Jordan Journal of Mathematics and Statistics (JJMS) 3(1), 2010, pp. 1-20

SUBSETTING AND IDENTIFICATION OF OPTIMAL MODELS IN GENERALIZED BILINEAR TIME SERIES MODELLING *

J. F. OJO AND D. K. SHANGODOYIN

ABSTRACT: Significant efforts have been made to study the theory of bilinear time series models, especially simple bilinear (BL) models. Much less efforts, however, have been made to identify optimal models in generalized bilinear models. Focus on optimal model identification; this study attempts to fill this gap. Full and subset generalized bilinear (SGBL) models are proposed and shown to be robust in achieving stationarity for all non-linear series. The parameters of the proposed models are estimated using robust nonlinear least square method and Newton-Raphson iterative method, and statistical properties of the derived estimates are investigated. An algorithm is proposed to eliminate redundant parameters from full order generalized bilinear models..

1. Introduction

Building probability models for time series data is an important activity that enables a statistician to understand the underlying random mechanisms generating the series. Better still, it provides invaluable assistance in forecasting the future. Linear time series, such as the autoregressive (AR) models, have been widely and successfully used in many fields. This is mainly because these models can be easily analyzed and provide fairly good approximations of the underlying random mechanisms of numerous real-life time series.

Nevertheless, in some situations linear time series models may be insufficient in explaining the underlying random mechanisms. This is, for instance, the case with sunspot data and the Canadian lynx data set. Linear time series models cannot adequately describe them, and the test proposed by Subba Rao and Gabr (1980) does confirm that linear Gaussian models fail to describe the above series. Thus a natural alternative that suggests itself is nonlinear models. Undoubtedly, the nonlinear time

2000 *Mathematics Subject Classification:* Primary: 37M10, Secondary: 46N30

Keywords: Non-linear Least Square, Newton-Raphson, Algorithm, Parameters and Stationarity.

Received on: March, 18, 2009 Accepted on: Feb. 8, 2010

series models are more complex than linear ones for several reasons. These are difficult parameter estimation of these models; intricate studying of statistical properties of most nonlinear models and sampling distribution of the estimates; and lastly, difficult evaluation of optimal forecasts for several steps in the future from these models. Yet despite these problems it seems reasonable to expect that in many situations nonlinear time series model should work better than a linear time series one.

Special nonlinear models considered by Granger and Andersen (1978) and Subba Rao (1981) are known as bilinear (BL) time series models. Providing a good fit, this class of time series has been found useful in many areas of biological sciences, ecology and engineering (e.g., Bruni et al. 1974). Thus many researchers have studied various bilinear models (e.g., Pham and Tran 1981, Gabr and Subba Rao 1981, Rao et al. 1983, Liu 1992, Cathy 1997, Gonclaves et al. 2000, Shangodoyin and Ojo 2003, Wang and Wei 2004, Boonchai and Eivind 2005, Bibi 2006, Doukhan et al. 2006, Drost et al. 2007, Usoro and Omekara 2008). This wide use and usefulness notwithstanding, optimal models have not yet been identified for this class of models. Another problem is that various models from this class could not achieve stationarity for all nonlinear series. Rao et al. (1983) gave a set of sufficient conditions for the existence of a strictly stationary stochastic process conforming to the following bilinear model:

$$X_t = \sum_{i=1}^{p} a_j X_{t-j} + \sum_{i=1}^{p}\sum_{j=1}^{q} b_{ij} X_{t-i} e_{t-j} + e_t$$, denoted as BL(p, 0, p, q)

where p is the order of the autoregressive component, and p, q is the order of the nonlinear component. $a_1, a_2, \ldots, a_p$ are the parameters of the autoregressive component and $b_{11}, b_{22}, \ldots, b_{pq}$ are the parameters of the nonlinear component

In this paper, we extend the work of Rao et al. (1983) to the proposed generalized bilinear models which are capable of achieving stationarity for all nonlinear series; this is an important improvement over other bilinear time series models. Many models also characterize the proposed models, as we shall see in section 2.

In addition, bilinear time series are characterized by too many parameters, some of which are close to zero. In the proposed models, we address this problem by employing the concept of subsetting. Subsetting helps remove these redundant parameters, thereby leading to so-called subset bilinear models. Gabr and Subba Rao (1981) worked on subset bilinear models and tested all the subsets of the best order of the full bilinear model before selecting the best subset. In this paper, subsetting concept is introduced to the proposed generalized bilinear model to determine its usefulness in achieving a better model.

2. Proposed generalized bilinear time series models

We define generalized bilinear (BL) time series models as follows:

Model 1(M1)

$\psi(B)X_t = \phi(B)\nabla^d X_t + \sum_{k=1}^{r}\sum_{l=1}^{s} b_{kl} X_{t-k} e_{t-l} + e_t$, denoted as BL (p, d, 0, r, s)

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 ..... - \phi_p B^p$ and

$$X_t = \psi_1 X_{t-1} + ... + \psi_{p+d} X_{t-p-d} + b_{11} X_{t-1} e_{t-1} + ... + b_{rs} X_{t-r} e_{t-s} + e_t \qquad (2.1)$$

$\phi_1,...,\phi_p$ are the parameters of the autoregressive component; $b_{11},.....,b_{rs}$ are the parameters of the nonlinear component and $\phi(B)$ is the autoregressive operator.

Model 2 (M 2)

$\psi(B)X_t = \phi(B)\nabla^d X_t + \theta(B)e_t + \sum_{k=1}^{r}\sum_{l=1}^{s} b_{kl} X_{t-k} e_{t-l}$, denoted as BL (p, d, q, r, s)

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 ..... - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 ..... - \theta_q B^q$ and

$$X_t = \psi_1 X_{t-1} + ... + \psi_{p+d} X_{t-p-d} + e_t - \theta_1 e_{t-1} - ..... - \theta_q e_{t-q} + b_{11} X_{t-1} e_{t-1} + ... + b_{rs} X_{t-r} e_{t-s} \qquad (2.2)$$

$\phi_1,...,\phi_p$ are the parameters of the autoregressive component; $\theta_1,...\theta_q$ are the parameters of the associated error process; $b_{11},.....,b_{rs}$ are the parameters of the non-linear component and $\theta(B)$ is the moving average operator.

The associated generalized subset bilinear models are denoted as SBL (p, d, 0, r, s) and SBL (p, d, q, r, s).They are represented by M3 and M4 respectively.

In the models above, *p* is the order of the autoregressive component; *q* is the order of the moving average process; *r*, *s* is the order of the nonlinear component and $\psi(B) = \nabla^d \phi(B)$ is the generalized autoregressive operator; ∇^d is the differencing operator and *d* is the degree of consecutive differencing required to achieve stationarity. e_t are independently and identically distributed as $N(0, \sigma_e^2)$ and the models are assume to be invertible.

In model 2, elements of $2^q - 1$ subsets when q = 3 were used to characterize the model as follows: M2: {F1= BL (p, 1, (1), r, s), F2= BL (p, 1, (1, 2) r, s) and F3= BL (p, 1, (1, 2, 3) r, s).The same thing was done for SBL (p, d, q, r, s). This is necessary

to see whether we shall achieve optimality in the full model other than testing all elements of the subsets. For M1, M2, M3 and M4 and at different levels of *t* optimal models are of great necessity especially in the forecasting of future values.

