Statistica Sinica 4(1994), 107-125

THE ASYMPTOTIC DISTRIBUTIONS OF RESIDUAL AUTOCORRELATIONS AND RELATED TESTS OF FIT FOR A CLASS OF NONLINEAR TIME SERIES MODELS

Sun Young Hwang, I. V. Basawa and J. Reeves

University of Georgia

Abstract: We derive the joint asymptotic distribution of the sample autocorrelation functions $\hat{\rho}_R(i)$ of lag i, $i = 1, \ldots, l$, based on the residuals $\{R_t\}$, $t = 1, 2, \ldots, n$, obtained after fitting a nonlinear time series model. Tests of fit based on the residuals and on the autocorrelations of residuals are also presented. Some simulation results are reported.

Key words and phrases: Nonlinear time series, diagnostic testing, residual autocorrelations, asymptotic distributions, conditional least squares.

1. Introduction

In a seminal paper, Box and Pierce (1970) obtained the joint limit distributions of sample autocorrelation functions (ACF) of the residuals after fitting a linear time series model. As shown by Box and Pierce (1970), the ACF of the residuals can be used as a diagnostic tool; and portmanteau type tests based on the residual ACF's can be developed to test model adequacy. See Brockwell and Davis (1991) for a general discussion of the problem of diagnostic testing. Box and Pierce (1970) and Durbin (1970) showed that the variances of the ACF's based on the residuals from a linear autoregressive process are different from those of the ACF's based on a sample of independent and identically distributed observations. Furthermore, the residual ACF's were shown to be correlated even for large n. The study of the asymptotic covariance matrix of the residual ACF's enabled Box and Pierce (1970) to make a detailed investigation of the problem of diagnostic checking of linear time series models.

In this paper, we obtain the joint limit distribution of the residual ACF's obtained from fitting a **nonlinear** autoregressive process. Li (1992) has recently discussed the limit distribution of the residual ACF's for nonlinear models with fixed coefficients. The nonlinear autoregressive model that we consider includes standard models such as threshold autoregression, exponential autoregression and random coefficient autoregressive models. Our results reduce to those of Box and Pierce (1970) in the special case of a linear autoregressive model and to those of Li (1992) for nonlinear models with fixed coefficients. In this sense, our paper can be viewed as a generalization of the paper by Box and Pierce (1970) and by Li (1992). The covariance structure of the residual ACF's for the nonlinear model is seen to be much more complicated than in the case of a linear model. However, there are some qualitative similarities between the results for the linear and nonlinear models. In both models, the ACF's with lower lags have asymptotic variances significantly different from unity, whereas for higher lags, they quickly converge to unity.

The paper is organized as follows. The nonlinear model and some special cases are presented in Section 2. The consistency and asymptotic normality of the conditional least squares estimators of the model parameters are discussed in Section 3. Section 4 contains the main result of the paper, viz., the joint limit distribution of the residual ACF's. Some tests of fit based on the residuals and their ACF's are derived in Section 5. Finally, some simulation results are presented in Section 6.

2. The Model and Examples

Consider a pth-order nonlinear autoregressive time series $\{Y_t, t \geq 1\}$ defined by

$$Y_t = H(\tilde{Y}_{t-1}, Z_t; \theta) + \varepsilon_t, \tag{2.1}$$

where $\{\varepsilon_t\}$ is an unobserved sequence of i.i.d. random errors with mean zero and variance σ_ε^2, H is a known function, $\tilde{Y}_{t-1} = (Y_{t-1}, \ldots, Y_{t-p})^T$, and θ is an unknown vector parameter, $\theta \in \Theta \subset R^J$. The unobserved random vectors $\{Z_t\}$ are assumed to be i.i.d. with mean zero, and independent of $\{\varepsilon_t\}$. Typically, $\{Z_t\}$ represents uncertainty about the parameter θ and H depends on Z_t and θ via the "random parameter" $(\theta + Z_t)$. The model in (2.1) covers several standard nonlinear models in addition to linear models. Note also that random coefficient linear and nonlinear models are also included in the class of processes given by (2.1). We now consider some examples of (2.1).

Examples

Ex. 1. Random coefficient autoregressive (RCA) processes

$$H(\tilde{Y}_{t-1}, Z_t; \theta) = (\theta_1 + Z_{t1})Y_{t-1} + \cdots + (\theta_p + Z_{tp})Y_{t-p}, \tag{2.2}$$

where $\{Z_{tj}, j = 1, \ldots, p\}$ are i.i.d. scalar random variables with mean zero. (See Nicholls and Quinn (1982) and Feigin and Tweedie (1985).) If we set $Z_{tj} \equiv 0$, $j = 1, \ldots, p$, in the above model, we get standard linear autoregressive processes.

Ex. 2. Threshold autoregressive (TAR) processes

The first-order threshold autoregressive process TAR(1) is defined by (2.1) with

$$H(\tilde{Y}_{t-1}, Z_t; \theta) = \theta_1 Y_{t-1}^+ + \theta_2 Y_{t-1}^-, \tag{2.3}$$

where $Y_{t-1}^+ = Y_{t-1} I(Y_{t-1} \geq 0)$, $Y_{t-1}^- = Y_{t-1} I(Y_{t-1} < 0)$ and where $I(\cdot)$ denotes the indicator function of the event enclosed. The model in (2.3) can easily be generalized to higher order TAR models. For instance, if R_i, $i = 1, \ldots, k$, denote k disjoint regions with $R = \bigcup_{i=1}^k R_i$, and I_{ij} denotes the indicator function of the event $(Y_{t-j} \in R_i)$, the TAR(p) model is specified by (2.1) with

$$H(\tilde{Y}_{t-1}, Z_t; \theta) = \sum_{i=1}^{k} \sum_{j=1}^{p} \theta_{ij} Y_{t-j} I_{ij}. \tag{2.4}$$

(See Tong (1990) for these and related models.)

Ex. 3. Random coefficient threshold autoregressive (RCTAR) processes

Replacing θ_1 and θ_2 in (2.3) by $Z_{t1} + \theta_1$ and $Z_{t2} + \theta_2$ we get a RCTAR(1) model. Also, replacing θ_{ij} in (2.4) by $Z_t(i,j) + \theta_{ij}$ we can extend the TAR(p) model to a RCTAR(p) model. (See Hwang and Basawa (1991), and Brockwell et al. (1992).)

