Interstat journals of Statistics
ISSN d0000056 online
http://interstat.statjournals.net

Parameter estimation for pure diagonal bilinear time series: An algorithm for maximum likelihood procedure

Bouzaachane Khadija

Mohammadia Engineering School, Department of EGT, LERMA
Avenue Ibn Sina P.B 765, Agdal Rabat, Morocco
Email: bouzaachane@netcourrier.com

Harti Mostafa

Math Info Department, Faculty of science Dhar Mahrez,
P.B 1796, Atlas, Fez, Morocco
Email: mharti@fsdmfes.ac.ma

Benghabrit Youssef

ENSAM, Marjane II, Meknes Ismailia, P.B 4024,
Ben M'Hamed, Meknes, Morocco
Email: you_benghabrit@yahoo.fr

Abstract

In this work, a new estimate algorithm for the parameters of a pure diagonal bilinear model is proposed. This algorithm turns out to be very reliable in estimating the true parameters's values of a given model. It combines maximum likelihood method, Kalman filter algorithm and simulated annealing.
Simulation results demonstrate that the algorithm is succeeds and promising.

Key-words : Time Series, Pure Diagonal Bilinear Model, Maximum LikeLihood, Kalman Filter, Simulated Annealing.

1 Introduction

The bilinear time series models have attracted considerable attention during the last years. They have found a variety of applications including those in economy, biology, software interfailure, signal processing...etc
An overview of models and their application can be found in Granger and Anderson (1978), Hristova(2005), Bouzaachane et al.(2006 a)...

In the most general form, a discrete-time process $(X_t)_{t\in\mathcal{Z}}$ that can be fitted with a bilinear model satisfies the following equation :

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=1}^{q} c_j e_{t-j} + \sum_{k=1}^{P}\sum_{l=1}^{Q} b_{kl} X_{t-k} e_{t-l} + e_t \tag{1}$$

where $(a_i, 1 \le i \le p)$, $(c_j, 1 \le j \le q)$ and $(b_{kl}, 1 \le k \le P, 1 \le l \le Q)$ are the coefficients of the model, and where e_t a sequence of independent and identically distributed (i.i.d.) Gaussian process with zero-mean and finite variance σ^2. In statistical literature, the above model is often denoted as the BL(p,q,P,Q) model. This model has been investigated in many works (see, for example, Subba Rao (1981), Guegan (1981), Bhaskara Rao et al.(1983), Liu et Brockwell (1988), Liu (1990)). Most of them discuss the probabilistic aspects of the model. Nevertheless, the statistical problem of the estimation of the parameters for some simple models has been considered in Pham and Tran (1981), Subba Rao (1984), Bouzaachane et al.(2006 b). The purpose of this work, is to investigate new approach for estimating the parameters of the pure diagonal bilinear model of arbitrary order, a particular case of the above model (1). This approach was first introduced in linear time series models, especially for estimating the parameters of ARMA models Melard (1983).
It deals with maximum likelihood method, and Kalman filter algorithm. Indeed, the main idea is to express the concerned model by state space form, and then deduce the log-likelihood function, which can be computed with Kalman filter algorithm (see Kalman (1960, 1963)). To obtain the maximum of the log-likelihood function, we used the simulated annealing method, which provide a global optimum regardless of the initial values. We have used two examples of pure bilinear models to examine the performance of the proposed method.
We mention that this algorithm was already used for the unit root bilinear model by Hristova (2005).
The remainder of the paper proceeds as follows. Section 2 lays out the pure diagonal bilinear model and its main property. Section 3 deals with the estimate algorithm of the parameters of the model of interest. In this section, we state, the definition of the state space form of the pure diagonal bilinear model, and the expression of log-likelihood function obtained by applying Kalman filter to the state space, and then maximize the likelihood using simulated annealing. In Section 4 we use two numerical examples to illustrate the estimation technique discussed in Section 3. The conclusion is provided in Section 5.

