The Annals of Statistics
1978, Vol. 6, No. 1, 224-231

LINEAR PREDICTION BY AUTOREGRESSIVE MODEL FITTING IN THE TIME DOMAIN

BY R. J. BHANSALI

University of Liverpool

Let $\{x_t\}$ be a purely nondeterministic stationary process satisfying all the assumptions made by Berk (1974), and $\{y_t\}$ be another purely nondeterministic stationary process. Assume that y_t is independent of x_t but has exactly the same statistical properties as that of x_t. Consider the linear prediction of future values of y_t on the basis of past values, using prediction constants estimated from a realisation of T observations of x_t by least-squares fitting of an autoregression of order k. By assuming that $k \to \infty$, $k^3/T \to 0$ as $T \to \infty$, the effect on the mean square error of prediction of estimating the autoregressive coefficients is determined. This effect is the same as for the case when the prediction constants are estimated by factorising a "windowed" estimate of the spectral density function of x_t.

1. There has recently been considerable interest in the time domain estimation of the spectral density function and of the linear predictor of a stationary time series by autoregressive model fitting (see Akaike (1969), Parzen (1974), Berk (1974)). Such an approach may be regarded as an alternative to the corresponding frequency domain approach of first estimating the spectral density function by the "windowed" method and then estimating the linear predictor by factorising the estimated spectrum (see Bhansali (1974)). Berk (1974) derived the asymptotic distribution of the autoregressive spectral estimates on the assumption that the fitted order k of the autoregression increases simultaneously with the sample size. Davidson (1965), Akaike (1970) and Yamamoto (1976), on the other hand, derived expressions for the asymptotic mean square error of prediction when the generating model is assumed to be a true finite order autoregression. However, the corresponding problem of determining the mean square error when linear prediction is carried out nonparametrically in the time domain by autoregressive model fitting does not seem to have been considered in the published literature. In this paper an approach similar to that of Berk (1974) is used to derive an expression for evaluating the mean square error of prediction up to h steps ahead ($h = 1, \cdots, \nu$) when the true order of the autoregression is assumed to be infinite. This expression shows that the nonparametric predictor given by time domain autoregressive model fitting is asymptotically equivalent to the corresponding linear predictor obtained by factorising a "windowed" estimate of the spectral density function.

The idea of fitting autoregression of order k such that $k \to \infty$ with the sample

Received March 1976; revised March 1977.

AMS 1970 *subject classifications*. Primary 62M20; Secondary 62M10.

Key words and phrases. Autoregressive process, time domain, Yule-Walker equations, time series.

size has also been used by Durbin (1959), (1960) to derive asymptotically efficient estimates of the parameters of the moving average and the autoregressive-moving average models. The results of this paper may be used to give an asymptotic justification of this procedure, which was not given by Durbin (see Hannan (1969)).

The notation used in this paper is the same as in Chapter 6 of Rao (1973). Hence for any two sequences of random variables $\{g_T\}$ and $\{h_T\}$ $g_T =_a g$ is used to denote the fact that $g_T - g$ tends to zero in probability as $T \to \infty$. Similarly a.Var $\{g_T\}$ is used to denote the variance of the asymptotic distribution of g_T as $T \to \infty$, and a.Cov $\{g_T, h_T\}$ is used to denote the covariance in the joint asymptotic distribution of g_T and h_T.

2. Estimated autoregressive coefficients. Suppose that $\{x_t\}$ $(t = 0, \pm 1, \cdots)$ is a purely nondeterministic stationary process satisfying the equation

$$\sum_{u=0}^{\infty} a(u)x_{t-u} = e_t\,, \quad a(0) = 1\,, \tag{2.1}$$

where the $a(u)$'s are real coefficients such that the polynomial

$$A(z) = \sum_{u=0}^{\infty} a(u)z^u$$

is bounded and bounded away from zero for $|z| \leqq 1$, and $\{e_t\}$ is a sequence of independent, identically distributed random variables with mean 0 and variance σ^2. Let $X_1, \cdots, X_T$ denote a realisation of $\{x_t\}$ and $c_u = \hat{a}_{ku}$ $(u = 1, \cdots, k)$ denote the kth order estimates of $a(u)$ obtained by minimising

$$(T-k)^{-1} \sum_{j=0}^{T-1-k} (X_{k+j-1} + c_1 X_{k+j} + \cdots + c_k X_{1+j})^2\,. \tag{2.2}$$