Ex. 4. Exponential Autoregression (EAR)

Take

$$H(\tilde{Y}_{t-1}, Z_t; \theta) = \sum_{j=1}^{p} \left\{\theta_{1j} + \theta_{2j} \exp(-\theta_{3j} Y_{t-j}^2)\right\} Y_{t-j}. \tag{2.5}$$

(See Tong (1990, p.129) for further details on this model.)

Ex. 5. Random coefficient exponential autoregression (RCEAR)

Replace θ_{1j} and θ_{2j} in (2.5) by $\theta_{1j} + Z_{t1}(j)$ and $\theta_{2j} + Z_{t2}(j)$ respectively to get an RCEAR(p) model. (See Hwang and Basawa (1991) for further details.)

3. The Consistency and Asymptotic Normality of the Conditional Least Squares Estimators

Define

$$M(\tilde{Y}_{t-1}; \theta) = E_\theta(Y_t \mid A_{t-1}) = E_\theta\Big(H(\tilde{Y}_{t-1}, Z_t; \theta) \mid A_{t-1}\Big), \tag{3.1}$$

where

$$A_{t-1} = \sigma(Y_{-p+1}, \ldots, Y_0, \ldots, Y_{t-1}), \quad t \geq 1,$$

is the σ-field generated by $\{Y_{-p+1}, \ldots, Y_0, \ldots, Y_{t-1}\}$. The conditional least squares (CLS) estimator $\hat{\theta}_n$ of θ is then obtained by minimizing $\sum_{t=1}^n U_t^2(\theta)$, where

$$U_t(\theta) = Y_t - M(\tilde{Y}_{t-1}; \theta) \tag{3.2}$$

denotes the prediction error. Consider the following regularity conditions:

(C.0) For each $\theta \in \Theta \subset R^J$, $\{Y_t, t \geq 1\}$ admits a unique initial distribution for $(Y_{-p+1}, \ldots, Y_0)$ so that $\{Y_t, t \geq -p+1\}$ is stationary and ergodic.

(C.1) The conditional expectation $M(\tilde{Y}_{t-1}; \theta)$ is a function of $\tilde{Y}_{t-1}$ and θ only, not depending on any nuisance parameters.

(C.2) The $J \times J$ matrix V defined by

$$V = E_\theta\Big[(\nabla M(\tilde{Y}_{t-1}; \theta))(\nabla M(\tilde{Y}_{t-1}; \theta))^T\Big]$$

exists and is positive definite for each $\theta \in \Theta$, where ∇M denotes the $(J \times 1)$ vector of first derivatives of M with respect to θ.

(C.3) There exist square integrable random variables

$$\alpha_i(\tilde{Y}_{t-1}; \theta), \quad \beta_{ij}(\tilde{Y}_{t-1}; \theta), \quad \gamma_{ijk}(\tilde{Y}_{t-1}; \theta),$$

and a positive constant c such that for all $\tilde{\theta}$ with $|\tilde{\theta} - \theta| < c$, we have

(i) $\left|\dfrac{\partial M(\tilde{Y}_{t-1}; \tilde{\theta})}{\partial \theta_i}\right| \leq \alpha_i(\tilde{Y}_{t-1}; \theta)$,

(ii) $\left|\dfrac{\partial^2 M(\tilde{Y}_{t-1}; \tilde{\theta})}{\partial \theta_i \partial \theta_j}\right| \leq \beta_{ij}(\tilde{Y}_{t-1}; \theta)$,

and

(iii) $\left|\dfrac{\partial^3 M(\tilde{Y}_{t-1}; \tilde{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_k}\right| \leq \gamma_{ijk}(\tilde{Y}_{t-1}; \theta)$,

$i, j, k = 1, \ldots, J$.

Remarks. Sufficient conditions for (C.0) for the type of processes considered in this paper are discussed, for instance, by Tong (1990) and the references therein. The condition (C.1) is satisfied in all the examples mentioned in Section 2. It is possible to relax (C.1) and permit $M(\tilde{Y}_{t-1}; \theta)$ to depend on unknown nuisance parameters. This would involve further regularity conditions in order to stabilize the Taylor expansion of $M(\tilde{Y}_{t-1}; \theta)$. For simplicity, we retain (C.1). Condition (C.2) implies that the elements of the derivative vector $\nabla M(\tilde{Y}_{t-1}; \theta)$ are linearly independent. Condition (C.3) is the usual Cramer type condition and is similar to the one imposed by Klimko and Nelson (1978). It can be shown that (C.3) is satisfied for all the examples mentioned in Section 2.

Theorem 3.1. *Under conditions* (C.0) *to* (C.3), *there exists a sequence of estimators* $\{\hat{\theta}_n\}$ *such that* $Q_n = \sum_{t=1}^{n} U_t^2(\theta)$ *attains a relative minimum at* $\hat{\theta}_n$ *and* $\hat{\theta}_n \overset{\text{a.s.}}{\to} \theta$, *as* $n \to \infty$.

Proof. Conditions (C.0) to (C.3) imply the relevant conditions (C.1 and C.2) of Tjostheim (1986) and hence the result follows from his Theorem 3.1.

Lemma 3.1. *Under* (C.0) *to* (C.3), *we have*

$$\sqrt{n}(\hat{\theta}_n - \theta) = V^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t(\theta)\nabla M(\tilde{Y}_{t-1};\theta) + o_p(1), \tag{3.3}$$

where V *is defined in* (C.2), $U_t(\theta) = Y_t - M(\tilde{Y}_{t-1};\theta)$ *and* $o_p(1)$ *is a term converging to zero in probability.*

Proof. First, note that we have

$$\sum_{t=1}^{n} U_t(\hat{\theta}_n)\nabla M(\tilde{Y}_{t-1};\hat{\theta}_n) = 0.$$

By the Taylor expansion of $U_t(\theta)$ at $\hat{\theta}_n$, we have

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\nabla M(\tilde{Y}_{t-1};\hat{\theta}_n)[U_t(\hat{\theta}_n) - U_t(\theta)]\\ &= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\nabla M(\tilde{Y}_{t-1};\hat{\theta}_n)\nabla M^T(\tilde{Y}_{t-1};\theta)(\hat{\theta}_n - \theta) + o_p(1).\end{aligned}$$

It is seen by (C.3) that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t(\theta)\nabla M(\tilde{Y}_{t-1};\hat{\theta}_n) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t(\theta)\nabla M(\tilde{Y}_{t-1};\theta) + o_p(1)$$

and

$$\frac{1}{n}\sum_{t=1}^{n}\nabla M(\tilde{Y}_{t-1};\hat{\theta}_n)\nabla M^T(\tilde{Y}_{t-1};\theta) = V + o_p(1).$$

Consequently,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t(\theta)\nabla M(\tilde{Y}_{t-1};\theta) = V\sqrt{n}(\hat{\theta}_n - \theta) + o_p(1),$$

which implies the result in (3.3), since V is positive definite.

