DEPARTMENT OF STATISTICS

The University of Wisconsin
Madison, Wisconsin

TECHNICAL REPORT NO. 173

August 1968

DISTRIBUTION OF RESIDUAL CORRELATIONS IN DYNAMIC/STOCHASTIC TIME SERIES MODELS

By

David A. Pierce
University of Wisconsin

This research was supported in part by the Wisconsin Alumni Research Foundation.

1. Introduction

An approach to the modeling of discrete time series such as commonly occur in economic situations and process control problems is discussed by Box and Jenkins [4] and involves the iterative use of a three stage procedure of identification, estimation, and diagnostic checking. Given a discrete time series $z_t, z_{t-1}, z_{t-2}, \ldots,$ which for example may be drawn from

(i) an integrated autoregressive - moving average (IARMA) process ([4], Chapters II and III)

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

where $\{a_t\}$ are white noise, that is, identically distributed random normal deviates, B is the backward shift operator $B\ z_t = z_{t-1}$, $\nabla = 1-B$, and $\phi(B) = 1-\phi_1 B-\ldots-\phi_p B^p$ and $\theta(B) = 1-\theta_1 B-\ldots-\theta_q B^q$ are polynomials in B of degree p and q such that $y_t = \nabla^d\ z_t$ is stationary; or

(ii) a dynamic/stochastic model, or a dynamic system with added noise ([4], Chapters IX and X)

$$z_t = \frac{\omega(B)}{\delta(B)}\ x_{t-b-1} + \frac{\theta(B)}{\phi(B)}\ a_t \tag{1.2}$$

where $\frac{\theta(B)}{\phi(B)}\ a_t$ is as in (1.1), $\{x_t\}$ represents a series of inputs to the system, b is the number of units of delay or dead time, and $\omega(B)$ and $\delta(B)$ are polynomials in B of degree u and v; then this three-stage procedure is as follows:

(a) a process of identification is used to find a particular subclass of models worth considering to represent the process $\{z_t\}$;

(b) a model in this subclass is fitted by efficient statistical methods (generally nonlinear least-squares procedures);

(c) an examination of the adequacy of the fit is made.

The stage (c) of diagnostic checking includes both tests of fit of the models (1.1) and (1.2) and means of indicating any needed model modifications. An important technique in this process is the study of the residuals $\{\hat{a}_t\}$ calculated from the parameter estimates $(\hat{\underline{\phi}},\hat{\underline{\theta}})$ or $(\hat{\underline{\phi}},\hat{\underline{\theta}},\hat{\underline{\delta}},\hat{\underline{\omega}})$ after fitting the model; and on the assumption that the $\{\hat{a}_t\}$ from a correctly fitted model should exhibit approximately the same behavior as the white noise $\{a_t\}$, and thus be nearly uncorrelated, it would be expected that useful indicators of model inadequacy could be provided by the residual autocorrelations

$$\hat{r}_k = \frac{\sum \hat{a}_t \hat{a}_{t-k}}{\sum \hat{a}_t^2} \tag{1.3}$$

for both models (1.1) and (1.2); and for model (1.2) also the residual cross correlations

$$r_k^* = \frac{\sum x_{t-k} \hat{a}_t}{\sqrt{\sum x_t^2 \sum \hat{a}_t^2}} . \tag{1.4}$$

In particular Box and Jenkins [4] have shown that unusually large cross correlation between $\{\hat{a}_t\}$ and $\{x_t\}$ indicates inadequacy in the dynamic parameters $\frac{\omega(B)}{\delta(B)}$, whereas marked autocorrelation of $\{\hat{a}_t\}$ without significant cross correlation suggests lack of fit in the stochastic or noise parameters $\frac{\theta(B)}{\phi(B)}$.

Since for any t, $\hat{a}_t \overset{p}{\to} a_t$ as the number n of observations in the series becomes infinite, it might be tempting to suppose that approximate large-sample tests of fit and diagnostic checks could be based on referring the <u>residual</u> correlations (1.3) and (1.4) to the distribution of the corresponding <u>white noise</u> correlations calculated from $\{a_t\}$ rather $\{\hat{a}_t\}$. For example it is known [1,2] that for large n the first m autocorrelations $\underline{r} = (r_1, r_2, \dots r_m)'$ for a white noise sequence $\{a_t\}$ are approximately distributed as $N(\underline{0}, \frac{1}{n}I)$, so that if the above supposition were warranted, standard errors of $\frac{1}{\sqrt{n}}$ could be attached to the $\hat{r}$'s and a "quality-control-chart" approach used to determine whether any of the residual correlations were unusually large. Also it might be supposed that the quantity

$$n \sum_{k=1}^{m} \hat{r}_k^2 \qquad (1.5)$$

would to a close approximation be distributed as χ_m^2, since this is true of the statistic $n \sum_{k=1}^{m} r_k^2$, so that large values of (1.5) would also place the model under suspicion.

However it has been shown by Box and Pierce [5] that, for the case of the IARMA models (1.1), such an approximation is invalid; that is, the large sample distributions of $\underline{r} = (r_1 \dots r_m)'$

and $\hat{\underline{r}} = (\hat{r}_1,\ldots,\hat{r}_m)'$ are not the same, and moreover are such that the consequence of erroneously assuming that the two distributions are the same is a serious underestimation of significance in diagnostic checking, that is, a tendency to overlook lack of fit when it exists (see section 2). However it is shown in [5] that when these factors are taken into account, valid diagnostic checks can still be proposed based on the residual autocorrelations.

The purpose of the present paper is therefore to obtain the distribution of the residual auto- and cross correlations for the dynamic models (1.2) as was done in [5] for the class of models (1.1), and to examine in the light of these results the use of residual correlations in diagnostic checking and tests of fit of these models. It will be shown that both the residual autocorrelations $\{\hat{r}_k\}$ and the residual cross correlations $\{\hat{r}_k^*\}$ have the same properties relative to their white noise counterparts as do the residual autocorrelations in the IARMA processes (1.1). However in obtaining their distributions it is assumed that the input $\{x_t\}$ is in a special form as follows.

Input whitening.

In the physical operation of many processes or systems of the type we are considering, x_t is a known input which is generally set at specific controlled values in order to influence the output z_t in some manner [such as maximizing z_t or maintaining it at some target value]. If a pair of series $\{x_t\}$ and $\{z_t\}$ have been observed under these conditions there is little basis for the applications of such concepts as stationarity in their analysis. However by properly designing the experiment, that is by judiciously selecting the input $\{x_t\}$, this situation can be altered favorably. In particular suppose x_t is chosen randomly in such a way that it follows a known stationary process

$$\phi_x(B)\ x_t = \theta_x(B)\ \alpha_t^*$$

or

$$T(B)\ x_t = \frac{\phi_x(B)}{\theta_x(B)}\ x_t = \alpha_t^* \ , \tag{1.6}$$

where $\{\alpha_t^*\}$ is white noise, that is, the $\{\alpha_t^*\}$ are independently and identically distributed as $N(o,\sigma_\alpha^2)$. Then by letting

$$y_t = T(B)\ z_t \tag{1.7}$$

and

$$\frac{\theta'(B)}{\phi'(B)} = \frac{\phi_x(B)\ \theta(B)}{\theta_x(B)\ \phi(B)}$$

the model (1.2) can be written in the form

$$y_t = \frac{\omega(B)}{\delta(B)}\ \alpha_t + \frac{\theta'(B)}{\phi'(B)}\ a_t \tag{1.8}$$

where $\alpha_t = \alpha^*_{t-b-1}$. That is, y_t is the sum of two independent autoregressive - moving average processes where in addition the "noise" $\{\alpha_t\}$ generating one of them is known. It is assumed in the sequel that the input has been chosen in this way and that an appropriate input-whitening transormation has been performed on both the $\{x_t\}$ and $\{z_t\}$ series, prior to fitting and diagnostic checking, so that the dynamic/stochastic model discussed in this paper can be written in the form (1.8).

2. Review of Properties of Residual Autocorrelations in IARMA Processes

For future references [sections 3 and 5] the principal results regarding the large-sample distribution of the residual autocorrelations for autoregressive-moving average processes obtained by Box and Pierce [5] are now set forth.

2.1 Distribution of residual autocorrelations.

Suppose an IARMA model (1.1) is correctly identified and fitted to a series $\{z_t\}$, producing residuals $\{\hat{a}_t\}$ calculated from the parameter estimates $\hat{\phi}_1,\ldots,\hat{\phi}_p,\hat{\theta}_1,\ldots,\hat{\theta}_q$ through the relation

$$\hat{\phi}(B)\ \nabla^d\ z_t = \hat{\theta}(B)\ \hat{a}_t, \tag{2.1}$$

thus determining residual autocorrelations $\hat{\underline{r}} = (\hat{r}_1,\hat{r}_2,\ldots)'$ as in (1.3).

Linear expansion of $\hat{r}_k$.

In [5] it was shown that to $O_p(\frac{1}{n})$ the residual autocorrelations $\{\hat{r}_k\}$ from the IARMA model (1.1) have the same distribution as the $\{\hat{r}_k\}$ from a correctly fitted autoregressive model [with $y_t = \nabla^d z_t$]

$$\eta(B)\ y_t = a_t \tag{2.2}$$

where

$$\eta(B) = 1-\eta_1 B-\dots -\eta_{p+q}B^{p+q} = \phi(B)\theta(B) \tag{2.3}$$

and moreover that the residual autocorrelations from the latter can to the same order be approximated by a first order Taylor expansion about the white noise autocorrelations $\underline{r} = (r_1, r_2, \dots)'$ where

$$r_k = \frac{\sum a_t\ a_{t-k}}{\sum a_t^2} \tag{2.4}$$

to obtain

$$\hat{r}_k = r_k + \sum (\eta_j - \hat{\eta}_j)\ \frac{\partial \hat{r}_k}{\partial \hat{\eta}_j}\ .$$

Approximation of the derivatives.

It was also established to $O_p(\frac{1}{\sqrt{n}}$ that

$$\frac{\partial \hat{r}_k}{\partial \eta_j} = \frac{\sum_{i=0}^{p+q} \eta_i (\rho_{k-i+j}+\rho_{k+i-j})}{-\sum_{i=0}^{p} \sum_{j=0}^{p} \eta_i \eta_j \rho_{i-j}} \tag{2.5}$$

$$= \xi_{k-j} \tag{2.6}$$

where ρ_ν is the lag-ν theoretical autocorrelation of the process $\{y_t\}$, the $\{\xi_\nu\}$ are the coefficients in

$$\xi(B) = 1+\xi_1 B+\xi_2 B^2+\ldots \quad = [\eta(B)]^{-1}, \tag{2.7}$$

and (2.6) utilizes the dual relations

$$\eta(B)\rho_k = \eta(B)\xi_k = 0,\ k > 0 \tag{2.8}$$

characteristic of the class of autoregressive processes (2.2). Thus it is seen that

$$\hat{r}_k = r_k+\sum(\eta_j-\hat{\eta}_j)\xi_{k-j}+O_p(\tfrac{1}{n}) \tag{2.9}$$

or

$$\underline{\hat{r}} \approx \underline{r} + X(\underline{\eta}-\underline{\hat{\eta}}) \tag{2.10}$$

where

$$X = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \xi_1 & 1 & \ddots & \vdots \\ \xi_2 & \xi_1 & \ddots & 0 \\ \cdot & \cdot & & 1 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \xi_{m-1} & \xi_{m-2} & \cdots & \xi_{m-p-q} \end{bmatrix} \tag{2.11}$$

is an $m\times(p+q)$ matrix. The approximation (2.10) works best when m is small relative to m yet sufficiently large that the $\{\xi_\nu\}$ for $\nu > m-p-q$ are essentially zero.