Following Berk (1974) define the vectors $\hat{\mathbf{a}}(k)' = [\hat{a}_{k1}, \cdots, \hat{a}_{kk}]$, $\mathbf{a}(k)' = [a(1), \cdots, a(k)]$, $\mathbf{X}_j(k)' = [X_j, X_{j+1}, \cdots, X_{j-k+1}]$, and let $c_1 = a_{k1}, \cdots, c_k = a_{kk}$ be the values that minimise

$$E[\{x_t + \sum_{u=1}^{k} c_u x_{t-u}\}^2]\,,$$

with minimum $\sigma^2(k)$. We will also need the $k \times k$ matrix of sample covariances

$$\hat{\mathbf{R}}(k)' = \frac{1}{T-k} \sum_{j=k}^{T-1} \mathbf{X}_j(k)\mathbf{X}_j(k)'\,,$$

and an infinite dimensional matrix

$$\mathbf{R} = [R(u-v)]\,, \qquad u, v = 1, 2, \cdots$$

with $R(u-v) = E(x_{t-u}x_{t-v})$. A $k \times k$ submatrix of $\mathbf{R}$ will be denoted by $\mathbf{R}(k)$.

Although Berk (1974) did not explicitly consider the asymptotic distribution of the estimated autoregressive coefficients, this could be derived from his results.

THEOREM 1. *Assume that the conditions stated in Theorem* 2 *of Berk* (1974) *are satisfied. In particular, assume that*:

(i) $A(e^{i\lambda})$ *is nonzero*, $-\pi < \lambda \leqq \pi$;
(ii) $E(e_t^4) < \infty$;

(iii) *The choice of k in terms of T is such that* $k^3/T \to 0$;
(iv) *The choice of k in terms of T is such that* $T^{\frac{1}{2}}(|a_{k+1}| + |a_{k+2}| + \cdots) \to 0$.

Then the joint asymptotic distribution of

$$T^{\frac{1}{2}}\{\hat{a}_{ku} - a(u)\} \quad \text{and} \quad T^{\frac{1}{2}}\{\hat{a}_{kv} - a(v)\}, \qquad u, v = 1, \cdots, k,$$

is bivariate normal with zero means and covariance structure

$$(2.3) \qquad \lim_{T\to\infty} T\,\text{a.Cov}\,\{\hat{a}_{ku}, \hat{a}_{kv}\} = \sigma^2 c_{uv} = \sum_{p=0}^{u-1} a(p)a(p-u+v), \quad 1 \leqq u \leqq v \leqq k,$$

where c_{uv} *denotes the term in the* uth *row and the* vth *column of* $\mathbf{R}^{-1}$.

PROOF. Suppose that the constants γ_{jk} of Theorem 2 of Berk (1974) take the following values:

$$\gamma_{jk} = 1 \quad j = u$$
$$= 0 \quad j \neq u$$

and let $\boldsymbol{\gamma}(k)' = [\gamma_{1k}, \gamma_{2k}, \cdots, \gamma_{kk}]$.

Then it follows from his Theorem 2 that

$$T^{\frac{1}{2}}\{\hat{a}_{ku} - a(u)\} =_a \frac{T^{\frac{1}{2}}}{(T-k)} \boldsymbol{\gamma}(k)'\mathbf{R}(k)^{-1} \sum_{j=k}^{T-1} \mathbf{X}_j(k)e_{j+1}, \qquad u = 1, \cdots, k.$$

Proof of the present theorem follows by repeating the argument given in Berk's Theorem 4, and introducing an infinite dimensional analogue of the matrix $\mathbf{U}$ defined by Wise (1955).

3. One-step prediction. Consider another stationary, purely nondeterministic process $\{y_t\}$ which is independent of $\{x_t\}$ but satisfies the same equation as $\{x_t\}$

$$\sum_{u=0}^{\infty} a(u)y_{t-u} = e_t, \quad a(0) = 1.$$

Suppose that it is desired to obtain the linear, least-squares predictor of the future values y_{n+h}; $h = 0, 1, \cdots, \nu - 1$, say of y_t from its known past. If the $a(u)$'s are known and the complete history of $\{y_t\}$ is available, then the one-step predictor is given by

$$(3.1) \qquad \hat{y}_n = -\sum_{u=1}^{\infty} a(u)y_{n-u}$$

with the corresponding mean square error of prediction

$$E[\{\hat{y}_n - y_n\}^2] = \sigma^2.$$

In practice, the $a(u)$'s are rarely known a priori: if these have been estimated by least squares, using a realisation of T observations of x_t in the manner described above, y_n will be estimated by

