J. R. Statist. Soc. B (1981),
43, *No.* 2, *pp.* 244–255

On the Theory of Bilinear Time Series Models

By T. Subba Rao

University of Manchester Institute of Science and Technology, UK

[Received November 1979. Final revision November 1980]

Summary

The theory of bilinear time series models is considered in this paper. The sufficient conditions for asymptotic stationarity of the bilinear time series models are derived, and the expressions for the variance and covariance are obtained. The conditions for the invertibility of the model are also included. The estimation of the parameters of the scalar bilinear time series model is considered. The bilinear models are fitted to sunspot numbers and also to a P-wave of a nuclear explosion. The forecasting of sunspot numbers is also considered.

Keywords: BILINEAR TIME SERIES MODELS; CONDITIONS FOR STATIONARITY AND INVERTIBILITY; ESTIMATION; FORECASTING; PRESSURE WAVE; SUNSPOT DATA; VOLTERRA SERIES; YULE–WALKER EQUATIONS

1. Introduction

THE classical theory of time series analysis has been well developed over the past two decades, and excellent accounts of this theory are available, for example in Hannan (1962, 1970), Box and Jenkins (1970) and many other books. An important assumption that is made in the classical theory is that the structure of the series can be described by a linear model such as an autoregressive, moving-average or mixed autoregressive moving-average model.

The assumption of linearity is often a very dubious one. The theory of Volterra (1930) and Wiener (1958) on functional series representation has provided great stimulus to the development of non-linear models, but unfortunately Wiener's representation is too general and the statistical estimation of the Wiener kernels is unwieldy. In view of this, several authors (Ozaki and Oda, 1977; Jones, 1978; Haggan and Ozaki, 1980; Tong and Lim, 1980) have recently discussed certain more specific types of non-linear models.

A particular class of non-linear models which have been extensively discussed in the control theory literature is the *bilinear* models. (See, for example, Ruberti, Isidori and d'Alessandro, 1972; Mohler, 1973.) The interesting feature of a bilinear system is that though it is non-linear, its structural theory is analogous to that of linear systems. (See Ruberti, Isidori and d'Alessandro, 1972 and references therein.)

The theory developed so far has dealt with the structural theory of deterministic bilinear differential equations, and only recently have attempts been made to extend these results to models where the input is a random function (see Granger and Andersen, 1978a).

Let $\{X(t)\}$ be a discrete parameter time series, satisfying the difference equation

$$X(t)+\sum_{j=1}^{p} a_j X(t-j) = \sum_{j=0}^{r} c_j e(t-j)+\sum_{l=1}^{m}\sum_{l'=1}^{k} b_{ll'} X(t-l)\, e(t-l'), \tag{1.1}$$

where $\{e(t)\}$ is an independent white noise process and $c_0 = 1$. We define the model (1.1) as a bilinear time series model BL(p,r,m,k) and the process $\{X(t)\}$ as a bilinear process. In their monograph Granger and Andersen (1978a) have considered the statistical properties of the model BL$(1,0,1,1)$. The autoregressive-moving average model ARMA(p,r) can be obtained from (1.1) by setting $b_{ll'} = 0$ for all l and l'.

In this paper the object is to study systematically the general theory of the bilinear time series models BL$(p,0,p,1)$ and BL$(p,0,p,q)$. In Section 2 we consider the state space representations of these models. The conditions for stationarity and the expressions for the covariance are given in Section 3. The invertibility condition is derived in Section 4. The estimation of the parameters of the bilinear time series model is considered in Section 5. In the final section the fitting of bilinear time series models to sunspot data, and to the P wave of a seismological time series, is considered. The forecasting of sunspot data is also considered in the same section, and the forecasts are compared with the forecasts obtained from the best linear model.

2. VECTOR FORM OF THE BILINEAR MODELS

It is well known that the linear autoregressive moving average models can be written in the form of a first-order vector difference equation (see Anderson, 1971; Priestley, 1978, 1980) and this vector form is known as the state space form. It is convenient to study the properties of the process when the model is in the state space form because of the Markovian nature of the model (Akaike, 1974). We shall represent the bilinear models in the state space form and various properties are derived in the following sections.

Consider the bilinear model BL $(p,0,p,1)$, i.e.

$$X(t)+\sum_{j=1}^{p} a_j X(t-j) = e(t)+\left(\sum_{l=1}^{p} b_{l_1} X(t-l)\right)e(t-1). \tag{2.1}$$

Let us define the matrices

$$\mathbf{A} = \begin{pmatrix} -a_1 & -a_2 & \dots & & -a_p \\ 1 & 0 & \dots & & 0 \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{21} & b_{31} & \dots & b_{p1} \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \tag{2.2}$$

and $\mathbf{C}' = (1,0,0,\dots,0)$, $\mathbf{H}' = (1,0,\dots,0)$, and let $\mathbf{x}'(t) = (X(t), X(t-1),\dots, X(t-p+1))$. With this notation, we can write the model (2.1) in the form

$$\begin{aligned} \mathbf{x}(t) &= \mathbf{A}\mathbf{x}(t-1)+\mathbf{B}\mathbf{x}(t-1)\,e(t-1)+\mathbf{C}e(t), \\ X(t) &= \mathbf{H}'\mathbf{x}(t). \end{aligned} \tag{2.3}$$

We define this model (2.3) as a vector form of the bilinear model BL $(p,0,p,1)$ and denote it by VBL (p) (the initial letter emphasizing the fact that (2.3) is written in the vector form).

Suppose we have the bilinear model BL $(p,0,p,q)$. We define the matrix $\mathbf{A}$, and the vectors $\mathbf{C}$, $\mathbf{H}$ and $\mathbf{x}(t)$ as before. Define the matrices

$$\mathbf{B}_j = \begin{pmatrix} b_{1j} & b_{2j} & \dots & b_{pj} \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix}, \quad (j=1,2,\dots,q). \tag{2.4}$$

Then the vector form of the bilinear model BL $(p,0,p,q)$ is (VBL (p,q)).

$$\begin{aligned} \mathbf{x}(t) &= \mathbf{A}\mathbf{x}(t-1)+\sum_{j=1}^{q} \mathbf{B}_j\,\mathbf{x}(t-1)\,e(t-j)+\mathbf{C}e(t), \\ X(t) &= \mathbf{H}'\mathbf{x}(t). \end{aligned} \tag{2.5}$$

In the following sections, we consider the conditions for stationarity and invertibility for the bilinear model BL $(p,0,p,1)$. For convenience, we use the state space form of the model, namely VBL (p) given by (2.3) for these derivations.

