*The Annals of Statistics*
1982, Vol. 10, No. 1, 114–120

# DIAGNOSTIC TESTS FOR MULTIPLE TIME SERIES MODELS

BY D. S. POSKITT AND A. R. TREMAYNE

*University of York, England*

This paper is concerned with the development and application of diagnostic checks for vector linear time series models. A hypothesis testing procedure based upon the score, or Lagrangean multiplier, principle is advocated and the distributions of the test statistic both under the null hypothesis and under a Pitman sequence of alternatives are discussed. Consideration of alternative models with singular sensitivity matrices when the null hypothesis is true leads to an interpretation of the score test as a pure significance test and to a notion of an equivalence class of local alternatives. Portmanteau tests of model adequacy are also investigated and are seen to be equivalent to score tests.

**1. Introduction.** The problems of estimation and diagnostic checking of scalar linear time series models have received wide attention in the statistical literature. Parallel developments for the multiple time series case have not been so readily forthcoming, which may not be surprising in view of the inherently greater complexity of these models. The estimation theory for vector linear time series models has been considered by a number of authors, for example Dunsmuir and Hannan (1976), Kohn (1978) and Nicholls (1977), but hypothesis testing procedures have been discussed in rather less detail. The purpose of this paper is to investigate the general structure, interpretation and use of the score, or Lagrangean multiplier, test principle of Rao (1948) and Silvey (1959) in multiple autoregressive-moving average, ARMA, models. Application of these tests to univariate time series models has been considered by Godfrey (1979), Newbold (1980) and Poskitt and Tremayne (1980).

The general $v$ component ARMA$(p, q)$ model may be written as

$$\sum_{j=0}^{p} \mathbf{A}_j \mathbf{x}(t-j) = \sum_{j=0}^{q} \mathbf{M}_j \boldsymbol{\varepsilon}(t-j), \tag{1.1}$$

where $\mathbf{A}_0 = \mathbf{M}_0 = \mathbf{I}_v$. The innovation $\{\boldsymbol{\varepsilon}(t)\}$ is assumed to be a sequence of stationary martingale differences so that, almost surely, $E(\boldsymbol{\varepsilon}(t) \mid \mathscr{F}_{t-1}) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}(t)\boldsymbol{\varepsilon}'(t) \mid \mathscr{F}_{t-1}) = \boldsymbol{\Phi}$, a positive definite symmetric matrix, $\mathscr{F}_\tau$ being the Borel field generated by all past history of $\boldsymbol{\varepsilon}(t)$, $t \le \tau$. As pointed out in Hannan and Heyde (1972), this represents a natural relaxation of the usual independent identically distributed assumption for $\{\boldsymbol{\varepsilon}(t)\}$. Employing the notation and conventions of Neudecker (1969) we set $\boldsymbol{\alpha} = \mathrm{vec}(\mathbf{A}_1 : \cdots : \mathbf{A}_p)$, $\boldsymbol{\mu} = \mathrm{vec}(\mathbf{M}_1 : \cdots : \mathbf{M}_q)$ and $\boldsymbol{\theta}' = (\boldsymbol{\alpha}' : \boldsymbol{\mu}')$. The $n = (p+q)v^2$ elements of $\boldsymbol{\theta}$ plus the $v(v+1)/2$ parameters of $\boldsymbol{\Phi}$ are assumed to be such that the structure (1.1) lies within the equivalence class considered by Dunsmuir and Hannan (1976) and Deistler, Dunsmuir and Hannan (1978) and is "simply identified." The permissible parameter space of the structural parameter $\boldsymbol{\theta}$ will be denoted $\Theta$.

In order to facilitate discussion, consider the general problem of testing null hypotheses concerning the true parameter, $\boldsymbol{\theta}^\dagger$, that can be represented in the form $\mathbf{h}(\boldsymbol{\theta}^\dagger) = \mathbf{0}$ where the $f < n$ vector-valued function $\mathbf{h}$ is twice continuously differentiable. No prior restrictions will be placed on the parameters of $\boldsymbol{\Phi}$, though the methods described here can be generalised to do this without undue difficulty. The score test statistic of $\mathbf{h}(\boldsymbol{\theta}^\dagger) = \mathbf{0}$ is

Received September 1980; revised April 1981.

*AMS* 1970 *subject classifications*. Primary 62M10; secondary 62F05.