Theorem 3.2. *Under conditions* (C.0) *to* (C.3) *we have*

$$\sqrt{n}(\hat{\theta}_n - \theta) \overset{d}{\to} N(0, V^{-1}AV^{-1}), \tag{3.4}$$

where

$$A = E_\theta\Big[U_t^2(\theta)(\nabla M(\tilde{Y}_{t-1};\theta))(\nabla M(\tilde{Y}_{t-1};\theta))^T\Big].$$

Proof. First, note that $\{U_t(\theta)\nabla M(\tilde{Y}_{t-1};\theta)\}$ forms a sequence of stationary and ergodic martingale differences with zero mean and covariance matrix A. The result then follows from Lemma 3.1 and an application of Billingsley's (1961) central limit theorem for martingales.

4. The Joint Asymptotic Distribution of the Residual Autocorrelations

Let

$$R_t = U_t(\hat{\theta}_n) = Y_t - M(\tilde{Y}_{t-1};\hat{\theta}_n), \tag{4.1}$$

denote the estimated prediction error which we shall also refer to as the residual. The sample autocorrelation function $\hat{\rho}_R(i)$ of lag i based on the residuals $\{R_t, t = 1,\ldots,n\}$ is defined by

$$\hat{\rho}_R(i) = \sum_{t=1}^{n-i}(R_t - \bar{R})(R_{t+i} - \bar{R})\Big/\sum_{t=1}^{n}(R_t - \bar{R})^2, \tag{4.2}$$

where $\bar{R} = n^{-1}\sum_{t=1}^{n} R_t$, and $i = 1,\ldots,l$, $l < n$. In this section, we shall derive the joint limit distribution of $\{\hat{\rho}_R(i), i = 1,\ldots,l\}$, as $n \to \infty$, for some fixed l.

First, we need some preliminary results.

Lemma 4.1. *Under* (C.0) *to* (C.3), *we have*

(i) $\dfrac{1}{\sqrt{n}}\sum_{t=1}^{n} R_t = O_p(1)$, i.e., *bounded in probability,*

(ii) $\dfrac{1}{n}\sum_{t=1}^{n}(R_t - \bar{R})^2 \xrightarrow{p} \sigma_u^2$,

where $\sigma_u^2 = EU_t^2(\theta)$.

(iii) $\dfrac{1}{\sqrt{n}}\sum_{t=1}^{n-i}(R_t - \bar{R})(R_{t+i} - \bar{R}) = \dfrac{1}{\sqrt{n}}\sum_{t=1}^{n-i} R_t R_{t+i} + o_p(1)$,

for $i = 1,\ldots,l < n$.

(iv) $\dfrac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t(\theta)\{M(\tilde{Y}_{t-1};\hat{\theta}_n) - M(\tilde{Y}_{t-1};\theta)\} \xrightarrow{p} 0$.

Proof. The proofs follow standard arguments. As an illustration, we shall verify the result in (i) and omit the proofs of (ii)-(iv). Consider

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(R_t - U_t(\theta))\\ =& -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[M(\tilde{Y}_{t-1};\hat{\theta}_n) - M(\tilde{Y}_{t-1};\theta)\right]\\ =& -\frac{1}{n}\sum_{t=1}^{n}\nabla M^T(\tilde{Y}_{t-1};\theta_n^*)\sqrt{n}(\hat{\theta}_n - \theta),\end{aligned}$$

where θ_n^* is between $\hat{\theta}_n$ and θ.
From (C.3), it follows that

$$\left|\nabla M(\tilde{Y}_{t-1};\theta_n^*)\right| \leq \sum_{i=1}^{J}\alpha_i(\tilde{Y}_{t-1};\theta), \tag{4.3}$$

for all sufficiently large n.
Hence, from Lemma 3.1,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}R_t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}U_t(\theta) + O_p(1).$$

It also follows from the central limit theorem for martingales (see, Billingsley (1961)) that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}U_t(\theta) \xrightarrow{d} N(0,\sigma_u^2), \tag{4.4}$$

where $\sigma_u^2 = EU_t^2(\theta)$. Consequently, $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}R_t = O_p(1)$.
For the main theorem, the following additional condition (C.4) is assumed to be satisfied.
(C.4) The stationary and ergodic sequence $\{U_t(\theta)\}$ has a **finite fourth moment.**

Theorem 4.1. *Under* (C.0) *to* (C.4), *we have*

$$\sqrt{n}[\hat{\rho}_R(1),\ldots,\hat{\rho}_R(l)]^T \xrightarrow{d} N_l(0,\Sigma), \tag{4.5}$$

where Σ *is the* $(l\times l)$ *matrix with the* (i,j)th *element given by*

$$\begin{aligned}\Sigma_{ij} =& \sigma_u^{-4}E\Big[U_t^2(\theta)\Big\{U_{t-i}(\theta) - \boldsymbol{m}_i^T V^{-1}\nabla M(\tilde{Y}_{t-1};\theta)\Big\}\\ &\cdot\Big\{U_{t-j}(\theta) - \boldsymbol{m}_j^T V^{-1}\nabla M(\tilde{Y}_{t-1};\theta)\Big\}\Big],\end{aligned} \tag{4.6}$$

with

$$m_i = E\Big[U_t(\theta)\nabla M(Y_{t+i-1};\theta)\Big] : J \times 1 \ vector,\ i = 1,\dots,l,$$

and

$$\sigma_u^2 = EU_t^2(\theta).$$

Proof. We first show that, for each $i = 1,\dots,l$,

$$\sqrt{n}\hat{\rho}_R(i) = \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n-i}(R_t-\bar{R})(R_{t+i}-\bar{R})}{\frac{1}{n}\sum_{t=1}^{n}(R_t-\bar{R})^2} \xrightarrow{d} N(0,\Sigma_{ii}), \tag{4.7}$$

and then derive the joint limit distribution of $(\hat{\rho}_R(1),\dots,\hat{\rho}_R(l))$ using the Cramer-Wold device (see, for instance, Serfling (1980, p.18)). By (ii) and (iii) of Lemma 4.1, we have

$$\sqrt{n}\hat{\rho}_R(i) = \sigma_u^{-2}\frac{1}{\sqrt{n}}\sum_{t=1}^{n-i}R_tR_{t+i} + o_p(1). \tag{4.8}$$