3. Stationarity and Convergence of Generalized Bilinear Models

In this section, we give a sufficient condition for the existence of strictly stationary process and convergence conforming to the bilinear model (2.1). This we do through the following theorem.

Theorem

Let $\{e_t, t \in Z\}$ be a sequence of independent identically distributed random variables defined on a probability space (Ω, F, P) such that E $e_t = 0$ and $Ee_t^2 = \sigma^2 < \infty$ .Let Ψ , $B_1, B_2, \ldots, B_q$ be q+1 matrices each of order p x p and

$$\Gamma_1 = \Psi \otimes \Psi + \sigma^2 (B_1 \otimes B_1),$$

$$\Gamma_j = \sigma^2 [B_j \otimes (\Psi^{j-1} B_1 + \Psi^{j-2} B_2 + ... + \Psi B_{j-1})$$

$$+ (\Psi^{j-1} B_1 + \Psi^{j-2} B_2 + ... + \Psi B_{j-1}) \otimes B_j$$

$$+ (B_j \otimes B_j)], \quad j = 2,3,\ldots\ldots s.$$

Suppose all the eigenvalues of the matrix

$$\underset{p^2q \times p^2q}{L} = \begin{pmatrix} \Gamma_1 & \Gamma_2 & ..... & \Gamma_{q-1} & \Gamma_q \\ I_{p^2} & 0 & ..... & 0 & 0 \\ 0 & I_{p^2} & & 0 & 0 \\ 0 & 0 & & I_{p^2} & 0 \end{pmatrix}$$

have moduli less than unity, i.e, $\rho(L) = \lambda < 1$. Let $\underset{px1}{C}$ be a given column vector. Then there exists a vector valued strictly stationary process $\{X_t, t \in Z\}$ conforming to the vector form of generalized bilinear model $\mathbf{X}_t = \Psi \mathbf{X}_{t-1} + \sum_{j=1}^{s} \mathbf{B}_j \mathbf{X}_{t-j} e_{t-j} + \mathbf{C} e_t$ for every t in Z.

Proof of theorem is given in the Appendix.

Description of Algorithm for Fitting Full and Subset Generalized Bilinear Models

For the sake of simplicity, we will break the algorithm down into the following steps:

Step 1
Fit various order of autoregressive model of the form
$X_t = \psi_1 X_{t-1} + ... + \psi_{p+d} X_{t-p-d} + e_t$

Step 2
Choose the model for which Akaike Information Criterion (AIC) is minimum among various order fitted in step 1.

Step 3
Fit possible subsets of chosen model in step 2 using $2^q - 1$ subsets approach Hagan and Oyetunji (1980).

Step 4
Choose the model for which AIC is minimum among the fitted models in step 3 to have the best subset model.

Step 5
Fit various order of the generalized bilinear model of the form $X_t = \psi_1 X_{t-1} + ... + \psi_{p+d} X_{t-p-d} + b_{11} X_{t-1} e_{t-1} + ... + b_{rs} X_{t-r} e_{t-s} + e_t$

and choose the model for which AIC is minimum

Step 6
Fit possible subsets of chosen model in step 5 using $2^q - 1$ subsets approach Shangodoyin and Ojo (2003).

Step 7
The model with the minimum AIC is the best subset generalized bilinear model.

Estimation of the Parameters of Generalized Bilinear Models Proposed

The joint density function of $(e_m, e_{m+1}, ..., e_n)$ where m = max (r, s) is given by

$$\frac{1}{(2\pi\sigma_e^2)^{(n-m+1)/2}} \exp\left(\frac{1}{-2\sigma_e^2} \sum_m^n e_t^2\right)$$

Since the Jacobian of transformation from $(e_m, e_{m+1},...., e_n)$ to $(X_m, X_{m+1},...., X_n)$ is unity, the likelihood function of $(X_m, X_{m+1},...., X_n)$ is the same as the joint density function of $(e_m, e_{m+1},...., e_n)$. Maximising the likelihood function is the same as minimizing the function $Q(G)$, where

$$Q(G) = \sum_{i=m}^{n} e_{t,}^2 \tag{3.1}$$

with respect to the parameter $G^{'} = (\psi_1,...., \psi_p; \theta_1, \theta_2,...., \theta_q; B_{11},...., B_{rs})$

Then the partial derivatives of Q(G) are given by

$$\frac{dQ(G)}{dG_i} = 2\sum_{t=m}^{n} e_t \frac{de_t}{dG_i} \qquad \text{(i = 1, 2,…..,R)} \tag{3.2}$$

$$\frac{d^2Q(G)}{dG_i dG_j} = 2(\sum_{t=m}^{n} e_t \frac{de_t}{dG_i}\frac{de_t}{dG_j} + \sum_{t=m}^{n} e_t \frac{d^2e_t}{dG_i dG_j})$$

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

$$\frac{de_t}{d\psi_i} + \sum_{j=1}^{s} W_t(t)\frac{de_{t-j}}{d\psi_i} = \begin{cases} 1, & \text{if i = 0} \\ X_{t-i}, & \text{if i = 1, 2, …, p} \end{cases} \tag{3.3}$$

$$\frac{de_t}{d\theta_i} + \sum_{j=1}^{s} W_t(t)\frac{de_{t-j}}{d\theta_i} = e_{t-i,} \qquad \text{if i = 1, 2,….., q} \tag{3.4}$$

$$\frac{de_t}{dB_{kmi}} + \sum_{j=1}^{s} W_j(t)\frac{de_{t-j}}{dB_{kmi}} = -X_{t-k}e_{t-m} \quad \text{(k=1,2,…,r ; } m_i = 1,2,\ldots,s) \tag{3.5}$$

$$\frac{d^2e_t}{d\psi_i d\psi_i^{'}} + \sum_{j=1}^{s} W_j(t)\frac{d^2e_{t-j}}{d\psi_i d\psi_i^{'}} = 0 \quad \text{(i, i' = 0, 1, 2, …, p)} \tag{3.6}$$

$$\frac{d^2e_t}{d\theta_i d\theta_i^{'}} + \sum_{j=1}^{s} W_j(t)\frac{d^2e_{t-j}}{d\theta_i d\theta_i^{'}} = 0 \qquad \text{(i, i' = 0, 1, 2, …, q)} \tag{3.7}$$

$$\frac{d^2e_t}{d\psi_i dB_{kmi}} + \sum_{j=1}^{s} W_j(t)\frac{d^2e_{t-j}}{dB_{kmi}d\phi_i} + X_{t-k}\frac{d^2e_{t-mi}}{d\psi_i} = 0$$

(i=0,1,2,…,p ; k_i =1,2,…,r; m_i=1,2,…,s) (3.8)

$$\frac{d^2e_t}{d\theta_i dB_{kmi}} + \sum_{j=1}^{s} W_j(t)\frac{d^2e_{t-j}}{dB_{kmi}d\theta_i} + X_{t-k}\frac{d^2e_{t-mi}}{d\theta_i} = 0$$

