PERFORMANCE OF BILINEAR AUTOREGRESSIVE MOVING AVERAGE MODELS: USING DEMOGRAPHIC TIME SERIES DATA

Samuel Zewdie, Ermias Buli

Department of Statistics, Adama Science and Technology University

ABSTRACT

This paper investigates the performance of bilinear time series autoregressive moving average (ARMA) models i.e. BL (p, 0, r, s) and BL (p, q, r, s). The full bilinear model was fitted to monthly data on number of marriages conducted at Addis Ababa City Municipality for 19 years. The estimation of the parameters and residual variance of BL (p, o, r, s) was compared with BL (p, q, r, s). In the series, having considered the best subset autoregressive model AR (p) and the best subset autoregressive moving average model ARMA (p, 1) as our initial values in fitting BL (8, 0, 5, 3) and BL (6, 1, 5,3); it was found out that the residual variance attached to BL (6, 1, 5, 3) was smaller than BL (8, 0, 5, 3) meaning that the bilinear time series with ARMA as the initial value performed better than bilinear time series with AR. The non-linearity of the marriage data used made us compare the performance of the bilinear time series with the linear.

KEYWORDS: Bilinear time series, Yule Walker Equation, Autoregressive, Moving average, Autoregressive moving average.

1. INTRODUCTION

Building probability models for time series data is an important activity, which enables the statistician to understand the underlying random mechanism generating the series. They also provide invaluable assistance in forecasting. Linear models have been used quite successfully for analyzing time series data. There are situations when it is felt that linear time series models may not be adequate in explaining the underlying random mechanism for instance, the sunspot data and Canadian lynx data. Jones (1978), Granger and Andersen (1978a), Haggan and Ozaki (1980), Priestley (1980), Tong and Lim (1980) and Subba Rao (1981) have considered particular types of nonlinear time series models. The nonlinear models considered by Granger and Andersen (1978a) and Subba Rao (1981) are known as bilinear time series models. This class of time series has been found to provide a better fit as well useful in many areas including biological sciences, ecology and engineering (Mohler (1973), Bruni, Dupillo and Roch (1974)).

Undoubtedly, the nonlinear time series models are more complex than linear models and they lead to several problems, which sometimes may be difficult to solve. These problems include:

1. Estimation of the parameter of these models is more difficult.
2. Statistical properties of quite a number of nonlinear models are quite difficult to investigate.

3. Study of the sampling distribution of the estimates can sometimes be quite complicated.
4. Lastly, evaluation of optimal forecasts for several steps in the future from these models is not easy, though not impossible.

In view of these problems involved, it seems reasonable to expect that when a nonlinear time series model is fitted, its performance must be significantly better than a linear time series model. One such model for which it may be possible to obtain optimal forecasts for several steps ahead and can perform better than a linear model is the bilinear time series model. Some theoretical considerations of this model have been reported in the papers by Granger and Anderson (1978a), Subba Rao (1981), Quinn (1982), and W.K. LI (1984).

A time series X_t is a bilinear process of order (p, q, r, s) if it satisfies the model

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + e_t - \sum_{j=1}^{q} \theta_j e_{t-j} + \sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij} X_{t-i} e_{t-j} \qquad (1.1)$$

where $\{e_t\}$is a sequence of independent, identically distributed random variables with mean zero and variance σ^2. We assume also that the model is invertible and we have a realization $\{x_1, x_2 \ldots x_n\}$ on the time series $\{x_t\}$. We denote (1.1) by BL (p, q, r, s). This model was considered by Granger and Andersen (1978a). We therefore investigate the performance of model (1.1) and

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij} X_{t-i} e_{t-j} + e_t \qquad (1.2)$$

and see which of the models performs better. We denote (1.2) by BL (p, 0, r, s)

A major problem with bilinear time series modelling is the problem of model selection. The problem is considerably more difficult than the autoregressive moving average case (i.e. r = s = 0). In the linear ARMA situation a preliminary identification on the orders of p and q can be done using sample autocorrelations and partial autocorrelation (Box–Jenkins, 1970). With the presence of bilinear term in (1.1) the usual Box-Jenkins procedure of identification cannot be applied. Thus, most authors resort to the use of AIC in selecting the right combination of p, q, r and s.

