© Journal of the American Statistical Association
December 1970, Volume 65, Number 332
Theory and Methods Section

Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models

G. E. P. BOX and DAVID A. PIERCE*

Many statistical models, and in particular autoregressive—moving average time series models, can be regarded as means of transforming the data to white noise, that is, to an uncorrelated sequence of errors. If the parameters are known exactly, this random sequence can be computed directly from the observations; when this calculation is made with estimates substituted for the true parameter values, the resulting sequence is referred to as the "residuals," which can be regarded as estimates of the errors.

If the appropriate model has been chosen, there will be zero autocorrelation in the errors. In checking adequacy of fit it is therefore logical to study the sample autocorrelation function of the residuals. For large samples the residuals from a correctly fitted model resemble very closely the true errors of the process; however, care is needed in interpreting the serial correlations of the residuals. It is shown here that the residual autocorrelations are to a close approximation representable as a *singular* linear transformation of the autocorrelations of the errors so that they possess a singular normal distribution. Failing to allow for this results in a tendency to overlook evidence of lack of fit. Tests of fit and diagnostic checks are devised which take these facts into account.

1. INTRODUCTION

An approach to the modeling of stationary and non-stationary time series such as commonly occur in economic situations and control problems is discussed by Box and Jenkins [4, 5], building on the earlier work of several authors beginning with Yule [19] and Wold [17], and involves iterative use of the three-stage process of identification, estimation, and diagnostic checking. Given a discrete time series $z_t, z_{t-1}, z_{t-2}, \cdots$ and using B for the backward shift operator such that $Bz_t = z_{t-1}$, the general autoregressive—integrated moving average (ARIMA) model of order (p, d, q) discussed in [4, 5] may be written

$$\phi(B)\nabla^d z_t = \theta(B)a_t \tag{1.1}$$

where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots \theta_q B^q$, $\{a_t\}$ is a sequence of independent normal deviates with common variance σ_a^2, to be referred to as "white noise," and where the roots of $\phi(B) = 0$ and $\theta(B) = 0$ lie outside the unit circle. In other words, if $w_t = \nabla^d z_t = (1-B)^d z_t$ is the dth difference of the series z_t, then w_t is the stationary, invertible, mixed autoregressive (AR)—moving average (MA) process given by

$$w_t = \sum_{i=1}^{p} \phi_i w_{t-i} - \sum_{j=1}^{q} \theta_j a_{t-j} + a_t,$$

and permitting $d > 0$ allows the original series to be (homogeneously) nonsta-

* G. E. P. Box is professor of statistics, University of Wisconsin. David A. Pierce is on leave from the Department of Statistics, University of Missouri, Columbia, as statistician, Research Department, Federal Reserve Bank of Cleveland. This work was supported jointly by the Air Force Office of Scientific Research under Grant AFOSR-69-1803 and by the U. S. Army Research Office under Grant DA-ARO-D-31-124-G917.

tionary. In some instances the model (1.1) will be appropriate after a suitable transformation is made on z; in others z may represent the noise structure after allowing for some systematic model.

This general class of models is too rich to allow immediate fitting to a particular sample series $\{z_t\} = z_1, z_2, \cdots, z_n$, and the following strategy is therefore employed:

1. A process of identification is used to find a smaller subclass of models worth considering to represent the stochastic process.
2. A model in this subclass is fitted by efficient statistical methods.
3. An examination of the adequacy of the fit is made.

The object of the third or diagnostic checking stage is not merely to determine whether there is evidence of lack of fit but also to suggest ways in which the model may be modified when this is necessary. Two basic methods for doing this are suggested:

Overfitting. The model may be deliberately overparameterized in a way it is feared may be needed and in a manner such that the entertained model is obtained by setting certain parameters in the more general model at fixed values, usually zero. One can then check the adequacy of the original model by fitting the more general model and considering whether or not the additional parameters could reasonably take on the specified values appropriate to the simpler model.

Diagnostic checks applied to the residuals. The method of overfitting is most useful where the nature of the alternative feared model is known. Unfortunately, this information may not always be available, and less powerful but more general techniques are needed to indicate the way in which a particular model might be wrong. It is natural to consider the stochastic properties of the residuals $\hat{a} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_n)'$ calculated from the sample series using the model (1.1) with estimates $\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p$; $\hat{\theta}_1, \hat{\theta}_2, \cdots \hat{\theta}_q$ substituted for the parameters. In particular their autocorrelation function

$$\hat{r}_k = \sum \hat{a}_t \hat{a}_{t-k} / \sum \hat{a}_t^2 \tag{1.2}$$

may be studied.

Now if the model were appropriate and the a's for the particular sample series were calculated using the *true* parameter values, then these a's would be uncorrelated random deviates, and their first m sample autocorrelations $r = (r_1, r_2, \cdots, r_m)'$, where m is small relative to n and

$$r_k = \frac{\sum a_t a_{t-k}}{\sum a_t^2}, \tag{1.3}$$

would for moderate or large n possess a multivariate normal distribution [1]. Also it can readily be shown that the $\{r_k\}$ are uncorrelated with variances

$$V(r_k) = \frac{n-k}{n(n+2)} \approx 1/n, \tag{1.4}$$

from which it follows in particular that the statistic $n(n+2)\sum_{k=1}^{m}(n-k)^{-1}r_k^2$ would for large n be distributed as χ^2 with m degrees of freedom; or as a further approximation,

$$n\sum_{k=1}^{m} r_k^2 \sim \chi_m^2. \tag{1.5}$$

It is tempting to suppose that these same properties might to a sufficient approximation be enjoyed by the $\hat{r}$'s from the *fitted* model; and diagnostic checks based on this supposition were suggested by Box and Jenkins [4] and Box, Jenkins, and Bacon [6]. If this assumption were warranted, approximate standard errors of $1/\sqrt{n}$ [or more accurate standard errors of $\sqrt{n-k}/n(n+2)$] could be attached to the $\hat{r}$'s and a quality-control-chart type of approach used, with particular attention being paid to the $\hat{r}$'s of low order for the indication of possible model inadequacies. Also it might be supposed that Equation (1.5) with $\hat{r}$'s replacing r's would still be approximately valid, so that large values of this statistic would place the model under suspicion.

It was pointed out by Durbin [10], however, that this approximation is invalid when applied to the residual autocorrelations from a fitted autoregressive model. For example, he showed that $\hat{r}_1$ calculated from the residuals of a first order autoregressive process could have a much smaller variance than r_1 for white noise.

The present paper therefore considers in some detail the properties of the $\hat{r}$'s and in particular their covariance matrix, both for AR processes (Sections 2 and 3) and for MA and ARIMA processes (Section 5). This is done with the intention of obtaining a suitable modification to the above diagnostic checking procedures (Sections 4 and 5.3)

The problem of testing fit in time series models has been considered previously by several authors. Quenouille [14][1] developed a large-sample procedure for AR processes based on their sample partial autocorrelations, which possesses the same degree of accuracy as the present one.[2] Quenouille's test was subsequently extended [3, 15, 18] to cover MA and mixed models. Whittle [16] proposed tests based on the likelihood ratio and resembling the overfitting method above. The present procedure (a) is a unified method equally applicable to AR, MA, and general ARIMA models, (b) is motivated by the intuitive idea that the residuals from a correct fit should resemble the true errors of the process, and (c) can be used to suggest particular modifications in the model when lack of fit is found [5].