(i=1,2,…,q ; k_i =1,2,…,r; m_i=1,2,…,s) (3.9)

$$\frac{d^2e_t}{d\psi_i d\theta_i} + \sum_{j=1}^{s} W_j(t)\frac{d^2e_{t-j}}{d\psi_i d\theta_i} = 0 \qquad (3.10)$$

$$\frac{d^2e_t}{dB_{kmi}dB'_{kmi}} + \sum_{j=1}^{s} W_j(t)\frac{d^2e_{t-j}}{dB_{kmi}dB'_{kmi}} + X'_{t-k}\frac{d^2e_{t-mi}}{dB_{kmi}} = -X_{t-k}\frac{de_{t-m}}{dB'_{kmi}},$$

(k, k' =1,2,…,r ; m_i m_i' = 1,2,…,s) (3.11)

$$W_j(t) = \sum_{j=1}^{s} B_{ij}X_{t-j}$$

We assume $e_t = 0$ (t = 1, 2, …, m-1) and also

$$\frac{de_t}{dG_i} = 0, \frac{d^2e_t}{dG_i dG_j} = 0, \quad \frac{d^3e_t}{d^2G_i d^2G_j} = 0, \qquad \text{(i, j = 1, 2, …, R; t = 1, 2, …, m-1)}$$

From $e_t = 0$ (t = 1, 2, …, m-1), $\frac{de_t}{dG_i} = 0, \frac{d^2e_t}{dG_i dG_j} = 0$, $\frac{d^3e_t}{d^2G_i d^2G_j} = 0$, and

$\frac{de_t}{dB_{kmi}} + \sum_{j=1}^{s} W_j(t)\frac{de_{t-j}}{dB_{kmi}} = -X_{t-k}e_{t-m}$ (k=1,2,…,r ; m_i =1,2,…,s), it follows that the second order derivatives with respect to ψ_i (i = 0, 1, 2, …, p) and θ_i (i = 0, 1, 2, …, q) are zero. For a given set of values {ϕi}, { θ_i } and {B_{ij} } one can evaluate the first and second order derivatives using the recursive equations 3.3, 3.4, 3.5 and 3.11.
Now let:

$$\mathbf{V}'(\mathbf{G}) = \frac{dQ(\mathbf{G})}{d\mathbf{G}_1}, \frac{dQ(\mathbf{G})}{d\mathbf{G}_2}, ..., \frac{dQ(\mathbf{G})}{d\mathbf{G}_k}$$

and let $\mathbf{H}(\mathbf{G}) = [d^2Q(\mathbf{G})/d\mathbf{G}_i d\mathbf{G}_j]$ be a matrix of second partial derivatives as in Krzanowski (1998). Expanding $\mathbf{V}(\mathbf{G})$, near $G = \hat{G}$ in a Taylor series, we obtain

$$V(\hat{G})_{\hat{G}=G} = 0 = V(G) + H(G)(\hat{G} - G) \qquad (3.12)$$

Rewriting this equation we get $\hat{G} - G = -H^{-1}(G)V(G)$, and thus obtain an iterative equation given by $G^{(k+1)} = G^{(k)} - H^{-1}(G^{(k)})V(G^{(k)})$ where $G^{(k)}$ is the set of estimates obtained at the k^{th} stage of iteration. The estimates obtained by the above iterative equations usually converge. For starting the iteration, we need to have good

sets of initial values of the parameters. This is done by fitting the best subset of the linear part of the bilinear model.

Estimation of the Parameters of Generalized Subset Bilinear Model Proposed

In the estimation procedure to be discussed in this section we assume that the sets of integers $\{k_1, k_2, ..., k_l\}$ and $\{(r_1, s_1), (r_2, s_2), ..., (r_m s_m)\}$ are fixed and known. Proceeding as in Subba Rao (1981), we can show that maximizing the likelihood function of $(X_{m1}, X_{m1+1}, ..., X_N)$ is the same as minimizing the function $Q(\theta) = \sum_{t=m1}^{N} e_t^2$ with respect to the parameters $(\psi_{k_1}, \psi_{k_2}, ..., \psi_{k_l}; b_{r_1 s_1}, ..., b_{r_m s_m})$.

The partial derivatives of $Q(\theta)$ are

$$G_i = \frac{dQ(\theta)}{d\theta_i} = 2\sum_{t=m_1}^{N} e_t \frac{de_t}{d\theta_i},$$

$$h_{ij} = \frac{d^2 Q(\theta)}{d\theta_i d\theta_j} = 2\sum_{t=m_1}^{N}\left(\frac{de_t}{d\theta_i}\right)\left(\frac{de_t}{d\theta_j}\right) + 2\sum_{t=m_1}^{N} e_t . \frac{d^2 e_t}{d\theta_i d\theta_j},$$

where the partial derivatives satisfy the recursive equations

$$\frac{de_t}{d\psi_{k_r}} = X_{t-k_r} - \sum_{j=1}^{m} b_{r_j s_j} X_{t-r_j} \frac{de_{t-s_j}}{\psi_{k_r}}, (r = 1,2,3,..., l)$$

$$\frac{de_t}{db_{r_q s_q}} = -X_{t-r_q} e_{t-s_q} - \sum_{j=1}^{m} b_{r_j s_j} X_{t-r_j} \frac{de_{t-s_j}}{db_{r_q s_q}},$$

In the calculation of these partial derivatives, we set $e_1 = e_2 = ... = e_{mo} = 0$ and

$$\frac{de_1}{d\theta_i} = \frac{de_2}{d\theta_i} = ... = \frac{de_{mo}}{d\theta_i} = 0, (i = 1,2,..., R)$$

Let $G^T(\theta) = (G_1, G_2, ..., G_R)$ and $H(\theta) = (h_{ij})$.

In evaluating the second order partial derivatives we approximate

$$h_{ij} = 2\sum_{t=m_1}^{N}\left(\frac{de_t}{d\theta_i}\right)\left(\frac{de_t}{d\theta_i}\right)$$

as is done in Marquardt algorithm. Expanding $G(\hat{\theta})$ near $\hat{\theta} = \theta$ in a Taylor series, we obtain $0 = G(\theta) + H(\theta)(\hat{\theta} - \theta)$.

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

$$\theta^{(k+1)} = \theta^{(k)} - H^{-1}(\theta^{(k)})G(\theta^{(k)})$$

$$\theta^{(k)} = \theta^{(k+1)} + H^{-1}(\theta^{(k)})G(\theta^{(k)}) \qquad (3.13)$$

where $\theta^{(k)}$ is the set of estimates obtained at the k[th] stage of iteration. For starting the iteration, we need to have good sets of initial values of the parameters. This is done by fitting the best subset of the linear part of the bilinear model.

4. Numerical example: The Wolfer sunspot data

To present the application of the models proposed, we will use a real time series dataset, the Wolfer sunspot, available in Box et al. (1994). The scientists track solar cycles by counting sunspots – cool planet-sized areas on the Sun where intense magnetic loops poke through the star's visible surface. It was Rudolf Wolf who devised the basic formula for calculating sunspots in 1848; these sunspot counts are still continued.

As the Wolfer sunspot data set represent a non-stationary series, the bilinear models proposed in this paper may be applied. The Wolfer sunspot data set, available in Box et al. (1994), in this paper is considered at three levels, namely for *t* = 50, 150 and 250. For the fitted model below we have used the algorithm and the estimation technique in the previous section.