Now,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n-i}R_tR_{t+i}\\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n-i}\Big[U_t(\theta)-\Big\{M(\tilde{Y}_{t-1};\hat{\theta}_n)-M(\tilde{Y}_{t-1};\theta)\Big\}\Big]\\
&\quad\times\Big[U_{t+i}(\theta)-\Big\{M(\tilde{Y}_{t+i-1};\hat{\theta}_n)-M(\tilde{Y}_{t+i-1};\theta)\Big\}\Big]\\
&= \begin{cases}
\frac{1}{\sqrt{n}}\sum_{t=1}^{n-i}U_t(\theta)\Big[U_{t+i}(\theta)-\Big\{M(\tilde{Y}_{t+i-1};\hat{\theta}_n)-M(\tilde{Y}_{t+i-1};\theta)\Big\}\Big]\\
-\frac{1}{\sqrt{n}}\sum_{t=1}^{n-i}U_{t+i}(\theta)\Big\{M(\tilde{Y}_{t-1};\hat{\theta}_n)-M(\tilde{Y}_{t-1};\theta)\Big\}\\
+\frac{1}{\sqrt{n}}\sum_{t=1}^{n-i}\Big\{M(\tilde{Y}_{t-1};\hat{\theta}_n)-M(\tilde{Y}_{t-1};\theta)\Big\}\Big\{M(\tilde{Y}_{t+i-1};\hat{\theta}_n)-M(\tilde{Y}_{t+i-1};\theta)\Big\}.
\end{cases}
\end{aligned} \tag{4.9}$$

Using the result (iv) in Lemma 4.1, the second term in (4.9) goes to zero in probability and it follows via (4.3) that the third term is of order $O_p(n^{-1/2})$.

Consequently,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n-i} R_t R_{t+i} \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n-i} U_t(\theta)\Big[U_{t+i}(\theta) - \nabla M^T(\tilde{Y}_{t+i-1};\theta_n^*)(\hat{\theta}_n - \theta)\Big] + o_p(1) \qquad (4.10)\\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n-i} U_t(\theta)\Big[U_{t+i}(\theta) - \boldsymbol{m}_i^T\sqrt{n}(\hat{\theta}_n - \theta)\Big] + o_p(1). \qquad (4.11)
\end{aligned}$$

Note that (4.11) follows by noting that

$$\boldsymbol{m}_i = E\Big[U_t(\theta)\nabla M(\tilde{Y}_{t+i-1};\theta)\Big]$$

and by applying the ergodic theorem. We then have, by Lemma 3.1,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n-i} R_t R_{t+i} \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n-i} U_t(\theta)U_{t+i}(\theta) - \boldsymbol{m}_i^T V^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t(\theta)\nabla M(\tilde{Y}_{t-1};\theta) + o_p(1). \qquad (4.12)
\end{aligned}$$

By changing the subscript in the second term in (4.12), we have,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n-i} R_t R_{t+i} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-i} U_{t+i}(\theta)\Big[U_t(\theta) - \boldsymbol{m}_i^T V^{-1}\nabla M(\tilde{Y}_{t+i-1};\theta)\Big] + o_p(1). \qquad (4.13)$$

Note that

$$\Big\{U_{t+i}(\theta)\Big[U_t(\theta) - \boldsymbol{m}_i^T V^{-1}\nabla M(\tilde{Y}_{t+i-1};\theta)\Big]\Big\}$$

is a sequence of zero mean martingale differences. Consequently,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n-i} R_t R_{t+i} \xrightarrow{d} N(0, v_i),$$

where

$$v_i = E\Big[U_{t+i}(\theta)\Big\{U_t(\theta) - \boldsymbol{m}_i^T V^{-1}\nabla M(\tilde{Y}_{t+i-1};\theta)\Big\}\Big]^2, \qquad (4.14)$$

which in turn, is identical to $\sigma_u^4\Sigma_{ii}$. Hence, we conclude, via (4.8),

$$\sqrt{n}\hat{\rho}_R(i) \xrightarrow{d} N(0, \Sigma_{ii}), \quad i = 1, \ldots, l.$$

In order to obtain the joint limit distribution of $(\hat{\rho}_R(1), \ldots, \hat{\rho}_R(l))$, we consider, for $\boldsymbol{a} = (a_1, \ldots, a_l)$, any real vector,

$$\sqrt{n}a_1\hat{\rho}_R(1) + \cdots + \sqrt{n}a_l\hat{\rho}_R(l) = \sigma_u^{-2}\frac{1}{\sqrt{n}}\sum_{i=1}^{l} a_i\Big\{\sum_{t=1}^{n-i} R_t R_{t+i}\Big\} + o_p(1). \quad (4.15)$$

Note that (4.15) follows from (4.8). Using similar arguments as in (4.9), (4.10) and (4.11), we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{l} a_i\Big\{\sum_{t=1}^{n-i} R_t R_{t+i}\Big\} \\
&= \sum_{i=1}^{l} a_i\Big[\frac{1}{\sqrt{n}}\sum_{t=1}^{n-i} U_t(\theta)U_{t+i}(\theta) - \boldsymbol{m}_i^T V^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n-i} U_t(\theta)\nabla M(\tilde{Y}_{t-1};\theta)\Big] + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t(\theta)\Big[\sum_{i=1}^{l} a_i\Big\{U_{t-i}(\theta) - \boldsymbol{m}_i^T V^{-1}\nabla M(\tilde{Y}_{t-1};\theta)\Big\}\Big] + o_p(1).
\end{aligned}$$

It may be noted that

$$\Big\{U_t(\theta)\Big[\sum_{i=1}^{l} a_i\Big(U_{t-i}(\theta) - \boldsymbol{m}_i^T V^{-1}\nabla M(\tilde{Y}_{t-1};\theta)\Big)\Big]\Big\}$$

forms a sequence of zero mean martingale differences. We then have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{l} a_i\Big\{\sum_{t=1}^{n-i} R_t R_{t+i}\Big\} \xrightarrow{d} N(0, \boldsymbol{a}^T\Gamma\boldsymbol{a}), \quad (4.16)$$

where $\Gamma = ((\Gamma_{ij})) : l \times l$ matrix with

$$\begin{aligned}
\Gamma_{ij} &= E\Big[U_t^2(\theta)\Big\{U_{t-i}(\theta) - \boldsymbol{m}_i^T V^{-1}\nabla M(\tilde{Y}_{t-1};\theta)\Big\} \\
&\quad \cdot \Big\{U_{t-j}(\theta) - \boldsymbol{m}_j^T V^{-1}\nabla M(\tilde{Y}_{t-1};\theta)\Big\}\Big] \\
&= \sigma_u^4\Sigma_{ij}, \quad i, j = 1, \ldots, l.
\end{aligned}$$

Consequently, using the Cramer-Wold device, we finally obtain

$$\sqrt{n}(\hat{\rho}_R(1), \ldots, \hat{\rho}_R(l))^T \xrightarrow{d} N(0, \Sigma),$$

where Σ is defined in (4.10).