$\underline{\hat{r}}$ as a linear transformation of $\underline{r}$.

A further result in [5] is that [to $O_p(\frac{1}{n})$]

$$\sum_k \hat{r}_k \, \xi_{k-j} = 0, \quad 1\leq j\leq p+q,$$

that is,

$$\underline{\hat{r}}' X = \underline{0} \, , \tag{2.12}$$

and from (2.10) and (2.12) it follows that

$$\underline{\hat{r}} = (I-Q) \, \underline{r}, \tag{2.13}$$

where

$$Q = X(X'X)^{-1} X' .$$

Thus to a close approximation for n sufficiently large,

$$\underline{\hat{r}} \sim N(\underline{0}, \frac{1}{n}[I-Q]) \tag{2.14}$$

which contrasts sharply to the large sample distribution

$$\underline{r} \sim N(\underline{0}, \frac{1}{n}I) \tag{2.15}$$

satisfied by the white noise autocorrelations [1,2].

Implications for diagnostic checking.

Since the elements $\{\xi_j\}$ of the X-matrix (2.11) are usually largest for the smallest values of j, it can be demonstrated that the biggest discrepancies between the distributions (2.14) and (2.15) [that is, the smallest variances and largest correlations in the residual autocorrelations] occur for the $\{\hat{r}_k\}$ of lowest lag k. Since it is precisely these quantities which are most apt to reveal existing lack of fit, it follows that treating $\hat{r}$'s as r's in diagnostic checking can cause a strong tendency to overlook significant model inadequacy. However if the $\hat{r}$'s are examined relative to their true (large-sample) distribution (2.14), these difficulties are overcome. In particular since the matrix I-Q is itempotent of rank m-p-q, the statistic

$$n \sum_{k=1}^{m} \hat{r}_k^2 \qquad (2.16)$$

will still possess a χ^2-distribution, only now with m-p-q rather than m degrees of freedom. Thus we reiterate the conclusion reached in [5] (which will shortly be seen to hold also in dynamic models), that while there exist sharp discrepancies between the distributions of $\hat{\underline{r}}$ and $\underline{r}$, these differences can be taken into account so that the residual correlations remain a powerful tool in diagnostic checking.

2.2 Alternative expansion for $\hat{r}_k$ is in the ARMA process.

The approximate linear expansion (2.9) for $\hat{r}_k$ is equally valid for the pure autoregressive and mixed autoregressive-moving average processes (2.2) and (1.1) respectively. For the ARMA process (1.1), however, with $\pi(B) = \phi(B)\ \theta^{-1}(B)$ and $\psi(B) = \pi^{-1}(B)$ we can write [with $y_t\ \nabla^d\ z_t$]

$$\pi(B)\ y_t = a_t,\ y_t = \psi(B)\ a_t; \qquad (2.17)$$

and an examination of equations (2.12) through (2.26) of [5] shows that replacing $\eta(B)$ by the infinite operator $\pi(B)$, and $\xi(B)$ by $\psi(B)$, causes no lose of validity in the resulting expression for $\hat{r}_k$ [the orthogonality relations (2.12) of the present paper no longer hold, however]; that is, for the mixed model (1.1) we have an alternative expansion to (2.9), namely

$$\hat{r}_k \doteq r_k + \sum (\pi_j - \hat{\pi}_j)\ \psi_{k-j} \qquad (2.18)$$

which will be very useful in the following section.

3. Distribution of Residual Autocorrelations

Suppose that two series $\{\alpha_t\}$ and $\{y_t\}$ are available [perhaps after an appropriate transformation as discussed at the end of section 1] for which a dynamic/disturbance model

$$y_t = \frac{\omega(B)}{\delta(B)}\alpha_t + \frac{\theta(B)}{\phi(B)} a_t \tag{3.1}$$

where

$$\omega(B) = \omega_o - \omega_1 B - \ldots - \omega_u B^u ,$$

$$\delta(B) = 1 - \delta_1 B - \ldots - \delta_v B^v ,$$

$\{\alpha_t\}$ and $\{a_t\}$ are both white noise [the former is the known input and the latter the unknown disturbances entering the system], and $\phi(B)$, $\theta(B)$ are as in (1.1), has been correctly identified. Then for any value

$$\dot{\underline{\lambda}} = (\dot{\underline{\omega}}, \dot{\underline{\delta}}, \dot{\underline{\phi}}, \dot{\underline{\theta}}) = (\dot{\omega}_0, \dot{\omega}_1, \ldots, \dot{\omega}_u, \dot{\delta}_1, \ldots, \dot{\delta}_v,$$

$$\dot{\phi}_1, \ldots, \dot{\phi}_p, \dot{\theta}_1, \ldots, \dot{\theta}_q)$$

in the parameter space, a quantity $\dot{a}_t$ is determined where

$$\dot{a}_t = a_t(\dot{\underline{\lambda}}) = \frac{\dot{\phi}(B)}{\dot{\theta}(B)} y_t - \frac{\dot{\phi}(B)}{\dot{\theta}(B)} \frac{\dot{\omega}(B)}{\dot{\delta}(B)} \alpha_t .$$

In particular setting $\dot{\underline{\lambda}} = \underline{\lambda}$ reproduces the model (3.1) while for $\dot{\underline{\lambda}} = \hat{\underline{\lambda}}$ the expression for the residuals $\{\hat{a}_t\}$ is obtained. Alternatively, with

$$\dot{V}(B) = \dot{V}_o + \dot{V}_1 B + \dot{V}_2 B^2 + \ldots = \frac{\dot{\omega}(B)}{\dot{\delta}(B)}$$

and

$$\dot{\pi}(B) = 1 - \dot{\pi}_1 B - \dot{\pi}_2 B^2 - \ldots = \frac{\dot{\phi}(B)}{\dot{\theta}(B)} ,$$

(3.1) becomes

$$y_t = V(B)\ \alpha_t + \frac{a_t}{\pi(B)} \quad . \tag{3.2}$$

and an expression equivalent to (3.2) is then

$$\begin{aligned} \dot{a}_t &= \dot{\pi}(B)\ y_t - \dot{\pi}(B)\ \dot{V}(B)\ \alpha_t \\ &= - \sum \dot{\pi}_i\ y_{t-i} + \sum \sum \dot{\pi}_i\ \dot{V}_j\ \alpha_{t-i-j} \end{aligned} \tag{3.3}$$

where $\dot{\pi}_o = -1$. It is assumed that the number n of observations in the two series is sufficiently large that there exists a number m where

(i) all π_j, V_j for $j \geq m$ are uniformly $o(\frac{1}{\sqrt{n}})$

(ii) $\frac{m}{n} = O(\frac{1}{\sqrt{n}})$.

Then in (3.3) and in all following expressions of a similar nature, the present order of approximation will be preserved by stopping the summations at i, j = m.

Based on (3.3) we can define an autocorrelation function $\dot{r}_k$ as

$$\dot{r}_k = \frac{\dot{c}_k}{\dot{c}_o} , \tag{3.4}$$

where $\dot{c}_k = \sum \dot{a}_t \dot{a}_{t-k}$ is n times the sample covariance of $\{\dot{a}_t\}$, so that from (3.3)

$$\dot{c}_k = \sum_t [\sum_i \dot{\pi}_i y_{t-i} - \sum_i \sum_j \dot{\pi}_i \dot{V}_j \alpha_{t-i-j}]\cdot[\sum_{i'} \dot{\pi}_{i'} - y_{t-i'-k}$$

$$- \sum_{i'} \sum_{j'} \dot{\pi}_{i'} \dot{V}_{j'} \alpha_{t-i'-j'-k}]$$

$$= \sum \sum \dot{\pi}_i \dot{\pi}_{i'} \overset{y}{c}_{k+i'-i}$$

$$+ \sum \sum \sum \sum \dot{\pi}_i \dot{\pi}_{i'} \dot{V}_j \dot{V}_{j'} \overset{\alpha}{c}_{i'+j'+k-i-j}$$

$$- \sum \sum \sum \dot{\pi}_i \dot{\pi}_{i'} \dot{V}_{j'} \overset{\alpha y}{c}_{i'+j'+k-i} - \sum \sum \sum \dot{\pi}_i \dot{\pi}_{i'} \dot{V}_j \overset{\alpha y}{c}_{i+j-k-i'}$$

$$\tag{3.5}$$

where

$$\overset{y}{c}_\nu = \sum y_t y_{t-\nu}, \quad \overset{\alpha y}{c}_\nu = \sum \alpha_{t-\nu} y_t,$$

and similarly for $\overset{\alpha}{c}_\nu$. Thus from (3.4) and (3.5) the autocorrelations $\dot{r}_k$ are determined as functions of $\dot{\underline{\pi}}$, $\dot{\underline{V}}$ and the various sample covariances of $\{\alpha_t\}$ and $\{y_t\}$. Our chief interest will be in the expressions obtained when $\dot{\underline{\pi}}$ and $\dot{\underline{V}}$ are set to the true and estimated parameter values, yielding respectively the white noise and residual autocorrelations $\dot{r}_k = r_k$ and $\dot{r}_k = \hat{r}_k$.

<u>Linear expansion of $\hat{r}_k$.</u>

Since $\hat{\pi}_j = \pi_j + O_p(\frac{1}{\sqrt{n}})$ and $\hat{V}_j = V_j + O_p(\frac{1}{\sqrt{n}})$, we can expand $\hat{r}_k$ about the true parameter values $(\dot{\underline{\pi}}, \dot{\underline{V}}) = (\underline{\pi}, \underline{V})$ to obtain

$$\hat{r}_k = r_k + \sum (\pi_i - \hat{\pi}_i)\, \hat{\mu}_{ik} + \sum (V_j - \hat{V}_j)\, \hat{\kappa}_{jk} + O_p(\tfrac{1}{n}) \qquad (3.6)$$

where

$$\hat{\mu}_{ik} = -\left.\frac{\partial \dot{r}_k}{\partial \dot{\pi}_i}\right|_{\wedge} = -\left[\sum a_t^2\right]^{-1} \left.\frac{\partial \dot{c}_k}{\partial \dot{\pi}_i}\right|_{\wedge}$$

$$\hat{\kappa}_{jk} = -\left.\frac{\partial \dot{r}_k}{\partial \dot{V}_j}\right|_{\wedge} = -\left[\sum \hat{a}_t^2\right]^{-1} \left.\frac{\partial \dot{c}_k}{\partial \dot{V}_j}\right|_{\wedge}$$

and "^" connotes that all derivatives are evaluated at the least-squares parameter estimates. Systematic differentiation of (3.5) show that

$$\hat{\mu}_{ik} = -[\sum \hat{a}_t^2]^{-1} \{\sum_{i'} \hat{\pi}_{i'} [\overset{y}{c}_{k+i-i'} + \overset{y}{c}_{k+i'-i}]$$

$$+ \sum_{i'} \sum_{j} \sum_{j'} \hat{\pi}_{i'} \hat{v}_j \hat{v}_{j'}$$

$$[\overset{\alpha}{c}_{i'+j'+k-i-j} + \overset{\alpha}{c}_{i+j'+k-i'-j}]$$

$$- \sum_{i'} \sum_{j} \hat{\pi}_{i'} \hat{v}_j [\overset{\alpha y}{c}_{i'+j+k-i} + \overset{\alpha y}{c}_{i+j+k-i'}]$$

$$- \sum_{i'} \sum_{j} \hat{\pi}_{i'} \hat{v}_j [\overset{\alpha y}{c}_{i'+j-k-i} + \overset{\alpha y}{c}_{i+j-k-i'}]\} \qquad (3.7)$$

and

$$\hat{\kappa}_{jk} = [-\sum \hat{a}_t^2]^{-1} \{\sum_{i} \sum_{i'} \sum_{j'} \hat{\pi}_i \hat{\pi}_{i'} \hat{v}_{j'}$$

$$[\overset{\alpha}{c}_{i'+j'+k-i-j} + \overset{\alpha}{c}_{i'+j+k-i-j'}]$$

$$- \sum_{i} \sum_{i'} \hat{\pi}_i \hat{\pi}_{i'} \overset{\alpha y}{c}_{i'+j+k-i}$$

$$- \sum_{i} \sum_{i'} \hat{\pi}_i \hat{\pi}_{i'} \overset{\alpha y}{c}_{i+j-k-i'}\}, \qquad (3.8)$$

where $\sum \hat{a}_t^2$ in these expressions is obtained by setting "·" = "^" and k = 0 in (3.5).