2. DISTRIBUTION OF RESIDUAL AUTOCORRELATIONS FOR THE AUTOREGRESSIVE PROCESS

In this section we obtain the joint large-sample distribution of the residual autocorrelations $\hat{\boldsymbol{r}} = (\hat{r}_1, \cdots, \hat{r}_m)'$ where $\hat{r}_k$ is given by (1.2), for an autoregressive process. This is done by first setting forth some general properties of AR processes, using these to obtain a set of linear constraints (2.9) satisfied by the $\{\hat{r}_k\}$, and then approximating $\hat{r}_k$ by a first order Taylor expansion (2.22) about the white noise autocorrelation r_k. Finally, these results are combined in matrix form to establish a linear relationship (2.27) between $\hat{\boldsymbol{r}}$ and $\boldsymbol{r}$ analogous to that between the residuals and true errors in a standard regression model, from which the distribution (2.29) of $\hat{\boldsymbol{r}}$ readily follows. Subsections 2.5–2.7 then discuss examples and applications of this distribution.

---

[1] See also [11].

[2] The authors are grateful to a referee for this observation.

2.1 The Autoregressive Process

The general AR process of order p,

$$\phi(B)y_t = a_t, \tag{2.1}$$

where B, $\phi(B)$, and $\{a_t\}$ are as in (1.1), can also be expressed as a moving average of infinite order by writing $\psi(B) = \phi^{-1}(B) = (1+\psi_1 B+\psi_2 B^2+ \cdots)$ to obtain

$$y_t = \psi(B)a_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}, \tag{2.2}$$

where $\psi_0 = 1$. By equating coefficients in the relation $\psi(B)\cdot\phi(B) = 1$, it is seen that the ψ's and ϕ's satisfy the relation

$$\psi_\nu = \begin{cases} \phi_1\psi_{\nu-1} + \cdots + \phi_{\nu-1}\psi_1 + \phi_\nu, & \nu \leq p \\ \phi_1\psi_{\nu-1} + \cdots + \phi_p\psi_{\nu-p}, & \nu \geq p. \end{cases} \tag{2.3}$$

Therefore by setting $\psi_\nu = 0$ for $\nu < 0$, we have

$$\psi_0 = 1; \quad \phi(B)\psi_\nu = 0, \qquad \nu \neq 0. \tag{2.4}$$

Suppose then we have a series $\{y_t\}$ generated by the model (2.1) or (2.2), where in general $y_t = \nabla^d z_t$ can be the dth difference ($d = 0, 1, 2, \cdots$) of the actual observations. Then for given values $\dot{\phi} = (\dot{\phi}_1, \cdots, \dot{\phi}_p)'$ of the parameters we can define

$$\dot{a}_t = a_t(\dot{\phi}) = y_t - \dot{\phi}_1 y_{t-1} - \cdots - \dot{\phi}_p y_{t-p} = \dot{\phi}(B)y_t \tag{2.5}$$

and the corresponding autocorrelation

$$\dot{r}_k = r_k(\dot{\phi}) = \frac{\sum \dot{a}_t\dot{a}_{t-k}}{\sum \dot{a}_t^2}. \tag{2.6}$$

Thus, in particular,

1. $a_t(\phi) = a_t$ as in (2.1), (2.2);
2. $a_t(\hat{\phi}) = \hat{a}_t$ are the residuals when (2.1) is fitted and least squares estimated $\hat{\phi}$ obtained; and
3. $r_k(\hat{\phi})$ and $r_k(\phi)$ are respectively the residual and white noise autocorrelations (1.2) and (1.3).

2.2 Linear Constraints on the $\hat{r}$'s

It is known that the residuals $\{\hat{a}_t\}$ above satisfy the orthogonality conditions

$$\sum_{t=p+1}^{n} \hat{a}_t y_{t-j} = 0, \qquad 1 \leq j \leq p. \tag{2.7}$$

Therefore if we let

$$\hat{\psi}(B) = \hat{\phi}^{-1}(B) = (1 - \hat{\phi}_1 B - \cdots - \hat{\phi}_p B^p)^{-1}, \tag{2.8}$$

then $y_t = \hat{\psi}(B)\hat{a}_t$, and from (2.7) we have

$$\begin{aligned} 0 &= \sum_t \sum_k \hat{\psi}_k \hat{a}_t \hat{a}_{t-k-j} \\ &= \sum_k \hat{\psi}_k \hat{r}_{k+j} \\ &= \sum \psi_k \hat{r}_{k+j} + O_p(1/n) \end{aligned} \tag{2.9}$$

where the symbol introduced in (2.9) denotes "order in probability" as defined in [13].

In leading up to (2.9) we have presumably summed an infinite number of autocorrelations from a finite series. However since $\{y_t\}$ is stationary we have $\psi_k \to 0$ as k becomes large; and unless ϕ is extremely close to the boundary of the stationarity region, this dying off of ψ_k is fast so that the summation can generally be stopped at a value of k much less than n. More precisely, we are assuming that n is larger than a fixed number N and for such n there exists a sequence of numbers m_n such that

(a) all ψ_j where $j \geq m_n - p$ are of order $1/\sqrt{n}$ or smaller, and
(b) the ratio m_n/n is itself of order $1/\sqrt{n}$.

Then in (2.9) and in all following discussion the error in stopping the summations at $k=m$ (we write m for m_n in the sequel) can to the present degree of approximation be ignored; and (b) also ensures that "end effects" (such as there being only $n-k$ terms summed in the numerator of $\hat{r}_k$ compared with n terms in the denominator) can also be neglected.

2.3 Linear Expansion of $\hat{r}_k$ about r_k

The root mean square error of $\hat{\phi}_j$, $1 \leq j \leq p$, defined by $\sqrt{E(\phi_j - \hat{\phi}_j)^2}$, is of order $1/\sqrt{n}$, and we can therefore approximate $\hat{r}_k$ by a first order Taylor expansion about $\hat{\phi} = \phi$ (evaluating the derivatives, however, at $\hat{\phi}$ rather than ϕ in order to obtain the simplification (2.12) below). Thus

$$\hat{r}_k = r_k + \sum_{j=1}^{p} (\phi_j - \hat{\phi}_j)\hat{\delta}_{jk} + O_p(1/n), \tag{2.10}$$

where

$$\hat{\delta}_{jk} = -\left.\frac{\partial \dot{r}_k}{\partial \dot{\phi}_j}\right|_{\dot{\phi}=\hat{\phi}}. \tag{2.11}$$