2. THE PROPOSED ESTIMATION TECHNIQUE

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

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + e_t - \sum_{j=1}^{q} \theta_j e_{t-j} + \sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij} X_{t-i} e_{t-j} \qquad (2.1)$$

$$e_t = X_t - \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} - \sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij} X_{t-i} e_{t-j} \qquad (2.2)$$

where $\{e_t\}$ are independent and each e_t is distributed N (0, σ_e^2). Here we assume the model 2.1 is invertible, and further assume we have a realization $\{x_1, x_2 \dots x_n\}$ on the time series $\{x_t\}$. The joint density function of $\{e_m, e_{m+1} \dots e_n\}$ where m = max (r, s) + 1 is given by

$$\frac{1}{(2\pi\sigma_e^2)^{(n-m+1)/2}} \exp(\frac{1}{-2\sigma_e^2}\sum_m^n e_t^2) \qquad (2.3)$$

Since the Jacobian of the transformation from $\{e_m, e_{m+1} \dots e_n\}$ to $\{x_m, x_{m+1} \dots x_n\}$ is unity, the likelihood function of $\{x_m, x_{m+1} \dots x_n\}$ is the same as the joint density function of $\{e_m, e_{m+1} \dots e_n\}$. Maximizing the likelihood function is the same as minimizing the function $Q(G)$, where

$$Q(G) = \sum_{i=m}^n e_{t,}^2 \qquad (2.4)$$

with respect to $G' = (\phi_0, \phi_1, \dots, \phi_p; \theta_1, \theta_2, \dots, \theta_q; B_{11}, \dots, B_{rs})$ For convenience, we shall write $G_1 = \phi_0, G_2 = \phi_1, \dots, G_R = B_{rs}$ where R = 1+p+q+rs. Then the partial derivatives of Q (G) are:

$$\frac{dQ(G)}{dG_i} = 2\sum_{t=m}^n e_t \frac{de_t}{dG_i} \qquad (i = 1, 2 \dots R) \qquad (2.5)$$

$$\frac{d^2Q(G)}{dG_i dG_j} = 2(\sum_{t=m}^n e_t \frac{de_t}{dG_i}\frac{de_t}{dG_j} + \sum_{t=m}^n e_t \frac{d^2 e_t}{dG_i dG_j}) \qquad (2.6)$$

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

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

$$\frac{de_t}{d\theta_i} + \sum_{j=1}^s W_t(t)\frac{de_{t-j}}{d\theta_i} = e_{t-i,} \qquad \text{if} \quad i = 1, 2, \dots, q \qquad (2.7)$$

$$\frac{de_t}{dB_{kmi}} + \sum_{j=1}^s W_j(t)\frac{de_{t-j}}{db_{kmi}} = -X_{t-k}e_{t-m} \quad (k=1,2,\dots,r \,;\, m_i = 1,2,\dots,s) \qquad (2.8)$$

$$\frac{d^2 e_t}{d\phi_i d\phi_i'} + \sum_{j=1}^s W_j(t)\frac{d^2 e_{t-j}}{d\phi_i d\phi_i'} = 0 \qquad (i, i' = 0, 1, 2, \dots, p) \qquad (2.9)$$

$$\frac{d^2 e_t}{d\theta_i d\theta_i'} + \sum_{j=1}^s W_j(t)\frac{d^2 e_{t-j}}{d\theta_i d\theta_i'} = 0 \qquad (i, i' = 0, 1, 2, \dots, q) \qquad (2.10)$$

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

$$(i=0,1,2,\dots,p \,;\, k_i = 1,2,\dots,r;\, m_i=1,2,\dots,s) \qquad (2.11)$$

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

(i=1,2,…,q ; k_i =1,2,…,r; m_i=1,2,…,s) (2.12)

$$\frac{d^2 e_t}{d\phi_i d\theta_i} + \sum_{j=1}^{s} W_j(t) \frac{d^2 e_{t-j}}{d\phi_i d\theta_i} = 0 \qquad (2.13)$$

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

(k, k' =1,2,…,r ; m_i, m_i' = 1,2,…,s) (2.14)

$W_j(t) = \sum_{j=1}^{s} B_{ij} x_{t-j}$, we assume $e_t = 0$ (t = 1, 2, …, m-1) and also

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

From these assumptions and 3.2.8 it follows that the second order derivatives with respect to ϕi (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 2.6, 2.7, 2.8 and 2.14. Now let

$$\mathbf{V}'(G) = \left(\frac{\partial Q(G)}{\partial G_1}, \frac{\partial Q(G)}{\partial G_2}, \ldots, \frac{\partial Q(G)}{\partial G_R} \right)$$

and let $\mathbf{H}(\mathbf{G}) = [\partial^2 Q(\mathbf{G}) / \partial G_i \partial G_j]$ be a matrix of second partial derivatives. Expanding **V**(**G**), near **G** = $\hat{\mathbf{G}}$ in a Taylor series, we obtain $[\mathbf{V}(\hat{\mathbf{G}})]_{\hat{\mathbf{G}}=\mathbf{G}} = 0 = \mathbf{V}(\mathbf{G}) + \mathbf{H}(\mathbf{G})(\hat{\mathbf{G}} - \mathbf{G})$. Rewriting this equation we get $\hat{\mathbf{G}} - \mathbf{G} = -\mathbf{H}^{-1}(\mathbf{G})\mathbf{V}(\mathbf{G})$, and thus obtain an iterative equation given by $\mathbf{G}^{(k+1)} = \mathbf{G}^{(k)} - \mathbf{H}^{-1}(\mathbf{G}^{(k)})\mathbf{V}(\mathbf{G}^{(k)})$ where $G^{(k)}$ is the set of estimates obtained at the k^{th} stage of iteration.