Remarks. (i) The existence of Σ_{ij} in (4.6) follows from (C.4). (ii) It should be noted that Σ_{ij} is free of "nuisance" parameters for linear time series. However, for nonlinear time series, Σ_{ij} may depend on "nuisance" parameters to be estimated.

(iii) For fixed coefficient models where $\{Z_t\}$ degenerates at zero, Σ_{ij} reduces to

$$\Sigma_{ij} = \delta_{ij} - \sigma_\varepsilon^{-2} \boldsymbol{m}_i^T V^{-1} \boldsymbol{m}_j, \tag{4.17}$$

where $\sigma_\varepsilon^2 = E\varepsilon_t^2$ and δ_{ij} is the Kronecker delta. This agrees with the result in Li (1992).

(iv) For the special case when $\{Y_t\}$ is linear AR(1),

$$Y_t = \theta Y_{t-1} + \varepsilon_t, \quad |\theta| < 1,$$

it can be shown that Σ_{ij} reduces to

$$\Sigma_{ij} = \begin{cases} \theta^{i+j} - \theta^{i+j-2}, & i \neq j \\ 1 - \theta^{2(i-1)}(1-\theta^2), & i = j. \end{cases} \tag{4.18}$$

The result in (4.18) agrees with that in Box and Pierce (1970). Furthermore, for linear AR(p), $p \geq 2$, it can be shown that

$$\Sigma_{ij} = \delta_{ij} - \eta_{ij}, \quad i, j = 1, \ldots, l,$$

with $\eta_{ij} = \sigma_\varepsilon^{-2} \boldsymbol{m}_i^T V^{-1} \boldsymbol{m}_j$ and η_{ij} satisfying the recurrsive equations:

$$\eta_{ij} = \sum_{v=1}^{p} \theta_v \eta_{i,j-v}, \quad \text{for} \quad j \geq p+1. \tag{4.19}$$

5. Tests of Fit

We now consider some tests of fit of the composite hypothesis: H_o: The sample $\{Y_t\}$, $t = 1, 2, \ldots, n$, comes from a stationary nonlinear time series given by (2.1). The tests are based on the estimated prediction errors $\{R_t\}$ and their sample autocorrelation functions (ACF), $\{\hat{\rho}_R(i), i = 1, \ldots, l\}$.

Tests based on the sample ACF

Consider the statistics

$$Q_n(l) = W_n^T \hat{\Sigma}^- W_n, \tag{5.1}$$

where $W_n = (\sqrt{n}\hat{\rho}_R(1), \ldots, \sqrt{n}\hat{\rho}_R(l))^T$, Σ^- is a generalized inverse of Σ given by (4.6), and $\hat{\Sigma}^-$ is a consistent estimate of Σ^-.

It can be shown, using Theorem 4.1 and standard arguments, that, as $n \to \infty$,

$$Q_n \xrightarrow{d} \chi_r^2, \quad \text{under} \quad H_0, \tag{5.2}$$

where $r = \text{rank}(\Sigma)$. The rejection region for the test of fit is therefore $\{Q_n \geq \chi^2_r\}$.

The computation of (5.1) requires the knowledge of Σ^- which may be complicated for some nonlinear models especially because Σ involves certain moments of the stationary distribution. It is therefore desirable to look for alternative simpler statistics. The portmanteau type statistic given below is useful in this context:

$$D_n(k) = n \sum_{i=1}^{k} \hat{\rho}^2_R(i). \tag{5.3}$$

If $\{\hat{\rho}_R(i)\}$, $i = 1, 2, \ldots$, were to behave like the ACF's of i.i.d. observations, one would expect D_n to have χ^2_{k-J} as the limiting distribution. Theorem 4.1, however, shows that the covariance matrix Σ of $\{\sqrt{n}\hat{\rho}_R(i), i = 1, \ldots, k\}$ is different from the one corresponding to the ACF's of i.i.d. observations. However, this difference is only pronounced for lower lags, mainly $i = 1$, and approaches zero for higher lags. One may therefore continue to use D_n as an approximate χ^2_{k-J} random variable. This is demonstrated in the simulation in the next section.

A test based on the prediction errors

A test statistic based on $\{R_t\}$ can be derived using the approach of Basawa (1987). Suppose we have observed the sample $\{Y_1, \ldots, Y_n, Y_{n+1}, \ldots, Y_{n+k}\}$. First, pretend that $\{Y_{n+i}, i = 1, \ldots, k\}$ are unknown and find their one-step predictors successively, i.e. find

$$M(\tilde{Y}_{n+i-1}; \theta) = E_\theta(Y_{n+i} | A_{n+i-1}), \quad i = 1, \ldots, k.$$

The corresponding prediction errors are

$$U_{n+i}(\theta) = Y_{n+i} - M(\tilde{Y}_{n+i-1}; \theta), \quad i = 1, \ldots, k,$$

and the estimated prediction errors are $R_{n+i} = U_{n+i}(\hat{\theta}_n)$, where $\hat{\theta}_n$ is the CLS estimator of θ based on $\{Y_1, \ldots, Y_n\}$. Even though it is possible to update $\hat{\theta}$, using the first $n + i - 1$ observations, we ignore this possibility since it will not affect the asymptotic distributions. We have

$$R_{n+i} = U_{n+i}(\theta) + (\nabla U_{n+i}(\theta))^T_{\theta^*}(\hat{\theta}_n - \theta), \tag{5.4}$$

where θ^* lies between θ and $\hat{\theta}_n$. From, Theorem 3.2, we have $\sqrt{n}(\hat{\theta}_n - \theta) = O_p(1)$, and, for all the examples mentioned in Section 2, it can be shown that $n^{-1/2}(\nabla U_{n+i}(\theta)) = o_p(1)$, for each $i = 1, \ldots, k$. Consequently, the second term on the right of (5.4) is $o_p(1)$. The limit distribution of any statistic based on $\{R_{n+i}, i = 1, \ldots, k\}$ will therefore be the same as that based on $\{U_{n+i}(\theta), i = 1, \ldots, k\}$.