Fitted Model M1, M2, M3and M4 at t=50

M1

$X_t = 0.314548X_{t-1} - 0.458429X_{t-2} - 0.302114X_{t-4} - 0.220568X_{t-5} - 0.386159X_{t-6} - 0.002758X_{t-1}e_{t-1} - 0.020647X_{t-1}e_{t-2} - 0.018189X_{t-1}e_{t-3} + 0.015317X_{t-2}e_{t-1} + e_t$

M2; F1

$X_t = 0.919317X_{t-1} - 0.567504X_{t-2} - 0.319398X_{t-5} - 0.813860e_{t-1} + 0.009533X_{t-1}e_{t-1} - 0.009665X_{t-1}e_{t-2} + e_t$

M2:F2

$X_t = 0.441064X_{t-1} + 0.324866X_{t-2} - 0.513187X_{t-3} - 0.00648e_{t-1} - 0.954833e_{t-2} + 0.002296X_{t-1}e_{t-1} - 0.005373X_{t-1}e_{t-2} - 0.011546X_{t-1}e_{t-3} + e_t$

M2:F3

$X_t = 0.872827X_{t-1} - 0.006059X_{t-2} - 0.528694X_{t-4} - 0.126321X_{t-5} - 0.003146X_{t-6} - 0.964519e_{t-1} - 0.704614e_{t-2} + 0.884318e_{t-3} + 0.006249 X_{t-1}e_{t-1} - 0.010149X_{t-1}e_{t-2} + e_t$

M3

$X_t = 0.314548X_{t-1} - 0.458429X_{t-2} - 0.302114X_{t-4} - 0.220568X_{t-5} - 0.386159X_{t-6} - 0.020351X_{t-1}e_{t-2} - 0.016095X_{t-1}e_{t-3} + 0.013653X_{t-2}e_{t-1} + e_t$

M4:S1

$X_t = 0.919317X_{t-1} - 0.567504X_{t-2} - 0.319398X_{t-5} - 0.813860e_{t-1} + 0.009533X_{t-1}e_{t-1} - 0.009665X_{t-1}e_{t-2} + e_t$

M4:S2

$X_t = 0.441064X_{t-1} + 0.324866X_{t-2} - 0.513187X_{t-3} - 0.000348e_{t-1} - 0.954833e_{t-2} - 0.012641X_{t-1}e_{t-3} + e_t$

M4:S3

$X_t = 0.872827X_{t-1} - 0.006059X_{t-2} - 0.528694X_{t-3} + 0.069292X_{t-4} - 0.126321X_{t-5} - 0.003146X_{t-6} - 0.964519e_{t-1} - 0.704614e_{t-2} + 0.884318 e_{t-3} + 0.006249X_{t-1}e_{t-1} + 0.010149X_{t-1}e_{t-2} + e_t$

Fitted Model M1, M2, M3and M4 at t=150

M1

$X_t = 0.412820X_{t-1} - 0.271125X_{t-2} - 0.270908X_{t-3} - 0.339150X_{t-5} - 0.293320X_{t-7} + 0.000325X_{t-1}e_{t-1} - 0.020870X_{t-1}e_{t-2} - 0.002425X_{t-1}e_{t-3} + 0.018075X_{t-2}e_{t-1} + 0.009283X_{t-2}e_{t-2} - 0.008691X_{t-2}e_{t-3} - 0.019234X_{t-3}e_{t-1} - 0.007737X_{t-3}e_{t-2} + e_t$

M2:F1

$X_t = 0.905337X_{t-1} - 0.509454X_{t-2} - 0.107803X_{t-3} - 0.007407X_{t-4} - 0.215636X_{t-5} - 0.152380X_{t-7} - 0.588895e_{t-1} + 0.002609X_{t-1}e_{t-1} - 0.014157X_{t-1}e_{t-2} + 0.005213X_{t-1}e_{t-3} - 0.009930X_{t-2}e_{t-1} + 0.005389X_{t-2}e_{t-2} - 0.011513X_{t-2}e_{t-3} - 0.023710X_{t-3}e_{t-1} - 0.008035X_{t-3} e_{t-2} + e_t$

M2:F2

$X_t = 0.667154X_{t-1} - 0.299180X_{t-3} - 0.199718X_{t-5} - 0.330045e_{t-1} - 0.404780e_{t-2} + 0.001675X_{t-1}e_{t-1} - 0.016864X_{t-1}e_{t-2} + 0.004643X_{t-1}e_{t-3} + 0.013121X_{t-2}e_{t-1} + 0.006690X_{t-2}e_{t-2} - 0.010201X_{t-2}e_{t-3} - 0.024090X_{t-3}e_{t-1} - 0.008779X_{t-3}e_{t-2} + e_t$

M2:F3

$X_t = 0.217421X_{t-1} + 0.172224X_{t-3} - 0.518088X_{t-4} - 0.218600X_{t-5} - 0.135334X_{t-6} - 0.269434X_{t-7} + 0.630377e_{t-1} - 0.119139e_{t-2} - 0.763971e_{t-3} + 0.002651X_{t-1}e_{t-1} - 0.002651X_{t-1}e_{t-1} - 0.015220X_{t-1}e_{t-2} + 0.001332X_{t-1}e_{t-3} + 0.010671X_{t-2}e_{t-1} + 0.007194X_{t-2}e_{t-2} - 0.008443 X_{t-2}e_{t-3} - 0.018346X_{t-3}e_{t-1} - 0.007363X_{t-3}e_{t-2} + e_t$

M3

$X_t = 0.412820X_{t-1} - 0.271125X_{t-2} - 0.270908X_{t-3} - 0.339150X_{t-5} - 0.293320X_{t-7} - 0.021157X_{t-1}e_{t-2} + 0.018650X_{t-2}e_{t-1} + 0.009595X_{t-2}e_{t-2} - 0.009477X_{t-2}e_{t-3} - 0.021380X_{t-3}e_{t-1} - 0.008220X_{t-3}e_{t-2} + e_t$

M4:S1

$X_t = 0.905337X_{t-1} - 0.509454X_{t-2} - 0.107803X_{t-3} - 0.007407X_{t-4} - 0.215636X_{t-5} - 0.152380X_{t-7} - 0.588895e_{t-1} + 0.004558X_{t-1}e_{t-1} - 0.005999X_{t-1}e_{t-2} - 0.011735X_{t-2}e_{t-3} - 0.013501X_{t-3}e_{t-1} + e_t$

M4:S2

$X_t = 0.667154X_{t-1} - 0.299180X_{t-3} - 0.199718X_{t-5} - 0.330045e_{t-1} - 0.404780e_{t-2} - 0.018617X_{t-1}e_{t-2} + 0.015930X_{t-2}e_{t-1} + 0.007861X_{t-2}e_{t-2} - 0.008252X_{t-2}e_{t-3} - 0.022099X_{t-3}e_{t-1} - 0.008791X_{t-3}e_{t-2} + e_t$

M4:S3

X_t =0.217421X_{t-1} + 0.172224X_{t-3} – 0.518088X_{t-4} – 0.218600X_{t-5}- 0.135334X_{t-6} – 0.269434X_{t-7} + 0.630377e_{t-1} - 0.763971e_{t-3} – 0.017434 $X_{t-1}e_{t-2}$ + 0.014963$X_{t-2}e_{t-1}$ + 009280$X_{t-2}e_{t-2}$ - 0.007589$X_{t-2}e_{t-3}$ – 0.019788$X_{t-3}e_{t-1}$ - 0.008451$X_{t-3}e_{t-2}$ + e_t