*Key words and phrases*. Multiple autoregressive-moving average models, score test, local alternatives, portmanteau statistics, equivalences.

structured in terms of the gradient vector of the criterion function

(1.2) $$\ell(\boldsymbol{\theta}) = T/2 \ln \det \mathbf{V},$$

$\mathbf{V} = T^{-1} \sum_{t=1}^{T} \mathbf{e}(t)\mathbf{e}'(t)$, evaluated at the point $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, the value that minimises $\ell(\boldsymbol{\theta})$ subject to the constraints. The residual $\mathbf{e}(t)$ is obtained from the equation

(1.3) $$\mathbf{e}(t) = \sum_{j=0}^{t-1} \mathbf{G}_j \mathbf{x}(t-j)$$

where

$$\mathbf{G}(z) = \sum_{j=0}^{\infty} \mathbf{G}_j z^j = \mathbf{M}^{-1}(z)\mathbf{A}(z), \qquad \mathbf{A}(z) = \sum_{j=0}^{p} \mathbf{A}_j z^j, \qquad \mathbf{M}(z) = \sum_{j=0}^{q} \mathbf{M}_j z^j.$$

That $\mathbf{e}(t)$ and hence $\mathbf{V}$ are functions defined on $\Theta$ and $\mathscr{F}_t$ is understood, though the arguments are not explicitly shown here and in what follows.

REMARK 1. If $\{\varepsilon(t)\}$ is Gaussian then, apart from asymptotically negligible terms, $\exp\{-\ell(\boldsymbol{\theta})\}$ is the likelihood maximised with respect to $\boldsymbol{\Phi}$, and $\hat{\boldsymbol{\theta}}$ is, following Whittle (1962), commonly referred to as the Gaussian estimator. For similar reasons the gradient vector of the criterion function $\mathbf{s}(\cdot) = \partial \ell(\cdot)/\partial\boldsymbol{\theta}$ henceforth will be called the score vector.

The score vector $\mathbf{s}(\boldsymbol{\theta})$ is given by $\sum_{t=1}^{T} \mathbf{d}'(t)\mathbf{V}^{-1}\mathbf{e}(t)$, where the derivative process $\mathbf{d}'(t) = \partial \mathbf{e}'(t)/\partial\boldsymbol{\theta}$. The Gaussian estimator $\hat{\boldsymbol{\theta}}$ is obtained as a solution to the equations

$$\mathbf{s}(\boldsymbol{\theta}) + \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\lambda} = \mathbf{0}, \qquad \mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$$

where $\mathbf{H}(\cdot) = \partial \mathbf{h}'(\cdot)/\partial\boldsymbol{\theta}$ is a $n \times f$ matrix of full column rank and the vector $\boldsymbol{\lambda}$ contains the $f$ Lagrangean multipliers; compare Aitchison and Silvey (1958). Set

$$\boldsymbol{\Omega}_T = T^{-1} \sum_{t=1}^{T} \mathbf{d}'(t)\mathbf{V}^{-1}\mathbf{d}(t).$$

It is assumed that $\boldsymbol{\Omega}_T$ converges in probability to a fixed matrix $\boldsymbol{\Omega}$ of rank $n-g$, $0 \le g < f$, whose range space is virtually disjoint from that of $\mathbf{H}'$. This ensures that the restricted estimator $\hat{\boldsymbol{\theta}}$ converges with probability one to a value $\boldsymbol{\theta}^o$ satisfying $\mathbf{h}(\boldsymbol{\theta}^o) = \mathbf{0}$; see Kohn (1978).

**2. Testing model misspecification.** The score test statistic for testing the null hypothesis $\mathbf{h}(\boldsymbol{\theta}^\dagger) = \mathbf{0}$ is given by

(2.1) $$S = T^{-1}\hat{\mathbf{s}}'\hat{\boldsymbol{\Omega}}^{-}\hat{\mathbf{s}} = T^{-1}\hat{\boldsymbol{\lambda}}'\hat{\mathbf{H}}'\hat{\boldsymbol{\Omega}}^{-}\hat{H}\hat{\boldsymbol{\lambda}}$$