To proceed further, we assume that $\{\varepsilon_t\}$, and $\{Z_t\}$ in (2.1) are normally distributed with zero mean and known variances $\sigma_\varepsilon^2, \sigma_z^2$. As before, denote

$$M(\tilde{Y}_{n+i-1};\theta) = E(Y_{n+i} \mid A_{n+i-1}),$$

and let

$$\tau_{n+i}^2 = \mathrm{Var}(Y_{n+i} \mid A_{n+i-1}). \tag{5.5}$$

For instance, for the RCA model in Ex. 1, it can be seen that

$$\tau_{n+i}^2 = \sigma_\varepsilon^2 + \sigma_z^2 \sum_{j=1}^{p} Y_{n+i-j}^2. \tag{5.6}$$

It will be assumed for simplicity that σ_ε^2 and σ_z^2 are known. Conditionally on A_{n+i-1}, $\{U_{n+i}(\theta), i = 1,\dots,k\}$ are independent $N(0,\tau_{n+i}^2)$. It is seen that

$$\sum_{i=1}^{k} U_{n+i}^2(\theta)\tau_{n+i}^{-2} \stackrel{d}{=} \chi_k^2, \quad \text{under } H_0. \tag{5.7}$$

Consider the statistic

$$\begin{aligned} W(n) &= \sum_{i=1}^{k} R_{n+i}^2\tau_{n+i}^{-2} = \sum_{i=1}^{k} U_{n+i}^2(\theta)\tau_{n+i}^{-2} + o_p(1) \\ &\stackrel{d}{\to} \chi_k^2, \quad \text{under } H_0. \end{aligned} \tag{5.8}$$

The rejection region of the test is therefore given by $\{\sum_{i=1}^k R_{n+i}^2\tau_{n+i}^{-2} \geq \chi_k^2\}$.

As seen above in (5.6), τ_{n+i}^2 typically depends on nuisance parameters σ_ε^2 and σ_z^2. Under the normality assumption one can estimate $\theta, \sigma_\varepsilon^2$ and σ_z^2 by the maximum likelihood (ML) method. The likelihood function based on the first n observations $\{Y_1,\dots,Y_n\}$ is given by

$$L_n(\theta,\sigma_\varepsilon^2,\sigma_z^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \prod_{i=1}^{n} \tau_i^{-1} \exp\left\{-\frac{1}{2}\tau_i^{-2}U_i^2(\theta)\right\} \tag{5.9}$$

where $U_i(\theta) = Y_i - E_\theta(Y_i|A_{i-1})$ and $\tau_i^2 = \mathrm{Var}_\theta(Y_i|A_{i-1})$. The ML estimates $\hat{\theta}, \hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_z^2$ can be obtained by maximizing $L_n(\theta,\sigma_\varepsilon^2,\sigma_z^2)$. The asymptotic properties of the ML estimators are discussed by Hwang and Basawa (1991).

In Table 5.1 below, we present some simulation results involving the distribution of the statistic $W(n)$ (Eq. 5.8) for the RCA model of Ex. 1. For these simulations, the true $\theta = .250$, and the variances of the Z and error distributions are .25 and 1.00, respectively, so the stability condition: $0 < \theta^2 + \sigma_z^2 < 1$, is satisfied. For each of the simulations, the process was run for n time periods ($n = 50, 100, 200$, and 400), MLE estimates of θ were obtained, and 1-step-ahead predictions of $Y(n+1)$, $Y(n+2)$, and $Y(n+3)$ were made. From this, the $W(n)$ statistic (Eq. 5.8) was calculated. This process was repeated randomly for 400 simulations for each n, once assuming that the variances of the Z and error distributions are known, and once assuming that these variances (and θ) are unknown.

Table 5.1 displays the results of these simulations. The mean and variance of $W(n)$ over the 400 simulations are given, along with the proportion of the simulations whose $W(n)$ values exceeded the 95th percentile [7.815] of a Chi-squared(3) distribution, since $W(n)$ should be distributed as a Chi-squared(3) random variable under the null hypothesis. As can be seen, the means, variances, and exceeding probability are close to the expected 3.0, 6.0, and 0.05 for all values of n when the σ_z^2 and the error variance are assumed known. For the unknown variances cases, the fit is not quite as good, but seems fair for $n \geq 200$. In both cases, as expected, the fit of $W(n)$ to the Chi-squared(3) distribution improves as n increases, and is very good for large n.

Table 5.1. Results of 400 simulations of $W(n)$ from RCA model

n	Known	Mean	Variance	Pr > 7.815
(Expected)		3.00	6.00	.0500
50	Yes	3.2143	8.1389	.0700
50	No	3.6503	17.5308	.0900
100	Yes	2.9555	5.3833	.0450
100	No	3.1283	6.9586	.0825
200	Yes	3.1081	6.9387	.0700
200	No	3.2763	8.8892	.0875
400	Yes	2.9775	6.0576	.0525
400	No	3.0266	6.3258	.0575

6. Some Simulation Results for a Threshold Autoregressive Model

A simulation study was carried out to compute the limiting variances of the residual ACF's of the following TAR(1) model.

$$Y_t = \theta_1 Y_{t-1}^+ + \theta_2 Y_{t-1}^- + \varepsilon_t,$$

where $\theta_1 < 1$, $\theta_2 < 1$ and $\theta_1\theta_2 < 1$.

