Journal of Mathematical Extension
Vol. 2, N0. 1,2 (2007–2008), 25-44

Discriminant Analysis in Bilinear Processes Using State Space Methodology

R. Chinipardaz
Shahid Chamran University

Abstract. This paper is concerned with discriminating between two bilinear time series processes. The classical methods to discrimination makes stationary and linearity assumptions to processes. However, in general, bilinear processes are not stationary nor linear. The method suggested in this paper is based on state space methodology and uses Cumulative Sum Control Chart. Classification of a series to H_1 or H_2 is treated like a sequential likelihood ratio test where optimal decision is to be made as soon as possible.
The method has been carried out to bilinear processes, but can be used to any adopted processes to state space models.

AMS Subject Classification: Primary 62H30; Secondary 62M10.
Keywords and Phrases: Bilinear Models, CUSUM, Kalman Filtering, State-Space.

1. Introduction

The problem of discriminating between two classes of time series is of great theoretical and practical interest. Shumway ([30]) and Shumway and Stoffer ([33]) summarized some of important areas with applications in various disciplines. Shumway ([30]) reviewed many different methods for stationary time series discrimination in both time domain and frequency domain approaches. Some more and newer references are [2, 11, 12, 21, 32] in frequency domain and [7,8,9,29] in time domain.

Suppose a realization of time series, $\mathbf{x} = (x_1, x_2, \ldots, x_T)$ has to be allocated to one of H_i models, $i = 1, 2$ where

$$H_i : \mathbf{x} \sim p_i(\mathbf{x})$$

and $p_i(\mathbf{x})$ is probability density function of $\mathbf{x}$ under H_i. The likelihood ratio-based discriminant rule is to classify $\mathbf{x}$ to H_1 if $p_1(\mathbf{x})/p_2(\mathbf{x}) \geqslant c$ and to H_2 otherwise. Threshold value of c is dependent on prior probabilities in Bayesian methodology and to one of two error probabilities in nonBayesians. The likelihood ratio approaches are optimal to discrimination and classification ([3,13]). In this approach usually Gaussian specific assumption has been made to processes in addition to stationarity and linearity assumptions.

However, in many practical problems, the series are non-stationary and/or nonlinear. For example, Shumway [32] showed the data set constructed by Blandford [5] were non-stationary. Kakizawa *et al*([21]) introduced some seismic data which are nonlinear (see also, e.g. [28]). Although various nonstationary time series processes have been developed, little attention has been paid to extend the methodology. For example, Priestley [27] was the first to give time-varying transfer function using

Cramer representation. Following this idea [10] established framework for time series with evolutionary spectrum as a locally stationary time series. Subsequently, Shumway ([32]) suggested a mixed frequency and time domain approach to discrimination of time series, (see also [29]).

Recently, Chinipardaz and Cox ([4]) have proposed the likelihood ratio discrimination based on kernel estimated density functions. In this paper a new approach has been suggested to bilinear models which are, in general, nonstatioanary and nonlinear. The method is based on a new look at the Cumulative Sum Control Chart (CUSUM) and is used to discriminate between two bilinear time series.

The paper has been organized as follows; in section two the log-likelihood function for state space models (SSM) is given using Kalman filter (KF) algorithm. The discrimination of Gaussian SSM is considered in third section. The new method is also suggested in this section. Section four is devoted to the discrimination between two moving average models of order one using both the classical method, as given in Chan *et al* (1996), and the suggested method given in the third section of the present paper. A comparison has also been made to show the performance of the new method. In the fifth section the bilinear models are viewed as SSM. Finally, in the last section, the discrimination of the two bilinear models with employing SSM is presented. A numerical simulation has been made to investigate the performance of the method.