2 Pure Diagonal Bilinear model and invertibility

Let $(X)_{1\le t\le N}$ be a discrete time process. A process of the form :

$$X_t = \sum_{i=1}^{p} b_{ii} X_{t-i} e_{t-i} + e_t \tag{2}$$

where $(b_{ii}, 1 \leq i \leq p)$ are the coefficients of the model, and where $(e_t)_{t\in\mathcal{Z}}$ is i.i.d. Gaussian process with zero mean and common finite variance σ^2, is called a pure diagonal bilinear process and denoted BL(0,0,p,p).

We are attracted to the BL(0,0,p,p) model because of its simple mathematical structure. In this work, our goal is to establish an estimate algorithm for the parameters of BL(0,0,p,p) model. So, to obtain the good estimators, it's still necessary to verify the condition of invertibility of the BL(0,0,p,p) in each step of this algorithm.

The notion of invertibility is very useful for statistical applications, such as the prediction of X_t given its past, or the use of algorithms for computing estimates of the parameters. Several definitions of this notion have been proposed in the literature. Granger and Andersen (1978), Guegan and Pham(1987), Pham and Tran (1981) and Subba Rao and Gabr (1984) have derived invertibility conditions for some particular stationary bilinear models. Liu (1990) has established the condition under what the general bilinear model (1) is invertible.

As far as the model (2) is concerned, a condition of invertibility of this model can, so, be deduced from the result of Liu (1990). This condition can be written as

$$p^2 \sum_{i=1}^{p} b_{ii}^2 \sigma^2 < 1. \tag{3}$$

The invertibility condition is fundamental for our algorithm. Indeed, to obtain the convergence of our algorithm, and consequently, the good estimations, we have implanted the invertibility condition in our algorithm such as, we can verify it in each iteration of our computational algorithm.

3 Estimate Algorithme

Let $\theta = (b_{11}, b_{22}, \ldots, b_{pp})$ be the unknown vector of parameters of the model BL(0,0,p,p) (2). In this contribution, we suggest estimating the vector of parameters θ, by means of the maximum likelihood method. We establish the log-likelihood function applying Kalman filter to state space form of the model BL(0,0,p,p).

3.1 State space representation and log-likelihood function

Let us define the $(p+1)\times(p+1)$ matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & 0 \\ \ldots & \ldots & \ldots & 1 & 0 \end{pmatrix},$$

and

$$\xi_t = \begin{bmatrix} e_t \\ e_{t-1} \\ \vdots \\ e_{t-p} \end{bmatrix}, v_t = \begin{bmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}',$$

with $(e_t) \sim \mathcal{N}(0, \sigma^2)$ and

$$H_{t-1} = [1, b_{11}X_{t-1}, b_{22}X_{t-2}, \dots, b_{pp}X_{t-p}]$$

is time-varying coefficients vector.
With this notation, we can rewrite the model (2) in the following state space form

$$\left\{ \begin{array}{lcll} \xi_{t+1} & = & A\xi_t + v_{t+1} & : \text{state equation} \\ X_t & = & H_{t-1}\xi_t & : \text{observation equation} \end{array} \right. \tag{4}$$

Given the set of observation $\mathcal{X}_t = (x_1, \dots, x_t)$ available at time t, the Kalman filter algorithm, noted FK, generate, recursively, an optimal forecast $\hat{\xi}_{t+1|t}$ of the state vector $\xi_{t+1}, t = 1, \dots, N$.

```
FK algorithm
    Step 1 : Initialization of the state vector $\hat{\xi}_{1|0}$ wich denotes
             a forcast of $\xi_1$ . The forcast of $X_1$ is given by
             $\hat{X}_{1|0} = H\hat{\xi}_{1|0}$.
    Step 2 : Iterate on $\hat{\xi}_{t+1|t}$ for $t = 2, \dots, N$.
    Step 3 : The forcast of $X_{t+1}$ is given by $\hat{X}_{t+1|t} = H\hat{\xi}_{t+1|t}$.
EndAlgorithm
```

The aim of this present work is to estimate the unknown parameters, $b_{ii}, i = 1, \dots, p$ by means of the maximum likelihood method. In the following we provide the expression of log-likelihood function derived using the Kalman filter algorithm applied to the state space form (4).