It can be shown that

$$\begin{aligned} \Sigma_{ij} &= \delta_{ij} - \eta_{ij} \\ V &= E_\theta \begin{bmatrix} (Y_{t-1}^+)^2 & 0 \\ 0 & (Y_{t-1}^-)^2 \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned}\eta_{ij} &= \sigma_\varepsilon^{-2}[E_\theta\varepsilon_t Y^+_{t+i-1}][E_\theta\varepsilon_t(Y^+_{t+j-1})^2]/E_\theta(Y^+_{t-1})^2 \\ &\quad + \sigma_\varepsilon^{-2}[E_\theta\varepsilon_t Y^-_{t+i-1}][E_\theta\varepsilon_t Y^-_{t+j-1}]/E_\theta(Y^-_{t-1})^2, \end{aligned} \tag{6.1}$$

where the expectation is taken under the stationary distribution of the process. Since the stationary moment structure is not explicity available as yet for the TAR(1) process, the expressions $E_\theta\varepsilon_t Y^+_{t+i-1}$ and $E_\theta(Y^-_{t-1})^2$ are replaced by their consistent estimators $n^{-1}\sum_{t=1}^n(Y_t - \theta_1 Y^+_{t-1} - \theta_2 Y^-_{t-1})Y^+_{t+i-1}$ and $n^{-1}\sum_{t=1}^n(Y^-_{t-1})^2$, respectively. Some results are summarized in Table 6.1. Here $\{\varepsilon_t\}$ are drawn from $N(0,1)$ and a total of 16 points of (θ_1,θ_2) are chosen with various δ-values, $\delta = 0, 0.1, 0.5, 0.6, 1.0$ and 2.25 where $\delta = |\theta_1 - \theta_2|$. 500 observations are simulated for each pair (θ_1,θ_2) using SAS (Statistical Analysis System) and the limiting variances Σ_{ii}, $i = 1,\ldots,8$, are calculated via (6.1). As in the case of the linear time series models, see Box and Pierce (1970), we find that for higher lags ($i = 6, 7, 8$, say), Σ_{ii} is very close to 1, and hence, only the smaller lag ACF's need to be carefully checked for the diagnostic purposes. (See also Li (1992) for similar simulation results.)

Table 6.1. Asymptotic variances of the residual autocorrelations of lag $i = 1,\ldots,8$ ($n = 500$)

δ	(θ_1,θ_2)	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$
0	(0.1, 0.1)	0.01241	0.98478	0.999968	0.998528	0.998228	0.999618	0.997218	0.998413
0	(0.5, 0.5)	0.26403	0.84074	0.944316	0.990857	0.985062	0.992734	0.998650	0.999354
0	(0.9, 0.9)	0.80469	0.83415	0.86283	0.97787	0.90915	0.92333	0.92717	0.93866
0.1	(0.1, 0.2)	0.02957	0.96980	0.997818	0.999036	0.997802	0.996474	0.996302	0.998988
0.1	(0.5, 0.6)	0.33595	0.78066	0.93517	0.96489	0.991905	0.987349	0.989888	0.993412
0.1	(0.8, 0.9)	0.71358	0.78458	0.85613	0.90031	0.93745	0.93716	0.94268	0.965505
0.5	(0.1, 0.6)	0.27399	0.85722	0.96272	0.97776	0.98265	0.994394	0.993935	0.997882
0.6	(0.3, 0.9)	0.60766	0.79660	0.89420	0.94024	0.96297	0.97048	0.981616	0.991414
1.0	(0.2, −0.8)	0.206099	0.963969	0.999093	0.999845	0.999826	0.994209	0.999485	0.996256
1.0	(0.5, −0.5)	0.10199	0.96512	0.991288	0.990060	0.991673	0.996973	0.998110	0.991251
1.0	(0.9, −0.1)	0.68044	0.88656	0.91696	0.94571	0.976153	0.988449	0.991028	0.992399
1.0	(−0.8, 0.2)	0.090527	0.970646	0.994836	0.997634	0.992266	0.998883	0.998548	0.994695
1.0	(−0.5, 0.5)	0.16492	0.96925	0.999815	0.998048	0.996525	0.996806	0.994282	0.999068
1.0	(−0.1, 0.9)	0.62588	0.84094	0.91487	0.96209	0.95942	0.981761	0.984655	0.990249
2.25	(−2.0, 0.25)	0.363910	0.904802	0.998202	0.999462	0.996091	0.999671	0.997317	0.996348
2.25	(0.25, −2.0)	0.397701	0.88561	0.998098	0.99181	0.999722	0.995834	0.997216	0.996857

In order to study the effect of smaller sample sizes on the asymptotic variance and the associated "power", seven different TAR models were used with sample sizes $n = 50, 100$ and 200. Each entry in the table below is the result from 250 simulations. Note that $\delta = |\theta_1 - \theta_2|$ and $\delta = 0$ indicates an AR(1) model. The sense in which we have used the term "power" is explained below.

Table 6.2 shows simulations for $\hat{\rho}(k)$, $(k = 1,2,3)$ as defined by Equation (4.2). The first value shown within each cell is the mean over the simulations of the asymptotic variance. These should in theory converge to the values denoted as Σ_{ii} in Equation (4.6). For $k > 1$, the variance is again very close to 1.00. For each simulation, in addition to estimating θ_1 and θ_2 by their ML estimators under TAR (and using these to calculate $\hat{\rho}$), we also obtained estimates of θ (and associated means and variances for $\hat{\rho}$) assuming that the true model was AR(1). Using the mean (always very near zero) and variances of $\hat{\rho}(k)$ as estimated under TAR, and assuming normality, we obtained a symmetric 95% confidence region for $\hat{\rho}(k)$. We then calculated the probability that a $\hat{\rho}(k)$ value estimated by the AR(1) process (assuming normality and using the mean and variance of $\hat{\rho}(k)$ estimated by AR(1) over the simulations) would fall in the rejection region of the TAR process. This is the value called "power" entered in the second column of Table 6.2. As expected, the power is approximately 5% for $\hat{\rho}(2)$ or $\hat{\rho}(3)$, since there is no difference in the behavior of $\hat{\rho}(k)$ between the TAR and AR(1) methods in this case. For $\hat{\rho}(1)$, power increases as n increases and is somewhat positively associated with increases in δ. This makes sense, since large values of δ are associated with models where there are two quite different θ's, so using AR(1) methods should fail in those cases.