3. EXPRESSIONS FOR THE COVARIANCES AND CONDITIONS FOR STATIONARITY

In this section we obtain the conditions for asymptotic stationarity of the time series $X(t)$

satisfying the model (2.3). We have

$$E(X(t)) = \mathbf{H}'E(\mathbf{x}(t)),$$
$$\operatorname{cov}(X(t)\,X(t+s)) = \mathbf{H}'E(\mathbf{x}(t)-E(x(t))\,(\mathbf{x}(t+s)-E(x(t+s))'\mathbf{H}.$$

In the following derivation we assume that the random variables $\{e(t)\}$ are independent and each $e(t)$ is distributed $N(0, 1)$, and obtain expressions for $\operatorname{cov}(\mathbf{x}(t)\,\mathbf{x}(t+s))$.

Let

$$\boldsymbol{\mu}(t) = E(\mathbf{x}(t)), \quad \mathbf{V}(t) = E(\mathbf{x}(t)\,\mathbf{x}'(t)),$$
$$\mathbf{S}(t) = E[\mathbf{x}(t)\,\mathbf{x}'(t)\,e(t)], \quad W(t) = E[\mathbf{x}(t)\,\mathbf{x}'(t)\,e^2(t)].$$

Taking expectations on both sides of (2.3) and noting $\mathrm{E}(\mathbf{x}(t)\,e(t)) = \mathbf{C}$, we obtain

$$\boldsymbol{\mu}(t+1) = \mathbf{A}\boldsymbol{\mu}(t)+\mathbf{BC} = \mathbf{A}^t\boldsymbol{\mu}(1)+\left(\sum_{j=0}^{t-1}\mathbf{A}^j\right)\mathbf{BC}. \tag{3.1}$$

If $\mathbf{B} = \mathbf{0}$ and $\boldsymbol{\mu}(1) = 0$ then $\boldsymbol{\mu}(t) = 0$ for all $t \geqslant 1$; and hence in this case no condition on the matrix $\mathbf{A}$ is necessary for the first-order stationarity. Otherwise, we proceed as follows.

Let the spectral radius of a matrix $\mathbf{A}$, $\rho(\mathbf{A})$, be

$$\rho(\mathbf{A}) = \max_i \{\,|\lambda_i(\mathbf{A})|\,\}, \tag{3.2}$$

where $\lambda_1(\mathbf{A})$ is the ith eigenvalue of $\mathbf{A}$, and it is known that $\rho(\mathbf{A}) \leqslant \|\mathbf{A}\|$ where $\|\mathbf{A}\|$ is any norm (Wilkinson, 1965). A *sufficient condition* for

$$\lim_{t\to\infty}\left[\mathbf{A}^t\mu(1)+\left(\sum_{j=0}^{t-1}\mathbf{A}^j\right)\mathbf{BC}\right]$$

to be finite is that $\rho(\mathbf{A}) < 1$. Under this condition the mean value $\boldsymbol{\mu}$ is then given by

$$\boldsymbol{\mu} = (\mathbf{I}-\mathbf{A})^{-1}\,\mathbf{BC}. \tag{3.3}$$

We now obtain the conditions for second-order stationarity. From the model (2.3) we have $E\{\mathbf{x}(t)\,e(t+1)\} = 0$, $E\{\mathbf{x}(t)\,e(t)\,e(t+1)\} = 0$. Also from (2.3) we obtain

$$\mathbf{V}(t) = \mathbf{AV}(t-1)\,\mathbf{A}'+\mathbf{AS}(t-1)\,\mathbf{B}'+\mathbf{BS}(t-1)\,\mathbf{A}'+\mathbf{BW}(t-1)\,\mathbf{B}'+\mathbf{CC}', \tag{3.4}$$

where

$$\mathbf{S}(t) = \mathbf{A}\boldsymbol{\mu}(t-1)\,\mathbf{C}'+\mathbf{BCC}'+\mathbf{C}\mu(t-1)\,\mathbf{A}'+\mathbf{CC}'\mathbf{B}', \tag{3.5}$$

$$\begin{aligned}\mathbf{W}(t) &= \mathbf{AV}(t-1)\,\mathbf{A}'+\mathbf{AS}(t-1)\,\mathbf{B}'+\mathbf{BS}(t-1)\,\mathbf{A}'+\mathbf{BW}(t-1)\,\mathbf{B}'+3\mathbf{CC}'\\ &= \mathbf{V}(t)+2\mathbf{CC}'.\end{aligned} \tag{3.6}$$

In obtaining the expression for $\mathbf{W}(t)$, we have made use of the fact that the random variables $\{e(t)\}$ are Gaussian with $E\{e(t)\} = 0$, $E\{e^2(t)\} = 1$, so that $E\{e^4(t)\} = 3$. The above derivation is still valid even if $e(t)$ is not Gaussian, but in this case $E\{e^4(t)\} = 3+K_4$, where K_4 is the fourth-order cumulant.

From (3.4) and (3.6) we have

$$\begin{aligned}\mathbf{V}(t) = \mathbf{AV}(t-1)\,\mathbf{A}'&+\mathbf{BV}(t-1)\,\mathbf{B}'+\mathbf{AS}(t-1)\,\mathbf{B}'\\ &+\mathbf{BS}(t-1)\,\mathbf{A}'+2\mathbf{BCC}'\mathbf{B}'+\mathbf{CC}'.\end{aligned} \tag{3.7}$$

We now assume that the process $\{\mathbf{x}(t)\}$ is first-order stationary so that $\boldsymbol{\mu}(t) = \boldsymbol{\mu}$ and this implies $\mathbf{S}(t) = \mathbf{S}$, where

$$\mathbf{S} = \mathbf{A}\boldsymbol{\mu}\mathbf{C}'+\mathbf{BCC}'+\mathbf{C}\boldsymbol{\mu}'\mathbf{A}'+\mathbf{CC}'\mathbf{B}'. \tag{3.8}$$

The expression (3.7) can now be written as

$$\mathbf{V}(t) = \mathbf{AV}(t-1)\,\mathbf{A}'+\mathbf{BV}(t-1)\,\mathbf{B}'+\Delta_1, \tag{3.9}$$

where

$$\Delta_1 = \mathbf{ASB}'+\mathbf{BSA}'+2\mathbf{BCC}'\mathbf{B}'+\mathbf{CC}'.$$

To find the conditions under which, as $t \to \infty$, $\mathbf{V}(t)$ tends to, say, $\mathbf{V}$, where $\mathbf{V}$ does not depend on t, we proceed as follows.