Now

$$\frac{\partial}{\partial \dot{\phi}_j}\left[\sum \dot{a}_t^2\right] = 0 \quad \text{at } \dot{\phi} = \hat{\phi}, \tag{2.12}$$

so that

$$\hat{\delta}_{jk} = -\left[\sum \hat{a}_t^2\right]^{-1} \left.\frac{\partial c_k}{\partial \dot{\phi}_j}\right|_{\dot{\phi}=\hat{\phi}} \tag{2.13}$$

where

$$\begin{aligned} \dot{c}_k &= \sum \dot{a}_t \dot{a}_{t-k} = \sum [\dot{\phi}(B) y_t][\dot{\phi}(B) y_{t-k}] \\ &= \sum_t \sum_{i=0}^{p} \sum_{j=0}^{p} \dot{\phi}_i \dot{\phi}_j y_{t-i} y_{t-k-j}, \end{aligned} \tag{2.14}$$

where in (2.14) and below, $\phi_0 = \dot{\phi}_0 = -1$. From (2.13) and (2.14) it follows that

$$\begin{aligned} \hat{\delta}_{jk} &= -\frac{\sum y_t^2}{\sum \hat{a}_t^2} \sum_{i=0}^{p} \hat{\phi}_i [r^{(y)}{}_{k-i+j} + r^{(y)}{}_{k+i-j}] \\ &= -\frac{\sum_{i=0}^{p} \hat{\phi}_i [r^{(y)}{}_{k-i+j} + r^{(y)}{}_{k+i-j}]}{\sum_{i=0}^{p} \sum_{j=0}^{p} \hat{\phi}_i \hat{\phi}_j r^{(y)}{}_{i-j}}, \end{aligned} \tag{2.15}$$

where

$$r_\nu^{(y)} = \frac{\sum y_t y_{t-\nu}}{\sum y_t^2}.$$

Let us approximate $\hat{\delta}_{jk}$ by replacing $\hat{\phi}$'s and $r^{(y)}$'s in (2.15) by ϕ's and ρ's (the theoretical parameters and autocorrelations of the autoregressive process $\{y_t\}$) and denote the result by δ_{jk}. That is,

$$\delta_{jk} = \frac{\sum_{i=0}^{p} \phi_i[\rho_{k-i+j} + \rho_{k+i-j}]}{-\sum_{i=0}^{p}\sum_{j=0}^{p} \phi_i\phi_j\rho_{i-j}}. \tag{2.16}$$

Now from Bartlett's formula [2, Equation (7)] we have

$$r_k^{(y)} = \rho_k + O_p(1/\sqrt{n}), \tag{2.17}$$

and as in the discussion preceding (2.10), $\hat{\phi}_j = \phi_j + O_p(1/\sqrt{n})$;
thus

$$\hat{\delta}_{jk} = \delta_{jk} + O_p(1/\sqrt{n}), \tag{2.18}$$

so that equation (2.10) holds when $\hat{\delta}_{jk}$ is replaced by δ_{jk}.

By making use of the recursive relation which is satisfied by the autocorrelations of an autoregressive process, namely

$$\rho_\nu - \phi_1\rho_{\nu-1} - \cdots - \phi_p\rho_{\nu-p} = \phi(B)\rho_\nu = 0, \qquad \nu \geq 1, \tag{2.19}$$

expression (2.16) can be simplified to yield

$$\delta_{jk} = \frac{\sum_{i=0}^{p} \phi_i\rho_{k-j+i}}{\sum_{i=0}^{p} \phi_i\rho_i}. \tag{2.20}$$

Thus δ_{jk} depends only on $(k-j)$, and we therefore write $\delta_{k-j} = \delta_{jk}$. Then it is straightforward to show that

(a) $\delta_0 = 1$

(b) $\delta_\nu = 0, \quad \nu < 0,$ and thus

(c) $\phi(B)\delta_\nu = \dfrac{\sum_{i=0}^{p} \phi_i[\phi(B)\rho_{\nu+i}]}{\sum_{i=0}^{p} \phi_i\rho_i} = 0, \qquad \nu \geq 1.$

Comparing (a), (b), and (c) with the corresponding results (2.4) for ψ_ν, we therefore have $\delta_\nu = \psi_\nu$, that is

$$\delta_{jk} = \psi_{k-j}, \tag{2.21}$$

whence, for $k = 1, 2, \cdots, m$,

$$\hat{r}_k = r_k + \sum_{j=1}^{p} (\phi_j - \hat{\phi}_j)\psi_{k-j} + O_p(1/n). \tag{2.22}$$

2.4 Representation of $\hat{r}$ as a Linear Transformation of r

We can now establish a relationship between the residual autocorrelations $\hat{r}$ and the white noise autocorrelations r. Let

$$X = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \psi_1 & 1 & & \vdots \\ \psi_2 & \psi_1 & \ddots & 0 \\ \vdots & \vdots & & 1 \\ \vdots & \vdots & & \vdots \\ \psi_{m-1} & \psi_{m-2} & \cdots & \psi_{m-p} \end{bmatrix} = [\mathbf{x}_1 \mid \mathbf{x}_2 \mid \cdots \mid \mathbf{x}_p]. \tag{2.23}$$

Then to $O_p(1/n)$ we can write (2.22) in matrix form as

$$\hat{\boldsymbol{r}} = \boldsymbol{r} + X(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}), \tag{2.24}$$

where from (2.9)

$$\hat{\boldsymbol{r}}'X = \mathbf{0}. \tag{2.25}$$

If we now multiply (2.24) on both sides by

$$Q = X(X'X)^{-1}X', \tag{2.26}$$

then using (2.25) we obtain

$$\hat{\boldsymbol{r}} = (I - Q)\boldsymbol{r}. \tag{2.27}$$

It is known [1] that $\boldsymbol{r}$ is very nearly normal for n moderately large. The vector of residual autocorrelations is thus approximately a linear transformation of a multi-normal variable and is therefore itself normally distributed. Specifically,

$$\boldsymbol{r} \sim N(\mathbf{0}, (1/n)I), \tag{2.28}$$

and hence

$$\hat{\boldsymbol{r}} \sim N(\mathbf{0}, (1/n)[I - Q]). \tag{2.29}$$

Note that the matrix $I-Q$ is idempotent of rank $m-p$, so that the distribution of $\hat{\boldsymbol{r}}$ has a p-dimensional singularity.

2.5 Further Consideration of the Covariance Structure of the $\hat{r}$'s

It is illuminating to examine in greater detail the covariance matrix of $\hat{\boldsymbol{r}}$, or equivalently the matrix Q. The latter matrix is idempotent of rank p, and its non-null latent vectors are the columns of X. Also,

$$X'X = \begin{bmatrix} \sum \psi_j^2 & \sum \psi_j\psi_{j-1} & \cdots & \sum \psi_j\psi_{j-p+1} \\ \sum \psi_j\psi_{j-1} & \sum \psi_j^2 & \cdots & \sum \psi_j\psi_{j-p+2} \\ \vdots & \vdots & & \vdots \\ \sum \psi_j\psi_{j-p+1} & \sum \psi_j\psi_{j-p+2} & \cdots & \sum \psi_j^2 \end{bmatrix} = \frac{\sigma_y^2}{\sigma_a^2} \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{bmatrix} \tag{2.30}$$

which when multiplied by σ_a^2 is the autocovariance matrix of the process itself. Let c^{ij} be the (ij)th element of $(X'X)^{-1}$ (given explicitly in [9]), and similarly q_{ij} for Q. If $\xi_j' = (\psi_{j-1}, \cdots, \psi_{j-p})$ denotes the jth row of X, then

$$\begin{aligned} q_{ij} &= \xi_i'(X'X)^{-1}\xi_j \\ &= \sum_{k=1}^{p}\sum_{l=1}^{p}\psi_{i-k}c^{kl}\psi_{j-l} \\ &= (-n)\operatorname{cov}[\hat{r}_i, \hat{r}_j] \quad \text{if } i \neq j. \end{aligned} \tag{2.31}$$