$$(3.2) \qquad \hat{y}_n(k) = \sum_{u=1}^{k} \hat{a}_{ku} y_{n-u}.$$

Let

$$u_n = y_n - \hat{y}_n(k) = w_n + z_n,$$

where

$$w_n = \{\hat{\mathbf{a}}(k)' - \mathbf{a}(k)'\}\mathbf{y}_{n-1}(k),$$
$$z_n = \sum_{u=k+1}^{\infty} a(u)y_{n-u},$$

and

$$\mathbf{y}_{n-1}(k) = [y_{n-1}, \cdots, y_{n-k}]'.$$

The asymptotic effect on the mean square error of prediction of estimating y_n using (3.2) rather than (3.1) may be determined by letting k and T tend to infinity simultaneously.

THEOREM 2. *Assume that the conditions* (i)—(iv) *given in Theorem* 1 *are satisfied. Then*

$$\lim_{T\to\infty} \text{a.Var}\left\{\left(\frac{T}{k}\right)^{\frac{1}{2}} w_n\right\} = \sigma^2. \tag{3.3}$$

PROOF. Let

$$y_t = \int_{-\pi}^{\pi} \exp(it\lambda)\, dS_y(\lambda),$$

where $S_y(\lambda)$ is a process with orthogonal increments. Then

$$\left(\frac{T}{k}\right)^{\frac{1}{2}} w_n = \int_{-\pi}^{\pi} \left(\frac{T}{k}\right)^{\frac{1}{2}} \boldsymbol{\alpha}'(k)\{\hat{\mathbf{a}}(k) - \mathbf{a}(k)\}\, dS_y(\lambda),$$

where

$$\boldsymbol{\alpha}'(k) = [e^{i(n-1)\lambda}, e^{i(n-2)\lambda}, \cdots, e^{i(n-k)\lambda}]. \tag{3.4}$$

Arguing as Berk does in proving his Theorem 2, we may show that

$$\left(\frac{T}{k}\right)^{\frac{1}{2}} w_n = I_1 + I_2 + I_3,$$

where

$$I_1 = \int_{-\pi}^{\pi} \left(\frac{T}{k}\right)^{\frac{1}{2}} \frac{\boldsymbol{\alpha}'(k)}{(T-k)} [\hat{\mathbf{R}}(k)^{-1} - \mathbf{R}(k)^{-1}] \sum_{j=k}^{T-1} \mathbf{X}_j(k)e_{j+1,k}\, dS_y(\lambda),$$

$$I_2 = \int_{-\pi}^{\pi} \left(\frac{T}{k}\right)^{\frac{1}{2}} \frac{\boldsymbol{\alpha}'(k)}{(T-k)} \mathbf{R}(k)^{-1} \sum_{j=k}^{T-1} \mathbf{X}_j(k)(e_{j+1,k} - e_{j+1})\, dS_y(\lambda),$$

$$I_3 = \int_{-\pi}^{\pi} \left(\frac{T}{k}\right)^{\frac{1}{2}} \frac{\boldsymbol{\alpha}'(k)}{(T-k)} \mathbf{R}(k)^{-1} \sum_{j=k}^{T-1} \mathbf{X}_j(k)e_{j+1}\, dS_y(\lambda),$$

and

$$e_{j+1,k} = \sum_{u=0}^{k} a(u)x_{j+1-u}.$$

For an arbitrary vector $\mathbf{x}$ let

$$||\mathbf{x}||_2 = (\mathbf{x}'\mathbf{x})^{\frac{1}{2}}$$

denote its Euclidean norm and for an $n \times n$ matrix $\mathbf{C} = [C_{ij}]$ $(i, j = 1, \cdots, n)$ define the matrix norms

$$||\mathbf{C}||_1 = \max_i \sum_{j=1}^{n} C_{ij},$$
$$||\mathbf{C}||_2 = \max(\mathbf{x}'\mathbf{C}'\mathbf{C}\mathbf{x})^{\frac{1}{2}}, \quad ||\mathbf{x}||_2 \leqq 1.$$

It then follows that

$$|I_1| < \frac{T^{\frac{1}{2}}}{T-k} \|\hat{\mathbf{R}}(k)^{-1} - \mathbf{R}(k)^{-1}\|_2 \|\sum_{j=k}^{T-1} \mathbf{X}_j(k) e_{j+1,k}\|_2 \int_{-\pi}^{\pi} \frac{\|\boldsymbol{\alpha}'(k)\|_2}{k^{\frac{1}{2}}} |dS_y(\lambda)| .$$

Hence, using (2.13), Lemma 3 and condition (iv) of Berk (1974) it follows that I_1 tends to zero in probability as $T \to \infty$. I_2 may similarly be shown to tend to zero in probability. Then (3.3) follows by taking conditional expectations of I_3 first with respect to the y's and then with respect to the x's.