Suppose we have observed a sample of size N. The approach will be to calculate the probability density, $f(X;\theta)$, of the process $X = \{X_t\}_{t=1}^N$. This approach requires specifying a particular distribution for the white noise process (e_t). So, we have assumed that (e_t) is Gaussian white noise, $e_t \sim \mathcal{N}(0, \sigma^2)$.
The likelihood function of the process $\{X_t\}_{t=1}^N$ can, then, be expressed by :

$$f(X;\theta) = f(x_1;\theta) \prod_{t=2}^{N} f(x_t|\mathcal{X}_{t-1};\theta), \tag{5}$$

where $\mathcal{X}_t = (x_1, \dots, x_t)$ and $f(x_t|\mathcal{X}_{t-1};\theta)$ denoted the recursively expressed probability density function of X_t given $\mathcal{X}_{t-1}$. The log likelihood function (denoted $L(X;\theta)$) can be found by taking logs of (5) :

$$L(X;\theta) = \log f(x_1;\theta) + \sum_{t=2}^{N} \log f(x_t|\mathcal{X}_{t-1};\theta). \tag{6}$$

The distribution of X_t conditional on $(X_1, X_2, \ldots, X_{t-1})$ is Gaussian with mean $\hat{X}_{t|t-1} = E[X_t|\mathcal{X}_{t-1}]$ and variance $\hat{M}_{t|t-1} = E[(X_t - \hat{X}_{t|t-1})^2]$ (see Hamilton (1994)). Hence, the log likelihood function of θ is :

$$L(X;\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{N}\log(\hat{M}_{t|t-1}) - \frac{1}{2}\sum_{t=1}^{n}\frac{(x_t - \hat{X}_{t|t-1})^2}{\hat{M}_{t|t-1}}. \tag{7}$$

Note that $\hat{X}_{t|t-1}$ and $\hat{M}_{t|t-1}$ can be computed by Kalman filter algorithm presented in the above. Hereafter, we are interesting in minimizing $l(X,\theta) = -L(X,\theta)$, which is clearly equivalent to maximizing $L(X,\theta)$.

3.2 Simulated annealing

Corana et al. (1987) provided a new global optimization algorithm for functions of continuous variables, derived from the simulated annealing algorithm introduced in combinatorial optimization.

This algorithm adopts an iterative random search procedure with adaptive moves along the coordinate directions. It permits uphill moves under the control of a probabilistic criterion, thus tending to avoid the first local minima encountered. Below, we give a brief description of the SA algorithm following the work of corona et al.(1987).

```
SA algorithm
    Step 0 :  Initialize :the vector parameters $\theta$ the step vector $v$ and the temperature $T$.
    Step 1 :  generate a random point $\theta'$ from the point $\theta$ by the rule
              $\theta' = \theta + r\nu_{m_h}d_h$
              $r \in [-1,1]$ is a uniformly distributed random number;
              $e_h$ is the vector of the $h^{\text{th}}$ coordinate direction and $\nu_{m_h} is$
              the component of the step vector $v,$ along the same direction.
    Step 2 :  If $l(X;\theta') \leq l(X;\theta)$ accept the new point
              Else  accept or reject the new point with acceptance probability $p$
              $p = \exp\left(\frac{l(X;\theta)-l(X;\theta')}{T}\right)$
              generate a uniformly distributed random number $p'$ in the range $[0,1]$
              If $p' < p,$ the point is accepted otherwise it is rejected.
              ( see Metropolis et al.(1953))
    Step 3 :  steps 1 to 2 are repeated for each coordinate direction $i, i = 1, \ldots, m.$
              (mis the dimension of the vector parameter)
    Step 4 :  steps 1 to 3 are repeated $N_s$ $times(N_s$ is the number of step variation)
              and the step vector $v$ is adjusted .
    Step 5 :  steps 1 to 4 are repeated $N_T$ $times$
              ($N_T$ is the number of temperature reduction);
              the temperature is reduced following the rule: $T' = R_T T$ with $R_T \in [0,1]$.
    Step 6 :  steps 1 to 5 are repeated until a termination criterion is satisfied.
End Algorithm
```