Let $\mathbf{D}$, $\mathbf{E}$ and $\mathbf{F}$ be three square matrices, each of order $p \times p$. Let d_{ij} be the element corresponding to the ith row and jth column of the matrix $\mathbf{D}$. Let $\mathbf{D}_{.j}$ $(j = 1,2,\ldots,p)$ be the jth column of $\mathbf{D}$. Define

$$\operatorname{vec}(\mathbf{D}) = \begin{pmatrix} \mathbf{D}_{.1} \\ \mathbf{D}_{.2} \\ \vdots \\ \mathbf{D}_{.p} \end{pmatrix}, \tag{3.10}$$

and the Kronecker product $\mathbf{D} \otimes \mathbf{E}$, which is of order $p^2 \times p^2$, as $\mathbf{D} \otimes \mathbf{E} = (d_{ij}\mathbf{E})$. Then we have (Neudecker, 1969), $\operatorname{vec}(\mathbf{DEF}) = (\mathbf{F}' \otimes \mathbf{D})\operatorname{vec}(\mathbf{E})$, $\operatorname{vec}(\mathbf{DE}) = (\mathbf{I} \otimes \mathbf{D})\operatorname{vec}(\mathbf{E})$. Using the above notation, we can write (3.9) as

$$\operatorname{vec}\{\mathbf{V}(t)\} = [\mathbf{A} \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{B}]\operatorname{vec}\{\mathbf{V}(t-1)\} + \operatorname{vec}\boldsymbol{\Delta}_1. \tag{3.11}$$

This is a first-order difference equation in $\operatorname{vec}\{\mathbf{V}(t)\}$, and the solution of this equation can be written in power series of $(\mathbf{A} \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{B})$. *For the solution of* $\operatorname{vec}\{\mathbf{V}(t)\}$ *to converge, it is sufficient that*

$$\rho[\mathbf{A} \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{B}] < 1. \tag{3.12}$$

This is the sufficient condition for the time series $\mathbf{x}(t)$ generated from (2.3) to be asymptotically second-order stationary. The condition (3.12) becomes

$$\rho[\mathbf{A} \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{B}\sigma_e^2] < 1 \quad \text{if } \sigma_e^2 \neq 1.$$

Assuming the condition (3.12) is satisfied, we obtain the expression for the variance and covariance of $\{\mathbf{x}(t)\}$. Let $\mathbf{V} = E(\mathbf{x}(t)\mathbf{x}'(t))$, then we have from (3.9)

$$\mathbf{V} = \mathbf{AVA}' + \mathbf{BVB}' + \boldsymbol{\Delta}_1, \tag{3.13}$$

which can be solved explicitly since the equation (3.13) is linear in $\mathbf{V}$. Here we do not need the explicit solution. From the model (2.3) we have

$$\begin{aligned} E(\mathbf{x}(t+1)\mathbf{x}'(t)) &= \mathbf{A}\mathbf{E}(\mathbf{x}(t)\mathbf{x}'(t)) + \mathbf{B}E(\mathbf{x}(t)\mathbf{x}'(t)e(t)) \\ &= \mathbf{AV} + \mathbf{BS}, \end{aligned} \tag{3.14}$$

and for $s > 1$,

$$E(\mathbf{x}(t+s)\mathbf{x}'(t)) = \mathbf{A}^{s-1}E(\mathbf{x}(t+1)\mathbf{x}'(t)) + \left(\sum_{j=0}^{s-2} \mathbf{A}^j\mathbf{BC}\right)\boldsymbol{\mu}'. \tag{3.15}$$

Let $\mathbf{C}(s) = E(\mathbf{x}(t+s)-\boldsymbol{\mu})(\mathbf{x}(t)-\boldsymbol{\mu})'$, then we can show that

$$\mathbf{C}(0) = \mathbf{AC}(0)\mathbf{A}' + \mathbf{BC}(0)\mathbf{B}' + \boldsymbol{\Delta}_2, \tag{3.16}$$

$$\mathbf{C}(1) = \mathbf{AC}(0) + \boldsymbol{\Delta}_3, \tag{3.17}$$

$$\mathbf{C}(s) = \mathbf{AC}(s-1) = \mathbf{A}^{s-1}\mathbf{C}(1) \quad (s = 2,3,\ldots), \tag{3.18}$$

where

$$\begin{aligned} \boldsymbol{\Delta}_2 &= \mathbf{B}\boldsymbol{\mu\mu}'\mathbf{B}' + \mathbf{A}\boldsymbol{\mu\mu}'\mathbf{A}' + \mathbf{ASB}' + \mathbf{BS}'\mathbf{A}' + 2\mathbf{BCC}'\mathbf{A}' + \mathbf{CC}' - \boldsymbol{\mu\mu}', \\ \boldsymbol{\Delta}_3 &= \mathbf{A}\boldsymbol{\mu\mu}' + \mathbf{BS} - \boldsymbol{\mu\mu}'. \end{aligned}$$

If we now suppose $\mathbf{A}$ and $\mathbf{B}$ are of the form (2.2), we obtain from (3.18)

$$\gamma(s)+a_1\,\gamma(s-1)+\ldots+a_p\,\gamma(s-p)=0, \quad s>1, \tag{3.19}$$

where $\gamma(s)=\operatorname{cov}(X(t+s)\,X(t))$. These equations are the same as the Yule–Walker equations for an ARMA $(p,1)$ and thus show that the bilinear model BL $(p,0,p,1)$ has the same covariance structure as an ARMA $(p,1)$.