Table 6.2. Results of 250 simulations of $\hat{\rho}(k)$; $k = 1,2,3$

$\hat{\rho}(1)$: Asymptotic variance/Power

δ	θ_1	θ_2	$n = 50$		$n = 100$		$n = 200$	
0.0	.5	.5	0.2049	0.0931	0.3116	0.0810	0.2317	0.0718
0.3	.6	.3	0.2657	0.1255	0.2250	0.1708	0.2188	0.2169
0.3	−.6	−.3	0.2332	0.0644	0.2084	0.0578	0.2328	0.0759
0.5	0	.5	0.1504	0.2120	0.1633	0.3038	0.1510	0.4596
0.6	.3	.9	0.5431	0.1114	0.5143	0.2367	0.5875	0.3677
1.0	.5	−.5	0.1771	0.5243	0.1740	0.7521	0.1701	0.9558
2.25	.25	−2.	0.3642	0.8788	0.3080	0.9995	0.3528	1.0000

$\hat{\rho}(2)$: Asymptotic variance/Power

δ	θ_1	θ_2	$n = 50$		$n = 100$		$n = 200$	
0.0	.5	.5	0.7477	0.0534	0.9226	0.0498	0.6643	0.0498
0.3	.6	.3	0.8492	0.0556	0.7452	0.0571	0.8197	0.0553
0.3	−.6	−.3	0.7281	0.0464	0.7922	0.0523	0.8571	0.0488
0.5	0	.5	0.8475	0.0554	0.9366	0.0503	0.9411	0.0537
0.6	.3	.9	0.8077	0.0619	0.8061	0.0608	0.8119	0.0912
1.0	.5	−.5	0.9337	0.0568	0.9444	0.0561	0.9181	0.0428
2.25	.25	−2.	0.7580	0.0340	0.8855	0.0611	1.0181	0.0495

$\hat{\rho}(3)$: Asymptotic variance/Power

δ	θ_1	θ_2	$n = 50$		$n = 100$		$n = 200$	
0.0	.5	.5	0.9582	0.0496	0.9813	0.0524	0.8953	0.0503
0.3	.6	.3	0.8852	0.0507	0.9127	0.0473	0.9307	0.0519
0.3	−.6	−.3	0.8712	0.0492	0.9993	0.0463	1.0208	0.0524
0.5	0	.5	0.9370	0.0525	0.8810	0.0522	1.1265	0.0508
0.6	.3	.9	0.9364	0.0546	0.6847	0.0785	1.0603	0.0687
1.0	.5	−.5	0.9192	0.0416	0.9171	0.0523	0.8260	0.0564
2.25	.25	−2.	0.8510	0.0599	0.8097	0.0763	0.9005	0.0922

Table 6.3 gives some computations for the statistic $Q_n(3)$ of Equation (5.1). The three entries in each cell are the simulated (over 250 simulations) mean, variance, and $\Pr\{Q_n >$ Chi-Squared $(3, .95)\}$ values of the $Q_n(3)$ statistic. The convergent distribution is theoretically Chi-squared with 3 d.f., so the three values within each cell should be 3.0, 6.0, and .05 respectively. The seven rows indicate the seven model configurations examined previously (presented in the same order here), while the four columns are the initial sample size ($n = 50, 100, 200, 400$) from which the parameters are estimated.

In each case, the distribution appears to get closer to Chi-squared(3) as n becomes larger. For 250 independent simulations, 95% C.I. for the three parameters are approximately (assuming Chi-squared(3) is true distribution):

Mean: [2.70, 3.30]

Variance: [4.95, 7.05]

Tail Area: [.023, .077].

Thus, configurations B-G are "indistinguishable" from Chi-squared(3) with respect to Mean or Tail criteria by $n = 200$, and by all three criteria by $n = 400$. For configuration A, the fit is clearly getting better as n increases, but a much larger sample size than $n = 400$ is needed for the result to be approximately correct. The parameter $\delta = |\theta_1 - \theta_2|$ seems to be a key determinant of rapidity of convergence. As δ increases (i.e. as model alphabetical order increases from A, B,. . .,G) the initial sample size (n) needed for the Chi-squared(3) approximation to work becomes smaller. For example, models F and G are approximately Chi-squared(3) by all three criteria even when $n = 50$, whereas model A needs an n well over 400 for convergence to be approximated.

Table 6.3. Results of 250 simulations of $Q_n(3)$
$Q_n(3)$
Mean variance tail

M	$n = 50$			$n = 100$			$n = 200$			$n = 400$		
A	8.557	99.426	0.336	5.740	67.334	0.212	4.393	21.025	0.140	3.602	12.990	0.096
B	4.320	22.585	0.132	3.604	12.724	0.112	3.135	8.446	0.072	3.129	6.834	0.076
C	4.696	39.482	0.148	3.519	17.980	0.068	3.042	9.233	0.060	2.614	5.110	0.032
D	3.372	8.505	0.056	3.306	9.058	0.104	3.109	7.120	0.044	2.934	7.121	0.060
E	3.121	9.204	0.088	2.704	5.633	0.040	2.878	4.516	0.036	3.032	6.439	0.052
F	3.289	7.611	0.052	3.234	7.001	0.068	3.193	7.152	0.060	3.317	6.518	0.056
G	2.829	5.259	0.036	2.834	4.517	0.036	3.121	5.886	0.044	3.056	5.416	0.032

Acknowledgements

I. V. Basawa's work was partially supported by a grant from the Office of Naval Research. We thank the referees for their constructive suggestions.

References

Basawa, I. V. (1987). Asymptotic distributions of prediction errors and related tests of fit for nonstationary processes. *Ann. Statist.* **15**, 46–58.

Billingsley, P. (1961). The Lindeberg-Levy theorem for martingales. *Proc. Amer. Math. Soc.* **12**, 788–792.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc.* **65**, 1509–1526.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd edition. Springer-Verlag, New York.

Brockwell, P. J., Liu, J. and Tweedie, R. L. (1992). On the existence of stationary threshold autoregressive moving-average processes. *J. Time Series Anal.* **13**, 95–107.

Durbin, J. (1970). Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica* **38**, 410–421.

Feigin, P. D. and Tweedie, R. L. (1985). Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments. *J. Time Series Anal.* **6**, 1–14.

Hwang, S. Y. and Basawa, I. V. (1991). Large sample inference for conditional exponential families with applications to nonlinear time series. To appear in *J. Statist. Plann. Inference.*

Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6**, 629–642.

Li, W. K. (1992). On the asymptotic standard errors of residual autocorrelations in nonlinear time series modelling. *Biometrika* **79**, 435–437.

Nicholls, D. F. and Quinn, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction. Lecture Notes in Statist.* **11**, Springer-Verlag, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley, New York.

Tjostheim, D. (1986). Estimation in nonlinear time series models. *Stoch. Process. Appl.* **21**, 251–273.

Tong, H. (1990). *Non-linear Time Series.* Oxford University Press, Oxford.

Department of Statistics, University of Georgia, Athens, GA 30602, U.S.A.

(Received July 1992; accepted August 1993)