3.3 Algorithm LKA

Our algorithm named LKA, conserves all the steps of SA algorithm except the step 1 and the step 2.
In the step 1 of our algorithm we implant the invertibility condition, So, it becomes :

```
Step 1 :  Do
              generate a random point $\theta' = (b'_{11}, b'_{22}, \dots, b'_{pp})$ from the point $\theta$,
              by the rule $\theta' = \theta + r\nu_{m_h} d_h, r \in [-1, 1]$
          Until $(p^2 \sum_{i=1}^{p} (b'_{ii}\sigma)^2 < 1)$.
```

The step 2 of our algorithm, where we compute the log-likelihood function of the given value of vector parameters, called the Kalman filter algorithm to alleviate this operation. Therefore, the step 2 can be rewritten as

```
Step 2 :  Called the Kalman filter algorithm
          compute $l(X;\theta')$ and $l(X;\theta)$
          If $l(X;\theta') \leq l(X;\theta)$ accept the new point
          Else  accept or reject the new point with acceptance probability $p$
              $p = \exp\left(\frac{l(X;\theta) - l(X;\theta')}{T}\right)$
              generate a uniformly distributed random number $p'$ in the range $[0, 1]$
              If $p' < p$, the point is accepted otherwise it is rejected.
```

To assess the performance of our algorithm, we have carried out some simulation experiments, illustrated in the next section.

4 Numerical Illustration

In this section, we use data simulated from two models to examine the performance of the technique of estimation we have discussed in the previous sections.
The first model is the pure diagonal bilinear BL(0,0,1,1). The second model, is the pure diagonal bilinear model BL(0,0,2,2). In both examples, we shall consider $(e_t)_{t \in \mathcal{Z}}$ as a Gaussian white noise process with zero mean and variance $\sigma^2 = 1$.

4.1 BL(0,0,1,1) model

Our first example involves data generated from a pure diagonal bilinear model defined by

$$X_t = b_{11} X_{t-1} e_{t-1} + e_t. \tag{8}$$

To investigate the quality of the LKA estimator, we conduct a series of simulation experiments for samples of size $N = 500$.
In each experiment we generate series of length $N = 500$, and then we applied our LKA algorithm to provide the estimation of b_{11}. We repeat this procedure 1000 times to compute

True values	Mean	Bias	MSE	T-statistic
0.1	0.1062	-0.0062	9.9670e-004	-0.1962
0.2	0.199318	6.82e-004	7.6654e-004	0.0246
0.3	0.29057	0.0094	0.0010	0.2914

Table 1: Mean, Bias, MSE and T-statistic of the LKA estimator for N=500

mean, bias, T-statistic and MSE of the LKA estimator. The results of this experiment are displayed in Table (1) for three different values of b_{11}, $b_{11} = 0.1, 0.2, 0.3$.

The numerical results presented in the table above, showed th,at the values of T-statistic exhibit the insignificance of the bias and the mean of LKA estimates are closer to true values of parameters.

Hence, we can conclude that our algorithm succeeds in this first experiment.

4.2 BL(0,0,2,2) model

Our second example involves data generated from a pure diagonal bilinear model defined by

$$X_t = b_{11}X_{t-1}e_{t-1} + b_{22}X_{t-2}e_{t-2} + e_t \qquad e_t \sim iidN(0,1)$$

Considering this model, we generate series of length $N = 500$, and then we applied our LKA algorithm to provide the estimations of b_{11} and b_{22}. We repeat this procedure 1000 times to compute mean, bias, T-statistic and MSE of the LKA estimators. The results of this experiment are displayed in Table (2), for values $b_{11} = 0.05$ and $b_{22} = 0.1$.