Fitted Model M1, M2, M3and M4 at t=250

M1

X_t =- 0.239576X_{t-2} - 0.361665X_{t-3} – 0.238746X_{t-4} – 0.325416X_{t-5}- 0.328627X_{t-6} - 0.209789X_{t-7} - 0.365561X_{t-8} + 0.000633$X_{t-1}e_{t-1}$ - 0.010392$X_{t-1}e_{t-2}$ + 0.007590$X_{t-1}e_{t-3}$ + 0.005443$X_{t-2}e_{t-1}$ + 0.000716$X_{t-2}e_{t-2}$ - 0.005326$X_{t-2}e_{t-3}$ - 0.013130$X_{t-3}e_{t-1}$ + e_t

M2:F1

X_t = - 0.239520X_{t-2} - 0.361508X_{t-3} – 0.238305X_{t-4} – 0.325330X_{t-5}- 0.328543X_{t-6} - 0.210276X_{t-7} - 0.364294X_{t-8} + 0.053173e_{t-1} + 0.000401$X_{t-1}e_{t-1}$ - 0.010942$X_{t-1}e_{t-2}$ + 0.007079$X_{t-1}e_{t-3}$ + 0.005241$X_{t-2}e_{t-1}$ + 0.001086$X_{t-2}e_{t-2}$ - 0.005791$X_{t-2}e_{t-3}$ – 0.013665 $X_{t-3}e_{t-1}$ + e_t

M2:F2

X_t =- 0.049537X_{t-1} - 0.032572X_{t-2} – 0.413331X_{t-3} – 0.226474X_{t-4}- 0.285681X_{t-5} - 0.319814X_{t-6} - 0.183383X_{t-7} - 0.307067X_{t-8} + 0.0944981e_{t-1} - 0.237065e_{t-2} + 0.000278$X_{t-1}e_{t-1}$ - 0.010284$X_{t-2}e_{t-2}$ + 0.007103$X_{t-1}e_{t-3}$ + 0.005506$X_{t-2}e_{t-1}$ + 0.001455$X_{t-2}e_{t-2}$ – 0.005354$X_{t-2}e_{t-3}$ – 0.014110$X_{t-3}e_{t-1}$ + e_t

M2:F3

X_t = - 0.712478X_{t-1} - 0.153047X_{t-2} + 0.032479X_{t-3} – 0.606080X_{t-4}- 0.351330X_{t-5} - 0.422284X_{t-6} - 0.407042X_{t-7} - 0.311950X_{t-8} + 0.809607e_{t-1} - 0.048903e_{t-2} - 0.673588e_{t-3} + 0.000174$X_{t-1}e_{t-1}$ - 0.012392$X_{t-1}e_{t-2}$ - 0.000523 + 0.008372$X_{t-2}e_{t-1}$ + 0.002290$X_{t-2}e_{t-2}$ – 0.004130$X_{t-2}e_{t-3}$ – 0.010699$X_{t-3}e_{t-1}$ + e_t

M3

X_t = -0.239576X_{t-2} + 0.361665X_{t-3} – 0.238746X_{t-4} – 0.325416X_{t-5}- 0.328627X_{t-6} – 0.209789X_{t-7} - 0.365561X_{t-8} - 0.009014$X_{t-1}e_{t-2}$ + 0.006529$X_{t-1}e_{t-3}$ + 0.004016$X_{t-2}e_{t-1}$ – 0.005079$X_{t-2}e_{t-3}$ - 0.012555$X_{t-3}e_{t-1}$ + e_t

M4:S1

X_t =- 0.239520X_{t-2} - 0.361508X_{t-3} – 0.238305X_{t-4} – 0.325330X_{t-5}- 0.328543X_{t-6} – 0.210276X_{t-7} - 0.364294X_{t-8} + 0.053173e_{t-1} - 0.006248 $X_{t-1}e_{t-2}$ + 0.005958$X_{t-1}e_{t-3}$ – 0.004798$X_{t-2}e_{t-3}$ - 0.011294$X_{t-3}e_{t-1}$ + e_t

M4:S2

X_t = - 0.049537X_{t-1} - 0.032572X_{t-2} – 0.413331X_{t-3} – 0.226474X_{t-4}- 0.285681X_{t-5} – 0.319814X_{t-6} - 0.183383X_{t-7} + 0.307067X_{t-8} + 0.094498e_{t-1} - 0.237065e_{t-2} – 0.005259$X_{t-1}e_{t-2}$ + 0.006275$X_{t-3}e_{t-3}$ – 0.004352$X_{t-2}e_{t-3}$ – 0.011644$X_{t-3}e_{t-1}$ + e_t

M4:S3

X_t = - 0.712478X_{t-1} - 0.153047X_{t-2} + 0.032479X_{t-3} – 0.606080X_{t-4}- 0.351330X_{t-5} – 0.422284X_{t-6} - 0.407042X_{t-7} - 0.311950X_{t-8} + 0.809607e_{t-1} - 0.048903e_{t-2} – 0.673588e_{t-3} - 0.005131$X_{t-1}e_{t-2}$ - 0.003221$X_{t-2}e_{t-3}$ – 0.007347$X_{t-3}e_{t-1}$ + e_t

The derived statistics from the above fitted models are given in table1, table 2 and table 3 below.

Table1. Goodness of fit of four generalized bilinear models at t = 50. Four models are compared, namely M1: BL(P, 1, 0, r, s), M2:{F1= BL(P, 1, [1], r, s), F2= BL(P, 1, [1,2], r, s), and F3= BL(P, 1, [1,2,3], r, s)} M3: SBL(P, 1, 0, r, s), M4:{S1= SBL(P, 1, [1], r, s), S2= SBL(P, 1, [1,2], r, s), and S3= SBL(P, 1, [1,2,3], r, s)}. All models are significant at P<0.001.

		Full Bilinear				Subset bilinear		
	M1		M2		M3		M4	
		F1	F2	F3		S1	S2	S3
Residual Variance	250.20	243.0	216.93	180.78	253.32	242.99	224.03	179.08
AIC(Akaike Inf. Criterion)	8.52	8.39	8.32	8.10	8.49	8.39	8.27	8.07
BIC (Bayesian Inf. Criterion)	8.68	8.47	8.43	8.18	8.59	8.47	8.31	8.17
R^2	0.58	0.60	0.64	0.69	0.57	0.60	0.63	0.70
Adjusted R^2	0.55	0.59	0.63	0.68	0.56	0.59	0.62	0.69
F(Statistic)	20.93	71.24	42.02	180.78	31.79	71.24	82.35	111.49

From table 1, M1, M2, M3 and M4 were compared at t = 50. The residual variance of M2 at F3 (180.78) is greater than the residual variance of M4 at S3 (179.08). Therefore, the impact of the elements of 2^3-1 is negligible and testing all possible subsets to have the best subsets is the best approach here. The optimal model here is M4 at S3.Comparing M1, M2, M3 and M4, M4 at S3 (179.08) emerge as the optimal model.