Approximation of the derivatives.

We have already remarked that the root mean square errors of $\{\hat{\pi}_j\}$ and $\{\hat{v}_j\}$ are $O(\frac{1}{\sqrt{n}})$, and since this is also true of the sample correlations of $\{\alpha_t\}$ and $\{y_t\}$, it follows that if μ_{ik} and κ_{jk} are the result of replacing the estimated parameters and covariances in (3.7) and (3.8) by the theoretical values, then

$$\hat{\mu}_{ik} = \mu_{ik} + O_p(\frac{1}{\sqrt{n}}) \ ,$$

$$\hat{\kappa}_{jk} = \kappa_{jk} + O_p(\frac{1}{\sqrt{n}} \ , \tag{3.9}$$

so that making the substitution in the linear expansion (3.6) preserves its accuracy; that is,

$$\hat{r}_k = r_k + \sum (\pi_i - \hat{\pi}_i)\ \mu_{ik} + \sum (V_j - \hat{V}_j)\ \kappa_{jk} + O_p(\frac{1}{n}). \tag{3.10}$$

In evaluating the approximations μ_{ik} and κ_{jk} in (3.10) to the derivatives in (3.6) it is first helpful to obtain an approximate form for $\frac{1}{n} \sum \hat{a}_t^2$, which converges in probability to the variance σ_a^2 of the white noise $\{a_t\}$. Now if

$$e_t = \frac{\theta(B)}{\phi(B)}\ a_t = \psi(B)\ a_t = \pi^{-1}(B)\ a_t$$

represents the noise in the dynamic model (3.1), then it is known (e.g., [4], Chapter II) that its theoretical autocovariances are given by

$$\overset{e}{\gamma}_k = \sigma_a^2 \sum \psi_j\ \psi_{j+k} \ . \tag{3.11}$$

Now since $\pi(B)\cdot\psi(B) = 1$ it can be shown that

$$\sum_i \sum_{i'} \pi_i \pi_{i'} \sum_j \psi_j \psi_{j+i-i'} = 1, \tag{3.12}$$

and combining (3.11) and (3.12) we obtain

$$\frac{1}{n}\sum \hat{a}_t^2 \approx \sigma_a^2 = \sum_i \sum_{i'} \pi_i \pi_{i'} \overset{e}{\gamma}_{i-i'} \tag{3.13}$$

$$= K^{-1}, \text{ say.}$$

Then with $\overset{\alpha}{\gamma}_k = E(\alpha_t \alpha_{t-k})$, $\overset{\alpha y}{\gamma}_k = E(\alpha_{t-k} y_t)$, and $\overset{y}{\gamma}_k = E(y_t y_{t-k})$ denoting the theoretical auto- and cross covariances of $\{\alpha_t\}$ and $\{y_t\}$, we have

$$\mu_{ik} = K \{\sum_{i'} \pi_{i'} [\overset{y}{\gamma}_{k+i-i'} + \overset{y}{\gamma}_{k+i'-i}]$$

$$+ \sum_{i'} \sum_j \sum_{j'} \pi_{i'} v_j v_{j'}$$

$$[\overset{\alpha}{\gamma}_{i'+j'+k-i-j} + \overset{\alpha}{\gamma}_{i+j'+k-i'-j}]$$

$$- \sum_{i'} \sum_j \pi_{i'} v_j [\overset{\alpha y}{\gamma}_{i'+j+k-i} + \overset{\alpha y}{\gamma}_{i+j+k-i'}]$$

$$- \sum_{i'} \sum_j \pi_{i'} v_j [\overset{\alpha y}{\gamma}_{i'+j-k-i} + \overset{\alpha y}{\gamma}_{i+j-k-i'}]\}$$

$$= K \{\sum_{i'} \pi_{i'} [\overset{y}{\gamma}_{k+i-i'} + \overset{y}{\gamma}_{k+i'-i}]$$

$$- \sum_{i'} \pi_{i'} (\sigma_\alpha^2 \sum_j v_j [v_{i'+j+k-i} + v_{i+j+k-i'}])\}$$

(3.14)

$$= K \sum_{i'} \pi_{i'} \, [\gamma^{e}_{k+i-i'} + \gamma^{e}_{k+i'-i}] \tag{3.15}$$

$$= \frac{\sum_{i'} \pi_i \, [\,{}^{e}_{k+i-i'} + {}^{e}_{k+i'-i}]}{\sum_{i} \sum_{i'} \pi_i \, \pi_{i'} \gamma^{e}_{i-i'}} \tag{3.16}$$

where (3.14) follows because

$$\gamma^{\alpha}_{k} = 0, \; k \neq 0; \; \gamma^{\alpha}_{o} = \sigma^2_{\alpha}; \; \gamma^{\alpha y}_{k} = \sigma^2_{\alpha} \, v_k;$$

equation (3.15) utilizes the easily verified relation

$$\gamma^{e}_{k} = \gamma^{y}_{k} - \sigma^2_{\alpha} \sum_{j} v_j \, v_{j+k} \; ;$$

and (3.16) follows from (3.15) and (3.13).

Now just as for an autoregressive process

$$\phi(B) \, e_t = a_t$$

the autovariances satisfy

$$\phi(B) \, \gamma^{e}_{k} = 0 \; [\text{compare } (2.8)],$$

it is also true for a mixed AR-MA process

$$e_t = \frac{\theta(B)}{\phi(B)} a_t = \frac{a_t}{\pi(B)}$$

that

$$\pi(B) \overset{e}{\gamma}_k = 0 \tag{3.17}$$

Therefore by comparing (3.16) with the corresponding formulas (2.5) and (2.6) in section 2, it follows that

$$\mu_{ik} = \psi_{k-i} \tag{3.18}$$

and thus to $O_p(\frac{1}{\sqrt{n}})$ the derivative with respect to π_j of $\hat{r}_k$ in a dynamic model is the same as that of $\hat{r}_k$ in the corresponding stochastic model. Moreover, to $O_p(\frac{1}{\sqrt{n}})$ the derivaties of $\hat{r}_k$ with respect to the dynamic parameters are approximately

$$\begin{aligned}
\kappa_{jk} &= K \{\sum_i \sum_{i'} \sum_{j'} \pi_i \pi_{i'} V_{j'} [\overset{\alpha}{\gamma}_{i'+j'+k-i-j} + \overset{\alpha}{\gamma}_{i'+j+k-i-j'}] \\
&\quad - \sum_i \sum_{i'} \pi_i \pi_{i'} [\overset{\alpha y}{\gamma}_{i'+j+k-i} + \overset{\alpha y}{\gamma}_{i\ +j-k-i'}]\} \\
&= K \sigma_\alpha^2 \{\sum \sum \pi_i \pi_{i'} [V_{i+j-k-i'} + V_{i'+j+k-i}] \\
&\quad - \sum \sum \pi_i \pi_{i'} [V_{i'+j+k-i} + V_{i+j-k-i'}]\} \\
&= 0
\end{aligned} \tag{3.19}$$

so that for large n we have the important result that the effect on $\hat{r}_k$ of changes in the parameters $\hat{V}_j$ (in a region containing the true parameter values) is negligible; that is, the residual autocorrelations $\hat{\underline{r}}$ are distributed independently of the dynamic parameters.

Distribution of $\hat{\underline{r}}$.

By substituting (3.18) and (3.19) into (3.10), the following expression is obtained:

$$\hat{r}_k = r_k + \sum (\pi_j - \hat{\pi}_j) \psi_{k-j} + O_p(\tfrac{1}{n}), \tag{3.20}$$

$1 \leq k \leq m$, where m is as in the discussion following (3.4). But equation (3.20) and the corresponding result (2.18) for the residual autocorrelations from an autoregressive - moving average model are identical. Therefore the large-sample distributions of $\hat{\underline{r}}$ in the two models

$$y_t = \frac{\omega(B)}{\delta(B)} \alpha_t + \frac{\theta(B)}{\phi(B)} a_t \tag{3.21}$$

and

$$\phi(B) \; y_t' = \theta(B) \; a_t \tag{3.22}$$

are the same and independent of any dynamic parameters in the model (3.21). Discussion of the distribution of the residual autocorrelations in dynamic models can therefore be referred to section 2.1 (or to [5] for greater detail), as all of the results obtained there apply equally in the present setting.

Dyamic model with white noise.

An important situation occurs when a dynamic model is fitted with the disturbances or noise assumed white [that is, $\phi(B) = \theta(B) = 1$ in (3.1)]. For this case, that is if a model of the form

$$y_t = \frac{\omega(B)}{\delta(B)} \alpha_t + a_t \tag{3.23}$$

is correctly identified and then fitted, it follows from (2.13) that to $O_p(\frac{1}{n})$, since $\underline{Q} = \underline{0}$,

$$\hat{\underline{r}} = \underline{r} \; . \tag{3.24}$$

Therefore we do have a situation where the supposition that for large n the residual autocorrelations "ought" to behave as though they were calculated from white noise will not lead us astray [It will everywhere else, as has been seen so far and as will be seen again in section 4].

4. Distribution of Residual Cross Correlations.

The derivation of the distribution of residual autocorrelations in dynamic models was considerably simplified by the fact that the linear expansion of $\hat{\underline{r}}$ was totally independent of the dynamic parameters and moreover identical to the expansion in ARMA models for which the distribution had already been obtained [5]. The situation is less straightforward for residual cross correlations $\hat{\underline{r}}^* = (\hat{r}_1, \hat{r}_2, \ldots)'$, where both the dynamic and stochastic parameters will be seen to play a role, but the same approach of approximating $\hat{r}_k^*$ by a first order Taylor expansion will be seen to lead to a representation of the residual cross correlations as a singular linear transformation of the corresponding white noise cross correlations. Furthermore it will be found that the degree or dimension of singularity in this transformation, and hence in the resulting covariance matrix of $\hat{\underline{r}}^*$, is the number (u+v+1) of estimated dynamic parameters in the model (3.1), irrespective of the number of parameters associated with the noise.