where the caret denotes evaluation of quantities at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The generalized inverse $\hat{\boldsymbol{\Omega}}^{-}$ becomes a regular inverse if $\boldsymbol{\Omega}$ is nonsingular. In order to determine the asymptotic null and non-null distributions of $S$, it is necessary to consider a sequence of true parameter points $\boldsymbol{\Theta}_T^\dagger$ tending to the fixed limit $\boldsymbol{\theta}^o$. Writing $\chi^2(\nu, \kappa)$ for the noncentral Chi squared distribution with degrees of freedom $\nu$ and noncentrality parameter $\kappa$, the statistic (2.1) is asymptotically distributed $\chi^2(f-g, 0)$ when the restrictions are true, whilst under the Pitman sequence of local alternatives $\sqrt{T}(\boldsymbol{\theta}_T^\dagger - \boldsymbol{\theta}^o) = \boldsymbol{\xi}$ for some $\boldsymbol{\xi}$, $\boldsymbol{\xi}'\boldsymbol{\xi} < \infty$, $S$ converges in distribution to $\chi^2(f-g, \kappa_o)$, $\kappa_o = \boldsymbol{\xi}'\mathbf{H}(\mathbf{H}'\boldsymbol{\Omega}^{-}\mathbf{H})^{-}\mathbf{H}'\boldsymbol{\xi}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^o$.

Adopting arguments employed by Dunsmuir and Hannan (1976) and Kohn (1978, 1979) to show the asymptotic normality of the score vector under both null and local alternatives, the proof of these results, which extend those of Silvey (1959), is then effected by utilising the theorems of Rao and Mitra (1971, Chapter 9). It also should be noted from Rao and Mitra (1971, Lemma 2.2.4(ii)) that $S$ is invariant with respect to the choice of generalized inverse, a point which is seen to be of importance in the applications discussed below.

Consider the practical situation in which a multiple ARMA$(p_o, q_o)$ has been fitted and where the true order of the observed process is $(p, q)$, $p \ge p_o$, and $q \ge q_o$. The restrictions implied by the null hypothesis, that the values chosen for $p_o$ and $q_o$ are correct, are of the simple exclusion type so that, using an obvious rearrangement and partitioning, $\mathbf{H}'$ may be expressed as $(\mathbf{0} : \mathbf{I}_f)$, $f = (r+s)v^2$, $r = p - p_o$ and $s = q - q_o$. In order to construct the statistic (2.1), the derivative processes must be determined. Firstly, differentiating with respect to $\alpha'$, we have

$$\frac{\partial \mathbf{e}(t)}{\partial \alpha'} = \mathbf{M}^{-1}(B)\frac{\partial \mathbf{A}(B)}{\partial \alpha'}\mathbf{x}(t) = \{\mathbf{x}'(t) \otimes \mathbf{M}^{-1}(B)\}\frac{\partial \operatorname{vec}\{\mathbf{A}(B)\}}{\partial \alpha'}.$$

It should be pointed out that the backward shift operator $B$ applies to the preceding component in the Kronecker product $\mathbf{x}'(t) \otimes \mathbf{M}^{-1}(B)$. This convention also holds in expressions (2.2) and (2.3) which follow and the same is true of $\partial\operatorname{vec}\{\mathbf{A}(B)\}/\partial\alpha'$. Writing

$$\operatorname{vec}\{\mathbf{A}(z)\} = \{\zeta_p'(z) \otimes \mathbf{I}_{v^2}\}\alpha + \operatorname{vec}\mathbf{I}_v$$

where $\zeta_p'(z) = (z, z^2, \cdots, z^p)$, we obtain

$$\frac{\partial \mathbf{e}(t)}{\partial \alpha'} = \zeta_p'(B) \otimes \mathbf{x}'(t) \otimes \mathbf{M}^{-1}(B). \tag{2.2}$$

Similarly,

$$\frac{\partial \mathbf{e}(t)}{\partial \mu'} = -\zeta_q'(B) \otimes \mathbf{e}'(t) \otimes \mathbf{M}^{-1}(B). \tag{2.3}$$

Expressions (2.2) and (2.3) imply that the elements of the sensitivity matrix

$$\Omega = \begin{bmatrix} \Omega_{\alpha\alpha} & \Omega_{\alpha\mu} \\ \Omega_{\alpha\mu}' & \Omega_{\mu\mu} \end{bmatrix}$$

are given by equations

$$\Omega_{\alpha\alpha} = (2\pi)^{-1}\int_{-\pi}^{\pi} \zeta_p\zeta_p^* \otimes \mathbf{KVK}^* \otimes (\bar{\mathbf{M}}\mathbf{V}\mathbf{M}')^{-1}\, d\omega$$

$$\Omega_{\alpha\mu} = -(2\pi)^{-1}\int_{-\pi}^{\pi} \zeta_p\zeta_q^* \otimes \mathbf{KV} \otimes (\bar{\mathbf{M}}\mathbf{V}\mathbf{M}')^{-1}\, d\omega \tag{2.4}$$