Table2. Goodness of fit of four generalized bilinear models at t = 150. Four models are compared, namely M1: BL(P, 1, 0, r, s), M2:{F1= BL(P, 1, [1], r, s), F2= BL(P, 1, [1,2], r, s), and F3= BL(P, 1, [1,2,3], r, s)} M3: SBL(P, 1, 0, r, s), M4:{S1= SBL(P, 1, [1], r, s), S2= SBL(P, 1, [1,2], r, s), and S3= SBL(P, 1, [1,2,3], r, s)}. All models are significant at P<0.001.

	Full Bilinear				Subset bilinear			
	M1	M2			M3	M4		
		F1	F2	F3		S1	S2	S3
Residual Variance	193.20	185.20	184.70	193.40	193.30	192.10	186.0	194.20
AIC(Akaike Inf. Criterion)	8.21	8.16	8.13	8.20	8.18	8.14	8.14	8.18
BIC (Bayesian Inf. Criterion)	8.36	8.32	8.21	8.36	8.30	8.25	8.22	8.30
R^2	0.61	0.62	0.63	0.61	0.61	0.61	0.62	0.60
Adjusted R^2	0.59	0.60	0.61	0.59	0.59	0.60	0.60	0.59
F(Statistic)	31.18	33.51	30.72	31.22	44.12	75.81	47.34	43.97

From table 2, M1, M2, M3 and M4 were compared at t = 150. In this table the residual variances of M2 at F1 (185.20), F2 (184.70) and F3 (193.40) are less than the residual variance of M4 at S1 (192.10), S2 (186.0) and S3 (194.20) respectively. With this, impact of the elements of 2^3-1 is pronounced and testing all possible subsets to have the best subsets is not necessary. The optimal model is M2 at F2 (184.70). Comparing M1, M2, M3 and M4, the optimal model is M2 at F2.

Table3. Goodness of fit of four generalized bilinear models at t = 250. Four models are compared, namely M1: BL(P, 1, 0, r, s), M2:{F1= BL(P, 1, [1], r, s), F2= BL(P, 1, [1,2], r, s), and F3= BL(P, 1, [1,2,3], r, s)} M3: SBL(P, 1, 0, r, s), M4:{S1= SBL(P, 1, [1], r, s), S2= SBL(P, 1, [1,2], r, s), and S3= SBL(P, 1, [1,2,3], r, s)}. All models are significant at P<0.001.

		Full Bilinear				Subset bilinear		
	M1		M2		M3		M4	
		F1	F2	F3		S1	S2	S3
Residual Variance	285.50	279.0	277.40	293.70	285.80	281.40	279.50	300.10
AIC(Akaike Inf. Criterion)	8.55	8.53	8.49	8.57	8.54	8.51	8.50	8.57
BIC (Bayesian Inf. Criterion)	8.65	8.62	8.60	8.62	8.61	8.56	8.55	8.61
R^2	0.55	0.56	0.57	0.54	0.55	0.55	0.56	0.53
Adjusted R^2	0.54	0.55	0.56	0.53	0.54	0.54	0.55	0.52
F(Statistic)	49.29	51.35	42.0	47.0	74.0	101.96	103.30	136.91

From table 3, M1, M2, M3 and M4 were compared at t = 250. In this table the residual variances of M2 at F1 (279.0), F2 (277.40) and F3 (293.7) are less than the residual variance of M4 at S1 (281.4), S2 (279.5) and S3 (300.1) respectively. With this, impact of the elements of 2^3-1 is pronounced and testing all possible subsets to have the best subsets is not necessary. The optimal model is M2 at F2 (277.4). Comparing M1, M2, M3 and M4, the optimal model is M2 at F2.

In the three tables, the optimal models identified gave us the highest *R*-squared as well as the adjusted *R*-squared. In addition, the optimal models identified have the minimum Akaike information criterion and Bayesian information criterion. Figures 1 and 2 in the appendix gave us a clear picture of the optimal models as we could see that the minimum point of the graph depict the lowest residual variance bringing our conclusion to the same as what was obtainable in tables 1, 2 and 3.

5. Conclusion

This study focused on generalized bilinear models that could handle all non-linear series. Bilinear models at different levels of sample sizes were considered using the non-linear real series. At each level, optimal models were identified. The optimal models identified attained stationarity. Furthermore, the subsetting concept introduced to the generalized bilinear model helped at arriving at optimal model for different sample sizes.

Moreover, estimation of parameters has witnessed a unique, consistent and convergent estimator that has prevented the models from exploding, thereby making stationarity possible. The introduction of the *d* factor in our models has made us to capture trend and seasonality in the data, which in turn helps arrive at stationarity easily for any time series data set.

Appendix

The proof of the theorem from section 3 for the sake of simplicity is carried out in the following steps:

Step 1

Let the process $\{\mathbf{S}_{n,t,} n, t \in Z\}$ be defined as follows: $S_{n,t} = Ce_t + (\Psi + B_1 e_{t-1})\mathbf{S}_{n-t,t-1} + B_2\mathbf{S}_{n-2,t-2}e_{t-2} + ... + B_s\mathbf{S}_{n-s,t-s}e_{t-s}$, if n>0 for every t in Z. We show that $\lim_{n\to\infty} \mathbf{S}_{n,t}$ exists almost surely for every t in Z. If $\mathbf{X}_t$ is the almost sure limit of $\{\mathbf{S}_{n,t,} n \geq 1\}$ for every t in Z, then it is obvious that the process $\{\mathbf{X}_t, t \in Z\}$ conforms to the bilinear model (2.1). It is also easy to check that for every fixed n in Z, $\{\mathbf{S}_{n,t,} t \in Z\}$ is a strictly stationary process.

Step 2

Let $\mathbf{s}_{n,t} = \mathbf{S}_{n,t} - \mathbf{S}_{n-1,t,} t \in Z$. We show that $E\left|(\mathbf{s}_{n,t})_i\right| \leq K\lambda^{n/2}$ for every $n \geq 0$ and i = 1, 2, ….,p, where K is a positive constant. Since $\lambda < 1$, this then implies that $\{\mathbf{S}_{n,t,} n \geq 1\}$ converges almost surely for every t in Z. (If $\{a_{n,} n \geq 1\}$ is a sequence of real numbers satisfying $|a_n - a_{n-1}| \leq K\lambda^n$ for every $n \geq 2$ for some positive constant K and $\lambda < 1$, then it is easy to show that $\{a_{n,} n \geq 1\}$ is a Cauchy sequence of real numbers.)

Step 3

First, we settle the question of integrability of the $\mathbf{s}_{n,t}$'s . Note that

$$\begin{aligned}\mathbf{s}_{n,t} &= \mathbf{S}_{n,t} - \mathbf{S}_{n-1,t} \\ &= (\Psi + B_1 e_{t-1})\mathbf{s}_{n-1,t-1} + B_2 \mathbf{s}_{n-2,t-2} e_{t-2} + ... + B_s \mathbf{s}_{n-s,t-s} e_{t-s} \\ &= Q_n(e_{t-1}, e_{t-2}, ..., e_{t-n})\mathbf{s}_{0,t-n} \quad = Q_n(e_{t-1}, e_{t-2}, ..., e_{t-n})Ce_{t-n},\end{aligned}$$

where $Q_n(e_{t-1}, e_{t-2}, ..., e_{t-n})$ is a matrix of order p x p and each entry of this matrix is a polynomial in $e_{t-1}, e_{t-2}, ..., e_{t-n}$ in which the power index of each e_{t-j} is either 0 or 1. Consequently, every entry in $Q_n(e_{t-1}, e_{t-2}, ..., e_{t-n})$ and hence in $\mathbf{s}_{n,t}$ is integrable. It is clear that distribution of $\mathbf{s}_{n,t}$ does not depend on t.