4.1 Residual cross correlations as a linear transformation of white noise cross correlations.

Suppose the series $\{y_t\}$ and $\{\alpha_t\}$ are available from a dynamic model (3.1) which we now write as

$$y_t = \frac{\omega(B)}{\delta(B)} \alpha_t + \frac{a_t}{\pi(B)} , \tag{4.1}$$

where $\pi(B) = \phi(B)\ \theta^{-1}(B)$. Then for any value $\dot{\underline{\lambda}} = (\dot{\underline{\pi}}, \dot{\underline{\omega}}, \dot{\underline{\delta}})$ of the parameters we can define [as in (3.3)]

$$\dot{a}_t = \dot{\pi}(B)\ y_t - \dot{\pi}(B)\ \frac{\dot{\omega}(B)}{\dot{\delta}(B)}\ \alpha_t, \tag{4.2}$$

$$\dot{c}_k^* = \sum \alpha_{t-k}\ \dot{a}_t , \tag{4.3}$$

and thus determine the <u>cross correlations</u> between $\{\alpha_t\}$ and $\{\dot{a}_t\}$ as

$$\dot{r}_k^* = (\sum \alpha_t^2 \sum \dot{a}_t^2)^{-\frac{1}{2}} \dot{c}_k^* \qquad (4.4)$$

so that setting $\dot{\underline{a}} = \hat{\underline{a}}$ and $\dot{\underline{a}} = \underline{a}$ in turn yields, respectively, the residual cross correlations

$$\hat{r}_k^* = \frac{\sum \alpha_{t-k} \hat{a}_t}{\sqrt{\sum \alpha_t^2 \sum \hat{a}_t^2}} \qquad (4.5)$$

and white noise cross correlations

$$r_k^* = \frac{\sum \alpha_{t-k} a_t}{\sqrt{\sum \alpha_t^2 \sum a_t^2}} \quad . \qquad (4.6)$$

<u>Linear expansion of $\hat{r}_k^*$.</u>

The residual cross correlations $\hat{r}_k^*$ as functions of $\hat{\underline{\lambda}} = (\hat{\underline{\omega}}, \hat{\underline{\delta}}, \hat{\underline{\pi}})$ can be approximated by means of a first order Taylor expansion about $\dot{\underline{\lambda}} = \underline{\lambda}$ to obtain

$$\hat{r}_k^* = r_k^* + \sum_{i=0}^{u} (\omega_i - \hat{\omega}_i)\, \kappa_{ik}^* + \sum_{j=1}^{v} (\delta_j - \hat{\delta}_j)\, \hat{\tau}_{jk}^*$$

$$+ \sum_{j=1}^{\infty} (\pi_j - \hat{\pi}_j)\, \hat{\mu}_{jk}^* + O_p(\tfrac{1}{n}) \qquad (4.7)$$

where as in previous expansions the accuracy to $O_p(\frac{1}{n})$ is a consequence of the size of the root mean square errors of $(\hat{\omega}_i, \hat{\delta}_j, \hat{\phi}_i, \hat{\theta}_j)$ and where

$$\hat{\kappa}_{ik}^* = - \frac{\partial \dot{r}_k}{\partial \dot{\omega}_i} \bigg|_{\wedge} = - [\sum \hat{a}_t^2 \sum \alpha_t^2]^{-\frac{1}{2}} \frac{\partial \dot{c}_k^*}{\partial \dot{\omega}_i} \bigg|_{\wedge}$$

$$\hat{\tau}_{jk}^* = - \frac{\partial \dot{r}_k^*}{\partial \dot{\delta}_j} \bigg|_{\wedge} = - [\sum \hat{a}_t^2 \sum \alpha_t^2]^{-\frac{1}{2}} \frac{\partial \dot{c}_k^*}{\partial \dot{\delta}_j} \bigg|_{\wedge}$$

$$\hat{\mu}_{jk}^* = - \frac{\partial \dot{r}_k^*}{\partial \dot{\pi}_j} \bigg|_{\wedge} = - [\sum \hat{a}_t^2 \sum \alpha_t^2]^{-\frac{1}{2}} \frac{\partial \dot{c}_k^*}{\partial \dot{\pi}_j} \bigg|_{\wedge}$$

where the "^" denotes that all derivatives are evaluated at the least squares parameter estimates. If $\hat{\lambda}_j$ represents any one of the parameters $(\hat{\omega}_i, \hat{\delta}_i, \hat{\pi}_i)$, then these derivatives can be determined from the relations

$$\frac{\partial \hat{r}_k^*}{\partial \hat{\lambda}_j} = (\sum \alpha_t^2 \sum \hat{a}_t^2)^{-\frac{1}{2}} \frac{\partial \hat{c}_k^*}{\partial \hat{\lambda}_j}$$

$$= \hat{K}^{-1} \frac{\partial}{\partial \hat{\lambda}_j} (\sum \alpha_{t-k} \hat{a}_t)$$

$$= \hat{K}^{-1} \sum_t \alpha_{t-k} \frac{\partial \hat{a}_t}{\partial \hat{\lambda}_j}$$

where $\hat{a}_t$ is as in (4.2). Thus

$$\hat{\kappa}^*_{ik} = \hat{K}^{-1} \sum \alpha_{t-k} \, [\hat{\pi}(B) \frac{1}{\hat{\delta}(B)} \, \alpha_{t-i}]$$

$$\hat{\tau}^*_{jk} = \hat{K}^{-1} \sum \alpha_{t-k} \, [-\hat{\pi}(B) \, \frac{\hat{\omega}(B)}{\hat{\delta}^2(B)} \quad \alpha_{t-j}] \qquad (4.8)$$

$$\hat{\mu}^*_{jk} = \hat{K}^{-1} \sum \alpha_{t-k} \, [-y_{t-j} + \frac{\hat{\omega}(B)}{\hat{\delta}(B)} \, \alpha_{t-j}]$$

Approximation of the derivatives.

Let κ^*_{ik}, τ^*_{jk}, and μ^*_{jk} be the result of replacing the parameter estimates $\hat{\underline{\lambda}} = (\hat{\underline{\omega}}, \hat{\underline{\delta}}, \hat{\underline{\pi}})$ in (4.7) by the true values $\underline{\lambda}$. Then, since

$$\frac{1}{n} \hat{K} = \sigma_a \, \sigma_\alpha + O_p(\frac{1}{\sqrt{n}}), \qquad (4.9)$$

it follows that

$$\hat{\kappa}^*_{ik} = \kappa^*_{ik} + O_p(\frac{1}{\sqrt{n}})$$

$$\hat{\tau}^*_{jk} = \tau^*_{jk} + O_p(\frac{1}{\sqrt{n}}) \qquad (4.10)$$

$$\hat{\mu}^*_{jk} = \mu^*_{jk} + O_p(\frac{1}{\sqrt{n}})$$

so that by making these substitutions in (4.6) the accuracy of the latter approximation is preserved; that is,

$$\hat{r}_k^* = r_k + \sum (\omega_i - \hat{\omega}_i)\ \kappa_{ik}^* + \sum (\delta_j - \hat{\delta}_j)\ \tau_{jk}^*$$

$$+ \sum (\pi_j - \hat{\pi}_j)\ \mu_{jk}^* + O_p(\tfrac{1}{n}) \qquad (4.11)$$

The expansion (4.11) can be simplified by observing [from (4.8) and (4.9)] that

$$\mu_{jk}^* = (n\ \sigma_a\ \sigma_\alpha)^{-1} \sum \alpha_{t-k}\ [-y_{t-j} + V(B)\ \alpha_{t-j}]$$

$$= \frac{1}{n\sigma_a\ \sigma_\alpha}\ [-c_{k-j}^{\alpha y} + \sum_i V_i\ c_{k-j-i}]$$

which again without affecting the degree of accuracy of (4.7) or (4.11) can be approximated by

$$\mu_{jk} = \frac{1}{\sigma_a\ \sigma_\alpha}\ [-\gamma_{k-j}^{\alpha y} + \sum V_j\ \gamma_{k-j-1}^{\alpha}]$$

$$= \frac{1}{\sigma_a\ \sigma_\alpha}\ [-V_{k-j}\ \sigma_\alpha^2 + V_{k-j}\ \sigma_\alpha^2] = 0 \qquad (4.12)$$

where the $\{c_\nu^{\alpha y}\}$, $\{\gamma_\nu^{\alpha y}\}$, etc. are as in the expressions following (3.5) and (3.13). Thus all terms containing $(\pi_j - \hat{\pi}_j)$ in the expansion (4.11) of $\hat{r}_k^*$ can be dropped, and it is seen that $\hat{\underline{r}}^*$ for large n does not depend on the noise parameter estimates $\hat{\pi}(B) = \hat{\theta}^{-1}(B)\ \hat{\phi}(B)$, which is interesting to compare with the fact that the autocorrelations $\hat{\underline{r}}$ in dynamic models do not depend on $\hat{V}(B) = \hat{\delta}^{-1}(B)\ \hat{\omega}(B)$ as was demonstrated in the last section. However it will now be seen that $\hat{\underline{r}}^*$ does depend on the true values $\pi(B)$ whereas $\hat{\underline{r}}$ was to a close approximation distributed independently of $\nabla(B)$ altogether.

It is necessary now to obtain the approximate derivatives of the linear expansion of $\hat{r}_k^*$ with respect to the dynamic parameters $\underline{\omega} = (\omega_o, \omega_1, \ldots, \omega_u)'$ and $\underline{\delta} = (\delta_1, \ldots, \delta_v)'$. From (4.7) and (4.8),

$$\kappa_{ik}^* = (n\ \sigma_\alpha\ \sigma_a)^{-1} \sum \alpha_{t-k}\ [\pi(B)\ \delta^{-1}(B)\ \alpha_{t-i}]$$

$$\tau_{jk}^* = (n\ \sigma_\alpha\ \sigma_a)^{-1} \sum \alpha_{t-k}\ [\pi(B)\ \delta^{-2}(B)\ \omega(B)\ \alpha_{t-j}]; \tag{4.13}$$

and by defining

$$\beta_t = \delta^{-2}(B)\ \pi(B)\ \alpha_t = \varepsilon(B)\ \pi(B)\ \alpha_t \tag{4.14}$$

where

$$\varepsilon(B) = 1 + \varepsilon_1 B + \varepsilon_2 B^2 + \ldots = \delta^{-2}(B), \tag{4.15}$$

equations (4.13) become

$$\kappa^*_{ik} = (n\ \sigma_\alpha\ \sigma_a)^{-1} \sum_t \alpha_{t-k}\ \delta(B)\ \beta_{t-i}$$

$$= \frac{1}{n\sigma_\alpha\ \sigma_a}\ \delta(B)\ c_{k-i}^{\alpha\beta} \tag{4.16}$$

and

$$\tau^*_{jk} = (n\ \sigma_\alpha\ \sigma_a)^{-1} \sum_t \alpha_{t-k}\ \omega(B)\ \beta_{t-j}$$

$$= \frac{1}{n\sigma_\alpha\ \sigma_a}\ \omega(B)\ c^{\alpha\beta}_{k-j} \tag{4.17}$$

where

$$c^{\alpha\beta}_k = \sum \alpha_{t-k}\ \beta_t$$

$$= \sum_t \alpha_{t-k}\ [\sum_i \sum_j \varepsilon_i\ \pi_j\ \alpha_{t-i-j}]$$

$$= \sum_i \sum_j \varepsilon_i\ \pi_j\ c^{\alpha}_{k-i-j}$$

Since $\frac{1}{n} c^{\alpha}_k = \gamma^{\alpha}_k + O_p(\frac{1}{\sqrt{n}})$ this may be approximated (upon dividing both sides by n) by

$$\gamma^{\alpha\beta}_k = \sum \sum \varepsilon_i\ \pi_j\ \gamma^{\alpha}_{k-i-j}$$

$$= \sigma^2_\alpha \sum_i \varepsilon_i\ \pi_{k-i} \tag{4.18}$$

$$= \sigma^2_\alpha\ \varepsilon(B)\ \pi_k$$

so that, denoting by κ_{ik} and τ_{jk} the results of substituting the approximation (4.18) for c_k into (4.16) and (4.17), our approximations to the derivatives (4.8) are finally

$$\kappa_{ik} = \frac{\sigma_\alpha}{\sigma_a} \delta(B)\ \varepsilon(B)\ \pi_{k-i}, \tag{4.19}$$

$$\tau_{jk} = \frac{\sigma_\alpha}{\sigma_a} \omega(B)\ \varepsilon(B)\ \pi_{k-j}. \tag{4.20}$$