and

$$\Omega_{\mu\mu} = (2\pi)^{-1}\int_{-\pi}^{\pi} \zeta_q\zeta_q^* \otimes \mathbf{V} \otimes (\bar{\mathbf{M}}\mathbf{V}\mathbf{M}')^{-1}\, d\omega$$

where $\mathbf{K}(z) = \mathbf{G}^{-1}(z)$ and, for convenience, the argument $z = e^{\iota\omega}$ is omitted; compare Dunsmuir and Hannan (1976, Section 4) and see also Nicholls (1976). It is convenient to introduce the notation $\Gamma_{\eta v}(z)$ for the cross-covariance generating function between any two processes $\{\eta(t)\}$ and $\{v(t)\}$. From (2.4) it is apparent that the elements of $\Omega$ appear as coefficients of powers of $z$ in the convolutions of $\Gamma_{xx}(z)$, $\Gamma_{x\varepsilon}(z)$ and $\Gamma_{\varepsilon\varepsilon}(z)$ with the multivariate inverse autocovariances of the moving average process $\mathbf{y}(t) = \mathbf{M}(B)\varepsilon(t)$.

If the null hypothesis is true, then (1.1) implies that

$$\mathbf{A}(z)\Gamma_{xx}(z) = \mathbf{M}(z)\Gamma_{\varepsilon x}(z), \tag{2.5}$$

where $\mathbf{A}(z)$ and $\mathbf{M}(z)$ are of degree $p_o$ and $q_o$ respectively. From (2.5)

$$\{\Gamma_{xx}'(z) \otimes \mathbf{I}_v\}\operatorname{vec}\mathbf{A}(z) = \{\Gamma_{\varepsilon x}'(z) \otimes \mathbf{I}_v\}\operatorname{vec}\mathbf{M}(z)$$

and, therefore,

$$\{\zeta_{p_o}'(z) \otimes \Gamma_{xx}'(z) \otimes \mathbf{I}_v\}\alpha + \{\Gamma_{xx}'(z) \otimes \mathbf{I}_v\}\operatorname{vec}\mathbf{I}_v = \{\zeta_{q_o}'(z) \otimes \Gamma_{\varepsilon x}'(z) \otimes \mathbf{I}_v\}\mu + \{\Gamma_{\varepsilon x}'(z) \otimes \mathbf{I}_v\}\operatorname{vec}\mathbf{I}_v.$$

Consequently, the vectors

$$\mathbf{a}_1' = ((\operatorname{vec}\mathbf{I}_v)' : \alpha' : \mathbf{0}_{v^2(r-1)}') \text{ and } \alpha_2' = ((\operatorname{vec}\mathbf{I}_v)' : \mu' : \mathbf{0}_{v^2(s-1)}')$$

are such that $\mathbf{a}_1'\Omega_{\alpha\alpha} = \mathbf{a}_2'\Omega_{\alpha\mu}'$ whenever $p > p_o$ and $q > q_o$. In addition, the generating equations $\mathbf{A}(z)\Gamma_{x\varepsilon}(z) = \mathbf{M}(z)\Phi$ may be similarly manipulated to show that $\mathbf{a}_1'\Omega_{\alpha\mu} = \mathbf{a}_2'\Omega_{\mu\mu}$

and hence the vector $\mathbf{a}' = (\mathbf{a}_1' : \mathbf{a}_2')$ annihilates $\mathbf{\Omega}$. Proceeding as in Poskitt and Tremayne (1981), it can be demonstrated that the rank of $\mathbf{\Omega}$ is $\{(p + q) - \min(r, s)\}v^2$ and thus $g = \min(r, s)v^2$. As in the univariate case, therefore, identifiability problems will be encountered when simultaneous extensions on the polynomials $\mathbf{A}(z)$ and $\mathbf{M}(z)$ are entertained; see Hannan (1970, Chapter 6).