Step 4

It is convenient to deal with the following processes. Define

$\mathbf{s}^*_{n,t} = Q_n(e_{t-1}, e_{t-2}, ..., e_{t-n})C$, if n > 0 for every t in Z. Equivalently, $\mathbf{s}_{n,t} = \mathbf{s}^*_{n,t} e_{t-n}$, n, $t \in Z$.

From the remark made regarding the $Q_n(.)$'s in step 3, it is obvious that every entry in $\mathbf{s}^*_{n,t}$ is square integrable. Further, it is easy to check that $\mathbf{s}^*_{n,t}$'s satisfy the following equation.

$$\mathbf{s}^*_{n,t} = (\Psi + B_1 e_{t-1})\mathbf{s}^*_{n-1,t-1} + B_2 \mathbf{s}^*_{n-2,t-2} e_{t-2} + ... + B_s \mathbf{s}^*_{n-s,t-s} e_{t-s} \qquad \text{(A1)}$$

for every n, t in Z. Also, the distribution of $\mathbf{s}^*_{n,t}$ does not depend on t, since the e_t's are independently identically distributed. Since $\mathbf{s}_{n,t} = \mathbf{s}^*_{n,t} e_{t-n}$ for all n and t in Z.

$$E\left|(\mathbf{s}_{n,t})_i\right| = E\left|(\mathbf{s}^*_{n,t})_i\right| \left|e_{t-n}\right| \le (E((\mathbf{s}^*_{n,t})_i)^2)^{1/2} (Ee^2_{t-n})^{1/2} \le \sigma (E((\mathbf{s}^*_{n,t})_i)^2)^{1/2}$$

for every i = 1, 2, ……,p. It suffices to obtain an upper bound for $E((\mathbf{s}^*_{n,t})_i)^2$ for every i = 1,2,…,p and n, t in Z. For this we evaluate $E(\mathbf{s}^*_{n,t} \otimes \mathbf{s}^*_{n,t}) = M_n$, say

Step 5

Let $D_1 = (\Psi + B_1 e_{t-1})\mathbf{s}^*_{n-1,t-1}$ and $D_i = B_i \mathbf{s}^*_{n-1,t-1} e_{t-i}$ for i = 2, 3, …, s.

$$\begin{aligned}\mathbf{s}^*_{n,t} \otimes \mathbf{s}^*_{n,t} &= (\sum_{i=1}^{s} D_i) \otimes (\sum_{i=1}^{s} D_i) = \{D_1 \otimes D_1\} + \{D_1 \otimes D_2 + D_2 \otimes D_2 + D_2 \otimes D_1\} \\ &+ \{D_1 \otimes D_3 + D_2 \otimes D_3 + D_3 \otimes D_3 + D_3 \otimes D_2 + D_3 \otimes D_1\} + ... \\ &+ \left\{\sum_{i=1}^{s-1} D_i \otimes D_s + D_s \otimes D_s + \sum_{i=1}^{s-1} D_s \otimes D_i\right\}. \qquad \text{(A2)}\end{aligned}$$

We evaluate the expectation of each expression within each set of brackets {}in (A2)

Step 6

We write $D_1 \otimes D_1 = ((\Psi + B_1 e_{t-1}) \otimes (\Psi + B_1 e_{t-1}))(\mathbf{s}^*_{n-1,t-1} \otimes \mathbf{s}^*_{n-1,t-1})$.

Since $\mathbf{s}^*_{n-1,t-1}$ is a function of $e_{t-2}, e_{t-3}, ..., e_{t-n}, \mathbf{s}^*_{n-1,t-1}$ and e_{t-1} are independently distributed. So,

$E(D_1 \otimes D_1) = ((\Psi \otimes \Psi + \sigma^2(B_1 \otimes B_1)) = \Gamma_1 M_{n-1}$.

Step 7

Expanding $\mathbf{s}^*_{n-1,t-1}$, we obtain

$$D_1 \otimes D_2 = ((\Psi + B_1 e_{t-1})(\Psi + B_1 e_{t-2}) \otimes B_2 e_{t-2})(\mathbf{s}^*_{n-2,t-2} \otimes \mathbf{s}^*_{n-2,t-2})$$
$$+ ((\Psi + B_1 e_{t-1}) \otimes B_2 e_{t-2})(B_2 \mathbf{s}^*_{n-3,t-3} e_{t-3} \otimes \mathbf{s}^*_{n-2,t-2}) + ... +$$
$$((\Psi + B_1 e_{t-1}) \otimes B_2 e_{t-2})(B_s \mathbf{s}^*_{n-1-s,t-1-s} e_{t-1-s} \otimes \mathbf{s}^*_{n-2,t-2}).$$

Therefore, $E(D_1 \otimes D_2) = \sigma^2((\Psi B_1) \otimes B_2) M_{n-2}$

In a similar fashion, we can show that $E(D_2 \otimes D_1) = \sigma^2(B_2 \otimes (\Psi B_1)) M_{n-2}$ and $E(D_2 \otimes D_2) = \sigma^2(B_2 \otimes B_2) M_{n-2}$.Consequently, the expected value of the entire expression in the second set of such brackets is $\sigma^2(B_2 \otimes (\Psi B_1) + (\Psi B_1) \otimes B_2 + B_2 \otimes B_2) M_{n-2} = \Gamma_2 M_{n-2}$.

Step 8: Pursuing ideas similar to those used in step 7, we can show that the expected value of the entire expression in the third set of such brackets in (A2) is $\sigma^2(B_3 \otimes (\Psi^2 B_1 + \Psi B_2) + (\Psi^2 B_1 + \Psi B_2) \otimes B_3 + B_3 \otimes B_3) = \Gamma_3 M_{n-3}$.

Step 9: The expectations of other expressions can be evaluated analogously. Finally, we obtain $M_n = E(\mathbf{s}^*_{n,t} \otimes \mathbf{s}^*_{n,t}) = \sum_{i=1}^{s} \Gamma_i M_{n-i}$ for all n.

Step 10

Since $M_n = E(\mathbf{s}^*_{n,t} \otimes \mathbf{s}^*_{n,t})$, we have

$$E((\mathbf{s}^*_{n,t})_i)^2 \le K' \lambda^n$$

where $\rho(L) = \lambda < 1$ and K' is a positive constant.

$\Gamma_1 = \Psi \otimes \Psi + \sigma^2(B_1 \otimes B_1)$,

For any two matrices $D = (d_{ij})$ and $P = (p_{ij})$ of orders $m \times n$ and $r \times s$ respectively, we denote the Kronecker product of D and P by $D \otimes P$ Neudecker (1969).

