p. 324) that the first few $\hat{r}_k$ should be given more careful scrutiny. Note that as $k$ becomes larger both $Sdk$ and $\sqrt{\hat{C}_k}$ approach the value $1/\sqrt{n}$.

**Acknowledgement**

I would like to thank Professor Howell Tong for helpful comments.

**References**


[Received February 1991. Revised January 1992]