It is interesting to note that for a homogeneous bilinear system obtained from (2.3) by putting $\mathbf{C}=\mathbf{0}$, we have

$$\boldsymbol{\mu}=\mathbf{0}, \quad \mathbf{S}=\mathbf{0} \quad \text{and} \quad \mathbf{V}(t)=\mathbf{A}\mathbf{V}(t-1)\,\mathbf{A}'+\mathbf{B}\mathbf{V}(t-1)\,\mathbf{B}'.$$

If $\mathbf{V}(1)=0$ (and $\mathbf{C}=\mathbf{0}$), we have $\mathbf{V}(t)=0$, $t\geqslant 1$. Proceeding as above we can show that a homogeneous bilinear system degenerates into a deterministic system if $\rho(\mathbf{A}\otimes\mathbf{A}+\mathbf{B}\otimes\mathbf{B})<1$ as $t\to\infty$, and if $\rho(\mathbf{A}\otimes\mathbf{A}+\mathbf{B}\otimes\mathbf{B})>1$, the system explodes.

If we now consider the model BL $(1,0,1,1)$ given by

$$X(t)+a_1\,X(t-1)=b_{11}\,X(t-1)\,e(t-1)+e(t),$$

the sufficient condition for the second-order stationarity of the process $X(t)$ is that $a_1^2+b_{11}^2<1$, and expressions for the covariances can be obtained from (3.18). The condition for stationarity and the expressions for variance and covariances agree with the results of Granger and Andersen (1978a).

In this section we obtained the conditions for the existence of second-order moments for the bilinear process satisfying the VBL (p) model. It must be noted that higher order moments need not always exist (Granger and Andersen, 1978a, p. 40).

4. INVERTIBILITY OF THE VBL (p) MODEL

For a time series model to be useful for forecasting purposes, it is necessary that it should be invertible. The invertibility of linear time series models has been discussed by Box and Jenkins (1970). Recently Granger and Andersen (1978c) have provided another definition of invertibility which can be applied to both linear and non-linear time series models. Their definition is as follows. Let $X(t)$ be a discrete parameter time series satisfying the model

$$X(t)=f\{X(t-j),e(t-j),j=1,2,\ldots,P)\}+e(t), \tag{4.1}$$

where the $\{e(t)\}$ are independent random variables. The random variables $\{e(t)\}$ are not observable. Let $\hat{e}(t)$ be an "estimate" of $e(t)$, and let the initial values of $\hat{e}(t)$ be set equal to zero. The model (4.1) is said to be invertible if

$$\lim_{t\to\infty} E\{e(t)-\hat{e}(t)\}^2\to 0, \tag{4.2}$$

when the model and the parameters are known completely. In this section, using this definition, we obtain a sufficient condition for the invertibility of the VBL (p) model. The condition obtained by Granger and Andersen (1978c) for the BL $(1,0,1,1)$ model is a special case of our result.

Consider the VBL (p) model given by (2.3) and let $X(t)=\mathbf{H}'\mathbf{x}(t)$. Let $\hat{e}(t)$ be an estimate of $e(t)$ satisfying the difference equation

$$X(t)=\mathbf{H}'\mathbf{A}\mathbf{x}(t-1)+\mathbf{H}'\mathbf{B}\mathbf{x}(t-1)\,\hat{e}(t-1)+\mathbf{H}'\mathbf{C}\hat{e}(t). \tag{4.3}$$

From (2.3) and (4.3) we have

$$\mathbf{H}'\mathbf{C}\xi_1(t)=-\mathbf{H}'\mathbf{B}\mathbf{x}(t-1)\,\xi_1(t-1), \tag{4.4}$$

where $\xi_1(t)=e(t)-\hat{e}(t)$. Assuming the process $\mathbf{x}(t)$ to be ergodic and proceeding as in Granger and Andersen (1978c), we can show that

$$E\{\xi_1^2(t)\}\leqslant[E\{\zeta^2(s)\}]^t, \tag{4.5}$$

where $\zeta(s) = \{\mathbf{H}'\mathbf{B}\mathbf{x}(s)\}/\mathbf{H}'\mathbf{C}$. Taking limits on both sides of (4.5),

$$\lim_{t\to\infty} E\{\xi_1^2(t)\} \leqslant \lim_{t\to\infty} [E\{\zeta^2(s)\}]^t. \tag{4.6}$$

The right-hand term of the inequality tends to zero as $t \to \infty$ if $E\{\zeta^2(s)\} < 1$, which implies that

$$\mathbf{H}'\mathbf{B}E\{\mathbf{x}(s)\,\mathbf{x}'(s)\}\,\mathbf{B}'\mathbf{H} < (\mathbf{H}'\mathbf{C})^2. \tag{4.7}$$

For a given **A**, **B**, **H** and **C**, one can evaluate explicitly $\mathbf{V} = E\{\mathbf{x}(s)\,\mathbf{x}'(s)\}$ which satisfies the equation (3.13). The condition (4.7) is a sufficient condition for the invertibility of the VBL(p) model. The condition given by Granger and Andersen (1978a, p. 74) for the bilinear model $X(t) = b_{11}\,X(t-1)\,e(t-1)+e(t)$ can be deduced from (4.7) by appropriate substitutions.

The conditions for the stationarity and invertibility of the model BL$(p,0,p,q)$ are not considered in this paper.