Since the elements of each column of X satisfy the recursive relation (2.4), we have $\phi(B)\xi_j = 0$, and hence

$$\phi(B)q_{ij} = 0, \tag{2.32}$$

where in (2.32) B can operate either on i or on j. This establishes an interesting recursive structure in the residual autocorrelation covariance matrix $(1/n)\cdot(I-Q)$ and provides an important clue as to how rapidly the covariances die out and the variances approach 1. Also, because of this property the entire covariance matrix is determined by specifying the elements

$$\begin{matrix} q_{11} & q_{12} & \cdots & q_{1p} \\ & q_{22} & \cdots & q_{2p} \\ & & \ddots & \vdots \\ & & & q_{pp} \end{matrix} \tag{2.33}$$

of Q, which are readily obtained by inverting the $X'X$ matrix (2.30).

2.6 Covariance Matrix of $\hat{r}$ for first and second order processes

Consider, for example, the first order autoregressive process $y_t = \phi y_{t-1} + a_t$, which in accordance with (2.2) we can write as

$$y_t = (1 - \phi B)^{-1}a_t = \sum_{j=0}^{\infty}\phi^j a_{t-j}. \tag{2.34}$$

For this process, $\psi_j = \phi^j$ and $(X'X)^{-1} = 1 - \phi^2$. From (2.31) the (ij)th element of Q is therefore $\phi^{i+j-2}(1-\phi^2)$, so that approximately the covariance matrix of the sample residual autocorrelations is

$$\sum\hat{r} = (1/n)(I - Q) = 1/n\begin{bmatrix} \phi^2 & -\phi + \phi^3 & -\phi^2 + \phi^4 & \cdots \\ -\phi + \phi^3 & 1 - \phi^2 + \phi^4 & -\phi^3 + \phi^5 & \cdots \\ -\phi^2 + \phi^4 & -\phi^3 + \phi^5 & 1 - \phi^4 + \phi^6 & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix} \tag{2.35}$$

For the second order process

$$y_t = (1 - \phi_1 B - \phi_2 B^2)^{-1}a_t = \psi(B)a_t, \tag{2.36}$$

we have

$$X = \begin{bmatrix} 1 & 0 \\ \psi_1 & 1 \\ \psi_2 & \psi_1 \\ \vdots & \vdots \end{bmatrix}, \qquad X'X = \frac{\sigma_y^2}{\sigma_a^2}\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix},$$

$$(X'X)^{-1} = \frac{\sigma_a^2}{\sigma_y^2(1-\rho_1^2)}\begin{bmatrix} 1 & -\rho_1 \\ -\rho_1 & 1 \end{bmatrix}, \qquad \sigma_y^2 = \frac{(1-\phi_2)\sigma_a^2}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]}.$$

Thus

$$q_{11} = 1-\phi_2^2, \qquad q_{12} = -\phi_1\phi_2(1+\phi_2), \qquad q_{22} = 1-\phi_2^2-\phi_1^2(1+\phi_2)^2,$$

from which Q and $\sum\hat{r} = 1/n(I-Q)$ may be determined using (2.32). In particular,

$$\left.\begin{aligned} V(\hat{r}_1) &= 1/n\cdot\phi_2^2, \\ V(\hat{r}_2) &= 1/n[\phi_2^2+\phi_1^2(1+\phi_2)^2], \quad \text{and} \\ V(\hat{r}_k) &= 1/n[1-\phi_1 q_{k,k-1}-\phi_2 q_{k,k-2}], \; k \geq 3. \end{aligned}\right\} \tag{2.37}$$

From these examples we can see a general pattern emerging. As in (2.33) the first p variances and corresponding covariances will be heavily dependent on the parameters $\phi_1, \cdots, \phi_p$ and in general can depart sharply from the corresponding values for white noise autocorrelations, whereas for $k \geq p+1$ a "1" is introduced into the expression for variances (as in (2.35) and (2.37)), and the recursion (2.32) ensures that as k increases the $\{\hat{r}_k\}$ behave increasingly like the corresponding $\{r_k\}$ with respect to both their variances and covariances.

2.7 The distribution of $n\sum_1^m \hat{r}_k^2$

We have remarked earlier that if the fitted model is appropriate and the parameters ϕ are exactly known, then the calculated a_t's would be uncorrelated normal deviates, their serial correlations r would be approximately $N(0, (1/n)I)$, and thus $n\sum_1^m r_k^2$ would possess a χ^2 distribution with m degrees of freedom. We now see that if m is taken sufficiently large so that the elements after the mth in the latent vectors of Q are essentially zero, then we should expect that to the order of approximation we are here employing, the statistic

$$n\sum_1^m \hat{r}_k^2, \tag{2.38}$$

obtained when estimates $\hat{\phi}$ are substituted for the true parameters ϕ in the model, will still be distributed as χ^2, only now with $m-p$ rather than m degrees of freedom. This result is of considerable practical interest because it suggests that an overall test of the type discussed in [4] can in fact be justified when suitable modifications coming from a more careful analysis are applied. Later we consider in more detail the use of this test, along with procedures on individual $\hat{r}$'s, in diagnostic checking.

3. MONTE CARLO EXPERIMENT

We have made certain approximations in deriving the distribution of the residual autocorrelations, and it is therefore of interest to investigate this distribution empirically through repeated sampling and to compare the results with (2.29). This was done for the first order AR process for $\phi = 0, \pm.1, \pm.3, \pm.5, \pm.7, \pm.9$. For given ϕ, $s = 50$ sets of $n = 200$ random normal deviates were generated on the computer using a method described in [7], with separate aggregates of deviates obtained for each parameter value. For the jth set a

series $\{y_t^{(j)}\}$ was generated using formula (2.34), $\hat{\phi}^{(j)}$ was estimated, $\{\hat{a}_t^{(j)}\}$ determined, and the quantities

$$\hat{r}_k^{(j)} = \frac{\sum \hat{a}_t^{(j)} \hat{a}_{t-k}^{(j)}}{\sum [\hat{a}_t^{(j)}]^2} \tag{3.1}$$

computed for $1 \leq k \leq m = 20$, $1 \leq j \leq s = 50$. This yielded sample variances and covariances

$$C_{kl} = \frac{1}{50} \sum_{j=1}^{50} \hat{r}_k^{(j)} \hat{r}_l^{(j)} \tag{3.2}$$