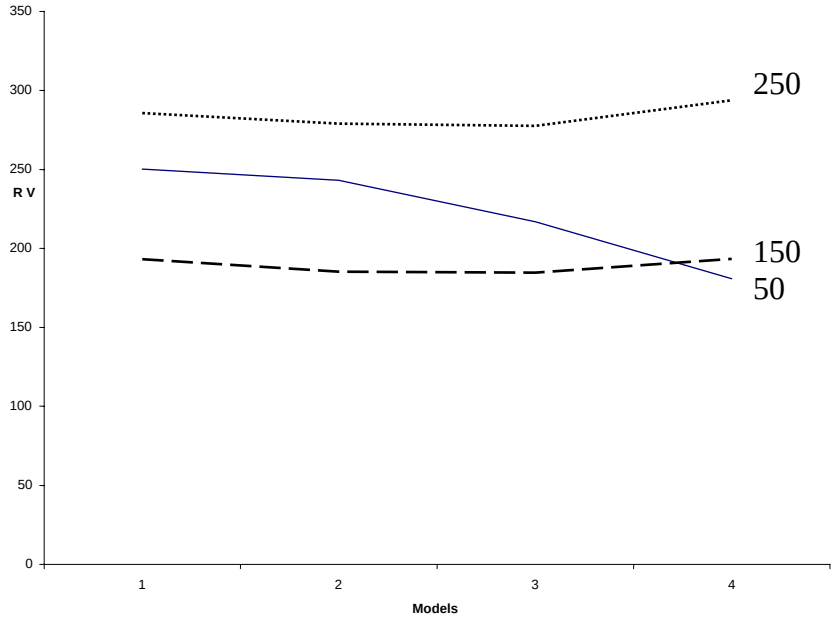

Figure 1: Residual Variance at t = 50, 150 and 250 for Generalized Bilinear Models

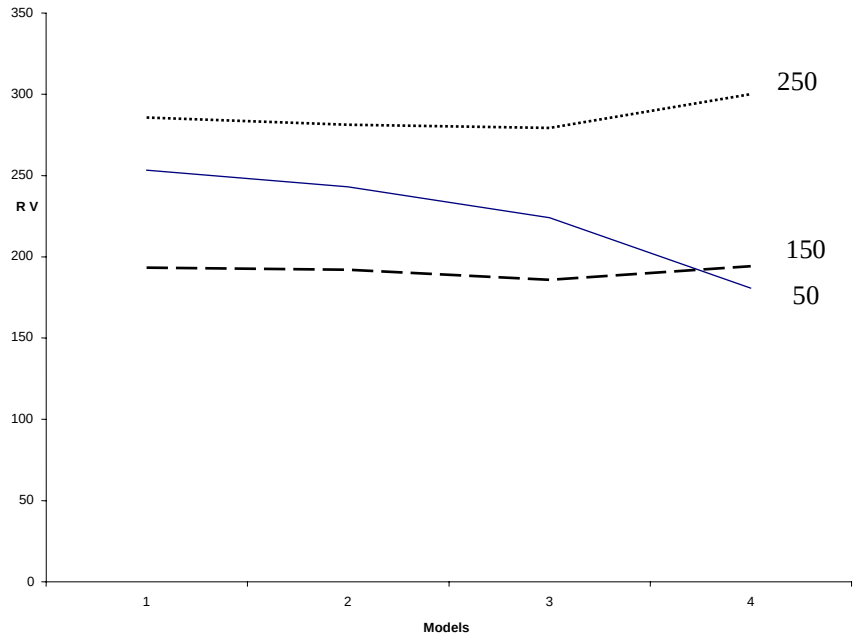

Figure 2: Residual Variance at t = 50, 150 and 250 for Generalized Subset Bilinear Models

6. References

[1] Bibi A. (2006). Evolutionary transfer functions of bilinear process with time varying coefficients. *Computer and Mathematics with Applications* 52, 3-4, 331-338.

[2] Box, G.E.P, Jenkins, G.M and Reinsel, G.C (1994). Time series analysis; forecasting and control. 3rd Edition. Prentice Hall, Inc.

[3] Boonaick K. Stensholt B.K. and Stensholt E. (2005). Multivariate bilinear time series: a stochastic alternative in population dynamics. *Geophysical Research Abstracts, Volume 7, 02219 @Europeans Geosciences Union 2005.*

[4] Bruni, C., Dupillo, G. and Koch, G. (1974). Bilinear Systems: An Appealing Class of Nearly Linear System in Theory and Application. *IEEE Trans. Auto Control Ac*-19, 334-338.

[5] Cathy, W. S. (1997). Detection of Additive Outliers in Bilinear Time Series. *Computational Statistics and Data Analysis* 24, 283-294.

[6] Doukhan, P., Latour, A., and Oraichi, D., (2006). A simple integer – valued bilinear time series model. *Adv. Appl. Prob.* 38, 559-578.

[7] Drost F.C., van den Akker R., Werker, B.J.M. (2007). Note on integer – valued bilinear time series models. *Econometrics and Finance group CentER, ISSN 0924-7815,* Tilburg University, The Netherlands.

[8] Gonclaves, E., Jacob P. and Mendes-Lopes N. (2000). A decision procedure for bilinear time series based on the Asymptotic Separation. *Statistics*, 333-348.

[9] Granger, C. W. J. and Anderson, A. P. (1978) *Introduction to Bilinear Time Series Models.* Gottingen, Germany: Vandenhoeck and Ruprecht.

[10] Haggan , V. and Oyetunji, O. B. (1980). On the Selection of Subset Autoregressive Time Series Models. *UMIST Technical Report No. 124* (Dept. of Mathematics).

[11] Krzanowski W.J. (1998). An Introduction to Statistical Modelling. Arnold, the UK.

[12] Liu, J. (1992). On stationarity and Asymptotic Inference of Bilinear time series models. *Statistica Sinica* 2(2), 479-494.

[13] Neudecker H. (1969). Some Theorems on Matrix Differentiation with Special References to Kronecker Matrix Products. *Jour. Amer. Statist. Ass.* 69, 953-963.

[14] Pham, T.D. and Tran, L.T. (1981) On the First Order Bilinear Time Series Model. *Jour. Appli. Prob.* 18, 617-627.

[15] Rao, M. B., Rao, T. S and Walker, A. M. (1983) On the Existence of some Bilinear Time Series Models. *Journal of Time Series Analysis* 4(2), 60-76.

[16] Shangodoyin, D. K. and Ojo, J. F (2003). On the Performance of Bilinear Time Series Autoregressive Moving Average Models. *Journal of Nigerian Statistical Association* 16, 1-12.

[17] Subba Rao, T. and Gabr, M. M. (1980). A Test for Linearity of Stationary Time Series. *Jour. Time Series Analysis* 1(2), 145-158.

[18] Gabr, M. M. and Subba Rao, T. (1981). The Estimation and Prediction of Subset Bilinear Time Series Models with Application. *Journal of Time Series Analysis* 2(3), 89-100.

[19] Subba Rao, T. (1981). On theory of Bilinear time Series Models .*Jour. R. Sta. Soc*. B. 43, 244-255.

[20] Usoro, A. E. and Omekara C. O. (2008). Lowere Diagonal Bilinear Moving Average Vector Models. *Advances in Applied Mathematical Analysis* 3(1), 49-54

[21] Wang H.B. and Wei B.C. (2004). Separable lower triangular bilinear model. *J. Appl. Probab.* 41(1), 221-235**.**

J. F. Ojo, Department of Statistics, University of Ibadan, Nigeria.

D. K. Shangodoyin, Department of Statistics, University of Botswana, Botswana.

E-mail address: *shangodoyink@mopipi.ub.bw.*