When (4.12), (4.19), and (4.20) are substituted for the derivatives in (4.7) or (4.11), we obtain

$$\begin{aligned}\hat{r}_k^* &= r_k^* + \sum (\omega_i - \hat{\omega}_i)\ \kappa_{ik} + \sum (\delta_j - \hat{\delta}_j)\ \tau_{jk} + O_p(\tfrac{1}{n}) \\ &= r_k^* + \frac{\sigma_\alpha}{\sigma_a} \sum_{i=0}^{u} (\omega_i - \hat{\omega}_i)\ \delta(B)\ \varepsilon(B)\ \pi_{k-i} \\ &\quad + \frac{\sigma_\alpha}{\sigma_a} \sum_{j=1}^{v} (\delta_j - \hat{\delta}_j)\ \omega(B)\ \varepsilon(B)\ \pi_{k-j} + O_p(\tfrac{1}{n}) \end{aligned} \tag{4.21}$$

<u>Linear constraints on the $\hat{r}^*$'s.</u>

To derive from (4.21) a representation of $\underline{\hat{r}}^*$ as a linear transformation of the white noise cross correlations $\underline{r}^*$ analogous to expressing the residuals of a linear regression as a function of the true errors [as in (2.13) for <u>auto</u>correlations], it is necessary establish the orthogonality relations

$$\sum \hat{r}^*\ \kappa_{ik} = 0, \quad 0 \le i \le u, \tag{4.22}$$

$$\sum \hat{r}_k^*\ \tau_{jk} = 0, \quad 1 \le j \le v, \tag{4.23}$$

where the κ's and τ's are given by (4.19) and (4.20). This will be done to the present order of approximation by once again making a linear expansion, this time of $\hat{a}_t$ itself and in terms of the parameters estimated in the model. Thus

$$\hat{a}_t = a_t + \sum_{i=1}^{p} (\hat{\phi}_i - \phi_i) \frac{\partial a_t}{\partial \phi_i} + \sum_{j=1}^{q} (\hat{\theta}_j - \theta_j) \frac{\partial a_t}{\partial \theta_j}$$

$$+ \sum_{i=0}^{u} (\hat{\omega}_i - \omega_i) \frac{\partial a_t}{\partial \omega_i} + \sum_{j=1}^{v} (\hat{\delta}_j - \delta_j) \frac{\partial a_t}{\partial \delta_j} + O_p(\frac{1}{n})$$

where now the derivatives are evaluated at the true parameter values. Comparing this to a linear regression model it is seen that to $O_p(\frac{1}{n})$,

$$\sum \hat{a}_t \frac{\partial a_t}{\partial \lambda_j} = 0 \tag{4.24}$$

where λ_j is any of the p+q+u+l+v parameters $\underline{\lambda} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \omega_o, \ldots, \omega_u, \delta_1, \ldots, \delta_v)$.
It will be sufficient to utilize the relations (4.24) for the dynamic parameters $\underline{\omega}$ and $\underline{\delta}$. Rewriting (4.2) as

$$\dot{a}_t = \frac{\dot{\phi}(B)}{\dot{\theta}(B)} Y_t - \frac{\dot{\phi}(B)}{\dot{\theta}(B)} \frac{\dot{\omega}(B)}{\dot{\delta}(B)} \alpha_t,$$

the appropriate derivatives are

$$\frac{\partial a_t}{\partial \omega_i} = \pi(B)\ \delta^{-1}(B)\ \alpha_{t-i}$$

$$= \pi(B)\ \delta(B)\ \varepsilon(B)\ \alpha_{t-i}$$

and

$$\frac{\partial a_t}{\partial \delta_j} = -\pi(B)\ \omega(B)\ \delta^{-2}(B)\ \alpha_{t-j}$$

$$= -\pi(B)\ \omega(B)\ \varepsilon(B)\ \alpha_{t-j}$$

Therefore, to $O_p(\frac{1}{n})$

$$0 = \sum \hat{a}_t \frac{\partial a_t}{\partial \omega_i}\ ,\ 0 \leq i \leq u$$

$$= \sum \hat{a}_t\ [\pi(B)\ \delta(B)\ \varepsilon(B)\ \alpha_{t-i}]$$

$$= \sum_t \hat{a}_t \sum_j \sum_k \sum_\ell \pi_j\ \delta_k\ \varepsilon_\ell\ \alpha_{t-i-j-k-\ell}$$

$$= \sum_j \sum_K \sum_\ell \pi_j\ \delta_j\ \varepsilon_\ell\ \hat{c}^*_{i+j+k+\ell}$$

$$= \sum_{j'} \sum_k \sum_\ell \delta_k\ \varepsilon_\ell\ \pi_{j'-i-k-\ell}\ \hat{c}^*_{j'}$$

(where $j' = i + j + k + \ell$)

$$= \sum_{j'} \hat{c}^*_{j'}\ \delta(B)\ \varepsilon(B)\ \pi_{j'-i}$$

$$= \sum \hat{r}^*_k\ \delta(B)\ \varepsilon(B)\ \pi_{k-i},\ 0 \leq i \leq u$$

(by changing j' to k and observing that dividing $\hat{c}^*_k$ by a constant does not change the zero-value of the expression),

and

$$0 = \sum \hat{a}_t \frac{\partial a_t}{\partial \delta_j}, \quad 1 \leq j \leq v$$

$$= \sum \hat{a}_t \; [\pi(B) \; \omega(B) \; \varepsilon(B) \; \alpha_{t-j}]$$

$$= \sum \hat{r}_k^* \; \omega(B) \; \varepsilon(B) \; \pi_{k-j}, \quad 1 \leq j \leq v$$

(by exactly analogous procedures).

$\underline{\hat{r}}^*$ as a linear function of $\underline{r}^*$.

The relations (4.22) and (4.23) having been established, we may proceed as follows. Let

$$\xi(B) = \delta(B) \; \varepsilon(B) \; \pi(B) = \delta^{-1}(B) \; \pi(B)$$

$$= 1 + \xi_1 \; B + \xi_2 \; B^2 + \ldots$$

and

$$\chi(B) = \omega(B) \; \varepsilon(B) \; \pi(B) = \omega(B) \; \delta^{-2}(B) \; \pi(B)$$

$$= \chi_o + \chi_1 \; B + \chi_2 \; B^2 + \ldots$$

Then, if

$$X = \frac{\sigma_\alpha}{\sigma_a} \left[\begin{array}{ccccc|ccccc} 1 & 0 & 0 & \cdots & 0 & \chi_o & 0 & 0 & \cdots & 0 \\ \xi_1 & 1 & 0 & \ddots & \vdots & \chi_1 & \chi_o & 0 & \ddots & \vdots \\ \xi_2 & \xi_1 & 1 & \ddots & 0 & \chi_2 & \chi_1 & \chi_o & \ddots & 0 \\ \cdot & \cdot & \cdot & & 1 & \cdot & \cdot & \cdot & \ddots & \\ \cdot & \cdot & \cdot & & \vdots & \cdot & \cdot & \cdot & & \chi_o \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & & \vdots \\ \xi_{m-1} & \xi_{m-2} & \xi_{m-3} & \cdots & \xi_{m-u-1} & \chi_{m-1} & \chi_{m-2} & \chi_{m-3} & \cdots & \chi_{m-v} \end{array} \right] \tag{4.25}$$

is an $[m \times (u+v+1)]$ matrix, then (4.21) can be written in vector form to $O_p(\frac{1}{n})$ as

$$\hat{\underline{r}}^* = \underline{r}^* + X \begin{pmatrix} \underline{\omega} - \hat{\underline{\omega}} \\ \underline{\delta} - \hat{\underline{\delta}} \end{pmatrix} \tag{4.26}$$

where from (4.22) and (4.23)

$$\hat{\underline{r}}^{*\prime} X = \underline{0} \tag{4.27}$$

so that multiplying both sides of (4.26) by

$$Q = X(X' X)^{-1} X'$$

it is easily seen that

$$\hat{\underline{r}}^* = (I - Q)\, \underline{r}^* \tag{4.28}$$

which except for the structure of the X-matrix is of the same form as (2.13) for the residual autocorrelations $\hat{\underline{r}}$. [Note that while X is scaled by the ratio $\frac{\sigma_\alpha}{\sigma_a}$, Q is independent of these parameters].

4.2. Distribution of $\underline{r}^*$.

Recall that once the relationship (2.13) between $\hat{\underline{r}}$ and $\underline{r}$ in an autoregressive process was established, the distribution of $\hat{\underline{r}}$ followed as a direct consequence of R. L. Anderson's [1], Bartlett's [3] and T. W. Anderson and Walker's [2] results on the large - sample distribution of the white noise autocorrelations, namely that

$$\underline{r} \sim N(\underline{0}, \frac{1}{n} I). \tag{4.29}$$

The author was unable to find in the literature corresponding results for the cross correlations $\underline{r}^*$ between two white noise sequences $\{\alpha_t\}$ and $\{a_t\}$, and this subsection is therefore devoted to obtaining the distribution of $\underline{r}^*$, which is found to reproduce (4.29).

Moments.

The white noise cross correlations $\{r_k^*\}$ are given by

$$r_k^* = \frac{\sum \alpha_{t-k}\, a_t}{\sqrt{\sum \alpha_t^2 \sum a_t^2}} \tag{4.30}$$

where k = 1, 2, ..., m and it is assumed that m is small relative to n. To evaluate the moments of (4.30) it is very helpful to make use of an argument due to Koopmans [8] and extended by Moran [9] and Jenkins [7]. Applied to the present situation the argument proceeds as follows. The ratio (4.30) is a homogeneous function of degree zero in $\underline{a} = (a_1, \ldots, a_n)$ and in $\underline{\alpha} = (\alpha_1, \ldots, \alpha_n)$ and thus it is independent of the lengths of these vectors $|\underline{a}| = \sqrt{\sum a_t^2}$, $|\alpha| = \sqrt{\sum \alpha_t^2}$, or any function of them, and instead depends only on the angles which $\underline{a}$ and $\underline{\alpha}$ make with the axes in n-dimensional space. Since the joint normal density functions of $\underline{a}$ and of $\underline{\alpha}$,

$$p(\underline{a}, \underline{\alpha}) = p_1(\underline{a})\ p_2(\underline{\alpha})$$

$$= \frac{1}{(2\pi)^n} e^{-\frac{1}{2\sigma_a^2} \sum a_t^2 - \frac{1}{2\sigma_\alpha^2} \sum \alpha_t^2}$$

are independent of these angles and depend only on the lengths, it follows that the distribution of the ratio r_k^* of

$$c_k^* = \sum \alpha_{t-k}\ a_t \tag{4.31}$$

and

$$d = (\sum \alpha_t^2 \sum a_t^2)^{1/2}, \tag{4.32}$$

given $d = d_o$, is independent of the value d_o; that is, r_k^* and d are independent. In particular, therefore,

$$E\ (r_k^* \cdot d)^p = E\ (r_k^*)^p\ E\ (d^p),$$

and thus

$$E\ (r_k^*)^p = \frac{E\ (\sum \alpha_{t-k}\ a_t)^p}{E\ (\sum \alpha_t^2 \sum a_t^2)^{p/2}}\ . \tag{4.33}$$

It is clear that these same arguments apply if several r^*'s are considered jointly; for example,

$$E\ (r_k^{*p}\ r_{k+j}^{*\ q}) = \frac{E[(\sum \alpha_{t-k}\ a_t)^p\ (\sum \alpha_{t-k-j}\ a_t)^q]}{E\ (\sum \alpha_t^2 \sum a_t^2)^{\frac{p+q}{2}}}. \tag{4.34}$$