To derive an expression for the asymptotic mean square error of predicting one step ahead, we note that

$$\text{(3.5)} \qquad \lim_{k\to\infty} E(z_n^2) = \sigma^2$$

$$\lim_{T\to\infty, k\to\infty} E(z_n w_n) = 0 .$$

Hence, to order T^{-1}, the asymptotic mean square error of prediction is given by

$$\text{(3.6)} \qquad \text{a.Var}\,\{u_n\} \sim \sigma^2 \left(1 + \frac{k}{T}\right) .$$

For two arbitrary stationary processes $\{y_t\}$ and $\{x_t\}$, Wahba (1969) considered prediction of y_t from a realisation of x_t by a distributed-lag model. An expression for the increase in the mean square error of prediction due to estimating the prediction constants was also given. In this paper, we have assumed that y_t is independent of x_t and has the same statistical properties as x_t. However, (3.6) is similar to the expression given by Wahba.

4. More than one step ahead prediction. The linear least-squares predictor of y_{n+h} $(h = 0, \cdots, \nu - 1)$ when the autoregressive coefficients are known may be written as

$$\text{(4.1)} \qquad \hat{y}_{n+h} = \mathbf{V}_{h+1} \boldsymbol{\chi}_{n-1} \qquad h = 0, 1, \cdots, \nu - 1 ,$$

where $\mathbf{V}_{h+1}$ is the first row of an infinite-dimensional matrix $\mathbf{A}^{h+1}$,

$$\mathbf{A} = \begin{bmatrix} -a(1) & -a(2) & -a(3) & \cdots \\ & I & & \end{bmatrix} ,$$

$\mathbf{I}$ denotes an infinite-dimensional identity matrix, and

$$\boldsymbol{\chi}_{n-1} = [y_{n-1}, y_{n-2}, \cdots]' .$$

The corresponding mean square error of prediction is given by

$$E[\{\hat{y}_{n+h} - y_{n+h}\}^2] = \sigma^2 \sum_{j=0}^{h} b^2(j) , \qquad h = 0, 1, \cdots, \nu - 1 ,$$

where the $b(j)$'s are the coefficients of the moving average representation of y_t. As in the last section, we will suppose that, in practice, the $a(u)$'s are estimated by $\hat{a}_{ku}$'s using a realisation of length T from x_t. y_{n+h} will be estimated by

$$\text{(4.2)} \qquad \hat{y}_{n+h}(k) = \hat{\mathbf{V}}_{h+1}(k) \mathbf{y}_{n-1}(k) ,$$

where $\hat{\mathbf{V}}_{h+1}(k)$ is the first row of $\hat{\mathbf{A}}^{h+1}(k)$, $\hat{\mathbf{A}}(k)$ is a $k \times k$ matrix similar to $\mathbf{A}$ but $a(u)$ is replaced by $\hat{a}_{ku}$ $(u = 1, \cdots, k)$ and $\mathbf{y}_{n-1}(k)$ is the same as in Section 3.

Let

$$u_{n+h} = y_{n+h} - \hat{y}_{n+h}(k) , \qquad h = 0, 1, \cdots, \nu - 1 ,$$

and

$$y_{n+h}(k) = \mathbf{V}_{h+1}(k)\mathbf{y}_{n-1}(k) .$$

Here $\mathbf{V}_{h+1}(k)$ is the first row of $\mathbf{A}^{h+1}(k)$ and $\mathbf{A}(k)$ is a $k \times k$ matrix similar to $\hat{\mathbf{A}}(k)$ but $\hat{a}_{ku}$ replaced by $a(u)$ $(u = 1, \cdots, k)$. We, therefore, have

$$y_{n+h} = w_{n+h} + z_{n+h}$$

where

$$w_{n+h} = y_{n+h}(k) - \hat{y}_{n+h}(k) ,$$

and

$$z_{n+h} = y_{n+h} - y_{n+h}(k) .$$

Following an approach taken by Yamamoto (1976), we get by Taylor series expansion