Parameters	True values	Mean	Bias	MSE	T-statistic
b_{11}	0.05	0.0499801	0.0000199	0.0000083	0.0068790
b_{22}	0.1	0.0999689	0.0000311	0.0003710	0.0016132

Table 2: Mean, Bias, MSE and T-statistic of the LKA estimator for N=500

One can see that the estimated values are reasonably close to the true values, and that the T-statistic reveal insignificance of the bias

5 Conclusion

In this paper, we consider the maximum likelihood estimation of the parameters of a pure diagonal bilinear time series model. The log-likelihood function constructed by the Kalman filter, is numerically maximized applying simulated annealing.

The results of our simulation study show that our estimation approach succeeds and it performs better.

References

[1] Bhaskara, R. M., Subba, R. T. et Walker, A. M., (1983), On the existence of some bilinear time series models, J. T. S. A. 4, pp. 95–110.

[2] Bouzaachane, K., Harti, M. et Benghabrit, Y., (2006 a), First-order superdiagonal bilinear time series for tracking software reliability, Interstat Journal, Feburary.

[3] Bouzaachane, K., Harti, M. et Benghabrit, Y., (2006 b), Parameter estimation for first-order superdiagonal bilinear time series : An algorithm for maximum likelihood procedure, Interstat Journal, Juillet.

[4] Corana, A., Marchesi, M., Martini C. and Ridella, S., (1987), Minimizing Multimodal functions of continuous variables with simulated annealing Algorithm, ACM Transactions on Mathematical software, 35, pp. 262–280.

[5] Granger, G. W. J and Anderson, A. P., (1978), An introduction to bilinear time series analysis, Vandehoeck and Ruprecht, Götting.

[6] Guegan, D., (1981), Etude d'un modèle non linéare, le modèle superdiagonal d'ordre un, C. R. Acad. sci. Paris ser, 1 293, pp. 95–98.

[7] Guegan, D. and Pham, T. D, (1987), Minimalit et inversibilit des modles bilinaires temps discret, C.R.A.S, 448, pp. 159–162.

[8] Hamilton, J. D., (1994), Time series analysis, Princeton: Princeton University Press.

[9] Kalman, R. E., (1960), A New Approach to Linear Filtering and Prediction Problems,Transaction of the ASME, Journal of Basic Engineering, 82, march.

[10] Kalman, R. E., (1963), New methods in Wiener filtering theory, In John L. Bogdanoff and Frank Kozin, eds. Proceeding of the first symposium of engineering applications of random function Function theory and probability, New York:Wiley, pp. 270–388.

[11] Liu, .J, (1990), A note on causality and invertibility of a general bilinear time series models, Adv. Appl. Prob, 22, pp. 247-250.

[12] Liu, J. and Brockwell, P. J., (1988), On the general bilinear time series models, J. Appl. Prob, 25, pp. 553–564.

[13] Melard, G., (1983), A fast algorithm for the exact likelihood of autoregressive-moving average models, Applied statistics, pp. 104-114.

[14] Metropolis, N., Rosenblut, A., Rosenblut, M., Teller, A. and Teller, E., (1953), Equation of state calculations by fast computing machines, Journal of Chemical Physics, 21, 1087-1090.

[15] Pham, D. T. and Tran, L. T., (1981), On the first order bilinear time series model, J.App. Prob, 18, pp. 617–627.

[16] Subba, R. T., (1981), On the theory of bilinear time series models, J.R.S.S, Serie B, Vol 43, 2, pp. 224–255.

[17] Subba, R. T. and Gabr, M. M., (1984), An introduction to bispectral analysis and bilinear time series models, Lecture note in statistic, N 24, springer, Berlin.