and sample correlations

$$R_{kl} = C_{kl} / \sqrt{C_{kk} C_{ll}} \, . \tag{3.3}$$

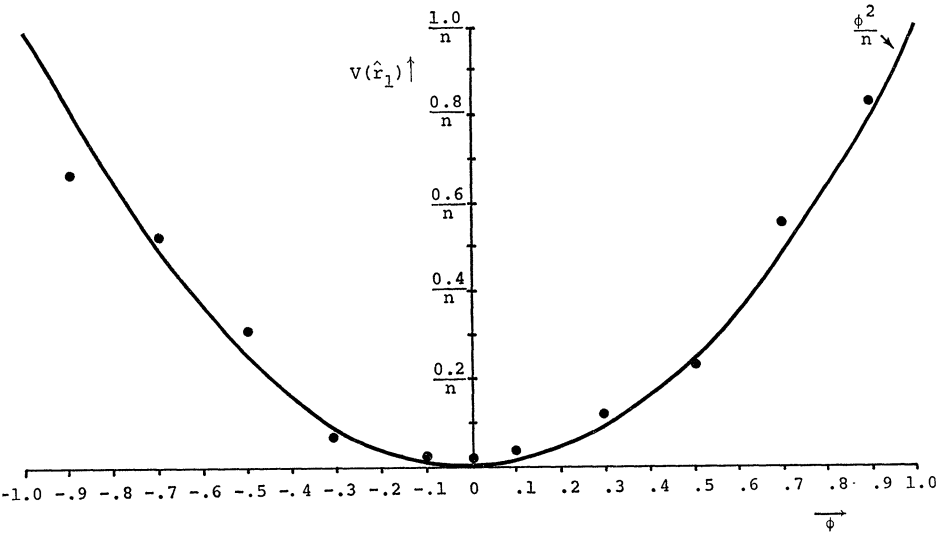
The results of this Monte Carlo sampling are set out in detail in [8] and in general confirm the adequacy of the approximations used. As an example of these calculations, Table 1 compares the empirical variances (3.2) of $\hat{r}_k$ and correlations (3.3) of $(\hat{r}_1, \hat{r}_k)$ with their theoretical counterparts obtained from (2.35). Allowing for the sampling error of the Monte Carlo estimates themselves, there is good agreement between the two sets of quantities, a phenomenon which occurred also for the other values of ϕ considered.

Since the large-sample variance ϕ^2/n of $\hat{r}_1$ departs the most from the common variance of $1/n$ for white noise autocorrelations, an examination of the empirical behavior of this quantity is of particular interest. Thus Figure 1 shows the sample variance of $\hat{r}_1$ for $\phi = 0, \pm.1, \pm.3, \pm.5, \pm.7, \pm.9$ in relation to the parabola $V(\hat{r}_1) = \phi^2/n$, with reasonable agreement between the two. (The coefficient of variation of the sample variance of $\hat{r}_k$ for $\phi \neq 0$ is approximately $\sqrt{2/s} = 1/5$, independent of k and n; at $\phi = 0$, $V(\hat{r}_1) = O(1/n^2)$.

Table 1. *THEORETICAL (AS IN (2.35)) AND EMPIRICAL (FROM MONTE-CARLO SAMPLING) VARIANCES AND CORRELATIONS OF SAMPLE RESIDUAL AUTOCORRELATIONS FROM FIRST-ORDER AR PROCESS WITH $\phi = .5$*

k	*Variance of $\hat{r}_k$ (multiplied by n)*		*Correlation between $\hat{r}_1$ and $\hat{r}_k$*	
	Theoretical	*Empirical*	*Theoretical*	*Empirical*
1	.250	.244	1.000	1.000
2	.813	.676	−.832	−.812
3	.953	.741	−.384	−.301
4	.988	.864	−.189	−.186
5	.997	1.240	−.094	−.366
6	.999	.967	−.047	−.221
7	1.000	.870	−.023	.083
8	1.000	1.203	−.012	−.148
9	1.000	.982	−.006	−.009
10	1.000	.881	−.003	−.080

Figure 1. THEORETICAL (LINE) AND EMPIRICAL (DOTS) VARIANCES OF $\hat{r}_1$



There are several additional comparisons which can be made based on certain functions of the $\hat{r}$'s. Thus we have seen that

$$\hat{l} = \sum \hat{\phi}^{k-1}\hat{r}_k = 0, \tag{3.4}$$

and in the course of our derivations we have had to make the approximation

$$l = \sum \phi^{k-1}\hat{r}_k = 0. \tag{3.5}$$

Some indication of the validity of this approximation is gained by examining the actual values of l from the sampling experiment, which were found to be distributed about zero with a variance of about one-hundredth that which would have been expected from the same linear form in white noise autocorrelations.

Of considerable importance because of its role in diagnostic checking is an examination of the quantity

$$n \sum_{k=1}^{m} \hat{r}_k^2 = 200 \sum_{k=1}^{20} \hat{r}_k^2, \tag{3.6}$$

which as in (2.38) should possess a χ^2—distribution with $\nu = m-1 = 19$ degrees of freedom. Such a distribution has a mean and variance of 19 and 38, respectively, with which the Monte Carlo values can be compared. When this was done, the overall or pooled empirical mean was found to be 18.1 and significantly different from 19. This difference is plausible, however, when it is realized that the statistic $n\sum_1^m \hat{r}_k^2$ possesses a χ^2_{m-p} distribution only insofar as the white noise autocorrelations $\boldsymbol{r} = (r_1, \cdots, r_m)'$ have a common variance of $1/n$; and from (1.4) it is seen that this approximation overestimates the true variance of a given r_k by a factor of $(n+2)/(n-k)$. In particular, for $n = 200$, $m = 20$, and a typical value of $k = 10$, the actual variance $V(r_k)$ is $190/202 \approx 94$ percent of the $1/n$ approximation. Since the residual autocorrelations $\hat{\boldsymbol{r}}$ are by (2.27) a linear transformation of $\boldsymbol{r}$, it is reasonable to expect that a comparable depression of

the variances of $\{\hat{r}_k\}$ would occur, and this would account for the discrepancy between the theoretical and empirical means of the statistic $200 \sum_1^{20} \hat{r}_k^2$ encountered above. (This phenomenon would also explain the tendency for the empirical variances themselves, such as those in Table 1, to take on values averaging about 5 percent lower than those based on the matrix $(1/n)(I-Q)$ of (2.29).)

4. USE OF RESIDUAL AUTOCORRELATIONS IN DIAGNOSTIC CHECKING

We have obtained the large sample distribution of the residual autocorrelations $\hat{r}$ from fitting the correct model to a time series, and we have discussed the ways in which this distribution departs significantly from that of the white noise autocorrelations r. It is desirable now to consider the practical implications of these results in examining the adequacy of fit of a model.

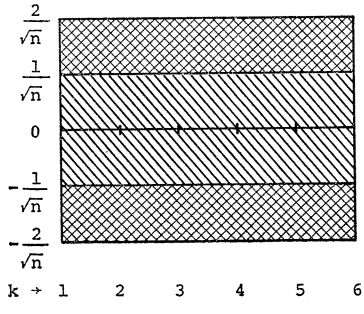
First of all it appears that even though the $\hat{r}$'s have a variance/covariance matrix which can differ very considerably from that of the r's, the statistic $n \sum_{k=1}^{m} \hat{r}_k^2$ will (since the matrix $I-Q$ is idempotent) still possess a χ^2-distribution, only now with $m-p$ rather than m degrees of freedom. Thus the overall χ^2-test discussed in Section 1 may be justified to the same degree of approximation as before when the number of degrees of freedom is appropriately modified.