2. State Space Models

Consider the following state space models (SSM) for a random process $\boldsymbol{\alpha}_t$,

$$\boldsymbol{\alpha}_t \;= G_t\boldsymbol{\alpha}_{t-1} + R_t\boldsymbol{\eta}_t, \tag{1}$$

where $\boldsymbol{\alpha}_t$ is a $m \times 1$ state vector, G_t is a $m \times m$ matrix, R_t is a $m \times g$ matrix and $\boldsymbol{\eta}_t$ is a $g \times 1$ vector of serially uncorrelated disturbances. Suppose we cannot observe the process $\{\boldsymbol{\alpha}_t\}$ directly, but instead we observe a related measurement process $\{\mathbf{y}_t\}$;

$$\mathbf{y}_t \;= Z_t\boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t \quad t = 1, 2, \ldots, T. \tag{2}$$

where $\mathbf{y}_t$ is the vector of N observed variables at time t, Z_t is an $N \times m$ matrix, the $m \times 1$ state vector and $\boldsymbol{\epsilon}_t$ a $N \times 1$ vector of serially uncorrelated disturbances. Thus $\mathbf{y}_t$, observable at time t, is a known linear transformation of $\boldsymbol{\alpha}_t$ plus a random noise $\boldsymbol{\epsilon}_t$. We want the best least square of $\boldsymbol{\alpha}_t$ base on previous observations $\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}$. For univariate observations design matrix Z_t reduces to a design vector $\mathbf{z}'_t$ and the covariance matrix and the covariance matrix of S_t to variance s_t. In the transition equation (1), G_t a $m \times m$ matrix, R_t is a $m \times g$ matrix and $\boldsymbol{\eta}_t$ a $g \times 1$ vector of serially uncorrelated disturbances. This notation of SSM is what is required in the present paper. A more general setup can be found in [17] and [18]. Let $E(\boldsymbol{\epsilon}_t) = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{\epsilon}_t) = S_t$, $E(\boldsymbol{\eta}_t) = \mathbf{0}, \mathrm{Var}(\boldsymbol{\eta}_t) = Q_t$.

Let $E(\boldsymbol{\alpha}_0) = \mathbf{a}_0, \mathrm{Var}(\boldsymbol{\alpha}_0) = P_0$ and assume $\boldsymbol{\epsilon}_t$, $\boldsymbol{\eta}_t$ are uncorrelated for all t_1, t_2.

The Kalman filter is used to optimally estimate the state vector, $\boldsymbol{\alpha}_t$, using the observed variables up to the present time t. Let the estimate of $\boldsymbol{\alpha}_t$ be $\mathbf{a}_t$, and the mean square error (MSE) matrix of $\mathbf{a}_t$ be P_t. Let the subscripts $t|t-1$ attached to a vector or matrix give its value at time t, given all information up to time $t-1$. Let $\mathbf{Y}_t = (\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1)$. The prediction equations are

$$\mathbf{a}_{t|t-1} = E\left(\mathbf{a}_t | \mathbf{Y}_{t-1}\right) = G_t \mathbf{a}_{t-1} \tag{3}$$

$$\begin{aligned} P_{t|t-1} &= E\left[\left(\mathbf{a}_t - \boldsymbol{\alpha}_t\right)\left(\mathbf{a}_t - \boldsymbol{\alpha}_t\right)' | \mathbf{Y}_{t-1}\right] \\ &= G_t P_{t-1} G_t' + R_t Q_t R_t'. \end{aligned} \tag{4}$$

The estimator of $\mathbf{y}_t$ is

$$\hat{\mathbf{y}}_{t|t-1} = Z_t \mathbf{a}_{t|t-1}.$$

The prediction error is

$$\mathbf{v}_t = \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1} = Z_t(\boldsymbol{\alpha}_t - \mathbf{a}_{t|t-1}) + \boldsymbol{\epsilon}_t,$$

which has MSE matrix

$$F_t = Z_t P_{t|t-1} Z_t' + S_t.$$

The updating equations are then

$$\mathbf{a}_t = \mathbf{a}_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1}\left(\mathbf{y}_t - Z_t \mathbf{a}_{t|t-1}\right) \tag{5}$$