5. Estimation of the Parameters of the Bilinear Time Series Model

We now consider the estimation of the parameters of the scalar bilinear time series model given by

$$X(t)+a_1\,X(t-1)+\ldots+a_p X(t-p)+a_0 = \sum_{l=1}^{p}\sum_{j=1}^{q} b_{lj}\,X(t-l)\,e(t-j)+e(t), \tag{5.1}$$

where the $\{e(t)\}$ are independent and each $e(t)$ is distributed $N(0,\sigma_e^2)$. Here we assume the model (5.1) is invertible, and further assume we have a realization $\{X(1), X(2), \ldots, X(n)\}$ on the time series $\{X(t)\}$. The joint density function of $\{e(m), e(m+1), \ldots, e(n)\}$, where $m = \max(p,q)+1$, is given by

$$\frac{1}{(2\pi\sigma_e^2)^{(n-m+1)/2}}\exp\left[-\frac{1}{2\sigma_e^2}\sum_{m}^{n} e^2(t)\right].$$

Since the Jacobian of the transformation from $\{e(m), e(m+1), \ldots, e(n)\}$ to $\{X(m), X(m+1), \ldots, X(n)\}$ is unity, the likelihood function of $\{X(m), X(m+1), \ldots, X(n)\}$ is the same as the joint density function of $\{e(m), e(m+1), \ldots, e(n)\}$. Maximizing the likelihood function is the same as minimizing the function $Q(\boldsymbol{\theta})$, where

$$Q(\boldsymbol{\theta}) = \sum_{t=m}^{n} e^2(t), \tag{5.2}$$

with respect to the parameters $\boldsymbol{\theta}' = (a_0, a_1, \ldots, a_p;\, b_{11}, \ldots, b_{pq})$. For convenience, we shall write $\theta_i = a_0$, $\theta_2 = a_1, \ldots, \theta_R = b_{pq}$ where $R = 1+p+pq$.

Then the partial derivatives of $Q(\boldsymbol{\theta})$ are given by

$$\begin{aligned} \frac{dQ(\boldsymbol{\theta})}{d\theta_i} &= 2\sum_{t=m}^{n} e(t)\frac{de(t)}{d\theta_i} \quad (i = 1,2,\ldots,R),\\ \frac{d^2 Q(\boldsymbol{\theta})}{d\theta_i\, d\theta_j} &= 2\sum_{t=m}^{n}\frac{de(t)}{d\theta_i}\,\frac{de(t)}{d\theta_j}+2\sum_{t=m}^{n} e(t)\frac{d^2 e(t)}{d\theta_i\, d\theta_j}, \end{aligned} \tag{5.3}$$

where these partial derivatives of $e(t)$ satisfy the recursive equations

$$\frac{de(t)}{da_i}+\sum_{j=1}^{q}\beta_j(t)\frac{de(t-j)}{da_i} = \begin{cases} 1, & \text{if } i = 0,\\ X(t-i), & \text{if } i = 1,2,\ldots,p, \end{cases} \tag{5.4}$$

$$\frac{de(t)}{db_{km_1}}+\sum_{j=1}^{q}\beta_j(t)\frac{de(t-j)}{db_{km_1}} = -X(t-k)\,e(t-m_1) \quad (k = 1,2,\ldots,p;\ m_1 = 1,2,\ldots,q), \tag{5.5}$$

$$\frac{d^2 e(t)}{da_i\, da_{i'}} + \sum_{j=1}^{q} \beta_j(t) \frac{d^2 e(t-j)}{da_i\, da_{i'}} = 0 \quad (i, i' = 0,1,2,\ldots,p),$$

$$\frac{d^2 e(t)}{da_i\, db_{km_1}} + \sum_{j=1}^{q} \beta_j(t) \frac{d^2 e(t-j)}{db_{km_1}\, da_i} + X(t-k) \frac{de(t-m_1)}{da_i} = 0$$

$$(i = 0,1,2,\ldots,p;\ k = 1,2,\ldots,p;\ m_1 = 1,2,\ldots,q),$$

$$\frac{d^2 e(t)}{db_{km_1}\, db_{k'm_1'}} + \sum_{j=1}^{q} \beta_j(t) \frac{d^2 e(t-j)}{db_{km_1}\, db_{k'm_1'}} + X(t-k') \frac{de(t-m')}{db_{km_1}} = -X(t-k) \frac{de(t-m)}{db_{k'm_1'}}$$

$$(k, k' = 1,2,\ldots,p;\ m_1, m_1' = 1,2,\ldots,q), \tag{5.6}$$

$$\beta_j(t) = \sum_{l=1}^{p} b_{lj} X(t-l).$$

We assume $e(t) = 0$ $(t = 1,2,\ldots,m-1)$ and also

$$\frac{de(t)}{d\theta_i} = 0, \quad \frac{d^2 e(t)}{d\theta_i\, d\theta_j} = 0 \quad (i, j = 1,2,\ldots,R;\ t = 1,2,\ldots,m-1).$$

From these assumptions and equation (5.5) it follows that the second-order derivatives with respect to a_i $(i = 0,1,2,\ldots,p)$ are zero. For a given set of values of $\{a_i\}$ and $\{b_{lj}\}$ one can evaluate the first- and second-order derivatives using the recursive equations (5.4), (5.5) and (5.6). Now let

$$\mathbf{G}'(\boldsymbol{\theta}) = \left[\frac{dQ(\boldsymbol{\theta})}{d\theta_1}, \frac{dQ(\boldsymbol{\theta})}{d\theta_2}, \ldots, \frac{dQ(\boldsymbol{\theta})}{d\theta_R}\right],$$

and let $\mathbf{H}(\boldsymbol{\theta}) = [d^2 Q(\boldsymbol{\theta})/d\theta_i\, d\theta_j]$ be a matrix of second-order partial derivatives. Expanding $\mathbf{G}(\hat{\boldsymbol{\theta}})$, near $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$ in a Taylor series, we obtain

$$[\mathbf{G}(\hat{\boldsymbol{\theta}})]_{\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}} = \mathbf{0} = \mathbf{G}(\boldsymbol{\theta}) + \mathbf{H}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}).$$

Rewriting this equation we get $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = -\mathbf{H}^{-1}(\boldsymbol{\theta})\,\mathbf{G}(\theta)$, and thus obtain the Newton–Raphson iterative equation

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \mathbf{H}^{-1}(\boldsymbol{\theta}^{(k)})\,\mathbf{G}(\boldsymbol{\theta}^{(k)}), \tag{5.7}$$

where $\boldsymbol{\theta}^{(k)}$ is the set of estimates obtained at the kth stage of iteration. The estimates obtained by the above iterative equations usually converge, although, as is well known, they may not correspond to the global minimum of $Q(\boldsymbol{\theta})$.