In order to enumerate the test statistic, a choice of generalized inverse must be made. Since there exist nonsingular submatrices $\mathbf{\Sigma}$ of order $(p_o + q_o + m)v^2$, $m = \max(r, s)$, of $\mathbf{\Omega}$, obtained by omitting $g$ rows and columns, a possible choice for $\mathbf{\Omega}^-$ is given by a matrix with elements of $\mathbf{\Sigma}^{-1}$ appropriately placed and zeros elsewhere; see Rao and Mitra (1971, Section 11.2). Choosing $\mathbf{\Sigma}$ by selecting those rows and columns of $\mathbf{\Omega}$ corresponding to ARMA$(p_o + r, q_o)$ when $r \geq s$ or ARMA$(p_o, q_o + s)$, $s \geq r$, alternatives, it follows from the invariance of $S$ with respect to $\mathbf{\Omega}^-$ that, amongst others, the score tests for testing ARMA$(p_o, q_o)$ against both ARMA$(p_o + m, q_o)$ and ARMA$(p_o, q_o + m)$, $m = r = s$, are asymptotically identical. Some related discussion in the context of univariate models is available in Poskitt and Tremayne (1980). A corollary of these results is that application of the score test principle to vector linear time series models leads to a procedure which may be viewed as a type of pure significance test (see Cox and Hinkley, 1974, Chapter 3), and that (2.1) is to be employed as a diagnostic check of the adequacy of a fitted ARMA $(p_o, q_o)$ specification.

As the value of the test statistic is not affected by the structure of the alternative model, once the value of $m$ has been determined, we investigate more closely the power of the score test procedure using the distribution of $S$ under a Pitman sequence of alternatives. Given $m$, the power of $S$ depends solely upon the noncentrality parameter $\kappa_o = \boldsymbol{\xi}'\mathbf{H}(\mathbf{H}'\mathbf{\Omega}^-\mathbf{H})^-\mathbf{H}'\boldsymbol{\xi}$. The value $\kappa_o$ resulting from a test of the null hypothesis against any identifiable alternative implicit in the choice of $\mathbf{\Sigma}$ is, however, equal to that obtained when testing against a nonidentifiable ARMA$(p_o + r, q_o + s)$ alternative. This follows from matrix manipulations which exploit the particular form of $\mathbf{H}$ and $\mathbf{\Omega}^-$ in the present context. Thus, values of $\boldsymbol{\xi}$ and of additional polynomial coefficients in the implied alternative can be chosen in such a way that the noncentrality parameter is equal irrespective of which $\mathbf{\Omega}^-$ is selected. In view of this and the comments of the last paragraph, an equivalence class of local alternatives converging to the null is defined.

Remark 2. As the asymptotic distribution of the likelihood ratio test statistic under a sequence of local alternatives is the same as that of $S$, it follows that the asymptotic power of the likelihood ratio procedure is also invariant with respect to members of the equivalence class of alternatives. Hence, the nonspecificity of the score test does not result in any loss of power, in that alternative specifications likely to be of interest to the practitioner will be locally equivalent to each other.

The general results obtained above imply that the ARMA$(p_o + r, q_o + s)$ alternative can be written locally in terms of only $(p_o + q_o + m)v^2$ parameters, rather than $(p_o + q_o + r + s)v^2$. An illustration of this is available by taking the null model as $\mathbf{A}(B)\mathbf{x}(t) = \mathbf{M}(B)\boldsymbol{\varepsilon}(t)$ and considering the particular member of the equivalence class of local alternatives

$$\mathbf{B}(B)\mathbf{A}(B)\mathbf{x}(t) = \mathbf{N}(B)\mathbf{M}(B)\boldsymbol{\varepsilon}(t).$$

Note this parameterisation requires that a subset of the roots of the alternative model be the same as those of the null. The matrix polynomials $\mathbf{B}(z) = \sum_{j=0}^{r} \mathbf{B}_j z^j$ and $\mathbf{N}(z) = \sum_{j=0}^{s} \mathbf{N}_j z^j$ are such that $\mathbf{B}_0 = \mathbf{N}_0 = \mathbf{I}_v$ and the other coefficients are assumed to be at most $O(T^{-1/2})$. Inverting $\mathbf{B}(z)$ gives

$$\mathbf{A}(B)\mathbf{x}(t) = \mathbf{B}^{-1}(B)\mathbf{N}(B)\mathbf{M}(B)\boldsymbol{\varepsilon}(t)$$

where the coefficients of $\mathbf{B}^{-1}(z) = \sum_{j=0}^{\infty} \boldsymbol{\beta}_j^1 z^j$ are obtained from the recurrence relations $\sum_{j=0}^{i} \mathbf{B}_j \boldsymbol{\beta}_{i-j}^1 = \mathbf{0}$, $\mathbf{B}_0^1 = \mathbf{I}_v$. Premultiplying $\mathbf{N}(z)$ by $\mathbf{B}^{-1}(z)$ and neglecting terms of $O(T^{-1})$ and smaller gives an extending polynomial of degree $m = \max(r, s)$, yielding an ARMA

$(p_o, q_o + m)$ local representation of the alternative. Similarly, an autoregressive representation arises naturally by inverting $\mathbf{N}(z)$.