3. ANALYSIS OF MONTHLY MARRIAGE DATA

The data considered is monthly number of marriages conducted at the Addis Ababa City Municipality for the years 1992 to 2011, giving 252 observations. The time plot of this data shows that it is nonlinear, encouraging.

Fitting of Full AR Models to the Data

The linear models are fitted to the first 176 observations, and we consider the choice of the order of the linear model. The linear models of all order up to AR(30) are fitted. The choice of the order is made on the basis of the information criterion of Akaike (1977), which is given by

$$\text{AIC}_{(p)} = N\ln \sigma_p^2 + 2p \quad \text{where} \quad \sigma_p^2 = \frac{1}{N} \sum_{t=1}^{N} (x_t - \sum_{i=1}^{p} \phi_i x_{t-i})^2$$

is the maximum likelihood estimate of the residual variance after fitting the AR(p). In practice, we specify a maximum lag L and fit successively AIC (1), AIC (2)… The minimum AIC is the best model for the data. In view of this, we found that AIC is minimum when p=8. The fitted model is

$$X_t = 2.0324X_{t-1} + 0.1263X_{t-2} - 0.2021X_{t-3} - 0.1363X_{t-4} + 1.1142X_{t-5} - 0.3213X_{t-6} - 0.0951X_{t-7} + 0.1714X_{t-8} + e_t$$

Fitting of Best Subset AR Models to the Data

The method described in chapter one was employed in fitting of best subset AR models. There are $2^8 - 1 = 255$ possible subsets. The choice of the order is made on the basis of minimum AIC and having considered the 255 possible subsets, it is found that AIC is minimum in the following model

$$X_t = 1.3066x_{t-1} - 0.5181x_{t-2} + 0.1868x_{t-8} + e_t$$

Fitting Of Full ARMA (P, 1) Models to the Data

Having considered the full AR (p) which later forms our initial values, introducing a MA term into our system we want to see whether there will be reduction in the residual variance. The linear models are fitted to the 176 observations and we consider the choice of the order of the model by using AIC. In view of this, it is found that AIC is minimum when p = 6 and q = 1 and the fitted model is:

$X_t = 2.057543x_{t-1} – 1.512057\ x_{t-2} + 0.264533x_{t-3} +0.330741x_{t-4}- 0.39.341x_{t-5} + 0.246004x_{t-6} + 0.781977e_{t-1} + e_t$

Fitting Of Best Subset ARMA (p, 1) Models to the Data

The method described in above was employed in fitting of best subset ARMA (p, 1) models. There are $2^6 – 1=63$ possible subsets. The choice of the order is made on the basis of minimum AIC and having considered the 63 subsets, it is found that AIC is minimum in the model

$$X_t = 2.092475x_{t-1} – 1.550083x_{t-2} + 0.381015x_{t-3} + 0.073620x_{t-6} – 0.823840e_{t-1} + e_t$$

Fitting of Full Bilinear Model to the Marriage Data with AR as Initial Value

We now employ the estimation procedure and algorithm for fitting bilinear models. It has been pointed out that $G^{(k)}$ is the set of estimates obtained at the k^{th} stage of iteration. Therefore, we need a good set of initial values of the parameters to start the iteration. For this study, we have chosen the coefficients of the AR part of above sections, which is equal to the corresponding best subset AR models. The bilinear models of all orders up to BL(8, 0, 5, 3) are fitted. The choice of the order is made on the basis of AIC. It is found that AIC is minimum when p=5 and q=3. The estimated values of the coefficient of the model are:

$\hat{a}_1$=1.3066, $\hat{a}_2$ = -0.5181, $\hat{a}_3$ =0.1868 and the values b_{lj} (l = 1,2,3,4,5 ; j = 1,2,3) are given by,

$$b_{lj} = \begin{bmatrix} 0.000459 & -0.025968 & 0.011788 \\ 0.019074 & 0.032467 & -0.051531 \\ -0.049079 & 0.012807 & 0.049275 \\ 0.027194 & -0.036035 & -0.008914 \\ 0.003068 & 0.015813 & -0.006886 \end{bmatrix}$$