$$P_t = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}. \tag{6}$$

The KF algorithm requires an estimate of the initial state vector and the variance of the error in this estimate, $\mathbf{a}_0$, $\mathbf{P}_0$, as well as a known state vector. The latter can be derived from the using model. For stationary time series models the usual initialization uses the unconditional mean and variance of the initial state, $\boldsymbol{\alpha}_t$, ([1,20]). For nonstationary time series models this approach is not available because unconditional means and variances change over time. Various approaches have been suggested to initialize the KF for nonstationary time series models (see [4,6,22]). In this paper following [4] the variance of the initial state is let to be large for nonstationary bilinear time series models. It is equivalent to diffuse or non-informative prior. Let

$$\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, S_t)\,,$$

$$\boldsymbol{\eta}_t \sim N(\mathbf{0}, Q_t),$$

$$\boldsymbol{\alpha}_0 \sim N(\mathbf{0}, Q_0),$$

assuming that $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are uncorrelated with the initial state vector $\boldsymbol{\alpha}_0$ make the mean and covariance structure of the model fully specified. Let parameters of the models be placed in a vector $\boldsymbol{\psi}$. Then the log-likelihood based on observations up to present time T can be written

$$L(\boldsymbol{\psi}) = \frac{-T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log|F_t| - \frac{1}{2}\sum_{t=1}^{T}\mathbf{v'}_t F_t^{-1}\mathbf{v}_t. \qquad (7)$$

where $\mathbf{v}_t$ and F_t are given as before. (See [18] for further details).

3. Discrimination of Gaussian State Space Models

Suppose an observed time series, $y_1, y_2, \ldots, y_T$, is to be allocated at time T to one of two Gaussian state space models, H_1 and H_2 described by their own measurement and transition equations as in the previous section. The log-likelihood ratio for the two models, giving the discriminant $DF_T(\mathbf{y})$, is from (7)

$$
\begin{aligned}
DF_T(\mathbf{y}) &= \log \frac{L_1(\boldsymbol{\psi})}{L_2(\boldsymbol{\psi})} \\
&= \frac{1}{2}\sum_{t=1}^{T} \log \frac{|F_{2,t}|}{|F_{1,t}|} - \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{v}'_{1,t}F_{1,t}^{-1}\mathbf{v}_{1,t} - \mathbf{v}'_{2,t}F_{2,t}^{-1}\mathbf{v}_{2,t}\right)
\end{aligned}
$$

where the suffices 1 and 2 refer to the models H_1 and H_2 respectively.

If necessary, classification of a series to H_1 or H_2 can be treated like a sequential probability ratio test where an optimal decision is to be made as soon as possible. However, in other situations the decision can be left until the end of the series has been reached. For the former case the theory of sequential probability ratio tests (e.g. [25]) cannot easily be applied to the problem. This is because the distribution of $DF_T(\mathbf{y})$ will usually be impossible to find analytically, hence the probabilities of misclassification cannot be calculated. However, if H_1 is the appropriate model rather than H_2, then the more observations used in the discrimination, the larger the value of $DF_T(\mathbf{y})$. Likewise if H_2 is the appropriate model, then the smaller the value of $DF_T(\mathbf{y})$. Indeed $DF_T(\mathbf{y})$ can be updated with every new observation so that

$$DF_{T+1}(\mathbf{y}) = DF_T(\mathbf{y}) + A_{T+1}$$

where

$$\begin{aligned} A_{T+1} &= \frac{1}{2}\log\frac{|F_{2,T+1}|}{|F_{1,T+1}|} \\ &\quad -\frac{1}{2}\left(\mathbf{v}'_{1,T+1}F^{-1}_{1,T+1}\mathbf{v}_{1,T+1} - \mathbf{v}'_{2,T+1}F^{-1}_{2,T+1}\mathbf{v}_{2,T+1}\right). \end{aligned}$$

Each value A_T can be considered as discriminant information in favour of H_1 or H_2 and $DF_T(\mathbf{y})$ is then the CUSUM of the A'_Ts. However, CUSUM methodology in the area of statistical process control is not appropriate since the A'_Ts are not independent. In practice a plot of the CUSUM $DF_T(\mathbf{y})$ may give overwhelming evidence in favour of H_1 or H_2, if it continually increases or decreases, respectively. When the situation is not so clear the percentage of the A'_Ts supporting H_1 can be noted. As soon as this reaches a certain level (95% say) the series is allocated to H_1 or at another level (5% say) allocated to H_2. Otherwise a further observation is taken. If the end of the series is reached without a decision being made, a final subjective decision will have to be made about H_1 and H_2 based on the A'_Ts.