If L denotes the likelihood function of $(X(m), X(m+1), \ldots, X(n))$, then we have, approximately,

$$\frac{1}{n} \frac{d^2 \log L}{d\boldsymbol{\theta}\, d\boldsymbol{\theta}'} = -\frac{1}{2\sigma_e^2} \frac{1}{n} \frac{d^2 Q(\boldsymbol{\theta})}{d\boldsymbol{\theta}\, d\boldsymbol{\theta}'}.$$

Let

$$-\frac{1}{n} \frac{d^2 \log L}{d\boldsymbol{\theta}\, d\boldsymbol{\theta}'}$$

converge (element-wise) stochastically to $\mathbf{I}(\boldsymbol{\theta})$, where $\mathbf{I}(\boldsymbol{\theta})$ is Fisher's information matrix. Then as $n \to \infty$,

$$\frac{1}{n} \frac{dQ(\boldsymbol{\theta})}{d\boldsymbol{\theta}\, d\boldsymbol{\theta}'}$$

converges stochastically to $2\sigma_e^2\, \mathbf{I}(\boldsymbol{\theta})$. Further as $n \to \infty$, it can be shown that $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ has approximately a multivariate normal distribution with mean $\mathbf{0}$ and variance covariance matrix equal to $\mathbf{I}^{-1}(\boldsymbol{\theta})$.

Remarks. The model (5.1) is different from the model BL $(p, 0, p, q)$ in the sense that (5.1) has an extra parameter a_0 which only affects the mean. It is found useful to include a constant a_0 when bilinear models are fitted to raw time series data.

5.1. *Initial Estimates*

To obtain a good set of estimates it is necessary that we should have a good set of initial values to start the iteration. We have tried two methods of estimating the initial values. The first method consists of fitting an AR (p) model (with constant a_0 present) and taking these coefficients as initial values for the autoregressive part of the bilinear model BL $(p, 0, p, q)$ and setting $b_{lj} = 0$ $(l = 1, 2, ..., p; j = 1, 2, ..., q)$. The second method is as follows (with constant a_0 included):

When a bilinear model of order BL $(p, 0, p, q-1)$ is fitted, we take for the initial values the estimates obtained from BL $(p, 0, p, q-1)$ or BL $(p-1, 0, p-1, q)$, which ever has the smallest residual sum of squares. If for the initial values the coefficients from the model BL $(p, 0, p, q-1)$ is chosen, the rest of the p coefficients are put equal to zero. Similarly if the model BL $(p-1, 0, p-1, q)$ is chosen, the rest of the $q+1$ coefficients are set equal to zero.

5.2. *Marquardt Algorithm*

An alternative to the Newton–Raphson technique is to use the Marquardt algorithm where only the first-order derivatives are necessary. For fitting the bilinear models, both these procedures are used, although in this paper only the results obtained from the Newton–Raphson technique are reported.

6. NUMERICAL ILLUSTRATIONS

Example 1. For our first illustration we consider the well-known Wolfer sunspot numbers for the years 1700–1955 (Walmeirer, 1961). Several linear and non-linear time series models have been fitted to these data. For example, to the sunspot numbers measured during the years 1770–1869, Box and Jenkins (1970) have fitted an autoregressive model of order 2, and to the residuals obtained from this AR (2) model Granger and Andersen (1978a, p. 86) have fitted a bilinear model BL $(1, 0, 1, 1)$ and found that there is 13·5 per cent reduction in the mean sum of squares of residuals. A recent statistical test of Subba Rao and Gabr (1981) has shown that this time series is highly non-linear. Recently, Tong and Lim (1980) have fitted threshold autoregressive models to the same data.

We now consider the choice of the order of the bilinear model. Using the estimation procedure described in the earlier section, the bilinear models of all orders up to BL $(5, 0, 4, 4)$ are fitted. The choice of the order is made on the basis of the information criterion of Akaike (1977), which is given by

$$\mathrm{AIC} = (N-M) \log \hat{\sigma}_e^2 + 2 \text{ (independent number of parameters)}, \tag{6.1}$$

where

$$\hat{\sigma}_e^2 = \frac{1}{N-M} \sum_{t=M+1}^{N} \hat{e}^2(t),$$

and $(N-M)$ is the number of observations used for calculating the likelihood function. For comparison of AIC values it is necessary that the likelihood function should be calculated over the same length of data for models of all orders. In view of this, for the sunspot numbers, we have omitted the first eight observations, $(M = 8)$, and the estimation is done for the next 238 observations. It is found that AIC is minimum when $p = 3$ and $q = 4$. The estimated values of the coefficients of the model are as follows: $\hat{a}_1 = -1{\cdot}93$, $\hat{a}_2 = 1{\cdot}46$, $\hat{a}_3 = -0{\cdot}27$, $\hat{a}_0 = -10{\cdot}9132$, and the values of $\hat{b}_{lj}$ $(l = 1, 2, 3; j = 1, 2, 3, 4)$ are given below in matrix form.

$$(\hat{b}_{lj}) = \begin{bmatrix} -0{\cdot}55 \times 10^{-2} & 0{\cdot}32 \times 10^{-2} & -0{\cdot}18 \times 10^{-2} & 0{\cdot}08 \times 10^{-2} \\ -0{\cdot}57 \times 10^{-2} & -0{\cdot}56 \times 10^{-2} & -0{\cdot}82 \times 10^{-2} & 0{\cdot}58 \times 10^{-2} \\ -0{\cdot}17 \times 10^{-2} & 0{\cdot}71 \times 10^{-2} & 0{\cdot}11 \times 10^{-1} & -0{\cdot}78 \times 10^{-3} \end{bmatrix}.$$

The values of $\hat{\sigma}_e^2$ and AIC are 143·86 and 1214·58 respectively.

The initial values of the parameters were obtained using the first method described in the previous section.

Forecasting of sunspot numbers

The performance of a time series model is judged on the basis of its forecasting performance. So it is natural to compare the forecasting performance of the fitted bilinear model with that of a fitted linear model.

We make use of the first 246 observations of the sunspot numbers for fitting the model, and then compare the forecasts with the true values for the next ten observations. The fitted linear and bilinear models are as follows. (The first ten observations are omitted, $M = 10$.)