The behaviour of the test statistic depends only on the fitted null model together with the value of $m$ and not on the particular form of any alternative being envisaged. In practice, a prior choice must be made for $m$ in the absence of knowledge about its true value, should it exist. If it is believed that the phenomena being modelled can, in reality, be characterised by a finite parameterisation, so that there is a true value of $m$, and if the identification stage of the model building process has been carefully conducted, it would generally seem undesirable, on philosophical grounds, to entertain large values of $m$. On the other hand, in the absence of a true value, the consistency of the test procedure would necessitate $m \to \infty$ with $m/T \to 0$ as $T \to \infty$, although it should be emphasized that the distributional results of this section do not require such an assumption.

**3. Portmanteau procedures.** In univariate time series modelling, it has been common practice to apply the portmanteau statistics of Box and Pierce (1970) and Pierce (1972) or the subsequent modifications of Ljung and Box (1978) as tests of model adequacy. Since the tests of this paper are also to be used as diagnostic checks, it would seem to be of interest to develop a multivariate version of the portmanteau statistic and to compare it with the score test. This can be achieved by considering the form of the score test statistic when a fitted ARMA$(p_o, q_o)$ specification is tested against the specific ARMA$(p_o, q_o + s)$ alternative.

The statistic (2.1) in this case is a function of the last $v^2m$, $m = s$, elements of the score vector pertaining to the extending coefficients of the alternative model and their conditional covariance matrix, denoted $\boldsymbol{\Omega}_{\mu_e\mu_e}$, which is the inverse of the lower right $m \times m$ submatrix of the inverse of (2.4). Consider the transformation $\mathscr{T}$ applied to power series and defined by the equation

$$(3.1) \qquad \mathscr{T}_{pq}[\boldsymbol{\Gamma}_{\eta v}(z)] = (2\pi)^{-1} \int_{-\pi}^{\pi} \zeta_p \zeta_q^* \otimes \boldsymbol{\Gamma}_{\eta v}\, d\omega$$

which generates a matrix with $\gamma_{\eta_k v_\ell}(j - i)$ in row $(i - 1)u + k$ and column $(j - 1)v + \ell$, $i = 1, \ldots, p, j = 1, \ldots, q, k = 1, \ldots, u$ and $\ell = 1, \ldots, v$, the argument $z = e^{i\omega}$ again being understood. Let $\mathbf{C}_k = T^{-1} \sum_{t=1}^{T} \mathbf{e}(t)\mathbf{e}'(t - k)$ and $\mathbf{V}^{1/2}$ be a lower triangular matrix such that $\mathbf{C}_o = \mathbf{V}^{1/2}(\mathbf{V}^{1/2})'$. Employing (2.3) in conjunction with (3.1) and natural modifications of Lemmas $A1$–$A4$ of Poskitt and Tremayne (1981), the subvector of scores needed, when $m$ is allowed to tend to infinity with $T$ such that $m/T \to 0$, can be shown to be equal to $T\mathbf{D}\mathbf{r}$ where $\mathbf{r} = \text{vec}(\mathbf{R}_1' : \cdots : \mathbf{R}_m')$, $\mathbf{R}_k = \mathbf{V}^{-1/2}\mathbf{C}_k(\mathbf{V}^{-1/2})'$ and

$$\mathbf{D} = \mathscr{T}_{mm}[z^q \mathbf{V}^{1/2} \otimes \mathbf{M}'^{-1}(z)(\mathbf{V}^{-1/2})'].$$

It follows that $T^{1/2}\hat{\mathbf{r}}$ possesses an asymptotic Gaussian distribution with covariance matrix $\mathbf{D}^{-1}\boldsymbol{\Omega}_{\mu_e\mu_e}\mathbf{D}'^{-1}$.