However, regarding the "quality-control-chart" procedure, that is the comparison of the $\{\hat{r}_k\}$ with their standard errors, some modification is clearly needed.

Figure 2 shows the straight-line standard error bands of width $1/\sqrt{n}$ associated with any set of white noise autocorrelations $\{r_k\}$. These stand in marked contrast to the corresponding bands for the residual autocorrelations $\{\hat{r}_k\}$, derived from their covariance matrix $(1/n)(I-Q)$ and shown in Figure 3 for selected first and second order AR processes. Since it is primarily the $\hat{r}$'s of small lags that are most useful in revealing model inadequacies, we see that the consequence of treating $\hat{r}$'s as r's in the diagnostic checking procedure can be a serious underestimation of significance, that is, a failure to detect lack of fit in the model when it exists. Of course, if the model would have been judged inadequate anyway, our conviction in this regard is now strengthened.

Suppose, for example, that we identify a series of length 200 as first order

Figure 2. STANDARD ERROR LIMITS FOR WHITE NOISE AUTOCORRELATIONS r_k



autoregressive and after fitting $\hat{\phi} = .5$. Suppose also that $\hat{r}_1 = .10$. Now the standard error of r_1 for white noise is $1/\sqrt{n} = .07$, so that $\hat{r}_1$ is well within the limits in Figure 2. Therefore if we erroneously regarded these as limits on $\hat{r}_1$ we would probably not conclude that this model was inadequate. However, if the true process actually were first order autoregressive (say with $\phi = .5$), the standard error of $\hat{r}_1$ would be $|\phi|/\sqrt{n} = .035$; since the observed $\hat{r}_1 = .10$ is almost three times this value, we should be very suspicious of the adequacy of this fit.

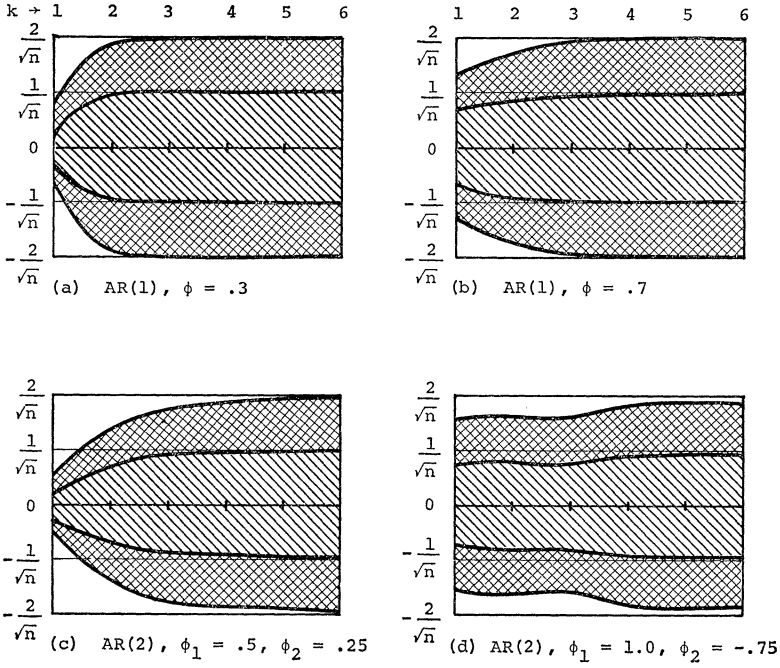
The situation is further complicated by the existence of rather high correlations between the $\hat{r}$'s, especially between those of small lags. For the first order process, the most serious correlation is

$$\rho[\hat{r}_1, \hat{r}_2] = -\frac{\phi}{|\phi|}\frac{1-\phi^2}{\sqrt{1-\phi^2+\phi^4}}$$

which, for example, approaches -1 as $\phi \to 0^+$ and is still as large as $-.6$ for $\phi = .7$. Correlation among the $\hat{r}$'s is even more prevalent in second and higher-order processes, where (as for variances) those involving lags up to $k = p$ can be particularly serious. From then on their magnitude is controlled by the recursive relationship (2.32); in particular, the closer ϕ is to the boundary of the stationarity region, the slower will be the dying out of $\text{cov}(\hat{r}_k, \hat{r}_l)$ or $\rho(\hat{r}_k, \hat{r}_l)$ although often in these situations the less serious will the initial correlations $\rho(\hat{r}_1, \hat{r}_2)$, $\rho(\hat{r}_2, \hat{r}_3)$, $\rho(\hat{r}_1, \hat{r}_3)$, etc., tend to be.

We have thus seen that the departure of the distribution of the residual autocorrelations $\hat{r}$ from that of white noise autocorrelations r is serious enough to

Figure 3. STANDARD ERROR LIMITS FOR RESIDUAL AUTOCORRELATIONS $\hat{r}_k$



warrant some modifications in their use in diagnostic checking. The residual autocorrelation function, however, remains a powerful device for this purpose.

5. DISTRIBUTION OF RESIDUAL AUTOCORRELATIONS FOR THE MOVING AVERAGE AND GENERAL ARIMA PROCESSES

In obtaining the distribution of $\hat{\boldsymbol{r}} = (\hat{r}_1, \cdots, \hat{r}_m)'$ for the pure autoregressive process in Section 2, considerable use was made of the recursive relation $\phi(B)\rho_k = 0$, which is not satisfied by moving average models $y_t = \theta(B)a_t$, or more generally by mixed models of the form (1.1) with $w_t = \nabla^d z_t$ denoting the stationary dth difference.

It is fortunate, therefore, that these models have in common with the pure AR models (2.1) an important property (derived in Section 5.1) because of which the distribution of their residual autocorrelations can be found as an immediate consequence of the autoregressive solution (2.29). This property is that if two time series, (a) the mixed autoregressive—moving average series (1.1), and (b) an autoregressive series

$$\pi(B)x_t = (1 - \pi_1 B - \cdots - \pi_{p+q}B^{p+q})x_t = a_t \tag{5.1}$$

are both generated from the *same set* of deviates $\{a_t\}$, and moreover if

$$\pi(B) = \phi(B)\theta(B), \tag{5.2}$$

then when these models are each fitted by least squares, their residuals, and hence also their residual autocorrelations, will be very nearly the same. Therefore if a mixed model of order (p, d, q) is correctly identified and fitted, its residual autocorrelations for n sufficiently large will be distributed as though the model had been of order $(p+q, d, 0)$ with the relations between the two sets of parameters given by (5.2). In particular the ψ's comprising the X-matrix (2.23) for the model (1.1) are the coefficients in $\psi(B) = [\phi(B)\theta(B)]^{-1}$.