The moments of $\underline{r}^*$ are now straightforward to obtain. Since

$$E(\sum \alpha_{t-k}\ a_t) = \sum E\ (\alpha_{t-k})\ E\ (a_t)$$
$$= 0\ ,$$

it follows that

$$E\ (r_k^*) = 0. \tag{4.35}$$

Moreover, from (4.34) with $p = q = 1$,

$$\text{Cov}\ (r_k^*,\ r_{k+j}^*) \propto E\ (\sum \alpha_{t-k}\ a_t)\ (\sum \alpha_{t-k-j}\ a_t)$$

$$= E\ (\sum_t \sum_s \alpha_{t-k}\ a_t\ \alpha_{s-k-j}\ a_s)$$

$$= \sum_t \sum_s E\ (\alpha_{t-k}\ \alpha_{s-k-j}) E(a_t\ a_s).$$

Since for $j \neq 0$ it is impossible for the relations

$$t = s$$
$$t-k = s-k-j$$

both to hold simultaneously, every term in the summand is zero. Consequently we have the important result that

$$\text{Cov}\ (r_k^*,\ r_{k+j}^*) = 0, \quad j \neq 0. \tag{4.36}$$

The derivation of nonvanishing moments of $\{r_k^*\}$ is facilitated by observing that

$$E\ (\sum \alpha_t^2 \sum a_t^2)^{p/2} = E(\sum \alpha_t^2)^{p/2}\ E\ (\sum a_t^2)^{p/2}$$

is $\sigma_\alpha^p\ \sigma_a^p$ times the square of the $(\frac{p}{2})^{\underline{th}}$ moment $\mu'_{p/2}$ of a χ_n^2 distribution. Specifically it is readily shown that

$$V(r_k^*) = E\left[(r_k^*)\right]^2 = \frac{n-k}{n^2} \simeq \frac{1}{n}\ . \tag{4.37}$$

Joint distribution.

From (4.37), (4.36), and (4.35), it is seen that for large n the white noise cross correlations (4.30) have $\underline{0}$ - mean and covariance matrix $\frac{1}{n}$ I. To show that jointly the statistics $\underline{r}^* = (r_1^*, \ldots, r_m^*)'$ will possess a multivariate normal distribution for large n, we apply the multivariate Central Limit Theorem (e.g., Cramer [6]) and Slutsky's [11] theorem to the cross correlations $\underline{r}^*$.

Write

$$\begin{aligned} x_k &= \sqrt{n}\, r_k^* \\ &= \frac{b_k}{d} \end{aligned} \qquad (4.38)$$

where

$$b_k = \frac{1}{\sqrt{n}} \sum \alpha_{t-k}\, a_t$$

and

$$d = \sqrt{\frac{1}{n} \sum \alpha_t^2 \, \frac{1}{n} \sum a_t^2} \, .$$

Then b_k is of the form

$$\sqrt{n} \quad \bar{y}$$

where

$$\bar{y} = \frac{1}{n} \sum y_i$$

is the mean of n independent and identically distributed variables $y_i = \alpha_{i-k} a_i$. Moreover, $\underline{b} = (b_1, \ldots, b_m)'$ has for all n mean $\underline{0}$ and covariance matrix $\sigma_a \sigma_\alpha I$. Hence by the multivariate Central Limit Theorem $\underline{b}$ has the limiting distribution $N(\underline{0}, \sigma_a \sigma_\alpha I)$. Furthermore,

$$d \overset{P}{\rightarrow} \sigma_a \sigma_\alpha$$

where "$\overset{P}{\rightarrow}$" denotes convergence in probability. Hence by Slutsky's theorem,

$$\underline{x} = \frac{1}{d} \underline{b} \qquad (4.39)$$

converges to the distribution $N(\underline{0}, I)$.

It is therefore shown that for large n

$$\underline{r}^* \sim N(\underline{0}, \frac{1}{n} I) \qquad (4.40)$$

which is identical to the limiting distribution (4.29) of white noise <u>auto</u>correlations.

4.3 Distribution of $\hat{\underline{r}}^*$.

The distribution (4.40) of the white noise cross correlations $\underline{r}^*$ having been established in the last section, we can now continue from where we had left off (4.28) in our discussion of the <u>residual</u> cross correlations $\hat{\underline{r}}^*$. Equation (4.28) now shows that approximately

$$\underline{\hat{r}}^* \sim N(\underline{0}, \frac{1}{n} [I - Q]), \tag{4.41}$$

which parallels the results obtained for residual autocorrelations in [5] and in the previous section. In particular the distribution of $\underline{\hat{r}}^* = (\hat{r}^*_1, \ldots, \hat{r}^*_m)'$ is contained in a subspace of dimension m-u-v-1; that is, the distribution has a (u+v+1)-dimensional singularity given by the constraints (4.22) and (4.23).

Distribution of $n \sum_1^m (\hat{r}^*_k)^2$.

Paralleling the observations made in sections 2 and 3 in connection with autocorrelations, it can be stated that if the fitted model is appropriate and the parameters $\lambda = (\underline{\omega}, \underline{\delta}, \underline{\phi}, \underline{\theta})$ are exactly known, then the caluclated "residuals" $\{a_t\}$ would be uncorrelated normal deviates or white noise, and from (4.40) their cross correlations $\underline{r}^*$ with the whitened input $\{a_t\}$ would be approximately $N(\underline{0}, \frac{1}{n} I)$. Thus under these conditions the statistic

$$n \sum_{k=1}^{m} (r^*_k)^2 \tag{4.42}$$

would possess a χ^2 distribution with m degrees of freedom. From (4.41) it is evident that if m is large enough so that the elements after the $m^{\underline{th}}$ in the latest vectors of Q [i.e., the columns of X] are essentially zero, yet m is still small relative to n, then the statistic

$$n \sum_{k=1}^{m} (\hat{r}_k^*)^2 \tag{4.43}$$

based on the <u>residual</u> cross correlations $\underline{\hat{r}}^*$, will still possess a χ^2 distribution [since the matrix I-Q is idempotent] only now with m-u-v-1 rather than m degrees of freedom. This result is of considerable practical importance in diagnostic checking as it shows that an over-all test of fit of the dynamic model (3.1) can be based on the statistic (4.43) simply by making an adjustment to the number of degrees of freedom in the χ^2-distribution of the statistic which would be appropriate had the model been correct and the parameters known exactly. In section 5 these ideas are explored further when the use of residual correlations in diagnostic checking is considered.

4.4. Examples.

The X-matrix (4.25) which determines the distribution of the residual cross correlations is of a considerably more complex nature than the corresponding matrix for autocorrelations, so that further understanding of the distribution of $\underline{\hat{r}}^*$ and how it departs from that of the white noise cross correlations $\underline{r}^*$ is best obtained by a consideration of the X and Q matrices occurring in some simple subclasses of the class of dynamic models (3.1). These will be considered under three headings depending on the structure of $\delta(B)$.

Case (i): $\delta(B) = 1$.

The situation is quite simplified if the transfer function $V(B)$ is finite, for then the operator $\chi(B)$ plays no role in the X-matrix (4.25), which now becomes simply

$$X = \frac{\sigma_\alpha}{\sigma_a} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\pi_1 & 1 & 0 & & \cdot \\ -\pi_2 & -\pi_1 & 1 & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & 0 \\ \cdot & \cdot & \cdot & & 1 \\ & & & & \vdots \\ -\pi_{m-1} & -\pi_{m-2} & -\pi_{m-3} & \cdots & -\pi_{m-u-1} \end{bmatrix} . \tag{4.44}$$

It is noteworthy that the degree of singularity in the covariance matrix $\frac{1}{n}$ (I-Q) is therefore determined by the <u>number</u> (u+1) of parameters $(\omega_o, \omega_1, \ldots, \omega_u)$ estimated but that the dynamic parameters themselves play no role, as X is made up entirely of coefficients of

$$\pi(B) = \theta^{-1}(B)\ \phi(B).$$

Thus

$$X'X = \begin{bmatrix} \sum\limits^{\infty} \pi_i^2 & \sum \pi_i \pi_{i-1} & \cdots & \sum \pi_i \pi_{i-u} \\ \sum \pi_i \pi_{i-1} & \sum \pi_i^2 & \cdots & \sum \pi_i \pi_{i-u+1} \\ \vdots & & & \vdots \\ \sum \pi_i \pi_{i-u} & \cdots & \cdots & \sum \pi_i^2 \end{bmatrix}. \qquad (4.45)$$

and from these results the Q and $\frac{1}{n}$ (I-Q) matrices can be determined.

White noise.

The situation is especially interesting if $\pi(B) = 1$, that is for the model

$$y_t = \omega(B)\ \alpha_t + a_t \qquad (4.46)$$

for which

$$X'X = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} = I_{u+1}\ .$$

We have

$$Q = X(X'X)^{-1}X' = \begin{array}{c} \begin{matrix} (u+1 & (m-u-1) \end{matrix} \\ \left| \begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array} \right| \end{array}$$

so that

$$I-Q = \left[\begin{array}{ccc|ccc} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ \hline & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{array}\right]$$

with zeros everywhere off the diagonal. Thus the variances of $\hat{r}_1^*, \ldots, \hat{r}_{u+1}^*$ are to the present order of approximation <u>zero</u> [that is, they are of order $\frac{1}{n^2}$]. This is comparable to the situation in which a first order AR model is fitted to a process where $\phi = 0$ (i.e., a white noise process); for in this case we would have from (2.14) for the lag 1 residual <u>auto</u>correlation [since $Q = \{q_{ij}\} = \{\phi^{i+j-2}(1-\phi^2)\}$]

$$V(\hat{r}_1) = \frac{\phi^2}{n} \text{ to } o(\frac{1}{n})$$

$$= O(\frac{1}{n^2}).$$

Higher-lagged $\hat{r}^*$'s are then distributed approximately as the corresponding r^*'s.

<u>Case (ii): $\delta(B) = 1 - \delta B$.</u>

The special case considered above where $\delta(B) = 1$ is somewhat artificial in dynamic systems for it means that the effect on the output of a given change in input does not build up over time but rather is instantaneous (that is, the <u>time constant</u>

[4] of the system is zero). The situation where $\delta(B)$ is a first order polynomial in B characterizes the first order dynamic models, which are of greater interest from a practical standpoint. Unfortunately, however, as soon as δ-parameters are included in the model the X-matrix on which the covariance matrix of $\hat{\underline{r}}^*$ is based becomes considerably more complicated, owing to the presence of the $\{\chi_j\}$ in its columns, where as in (4.25) these coefficients are determined from

$$\chi(B) = \omega(B)\ \delta^{-2}(B)\ \pi(B).$$

Thus we shall here discuss only in general terms the nature of the distribution of $\hat{\underline{r}}^*$ for these models.

First order dynamic model, white noise.