AR *model*

The sample mean of the 246 observations is subtracted from each observation and an autoregressive model is fitted to those deviations. The best AR model is found to be AR (9) and the fitted model is

$$\begin{aligned} &X(t)-1{\cdot}224X(t-1)+0{\cdot}488X(t-2)+0{\cdot}124X(t-3)-0{\cdot}166X(t-4)\\ &\quad+0{\cdot}150X(t-5)-0{\cdot}039X(t-6)+0{\cdot}036X(t-7)+0{\cdot}069X(t-8)\\ &\quad-0{\cdot}111X(t-9)=e(t). \end{aligned} \tag{6.2}$$

The mean sum of squares of residuals is 185·82 and the AIC value is 1253·053.

Bilinear time series model

As some coefficients of the full bilinear model fitted are small when compared to the other coefficients, it is natural to see if a reduction in mean sum of squares of residuals and AIC can be achieved by fitting a subset bilinear model. An algorithm for fitting a subset bilinear model has been developed by Mr M. M. Gabr and the author, and details will be reported elsewhere. The fitted bilinear model is

$$\begin{aligned} &X(t)-1{\cdot}209X(t-1)+0{\cdot}502X(t-2)-0{\cdot}173X(t-9)\\ &\quad=5{\cdot}891-0{\cdot}0098X(t-2)\,e(t-1)+0{\cdot}0103X(t-8)\,e(t-1)\\ &\qquad-0{\cdot}0048X(t-8)\,e(t-3)+0{\cdot}0016X(t-3)\,e(t-2)\\ &\qquad+0{\cdot}0014X(t-4)\,e(t-7)+e(t). \end{aligned} \tag{6.3}$$

The mean sum of squares of residuals is 141·18 and the AIC value is 1186·2.

As a further test we obtained the one-step-ahead forecasts for the two model (6·2) and (6·3). Suppose $\{X(t)\}$ is a discrete parameter time series, and we wish to predict $X(t_0+m)$ given the semi-infinite realization $\{X(s), -\infty < s \leqslant t_0\}$. Let this predictor be $\tilde{X}_{t_0}(m)$. Then it is well known that $E(X(t_0+m)-\tilde{X}_{t_0}(m))^2$ is minimum if and only if

$$\tilde{X}_{t_0}(m) = E(X(t_0+m)\,|\,X(s),\ s \leqslant t_0).$$

The values of $\tilde{X}_{t_0}(1)$ from the models (6·2) and (6·3) are given in Table 1.

The mean sum of squares of one-step-ahead forecast errors for the ten values for the bilinear model is 165·126 and for the linear model is 484·394. The reduction in the mean sum of squares of errors is very substantial.

Bilinear models have been fitted to other types of non-linear time series, and details will be discussed elsewhere. For a brief summary, see Subba Rao (1980).

Example 2. For our second illustration we consider a seismic record obtained from an underground nuclear explosion that was carried out in the USA on October 29th, 1966. The record is that of a P-wave (pressure wave) and this event is nowadays commonly known as

TABLE 1
One-step-ahead forecasts of sunspot numbers

t	247	248	249	250	251	252	253	254	255	256
True values $X(t)$	92·6	151·6	136·3	134·7	83·9	69·4	31·5	13·9	4·4	38·0
$\tilde{X}_t(1)$ from (9·2)	59·8	120·0	157·7	104·1	105·5	45·2	40·4	10·4	4·9	22·9
$\tilde{X}_t(1)$ from (9·3)	77·9	130·0	149·8	119·8	86·2	51·4	38·9	18·8	3·3	25·7

"Longshot". The digitized record is obtained by sampling at the rate of 16 observations per second, giving altogether 512 observations. The graph of the record is given in Fig. 1. Dargahi-Noubary, Laycock and Subba Rao (1978) have fitted a linear AR model to these data.

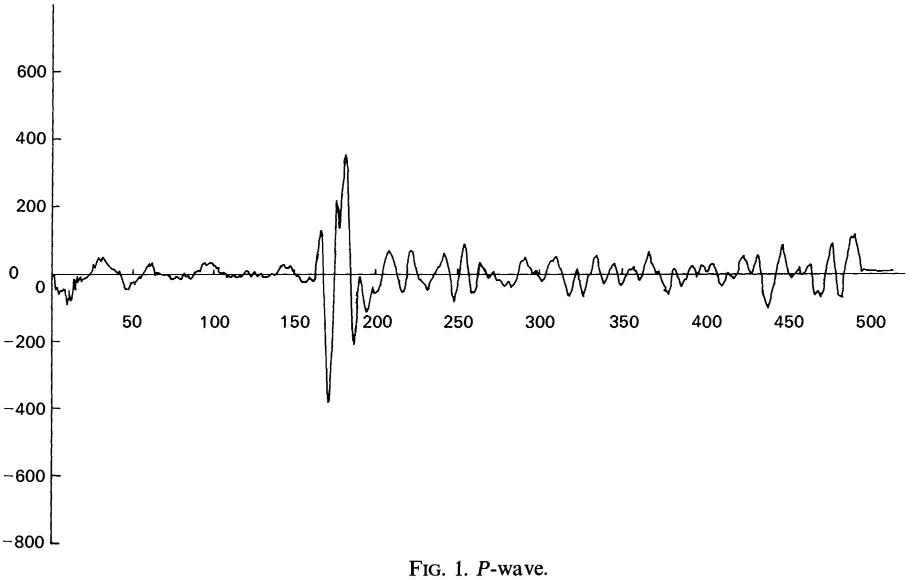

FIG. 1. *P*-wave.

The test proposed by Subba Rao and Gabr (1981) shows that this series is highly non-linear, suggesting that a non-linear model may be more appropriate.

The best linear AR model is found to be AR(3), and the model fitted is

$$X(t)-2{\cdot}002X(t-1)+1{\cdot}443X(t-2)-0{\cdot}300X(t-3)-0{\cdot}088 = e(t). \tag{6.4}$$

The values of $\hat{\sigma}_e^2$ and AIC are respectively 79·7 and 2227·5.