The covariance matrix of $T^{1/2}\hat{\mathbf{r}}$ approaches the idempotent matrix $\mathbf{I}_{v^2m} - \mathbf{E}'(\mathbf{E}\mathbf{E}')^{-1}\mathbf{E}$ of rank $v^2(m - p_o - q_o)$, where

$$\mathbf{E}' = [\mathbf{E}_1' : \mathbf{E}_2']$$

$$\mathbf{E}_1 = \mathscr{T}_{p_o m}[-\mathbf{K}(z)\mathbf{V}^{1/2} \otimes \mathbf{M}'^{-1}(z)(\mathbf{V}^{-1/2})']$$

and

$$\mathbf{E}_2 = \mathscr{T}_{q_o m}[\mathbf{V}^{1/2} \otimes \mathbf{M}'^{-1}(z)(\mathbf{V}^{-1/2})'].$$

Consequently

$$(3.2) \qquad S \doteqdot T\hat{\mathbf{r}}'(I_{v^2m} - \hat{\mathbf{E}}'(\hat{\mathbf{E}}\hat{\mathbf{E}}')^{-1}\hat{\mathbf{E}})^{-}\hat{\mathbf{r}}$$

is asymptotically distributed as a Chi squared variate with $v^2(m - p_o - q_o)$ degrees of freedom for large $m$. A consequence of the idempotence of the covariance matrix is that a legitimate choice of generalized inverse in (3.2) is provided by the identity matrix. This leads to a multivariate version of the Box-Pierce portmanteau statistic given, in an obvious notation, by

$$Q = T \sum_{k=1}^{m} \sum_{i=1}^{v} \sum_{j=1}^{v} \{\hat{r}_{ij}(k)\}^2 = T \sum_{k=1}^{m} \sum_{i=1}^{v} \sum_{j=1}^{v} \hat{r}_{ij}(k)\hat{r}_{ji}(-k) \tag{3.3}$$

and extends the results of Chitturi (1974). An alternative derivation of the multivariate portmanteau statistic is provided by Hosking (1980) and McLeod (1979) contains discussion of related topics.

In addition, for any finite $m \geq 1$, a portmanteau type test based on residual auto- and cross-correlations will be identical to a score test with the same degrees of freedom when the moving average alternative is of a factored form as discussed in Section 2. In this case, $\mathbf{D} = \mathbf{I}_m$ so that

$$Q = S = T\hat{\mathbf{r}}'\hat{\boldsymbol{\Omega}}_{\mu_e\mu_e}^{-1}\hat{\mathbf{r}}. \tag{3.4}$$

This provides a direct generalisation of the results of McLeod (1978) and Newbold (1980).

Bearing in mind the invariances previously discussed, the above arguments amplify the pure significance test characteristics of diagnostic checks based on the score principle.

A corollary of the equivalences between score and portmanteau procedures is that the distribution of portmanteau statistics both under the null hypothesis of correct specification of the fitted model and under a sequence of asymptotically equivalent local alternatives is known.

**4. Further considerations.** The portmanteau statistic (3.3) is easily computed once the appropriate residual auto- and cross-correlations have been evaluated, and it seems of practical interest to consider whether the computation of the score test statistic can be accomplished in a fairly simple manner. Appealing to the invariance properties of the score test, it follows that, in addition to (2.1) and (3.4), $S$ can also be expressed as

$$\hat{\mathbf{w}}'(\mathbf{I}_T \otimes \hat{\mathbf{V}}^{-1})\hat{\mathbf{Z}}(\hat{\mathbf{Z}}'(\mathbf{I}_T \otimes \hat{\mathbf{V}}^{-1})\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}'(\mathbf{I}_T \otimes \hat{\mathbf{V}}^{-1})\hat{\mathbf{w}}$$

where the derivative processes of (2.2) and (2.3) that define the $n \times vT$ matrix $\mathbf{Z}' = (\mathbf{d}'(1): \cdots : \mathbf{d}'(T))$ are chosen so as to render the sensitivity matrix of full rank. The $1 \times vT$ vector $\mathbf{w}' = (\mathbf{e}'(1): \cdots : \mathbf{e}'(T))$. This may be interpreted as the explained sum of squares from an auxiliary generalized least squares regression of the fitted residuals on the estimated derivatives.

Although consideration has only explicitly been given to the case of testing common polynomial degrees, the score test procedure can be used for other situations of interest in time series analysis. For example, the familiar ARMAX model can be obtained by writing the $v$ component vector ARMA model in partitioned form as

$$\begin{bmatrix} \mathbf{A}_{11}(B) & \mathbf{A}_{12}(B) \\ \mathbf{A}_{21}(B) & \mathbf{A}_{22}(B) \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{11}(B) & \mathbf{M}_{12}(B) \\ \mathbf{M}_{21}(B) & \mathbf{M}_{22}(B) \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_1(t) \\ \boldsymbol{\varepsilon}_2(t) \end{bmatrix},$$

imposing the restrictions $\mathbf{A}_{21}(z) = \mathbf{M}_{21}(z) = \mathbf{M}_{12}'(z) = \mathbf{0}$ and concentrating attention on the first block of equations whilst ignoring the second. Many phenomena of interest, however, may not be sufficiently well understood to make such an assumption without subjecting it to test. It should now be clear how tests concerning hypotheses of exogeneity and feedback can be constructed using the score test statistic of (2.1). See Pierce and Haugh (1977) for a review paper on causal interrelationships.