$$\hat{y}_{n+h}(k) = y_{n+h}(k) + \{\hat{\mathbf{a}}(k) - \mathbf{a}(k)\}'\mathbf{M}_{h+1}(k)\mathbf{y}_{n-1}(k) + \text{higher order terms}$$

where

$$\text{(4.3)} \qquad \mathbf{M}_{h+1}(k) = \frac{\partial \mathbf{V}_{h+1}(k)}{\partial \hat{\mathbf{a}}(k)}\bigg|_{\hat{\mathbf{a}}(k)=\mathbf{a}(k)} = \textstyle\sum_{j=0}^{h} b(j)\mathbf{A}^{h-j}(k) .$$

Let

$$\mathbf{M}_{h+1}(k)\mathbf{y}_{n-1}(k) = \textstyle\int_{-\pi}^{\pi} \mathbf{q}(k)\, dS_y(\lambda) ,$$

where

$$\mathbf{q}(k) = \mathbf{M}_{h+1}(k)\boldsymbol{\alpha}(k) .$$

Set

$$\xi_{n+h} = \{\hat{\mathbf{a}}(k)' - \mathbf{a}(k)'\}\mathbf{M}_{h+1}(k)\mathbf{y}_{n-1}(k) .$$

Then its asymptotic variance may be determined using the following result.

THEOREM 3. *Assume that the conditions* (i)—(iv) *stated in Theorem* 1 *are satisfied. Then*

$$\text{(4.4)} \qquad \lim_{T\to\infty} \text{a.Var}\left\{\left(\frac{T}{k}\right)^{\frac{1}{2}} \xi_{n+h}\right\} = \lim_{k\to\infty} k^{-1} \textstyle\int_{-\pi}^{\pi} \overline{\mathbf{q}(k)}'\mathbf{R}(k)^{-1}\mathbf{q}(k) f(\lambda)\, d\lambda = c , \quad \textit{say},$$

where $f(\lambda)$ *denotes the spectral density function of* x_t.

PROOF. The argument of Theorem 2 with $\boldsymbol{\alpha}(k)$ replaced by $\mathbf{q}(k)$ yields (4.4). To complete the proof we will show that as $k \to \infty$, c remains bounded.

Let m_{rs} $(r, s = 1, \cdots, k)$ denote the term in the rth row and the sth column of $\mathbf{M}_{h+1}(k)$ and let

$$H_r(\lambda) = \textstyle\sum_{s=1}^{k} m_{rs} \exp(-is\lambda) .$$

Then

$$c = \lim_{k\to\infty} k^{-1} \textstyle\int_{-\pi}^{\pi} \{\sum_{j=1}^{k} D_j(\lambda)D_j^*(\lambda)\} f(\lambda)\, d\lambda ,$$

where

$$D_j(\lambda) = \sum_{r=1}^{j} \frac{a_{j-1,j-r}}{\sigma^2(j-1)} H_r(\lambda) ,$$

and $D_j^*(\lambda)$ is similarly defined with $H_r(\lambda)$ replaced by its complex-conjugate $\overline{H_r(\lambda)}$. Now

$$|H_r(\lambda)| \leqq ||\mathbf{M}_{h+1}(k)||_1 \leqq \sum_{p=0}^{h} b(p)\{||\mathbf{A}(k)||_1\}^p$$

remains bounded for all $h = 0, 1, \cdots, \nu - 1$. Since the first derivative of $f(\lambda)$ exists for all λ, it follows that $D_j(\lambda)$ and $D_j^*(\lambda)$, and hence c, also remain bounded.

We have

$$\lim_{k\to\infty} Ez_{n+h}^2 = \sigma^2 \sum_{j=0}^{h} b^2(j) .$$

Also, as $k \to \infty$, z_{n+h} and w_{n+h} are asymptotically uncorrelated. These results, therefore, show that as $T \to \infty$, $k \to \infty$ the asymptotic mean square error of predicting more than one step ahead may be approximated by

$$\text{a.Var}\,(u_{n+h}) \sim \sigma_{h+1}^2 + \frac{k}{T}\,c . \tag{4.5}$$

In practice k and T will be finite. Provided that these are large, (4.5) may be approximated by

$$\sigma_{h+1}^2 + T^{-1}\,\text{tr}\,\{\mathbf{M}'_{h+1}(k)\mathbf{R}(k)^{-1}\mathbf{M}_{h+1}(k)\mathbf{R}(k)\} . \tag{4.6}$$