Fitting Of Full Bilinear Model with ARMA (P, 1) As Initial Values

As it has been stressed that $G^{(k)}$ is the set of estimates obtained at the kth stage of iteration, therefore making use of the initial value of ARMA which is the best subset; we now consider the choice of the bilinear model. The bilinear models of all up to BL (6, 1, 5, 3) are fitted. The choice of the order is made on the basis of the information criterion. It is found that AIC is minimum when p=5 and q=3. The estimated values of the coefficient of the model are as follows $\hat{a}_1$=2.092475 $\hat{a}_2$=-1.550083, $\hat{a}_3$ =0.381015, $\hat{a}_6$=0.073620, c_1=-0.823840 and values of b_{lj} (i=1, 2, 3, 4, 5; j= 1, 2, 3) are given by,

$$b_{lj} = \begin{bmatrix} -0.008157 & -0.000397 & 0.019829 \\ -0.005019 & 0.004150 & -0.047450 \\ -0.029462 & 0.000483 & 0.033112 \\ -0.007564 & -0.000718 & -0.021676 \\ -0.0021676 & 0.001504 & -0.009393 \end{bmatrix}$$

VALUES OF $\sigma_{\hat{e}}^2$ AND AIC FOR THE DIFFERENT MODELS

MODEL	**FULL (AR_8)**	**FULL ($ARMA_{6,1}$)**	**SUBSET (AR_3)**	**SUBSET ($ARMA_{3,1}$)**	**BILINEAR AR BL(8,0,5,3)**	**BILINEAR (ARMA) BL(6,1,5,3)**
$\sigma_{\hat{e}}^2$	224.07	220.1432	230.0242	226.8696	158.74488	157.02
AIC	5.072	5.473517	5.474023	5.479771	5.257106	5.783114

The above summary gives us the understanding that having the initial values to be the best ARMA model performed better than having the initial values to be the best AR. Therefore the model BL (p, 0, r, s) performed better than BL(p, q, r, s). The residual variance made us conclude thus. It is quite glaring from our analysis and the model above that bilinear model performed better than linear model looking at their residual variance.

4. SUMMARY AND CONCLUSIONS

Our method of estimation is a non-linear least square method which is being estimated iteratively which makes use of initial values. As a result we fit the best subset AR models to our data making use of algorithm described by Haggan and Oyetunji (1980) and the coefficient taken as our initial values. Also, because we want to compare BL(p, 0, r, s) with BL(p, q, r, s) we fit the best subset ARMA (p, 1) to our data and the coefficient taken as our initial values.

Having got our initial values, we fitted bilinear model to our data that is BL(8, 0, 5, 3) as well as BL(6, 1, 5, 3). The residual variance attached to BL (6, 1, 5, 3) is smaller than the residual variance attached to BL (8, 0, 5, 3) suggesting that BL (6, 1, 5, 3) perform better than BL (8, 0, 5, 3). It was glaring that bilinear model performed better than linear model because the data used had been tested for linearity and found out that it was non-linear which could be best estimated by bilinear model.

5. REFERENCES

Anderson T. W. (1977) Estimation for Autoregressive Moving Average Models in the Time and Frequency Domains. *The Ann. of Stat.* Vol 5 No. 5, 842-865

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

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

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.

Gabr, M. M. and T. Subba Rao (1981). The Estimation and Prediction of Subset Bilinear Time Series Models with Application. *Journal of Time Series Ana.* Vol2 No. 3.

Granger, C. W. J. and A. P. Anderson (1978a) *Introduction to Bilinear Time Series Models* Vandenhoeck and Puprecht.

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

Haggan, V. and T. Ozaki, (1980) Amplitude Dependent Exponential AR Model Fitting for Non-Linear Random Vibrations. *Proc. International Time Series Meeting, Nottinghamm,* ed. O. D. Anderson. North Holland.

Hannan, E. J. (1970) *Multiple Time Series*. John Wiley: New York

Jones, D. A. (1978) Non-linear Autoregressive Processes. *Proc. Royal Soc. London* (A)

Li W. K. (1984) On the Autocorrelation Structure and Identification of some Bilinear Time Series . *Journ. Of Time Series Analysis vol*. 5. No. 3.

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

Priestley, M. B. (1980) State-Dependent Models; a General Approach to Non-Linear Time Series Analysis. *Jour. Time Series Analysis* 1, 47-71.

Quinn, B. G. (1982), Stationarity and Invertibility of Simple Bilinear Models. *Stoch. Processes Appli.* 12, 225-230.

Rao, M. B., T. S. Rao and A. M. Walker (1983) On the Existence of some Bilinear Time Series Model. *Jour. Of Time Series Analysis* vol. 4, No. 2.

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

Tong H. and Lim, K. S. (1980) Threshold Autoregression, Limit Cycles and Cyclical Data. *Journal of Royal Statist. Soc*. B 42, 245-292.

Tuan Dinh Pham and Lanh Tat Tran (1981) On the First Order Bilinear Time Series Model. *Jour. Appli. Prob.* 18, 617-627.

Waldmeirer, M. (1961). *The Sunspot Activity in the Years 1610-1960,* Schulthese.