It will often be the case in practice that the researcher is seeking a parsimonious parameterisation of the process and may not wish to consider fitting a wide range of alternative models. In this vein, and by way of conclusion, the attractions of the score test procedure in time series analysis lie not only in its desirable asymptotic properties but also in its use as an aid to deciding whether or not a sufficiently liberal parametric representation has been adopted.

**Acknowledgement.** We are pleased to acknowledge the help of the reviewers whose comments have, we believe, altered the paper for the better, though the usual caveat holds.

## REFERENCES

AITCHISON, J. and SILVEY, S. D. (1958). Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29** 813–828.

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

CHITTURI, R. V. (1974). Distribution of residual autocorrelations in multiple autoregressive schemes. *J. Amer. Statist. Assoc.* **69** 928–934.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics.* Chapman and Hall, London.

DEISTLER, M., DUNSMUIR, W. and HANNAN, E. J. (1978). Vector linear time series models: corrections and extensions. *Adv. in Appl. Probab.* **10** 360–372.

DUNSMUIR, W. and HANNAN, E. J. (1976). Vector linear time series models. *Adv. in Appl. Probab.* **8** 339–364.

GODFREY, L. G. (1979). Testing the adequacy of a time series model. *Biometrika* **66** 67–72.

HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

HANNAN, E. J. and HEYDE, C. C. (1972). On limit theorems for quadratic functions of discrete time series. *Ann. Math. Statist.* **43** 2058–2066.

HOSKING, J. R. M. (1980). The multivariate portmanteau statistic. *J. Amer. Statist. Assoc.* **75** 602–608.

KOHN, R. (1978). Asymptotic properties of time domain Gaussian estimators. *Adv. in Appl. Probab.* **10** 339–359..

KOHN, R. (1979). Asymptotic estimation and hypothesis testing results for vector linear time series models. *Econometrica* **47** 1005–1030.

LJUNG, G. M. and BOX, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65** 297–303.

MCLEOD, A. I. (1978). On the distribution of residual autocorrelations in Box-Jenkins models. *J.R. Statist. Soc. B* **40** 296–302.

MCLEOD, A. I. (1979). Distribution of residual cross-correlation in univariate ARMA time series models. *J. Amer. Statist. Assoc.* **74** 849–855.

NEUDECKER, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. *J. Amer. Statist. Assoc.* **64** 953–963.

NEWBOLD, P. (1980). The equivalence of two tests of time series model adequacy. *Biometrika* **67** 463–465.

NICHOLLS, D. F. (1976). The efficient estimation of vector linear time series models. *Biometrika* **63** 381–390.

NICHOLLS, D. F. (1977). A comparison of estimation methods for vector linear time series models. *Biometrika* **64** 85–90.

PIERCE, D. A. (1972). Residual correlations and diagnostic checking in dynamic-disturbance time series models. *J. Amer. Statist. Assoc.* **67** 636–640.

PIERCE, D. A. and HAUGH, L. D. (1977). Causality in temporal systems: characterizations and a survey. *J. Econometrics* **5** 265–293.

POSKITT, D. S. and TREMAYNE, A. R. (1980). Testing the specification of a fitted autoregressive-moving average model. *Biometrika* **67** 359–363.

POSKITT, D. S. and TREMAYNE, A. R. (1981). An approach to testing linear time series models. *Ann. Statist.* **9** 974–986.

RAO, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proc. Camb. Phil. Soc.* **44** 50–57.

RAO, C. R. and MITRA, S. K. (1971). *Generalized inverse of matrices and its applications.* Wiley, New York.

SILVEY, S. D. (1959). The Lagrangian multiplier test. *Ann. Math. Statist.* **30** 389–407.

WHITTLE, P. (1962). Gaussian estimation in stationary time series. *Bull. Internat. Statist. Inst.* **39** 105–129.

DEPARTMENT OF ECONOMICS &
RELATED STUDIES,
UNIVERSITY OF YORK,
HESLINGTON,
YORK Y01 5DD
U.K.