5.1 Equality of Residuals in AR and ARIMA Models

Let w_t and x_t be as in (1.1) and (5.1); (5.2) then implies

$$w_t = \theta^2(B)x_t. \tag{5.3}$$

As in (2.5), define

$$\dot{a}_t^{\mathrm{AR}} = a_t^{\mathrm{AR}}(\dot{\boldsymbol{\pi}}) = \dot{\pi}(B)x_t = -\sum_{j=0}^{p+q} \dot{\pi}_j x_{t-j} \tag{5.4}$$

where $\dot{\pi}_0 = -1$, and now also

$$\dot{a}_t^* = a_t^*(\dot{\boldsymbol{\phi}}, \dot{\boldsymbol{\theta}}) = \dot{\phi}(B)\dot{\theta}^{-1}(B)w_t = [\sum_{i=0}^{p} \dot{\phi}_i B^i][\sum_{j=0}^{q} \dot{\theta}_j B^j]^{-1} w_t, \tag{5.5}$$

where $\dot{\phi}_0 = \dot{\theta}_0 = -1$. We will expand these quantities about the true parameter values and go through a least squares estimation in each case which is analogous to writing the linear regression model $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ as

$$\dot{\boldsymbol{e}} = \boldsymbol{y} - \dot{\boldsymbol{y}} = X(\boldsymbol{\beta} - \dot{\boldsymbol{\beta}}) + \boldsymbol{\varepsilon} = X\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \tag{5.6}$$

for fixed $\boldsymbol{\beta}$, and then performing the regression directly on $\dot{\boldsymbol{e}}$ rather than on $\boldsymbol{y}$. The equality of the residuals in the two cases depends heavily on the fact that the derivatives in each expansion involve the same autoregressive variable x_t.

Thus

$$\frac{\partial \dot{a}_t^{\mathrm{AR}}}{\partial \dot{\pi}_j} = -x_{t-j}, \qquad 1 \le j \le p+q, \text{ irrespective of } \dot{\pi};$$

$$\frac{\partial \dot{a}_t^{*}}{\partial \dot{\phi}_j} = -\dot{\theta}^{-1}(B)w_{t-j}, \qquad 1 \le j \le p$$

$$= -\theta(B)x_{t-j} \qquad \text{at } (\dot{\phi}, \dot{\theta}) = (\phi, \theta); \quad \text{and}$$

$$\frac{\partial \dot{a}_t^{*}}{\partial \dot{\theta}_j} = \dot{\phi}(B)\dot{\theta}^{-2}(B)w_{t-j}, \qquad 1 \le j \le q$$

$$= \phi(B)x_{t-j} \qquad \text{at } (\dot{\phi}, \dot{\theta}) = (\phi, \theta).$$

Then

$$\dot{a}_t^{\mathrm{AR}} = a_t^{\mathrm{AR}} + \textstyle\sum_{j=1}^{p+q} (\pi_j - \dot{\pi}_j)x_{t-j}, \tag{5.7}$$

and approximately

$$\begin{aligned}
\dot{a}_t^{*} &= a_t^{*} + \textstyle\sum_{i=1}^{p} (\phi_i - \dot{\phi}_i)\theta(B)x_{t-i} - \sum_{j=1}^{q} (\theta_j - \dot{\theta}_j)\phi(B)x_{t-j} \qquad (5.8)\\
&= a_t^{*} + \textstyle\sum_{i=1}^{p} (\phi_i - \dot{\phi}_i)x_{t-i} - \sum_{j=1}^{q} (\theta_j - \dot{\theta}_j)x_{t-j}\\
&\quad + \textstyle\sum_{i=1}^{p}\sum_{j=1}^{q} [\phi_i(\theta_j - \dot{\theta}_j) - \theta_j(\phi_i - \dot{\phi}_i)]x_{t-i-j}\\
&= a_t^{*} + \textstyle\sum_{i=1}^{p} (\phi_i - \dot{\phi}_i)x_{t-i} - \sum_{j=1}^{q} (\theta_j - \dot{\theta}_j)x_{t-j}\\
&\quad + \textstyle\sum_{i=1}^{p}\sum_{j=1}^{q} [\dot{\phi}_i(\theta_j - \dot{\theta}_j) - \dot{\theta}_j(\phi_i - \dot{\phi}_i)]x_{t-i-j} \qquad (5.9)\\
&= a_t^{*} + \textstyle\sum_{j=1}^{p+q} (\beta_j - \dot{\beta}_j)x_{t-j}.
\end{aligned}$$

Thus letting $\beta = (\beta_1, \cdots, \beta_{p+q})'$ and $\lambda = \begin{bmatrix}\phi\\ \theta\end{bmatrix}$, we see that

$$\beta = A\lambda, \tag{5.10}$$

where A is a $(p+q)$-square matrix whose elements involve λ but not the true parameter values λ. For example, if $p=q=1$, we would have

$$\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix} = \begin{bmatrix} 1 & -1\\ -\dot{\theta} & \dot{\phi}\end{bmatrix}\begin{bmatrix}\phi\\ \theta\end{bmatrix} \tag{5.11}$$

Now equations (5.7) and (5.9) can be written as

$$\dot{\boldsymbol{a}}^{\mathrm{AR}} = \boldsymbol{a} + X(\boldsymbol{\pi} - \dot{\boldsymbol{\pi}}) \tag{5.12}$$

$$\dot{\boldsymbol{a}}^{*} = \boldsymbol{a} + X(\beta - \dot{\beta}) \tag{5.13}$$

where the error in (5.13) is $O(|\beta - \dot{\beta}|^2)$, and where we have made use of the fact that, at $\dot{\pi} = \pi$, $\dot{\theta} = \theta$, and $\dot{\phi} = \phi$,

$$a_t^{\mathrm{AR}} = a_t^{*} = a_t. \tag{5.14}$$

Thus in (5.12) the sum of squares

$$\boldsymbol{a}'\boldsymbol{a} = \sum a_t^2 = \sum [a_t^{\mathrm{AR}}(\boldsymbol{\pi})]^2$$

is minimized as a function of $\boldsymbol{\pi}$ when

$$\boldsymbol{\pi} - \dot{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}} - \dot{\boldsymbol{\pi}} = (X'X)^{-1}X'\dot{\boldsymbol{a}}^{\mathrm{AR}}, \tag{5.15}$$

while in (5.13) if we write

$$\boldsymbol{a}^* = \boldsymbol{a} + X[A(\lambda - \dot{\lambda})] = \boldsymbol{a} + Z(\lambda - \dot{\lambda}),$$

then the sum of squares

$$\boldsymbol{a}'\boldsymbol{a} = \sum a_t^2 = \sum [a_t^*(\lambda)]^2$$

is minimized as a function of λ when

$$\lambda - \dot{\lambda} = \hat{\lambda} - \dot{\lambda} = (Z'Z)^{-1}Z'\dot{\boldsymbol{a}}^* = A^{-1}(\hat{\beta} - \dot{\beta});$$

that is,

$$\hat{\beta} - \dot{\beta} = (X'X)^{-1}X'\dot{\boldsymbol{a}}^*. \tag{5.16}$$

Then by setting $\dot{\boldsymbol{a}} = \boldsymbol{a}$ in (5.15) and (5.16), we have from (5.14) the important equality

$$\hat{\pi} - \pi = (X'X)^{-1}X'\boldsymbol{a} = \hat{\beta} - \beta; \tag{5.17}$$

and finally by setting "." = "$\hat{}$" in (5.12) and (5.13), it follows from (5.17) that to $O_p(1/n)$

$$\hat{\boldsymbol{a}}^{\mathrm{AR}} = \boldsymbol{a} + X(\pi - \hat{\pi}) = \boldsymbol{a} + X(\beta - \hat{\beta}) = \hat{\boldsymbol{a}}^*, \tag{5.18}$$

and thus (to the same order) $\hat{\boldsymbol{r}}^{\mathrm{AR}} = \hat{\boldsymbol{r}}^*$, as we set out to show.