Suppose the appropriate model for a given process can be written

$$y_t = \frac{\omega_o}{1 - \delta B}\alpha_t + a_t. \tag{4.47}$$

Then

$$\delta^{-1}(B) = 1 + \delta B + \delta^2 B^2 + \ldots$$

$$\delta^{-2}(B) = 1 + 2\delta B + 3\delta^2 B^2 + 4\delta^3 B^3 + \ldots,$$

so that, since $\pi(B) = 1$,

$$\xi(B) = 1 + \delta B + \delta^2 B^2 + \dots$$

$$\chi(B) = \omega_o [1 + 2 \delta B + 3 \delta^2 B^2 + \dots]$$

Since the value of ω_o does not enter into the matrix $Q = X(X'X)^{-1} X'$ and thus does not influence the distribution of $\underline{\hat{r}}^*$, there is no loss of generality in assuming $\omega_o = 1$. Likewise we can assume $\frac{\sigma_\alpha}{\sigma_a} = 1$. Then

$$X = \begin{bmatrix} 1 & 1 \\ \delta & 2\delta \\ \delta^2 & 3\delta^2 \\ \delta^3 & 4\delta^3 \\ \vdots & \vdots \end{bmatrix}, \qquad X' X = \begin{bmatrix} \frac{1}{(1-\delta^2)} & \frac{1}{(1-\delta^2)^2} \\ \frac{1}{(1-\delta^2)^2} & \frac{(1+\delta^2)}{(1-\delta^2)^3} \end{bmatrix},$$

$$|X' X| = \frac{\delta^2}{(1-\delta^2)^4},$$

and

$$(X' X)^{-1} = \frac{1}{\delta^2} \begin{bmatrix} 1-\delta^4 & -(1-\delta^2)^2 \\ -(1-\delta^2)^2 & (1-\delta^2)^3 \end{bmatrix} \tag{4.48}$$

Thus if $Q = X(X' X)^{-1} X' = \{q_{ij}\}$, and $\underline{\varepsilon}_j' = (\delta^{j-1}, j \delta^{j-1})$ is the $j^{\underline{th}}$ row of X, then the elements of Q are given by

$$q_{ij} = \underline{\varepsilon}_i' (X' X)^{-1} \underline{\varepsilon}_j \tag{4.49}$$

from which the matrix Q and thus the covariance matrix $\frac{1}{n}$ (I-Q) of $\underline{\hat{r}}^*$ can be determined. For example, marginally $(\hat{r}_1^*, \hat{r}_2^*, \hat{r}_3^*)$ are for large n approximately $N(\underline{0}, \Sigma)$ where

$$\Sigma = \frac{1}{n}\begin{bmatrix} \delta^4 & -2\,\delta^3\,(1-\delta^2) & 2\,\delta^3\,(1-\delta^2)\,(1-2\delta^2) \\ & \delta^2\,(1-7\delta^2+6\delta^4) & -\,2\,\delta\,(1-3\delta^2+3\delta^4)\,(1-\delta^2) \\ (\text{sym.}) & & 1-\delta^2\,(1-\delta^2)\,(4-11\delta^2+9\delta^4) \end{bmatrix} \tag{4.50}$$

The variances and covariances for higher-lagged $\hat{r}^*$'s can also be determined in a straightforward manner, but the formulas become progressively more involved; however, it is possible to draw some general conclusions. From the matrix (4.50) it is seen that $V(\hat{r}_1^*)$ and $V(\hat{r}_2^*)$ can be quite small, and moreover that the covariances and correlations of the residual cross correlations $\underline{\hat{r}}^*$ of small order can be very high. However for higher-lagged $\hat{r}^*$'s it is evident from (4.49) that the general element q_{ij} of Q is of order at most δ^{i+j-4} [since the elements of $\underline{\varepsilon}_i$ are multiples of δ^{i-1}]. Thus as i and/or j becomes larger there is a return of the elements of the matrix $\frac{1}{n}$ (I-Q) to the elements (zero or $\frac{1}{n}$) of the matrix $\frac{1}{n}$ I which is the large sample covariance matrix of the <u>white noise</u> cross correlations. Hence for this model the two-dimensional singularity in the distribution of $\underline{\hat{r}}^*$ is concentrated heavily on the first two or three lags with little effect on higher-lagged $\hat{r}^*$'s.

First order dynamic model, non-white noise.

Consider now the model

$$y_t = \frac{1}{1 - \delta B} \alpha_t + \frac{\theta(B)}{\phi(B)} a_t \qquad (4.51)$$

where in accordance with the discussion preceding (4.48) we have assumed that the unknown parameter ω_o is 1. From the general results of section 4.1 it follows that the presence of the noise parameters $\underline{\theta}$ and $\underline{\phi}$ this model will not cause any further singularity in the distribution of $\hat{\underline{r}}^*$ as compared with the model (4.47) where the noise was assumed white. In particular therefore it is significant that the statistic

$$n \sum_{k=1}^{m} (\hat{r}_k^*)^2 \qquad (4.52)$$

will for large n possess a χ^2 distribution with m-u-v-1 (here, m - 2) degrees of freedom regardless of which and how many parameters $(\underline{\phi}, \underline{\theta})$ are associated with the noise $\{a_t\}$.

However the distribution of $\hat{\underline{r}}^*$ itself is considerably affected by the presence of autocorrelation in the noise. The X-matrix still has two columns but they now consist of the coefficients of

$$\xi(B) = (1 + \delta B + \delta^2 B^2 + \ldots)(1 + \pi_1 B + \pi_2 B^2 + \ldots)$$

$$\chi(B) = (1 + 2 \delta B + 3 \delta^2 B^2 + \ldots)(1 + \pi_1 B + \ldots) \qquad (4.53)$$

where $\pi(B) = \frac{\phi(B)}{\theta(B)}$.

From (4.53) it is evident that the general effect of an increasing degree of autocorrelation in the noise [either through additional parameters $\{\theta_j\}$, $\{\phi_i\}$ or through the ϕ's lying closer to the boundary of the region of stationarity] will be to distribute the singularity of $\hat{\underline{r}}^*$ throughout more of the lags; that is, the restoration in the behavior of $\{\hat{r}_k^*\}$ to that of the $\{r_k^*\}$ for increasing k will be slower, but also the initial values (e.g., $\hat{r}_1^*$, $\hat{r}_2^*$) will behave less unlike the corresponding white noise cross correlations. [And thus a qualification to the remarks made concerning the χ^2-distribution of (4.52) is that for highly autocorrelated noise the values of m and n may have to be greater for the approximation to be equally feasible]. This phenomenon is entirely analogous to the effect on the distribution of the autocorrelations $\hat{\underline{r}}$ of letting $\phi \to 1$ in a first order AR process, as discussed in [5].

For models in which $\omega(B)$ is of first degree or higher in B, the situation is more complex but the same general considerations concerning the distribution of $\hat{\underline{r}}^*$ apply.

Case (iii): Higher order dynamic models.

To get an idea of the situation which obtains for more complex dynamic/stochastic models, let us finally consider briefly the second-order dynamic model with delay and (1, 0, 1) noise, given by

$$y_t = \frac{\omega_o - \omega_1 B}{1 - \delta_1 B - \delta_2 B^2} \alpha_t + \frac{1 - \theta B}{1 - \phi B} a_t \tag{4.54}$$

From the general results in section (4.1) we have

$$\xi(B) = (1 - \delta_1 B - \delta_2 B^2)^{-1} (1 - \theta B)^{-1} (1 - \phi B)$$

$$= 1 + \xi_1 B + \xi_2 B^2 + \ldots \tag{4.55}$$

$$\chi(B) = (1 - \delta_1 B - \delta_2 B^2)^{-2} (\omega_o - \omega_1 B) (1 - \theta B)^{-1} (1 - \phi B)$$

$$= \chi_o + \chi_1 B + \chi_2 B^2 + \ldots \tag{4.56}$$

and

$$X = \begin{bmatrix} 1 & 0 & \chi_o & 0 \\ \xi_1 & 1 & \chi_1 & \chi_o \\ \xi_2 & \xi_1 & \chi_2 & \chi_1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \xi_{m-1} & \xi_{m-2} & \xi_{m-1} & \xi_{m-2} \end{bmatrix}$$

where m is large enough so the ξ_{m-2}, ξ_{m-1}, ... and χ_{m-2}, χ_{m-1}, ... are of order $\frac{1}{\sqrt{n}}$ or less yet $\frac{m}{n}$ is itself of order $\frac{1}{\sqrt{n}}$.

The coefficients $\{\xi_j\}$ and $\{\chi_j\}$ and the resulting X, (X' X), Q, and $\frac{1}{n}$ (I - Q) matrices can conceptually be determined through (4.55) and (4.56), although such a procedure would be rather awkward to carry through. One suggestion might be to approximate $(\omega_o, \omega_1, \delta_1, \delta_2, \theta, \phi)$ by their estimated values and use these to determine an approximate X-matrix numerically by computer.

5. Diagnostic Checking in Dynamic Models.

The approximate joint distribution of the autocorrelations $\hat{\underline{r}}$ of the residuals $\{\hat{a}_t\}$ in fitted dynamic models has been obtained in section 3, and that of the cross correlations $\hat{\underline{r}}^*$ between the residuals and the whitened input $\{\alpha_t\}$ in section 4. In both cases it has been seen that these distributions can in some respects differ sharply from the distributions of the white noise correlations, that is, the correlations which would result if the "residuals" were calculated not from the estimated parameters but instead from the true parameter values. The properties of $\hat{\underline{r}}$ and $\hat{\underline{r}}^*$ therefore need to be carefully considered for their effective utilization in diagnostic checking.

5.1 Summary of the distribution of residual correlations.

We have seen that both $\hat{\underline{r}}$ and $\hat{\underline{r}}^*$ are singular linear transformation of the white noise correlations $\underline{r}$ and $\underline{r}^*$,

$$\hat{\underline{r}} = (I - Q)\ \underline{r} \tag{5.1}$$

$$\hat{\underline{r}}^* = (I - Q)\ \underline{r}^* \tag{5.2}$$

where Q is of the form

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

X is an $m \times \ell$ matrix of rank ℓ, and ℓ is the number (p+q) noise parameters in the model (3.1) for equation (5.1), and the number (u+v+1) of dynamic parameters for equation (5.2). Thus the residual correlations have for large n a singular normal distribution concentrated in an $(m-\ell)$ dimensional subspace of Euclidean m-space. Now it is easily seen that if the parameters $(\underline{\omega},\underline{\delta},\underline{\phi},\underline{\theta})$ are interior to their admissibility regions [that is, the roots of $\omega(x) = 0$, $\phi(x) = 0$, etc. lie outside the unit circle], then the X-matrix for either (5.1) or (5.2) is such that as we progress downward along any column (say the j^{th}), the elements $\{x_{kj}\}$, k = 1, 2, ... go to zero [which is of course why we have always been able to guarantee , for n sufficiently large, the existence of a number m such as in the discussion following (3.3)]. Thus the matrix Q will have its largest elements in the upper left-hand corner, that is for smallest values of i and j, with other $\{q_{ij}\}$ going to zero as either i or j increases. Since the residual correlation covariance matrices depend on I-Q, it is therefore seen that the departure of the distributions of $\hat{\underline{r}}$ and $\hat{\underline{r}}^*$ from those of $\underline{r}$ and $\underline{r}^*$ are greatest for the residual correlations of smallest lag; and specifically that this departure can consist of unusually small variances and very high correlations among the $\hat{r}$'s and $\hat{r}^*$s of low order or lag (Higher lagged residual correlations then behave increasingly like their white noise counterparts with respect to both their variances and covariances).

5.2 The over all χ^2 test.

Since the large sample covariance matrix $I - X(X' X)^{-1}X'$ of the quantities

$$\underline{x} = \sqrt{n}\, \hat{\underline{r}},\ \underline{x}^* = \sqrt{n}\, \hat{\underline{r}}^* \tag{5.3}$$

is idempotent of rank m-ℓ, where ℓ is the rank of X, it follows that the statistics

$$n \sum_{j=1}^{m} \hat{r}_j^2 \tag{5.4}$$

and

$$n \sum_{j=1}^{m} (\hat{r}_j^*)^2 \tag{5.5}$$

to the order of approximation we are here employing, possess χ^2 distributions with m-ℓ degrees of freedom. Moreover in the distribution of $\hat{\underline{r}}$, ℓ is the number (p+q) of stochastic parameters estimated and is independent of the dynamic model itself; while in the distribution of $\hat{\underline{r}}^*$, ℓ is the number (u+v+1) of dynamic parameters estimated and in turn is independent of the noise structure associated with the model.