If the given T observations were assumed to be a sample of an autoregressive process of order k, then the resulting asymptotic mean square error of prediction will be given by (3.6) and (4.6). The above results, therefore, show that the expressions valid for this case may still be used for the case when the true order is in fact infinite, provided that k is large enough to ignore the bias due to fitting a finite autoregression. Indeed, in practice, k may be chosen so as to minimise the expected one step prediction error (3.6). Since for a finite kth order autoregressive process, the asymptotic one step prediction error, to order T^{-1}, is also given by (3.6); this suggestion may be implemented by using the FPE-criterion introduced by Akaike (1970). We, however, note that the FPE criterion was originally suggested for estimating the order of a *finite* autoregressive process. Therefore, it may also be used for determining an optimal finite order approximation to a true infinite order autoregressive process. Parzen (1974) has suggested an alternative CAT-criterion for this purpose.

Bhansali (1974), (1977) gives expressions for the increase in the mean square error of prediction when the autoregressive coefficients are estimating by factorising a "windowed" estimate of the spectral density function. The expressions (2.3), (3.6) and (4.6) given above are similar to the corresponding results given there. If $2k$ equals the equivalent number of independent spectral estimates (see Bloomfield (1972)) then the two mean square errors of prediction will be the same. When this condition is satisfied, the asymptotic distribution of the autoregressive spectral estimates is also the same as that of the "windowed" estimate (see Hannan (1970), Berk (1974)).

Expressions (3.6) and (4.6) were also derived by Bhansali (1976) when studying the effect of mis-specifying the order of a finite autoregression, on the assumption that first T and then the order k of the fitted autoregression tends to infinity. The results given in this paper give conditions on the relative asymptotic rates of k and T for the bias in using a finite autoregression to vanish sufficiently quickly.

Acknowledgments. The author is grateful to Professor M. R. Sampford, and the Editor, for helpful suggestions on exposition.

REFERENCES

[1] AKAIKE, H. (1969). Power spectrum estimation through autoregressive model fitting. *Ann. Inst. Statist. Math.* **21** 407–419.

[2] AKAIKE, H. (1970). Statistical predictor identification. *Ann. Inst. Statist. Math.* **22** 203–217.

[3] BERK, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2** 489–502.

[4] BHANSALI, R. J. (1974). Asymptotic properties of the Wiener–Kolmogorov predictor, I. *J. Roy. Statist. Soc. B* **36** 61–73.

[5] BHANSALI, R. J. (1976). Effects of mis-specifying the order of an autoregressive process on the mean square error of prediction. Research Report CSS/76/10/3, Dept. Computational and Statist. Science, Univ. of Liverpool.

[6] BHANSALI, R. J. (1977). Asymptotic properties of the Wiener–Kolmogorov predictor, II. *J. Roy. Statist. Soc. B* **39** 66–72.

[7] BLOOMFIELD, P. (1972). On the error of prediction of a time series. *Biometrika* **59** 501–508.

[8] DAVIDSON, I. D. (1969). The prediction error of stationary Gaussian time series of unknown covariance. *IEEE Trans. Information Theory* **IT-11** 527–532.

[9] DURBIN, J. (1959). Efficient estimation of parameters in time series models. *Biometrika* **46** 306–316.

[10] DURBIN, J. (1960). The fitting of time series models. *Rev. Int. Statist. Inst.* **28** 233–244.

[11] HANNAN, E. J. (1969). The estimation of mixed moving average autoregressive systems. *Biometrika* **56** 579–593.

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

[13] PARZEN, E. (1974). Some recent advances in time series modelling. *IEEE Trans. Automatic Control* **AC-19** 723–730.

[14] RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

[15] WAHBA, G. (1969). Estimation of the coefficients in a multidimensional distributed lag model. *Econometrika* **37** 398–407.

[16] WISE, J. (1955). The autocorrelation function and the spectral density function. *Biometrika* **42** 151–159.

[17] YAMAMOTO, T. (1976). Asymptotic mean square prediction error for an autoregressive model with estimated coefficients. *Appl. Statist.* **25** 123–127.

DEPARTMENT OF COMPUTATIONAL AND STATISTICAL SCIENCE
UNIVERSITY OF LIVERPOOL
BROWNLOW HILL
P.O. BOX 147
LIVERPOOL L69 3BX, ENGLAND