The best bilinear time series model is found to be BL(6, 0, 5, 1) and the fitted model is

$$\begin{aligned} &X(t)-1{\cdot}983X(t-1)+1{\cdot}215X(t-2)+0{\cdot}300X(t-3)-0{\cdot}619X(t-4)\\ &+0{\cdot}246X(t-5)+0{\cdot}266\\ &=\{0{\cdot}042X(t-1)-0{\cdot}044X(t-2)-0{\cdot}018X(t-3)+0{\cdot}039X(t-4)\\ &-0{\cdot}018X(t-5)\}\,e(t-1)+e(t). \end{aligned} \tag{6.5}$$

The values of $\hat{\sigma}_e^2$ and AIC are respectively 63·2 and 2124·5. These two values are smaller than the corresponding values obtained from the linear model (6.4). The fitted bilinear model has more parameters than the linear model, however, this is compensated by the amount of reduction in the mean residual sum of squares and AIC value.

7. CONCLUSIONS

Bilinear models have been developed by control engineers to describe the input–output relationship of a deterministic non-linear system. The advantage of this type of model is that it is a finite parameter model which can approximate to a reasonable accuracy the general Volterra series expansion. The validity of this approximation for stochastic time series data are under investigation.

Granger and Andersen (1978a) have studied the properties of a simple bilinear time series model and thus have made a beginning in discussing the usefulness of these models in time series. In this paper the results have been extended to cover more general bilinear models. The theory of estimation of the parameters has also been considered and has been applied to sunspot numbers and seismological data.

So far it has been argued in time series literature that if the time series under consideration is non-Gaussian a transformation, such as logarithmic transformation, may make the time series Gaussian. If one wishes to obtain forecasts of the time series, these can be obtained from the transformed data. As pointed out by Granger and Andersen (1978b, p. 31), these forecasts are biased and lead to a higher mean square error. In fact, the author (Subba Roa, 1980) has shown that the forecasting performance of the bilinear model fitted to the original Canadian lynx data (see Campbell and Walker, 1977) is better than the linear and non-linear models fitted to the transformed Canadian lynx data. This shows that if a time series is found to be non-Gaussian, one should consider fitting a non-linear time series model to the original data. One such non-linear time series model is the bilinear time series model. There are many theoretical problems still to be solved, which the author hopes to solve in due course.

ACKNOWLEDGEMENTS

I am grateful to Professor M. B. Priestley for going through the manuscript and making many helpful suggestions. The author is also grateful to the referees and editors whose comments led to substantial improvement in the presentation. The derivation given in Section 3 for obtaining the expressions for variances and covariances is along the lines suggested by one of the referees, and this is an alternative to the one given originally by the author. Many thanks to Dr V. Haggan and Mr M. M. Gabr for computational help and for drawing the pictures.

REFERENCES

AKAIKE, H. (1974). Markovian representation of stochastic processes and its application to the analysis of autoregressive-moving average processes. *Ann. Inst. Statist. Math.*, **26**, 363–383.

—— (1977). On entropy maximisation principle. *Proc. Symp. on Applications of Statistics, Dayton, Ohio, June 1977* (P. R. Krishnaiah, ed.).

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* New York and London: Wiley.

BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series, Forecasting and Control.* San Francisco: Holden-Day.

CAMPBELL, M. J. and WALKER, A. M. (1977). A survey of statistical work on the Mackenzie River series of annual Canadian lynx trappings for the years 1821–1934, and a new analysis. *J. R. Statist. Soc.* A, **140**, 411–431.

Dargari-Noubary, G. R., Laycock, P. J. and Subba Rao, T. (1978). Non-linear stochastic models for seismic events with applications in event identification. *Geophys. J. Roy. Astro. Soc.*, **55**, 655–668.

GRANGER, C. W. J. and ANDERSEN, A. P. (1978a). *Introduction to Bilinear Time Series Models.* Göttingen: Vandenhoeck & Ruprecht.

—— (1978b). Non-linear time series modelling. In *Applied Time Series Analysis* (D. F. Findley, ed.). New York: Academic Press.

—— (1978c). On the invertibility of time series models. *Stoch. Processes and their Applic.*, **8**, 83–92.

HAGGAN, V. and OZAKI, T. (1980). Amplitude-dependent exponential AR model fitting for non-linear random vibrations. *Proc. International Time Series Meeting, Nottingham, U.K., March 1979* (O. D. Anderson, ed.).

HANNAN, E. J. (1962). *Time Series Analysis.* London: Methuen Monographs.

—— (1970). *Multiple Time Series.* New York and London: Wiley.

JONES, D. A. (1978). Non-linear autoregressive processes. *Proc. Roy. Soc. London* (A), **360**, 71–95.

MOHLER, R. R. (1973). *Bilinear Control Processes.* New York: Academic Press.

NEUDECKER, H. (1969). Some theorems on matrix differentiation with special references to Kronecker matrix products. *J. Amer. Statist. Ass.*, **69**, 953–963.

OZAKI, T. and ODA, H. (1977). Non-linear time series identification by AIC. *Prof. IFAC Workshop.* Oxford: Pergamon.

PRIESTLEY, M. B. (1978). Non-linear models in time series analysis. *The Statistician*, **27**, 159–176.

—— (1980). System identification, Kalman filtering and stochastic control. In *Directions in Time Series.* Institute of Mathematical Statistics, U.S.A.

RUBERTI, A., ISIDORI, A. and d'ALLESANDRO, P. (1972). *Theory of Bilinear Dynamical Systems.* Berlin: Springer-Verlag.

SUBBA RAO, T. (1980). In discussion of H. Tong and K. S. Lim. *J. R. Statist. Soc.* B, **42**, 245–292.

SUBBA RAO, T. and GABR, M. M. (1981). A test for linearity of stationary time series. *Jour. Time Series Analysis*, Vol. 1, No. 2.

TONG, H. and LIM, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *J. R. Statist. Soc.* B, **42**, 245–292.

VOLTERRA, V. (1930). *Theory of Functionals and an Integral and Integrodifferential Equations.* London: Dover.

WALDMEIRER, M. (1961). *The Sunspot Activity in the years 1610–1960.* Schulthese.

WIENER, N. (1958). *Non-linear Problems in Random Theory.* London: Wiley.

WILKINSON, J. H. (1965). *The Algebraic Eigenvalue Problem.* Oxford: Clarendon.