It can therefore be concluded that regardless of the particular dynamic/stochastic model fitted to a set of data, a useful over-all check of the adequacy of the fit of the model is provided by referring the statistics $n \sum \hat{r}^2$ and $n \sum (\hat{r}^*)^2$ to their appropriate χ^2-distributions. Such procedures would not be expected to be as sensitive as those based on individual $\hat{r}$'s and $\hat{r}^*$'s, but their applicability is universal in that the distributions of (5.4) and (5.5) are independent of the particular parameter values appropriate to the model.

5.3. Diagnostic procedures applied to individual residual correlations.

As was indicated in section 1, the underlying rationale in the use of residuals and of residual auto- and cross correlations in diagnostic checking is that for a correctly identified and fitted model the residuals $\{\hat{a}_t\}$ should resemble the white noise $\{a_t\}$ from which the series is assumed to have been generated. We have seen that in some important respects this resemblance breaks down, but that the consequences of this for the use of the overall criteria $n \sum \hat{r}^2$ or $n \sum \hat{r}^{*2}$ in diagnostic checking were slight, as only a modification in the number of degrees of freedom in their χ^2-distribution was required.

The story is different, however, when individual residual correlations are examined, for it has been seen that the variances and covariances of the $\{\hat{r}_k\}$ and $\{\hat{r}^*_k\}$, especially for small values of k, can differ greatly from those of the correlations $\{r_k\}$ and $\{r^*_k\}$ based on white noise. Thus if individual residual correlations are to be compared with their standard errors to determine whether they are unusually large, then modifications of the "quality-control-chart" procedure which would be appropriate for white noise correlations are now necessary.

The situation is very similar to that discussed in [5] for autoregressive residual autocorrelations, and figure 2, taken from that paper, illustrates the narrowing of the standard-error bands which occurs, contrasted to those of figure 1 for white-noise autocorrelations, which from (4.40) are identical to the standard error bands for white noise <u>cross</u> correlations. The particular AR examples in (a), (b), (c), and (d) of figure 2 are appropriate also for examining adequacy of the noise structure of dynamic models via the residual autocorrelations when the fitted noise is AR(1) or AR(2) with parameters as shown, which from (2.3) and the preceding discussion also includes fitted dynamic models with MA(1), MA(2), and mixed AR(1)-MA(1) noise. The standard error bands for $\hat{\underline{r}}$ with other noise structures, and those of the residual cross correlations $\hat{\underline{r}}^*$ for various dynamic and disturbance models, are all similar in appearance to those presented in this figure, since the singularity of their joint normal distributions is of the same nature in all cases, that is, it is such that there is a strong depression of the variances of the residual correlations of small lags. It is thus seen that a <u>failure to take account of this discrepancy between the white noise and the residual auto- and cross correlations can lead to a serious underestimation of significant inadequacy in both the dynamic and the noise components of the model</u>. This fact is especially important because if lack of fit <u>does</u> exist it is the $\hat{r}$'s and $\hat{r}^*$'s of

FIGURE 1. Standard error limits for white noise autocorrelations r_k

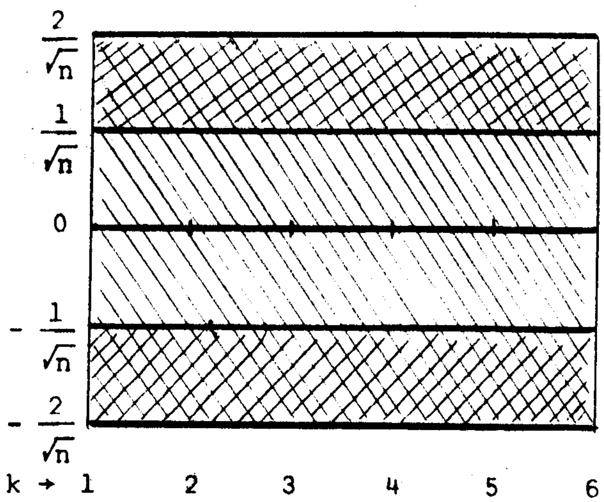
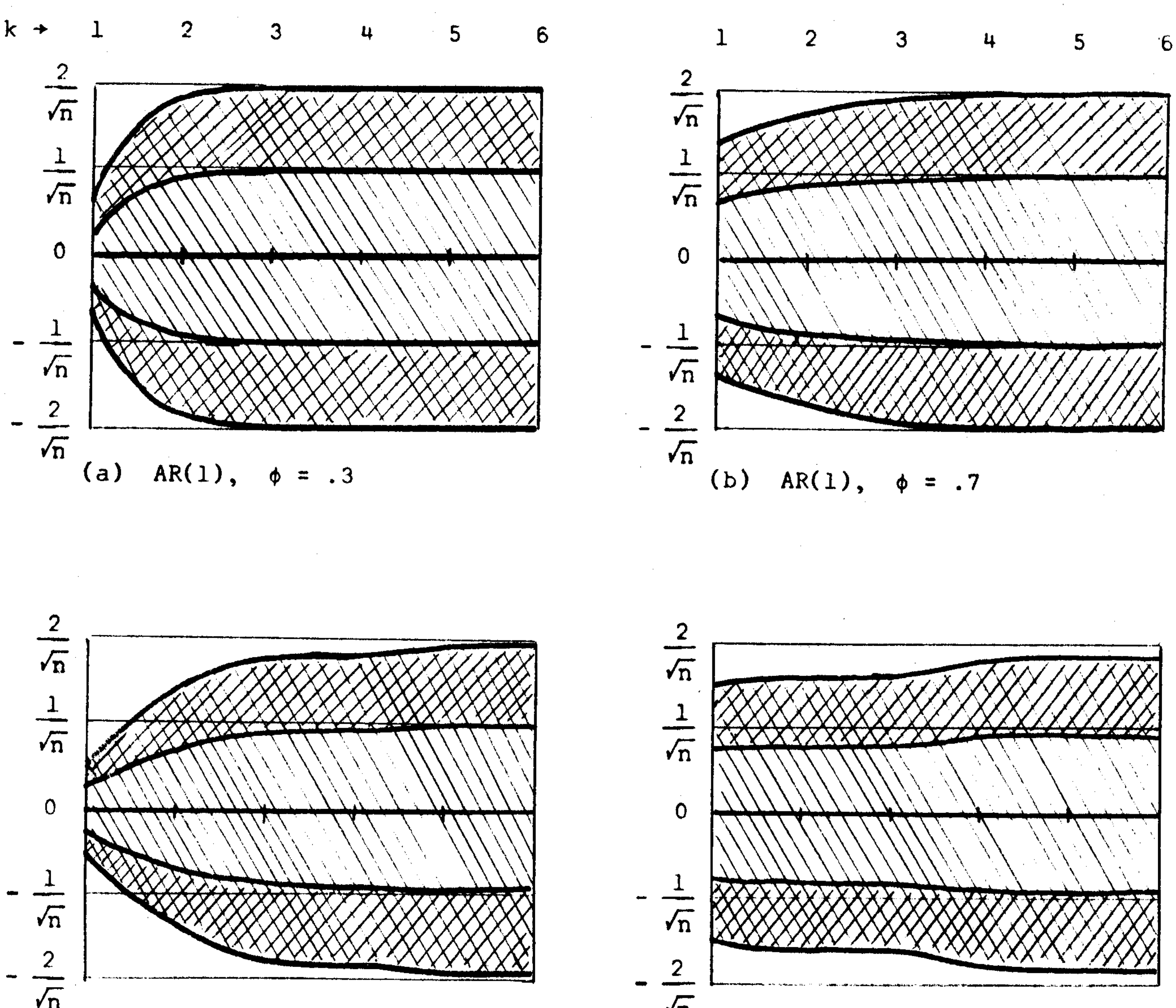


FIGURE 2. Standard error limits for residual autocorrelations $\hat{r}_k$



(a) AR(1), $\phi = .3$

(b) AR(1), $\phi = .7$

(c) AR(2), $\phi_1 = .5$, $\phi_2 = .25$

(d) AR(2), $\phi_1 = 1.0$, $\phi_2 = -.75$

lowest order which are most apt to reveal this.

The situation is further complicated by the presence of high correlation among the $\hat{r}$'s and $\hat{r}^*$'s. Since $\hat{r}_1^*$ and $\hat{r}_2^*$, for example, will usually be highly dependent, a decision as to whether one is significantly large should properly take account of the other, so that the construction of statistical tests based on several individual $\hat{r}^*$'s [as opposed, for example, to the over-all χ^2-test] is more difficult than if they were independent as would be the case for r_1^* and r_2^* for white noise. However, charts such as in figure 2 can still provide rough guidelines.

5.4 Conclusion: Use of residual correlations in diagnostic checking.

It has been demonstrated in this paper that the departure of the distributions of the residual correlations $\underline{\hat{r}}$ and $\underline{\hat{r}}^*$ in dynamic models from those of the white noise correlations $\underline{r}$ and $\underline{r}^*$ is of sufficient extent to warrant a careful consideration of their properties in connection with their use in diagnostic checking. Since the consequence of supposing that the $\hat{r}$'s and $\hat{r}^*$'s can be regarded as r's and r^*'s was found to be an <u>under</u>-estimation of significant model inadequacy, it follows that whenever diagnostic testing procedures based on this erroneous supposition do reveal lack of fit in the model, more sensitive procedures based on the singular distributions of $\underline{\hat{r}}$ and $\underline{\hat{r}}^*$ will also lead to this conclusion, and even more forcefully so. However it is essential to consider the true (even if approximate) distributions of the residual auto- and cross correlations in diagnostic checking if existing model inadequacy is always to be detected. When this is done these statistics remain useful and important tools in examining the adequacy of fit of stochastic and dynamic time series models.

6. References

[1] Anderson, R.L. (1942), Distribution of the serial correlation coefficient. AMS 13, 1-13.

[2] Anderson, T.W. and Walker, A.M. (1964), On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process. AMS 35, 1296-1303.

[3] Bartlett, M.S. (1946), On the theoretical specification and sampling properties of autocorrelated time series. JRSS(B) 8, 27.

[4] Box, G.E.P. and Jenkins, G.M. (1967), Statistical Models for Prediction and Control. Technical Reports #72, 77, 79, 94, 95, 99, 103, 104, 116, 121, and 122, Dept. of Statistics, University of Wisconsin, Madison, Wis. [To be published by Holden-Day].

[5] Box, G.E.P. and Pierce, D.A. (1968), Distribution of residual autocorrelations in integrated autoregressive-moving average time series models. Technical report #154, Dept. of Statistics, University of Wisconsin.

[6] Cramer, H. (1946), Mathematical Methods of Statistics. Princeton University Press.

[7] Jenkins, G.M. (1954), Tests of hypotheses in the linear autoregressive model. I. Null hypothesis distribution in the Yule Scheme. Biometrika 41, 405-419.

[8] Koopmans, T. (1942), Serial correlation and quadratic forms in normal variables. AMS 13, 14-33.

[9] Moran, P.A.P. (1948), Some theorems on time series. II. The significance of the serial correlation coefficient. Biometrika 35, 255-260.

[10] Pierce, D.A. (1968), Studies in the Modeling of Stochastic and Dynamic Systems with Special Reference to the Distribution of Residual Correlations. Ph.D. Thesis, University of Wisconsin, Madison, Wisc.

[11] Slutsky, E. (1925), Uber stochastische asymptoten und Grenzwerte. Metron 5 (#3), 3.