5.2 Monte Carlo Experiment

The equality (5.18) between the residuals from the autoregressive and mixed models depends on the accuracy of the expansion (5.8), that is, on the extent of linearity in the moving average model, between the true and estimated values θ and $\hat{\theta}$. It is therefore worthwhile to confirm this model-duality by generating and fitting pairs of series of the form (1.1) and (5.1) and comparing their residuals, or more to our purpose, their residual autocorrelations. This was done for $p+q=1$ and $p+q=2$ for series of length 200. Some indication of the close-

Table 2. RESIDUAL CORRELATIONS FROM FIRST ORDER AR AND MA TIME SERIES GENERATED FROM SAME WHITE NOISE ($n=200$)

k	$\phi=\theta=.1$		$\phi=\theta=.5$		$\phi=\theta=.9$	
	$\hat{r}_k^{AR}$	$\hat{r}_k^{MA}$	$\hat{r}_k^{AR}$	$\hat{r}_k^{MA}$	$\hat{r}_k^{AR}$	$\hat{r}_k^{MA}$
1	−.029	−.010	.003	−.005	−.048	−.057
2	.164	.169	.044	.045	.157	.151
3	.096	.099	−.098	−.096	.008	.009
4	−.050	−.049	.014	.021	−.126	−.127
5	−.003	−.006	.057	.058	.034	.035
6	−.143	−.144	.010	.012	−.091	−.090
7	−.023	−.026	−.004	.001	−.001	−.000
8	−.040	−.041	−.054	−.046	−.038	−.035
9	.010	.009	.052	.052	−.004	.000
10	−.049	−.049	−.065	−.067	.113	.116
$\hat{\phi}$ or $\hat{\theta}\rightarrow$	.159	.057	.543	.451	.922	.870

ness of the agreement is obtained from the few results for first order AR and MA processes shown in Table 2, where it is seen that the residual autocorrelation $\hat{r}_k^{AR}$ and $\hat{r}_k^{MA}$ are equal or nearly equal to the second decimal place.

A sampling experiment of the type described in Section 3 was also performed for the first order MA process. The results were very similar, which is to be expected in view of (5.18).

5.3 Conclusions

We have shown above that to a close approximation the residuals from any moving average or mixed autoregressive-moving average process will be the same as those from a suitably chosen autoregressive process. We have further confirmed the adequacy of this approximation by empirical calculation. It follows from this that we need not consider separately these two classes of processes; more precisely,

1. We can immediately use the AR result to write down the variance/covariance matrix of $\hat{\mathbf{r}}$ for any autoregressive-integrated moving average process (1.1) by considering the corresponding variance/covariance matrix of $\hat{\mathbf{r}}$ from the pure AR process
$$\pi(B)x_t = \theta(B)\phi(B)x_t = a_t. \tag{5.19}$$
2. All considerations regarding the use of residual autocorrelations in tests of fit and diagnostic checking discussed in Section 4 for the autoregressive model therefore apply equally to moving average and mixed models.
3. In particular it follows from the above that a "portmanteau" test for the adequacy of any ARIMA process is obtained by referring $n\sum_{k=1}^{m} \hat{r}_k^2$ to a χ^2 distribution with ν degrees of freedom, where $\nu = m - p - q$.

REFERENCES

[1] Anderson, R. L., "Distribution of the Serial Correlation Coefficient," *The Annals of Mathematical Statistics*, 13 (March 1942), 1–13.

[2] Bartlett, M. S., "On the Theoretical Specification and Sampling Properties of Autocorrelated Time Series," *Journal of the Royal Statistical Society*, Series B, 8 (April 1946), 27–41.

[3] ——— and Diananda, P. H., "Extensions of Quenouille's Tests for Autoregressive Schemes," *Journal of the Royal Statistical Society*, Series B, 12 (April 1950), 108–15.

[4] Box, G. E. P. and Jenkins, G. M., *Statistical Models for Prediction and Control*, Technical Reports #72, 77, 79, 94, 95, 99, 103, 104, 116, 121, and 122, Department of Statistics, University of Wisconsin, Madison, Wisconsin, 1967.

[5] ———, *Time Series Analysis Forecasting and Control*, San Francisco: Holden-Day, Inc., 1970.

[6] ——— and Bacon, D. W., "Models for Forecasting Seasonal and Non-Seasonal Time Series," in B. Harris, ed., *Spectral Analysis of Time Series*, New York: John Wiley & Sons, Inc., 1967.

[7] Box, G. E. P. and Muller, M. E., "Note on the Generation of Random Normal Deviates," *The Annals of Mathematical Statistics*, 29 (June 1958), 610–11.

[8] Box, G. E. P. and Pierce, D. A., "Distribution of Residual Autocorrelations in Integrated Autoregressive-Moving Average Time Series Models," Technical Report #154, Department of Statistics, University of Wisconsin, Madison, April, 1968.

[9] Durbin, J., "Efficient Estimation of Parameters in Moving Average Models," *Biometrika*, 46 (December 1959), 306–16.

[10] ———, "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38 (May 1970), 410–21.

[11] Grenander, U. and Rosenblatt, M., *Statistical Analysis of Stationary Time Series*, New York: John Wiley & Sons, Inc., 1957.
[12] Mann, H. B. and Wald, A., "On the Statistical Treatment of Linear Stochastic Difference Equations," *Econometrica*, 11 (July 1943), 173–220.
[13] ——— "On Stochastic Limit and Order Relationships," *The Annals of Mathematical Statistics*, 14 (September 1943), 217–26.
[14] Quenouille, M. H., "A Large-Sample Test for the Goodness of Fit of Autoregressive Schemes," *Journal of the Royal Statistical Society*, Series A, 110 (June 1947), 123–9.
[15] Walker, A. M., "Note on a Generalization of the Large-Sample Goodness of Fit Test for Linear Autoregressive Schemes," *Journal of the Royal Statistical Society*, Series B, 12 (April 1950), 102-7.
[16] Whittle, P., "Tests of Fit in Time Series," *Biometrika*, 39 (December 1952), 309–18.
[17] Wold, H., *A Study in the Analysis of Stationary Time Series*, Stockholm: Almquist and Wiksell, 1938.
[18] ———, "A Large-Sample Test for Moving Averages," *Journal of the Royal Statistical Society*, Series B, 11 (April 1949), 297–305.
[19] Yule, G. U., "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers," *Philosophical Transactions*, (A) 226 (July 1927), 267–98.