4. Discrimination between two MA(1) Models in Classical Method and New Method

Chan *et al,* ([7])obtained the classical method to discrimination between two moving average processes of order one. Consider

$$H_i : y_t = \theta_i \epsilon_{t-1} + \eta_t, \quad (i = 1, 2) \quad \eta_t \sim N(0, \sigma^2)$$

The discrimination leads to allocating $\mathbf{y}$ to H_1 if

$$\sum_{j=1}^{T} \frac{(\theta_1 - \theta_2)\left(\theta_1 + \theta_2 + 2\cos\frac{j\pi}{T+1}\right) z_j^2}{\left(1 + \theta^2{}_1 + 2\theta_1 \cos\frac{j\pi}{T+1}\right)\left(1 + \theta^2{}_2 + 2\theta_2 \cos\frac{j\pi}{T+1}\right)} \leqslant \sum_{j=1}^{T} \ln \frac{1 + \theta^2{}_2 + 2\theta_2 \cos\frac{j\pi}{T+1}}{1 + \theta^2{}_1 + 2\theta_1 \cos\frac{j\pi}{T+1}} \tag{1}$$

and to H_2 otherwise, where $z_j = (\frac{2}{T+1})^{\frac{1}{2}} \sum_{k=1}^{T} x_k \sin\frac{jk\pi}{T+1}$, (Chan *et al* [7]). Now, we review this example using the method given in the previous section. Define the state vector according to H_i as

$$\boldsymbol{\alpha}_{i,t} = \begin{bmatrix} y_t \\ \theta_i \eta_t \end{bmatrix}'.$$

The measurement equation is

$$y_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \boldsymbol{\alpha}_{i,t} \quad t = 1, 2, \ldots, T$$

where the state equation is

$$\boldsymbol{\alpha}_{i,t} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \boldsymbol{\alpha}_{i,t-1} + \begin{bmatrix} 1 \\ \theta_i \end{bmatrix} \eta_t.$$

The initial state vector is $\mathbf{0}$ with unconditional covariance matrix (see [18] for more detail)

$$P_{i,1|0} = \begin{bmatrix} 1+\theta_i^2 & \theta_i \\ \theta_i & \theta_i^2 \end{bmatrix}.$$

In this case $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $G_t = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $R_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. All matrices are time invariant, process is stationary and $f_{i,1} = ZP_{i,1|0}Z' = [P_{i,1|0}]_{1,1} = 1+\theta_i^2$, where $[P_{i,1|0}]_{k,l}$ stands for (k,l) th element of matrix $P_{i,1|0}$. The innovation

$$v_{i,t} = y_t - \frac{\theta_i}{f_{i,t-1}} v_{i,t-1}$$

has normal distribution with zero mean and variance $f_{i,t}$. After some algebraic manipulation

$$[P_{i,t|t-1}]_{j,k} = \begin{cases} 1\frac{\theta_i^{2t}}{1+\theta_i^2+\cdots+\theta_i^{2(t-1)}} & j=k=1 \\ \theta_i & otherwise \end{cases}$$

where

$$\begin{aligned} f_{i,t} &= 1 + \frac{\theta_i^{2t}}{1+\theta_i^2+\cdots+\theta_i^{2(t-1)}}, \\ v_{i,t} &= y_t + \sum_{k=1}^{T-1} \frac{(-\theta_i)^k y_{t-k}}{f_{i,t-1}\cdots f_{i,t-k}}, \end{aligned}$$

and

$$f_{i,t-1}\cdots f_{i,t-k} = \frac{1-\theta_i^{2t}}{1-\theta_i^{2(t-k)}}.$$

Then the updating $DF_{j-1}(\mathbf{y})$ after observing new data, y_j, is

$$DF_j(\mathbf{y}) = DF_{j-1}(\mathbf{y}) + A_j \tag{2}$$

where

$$A_j = \frac{1}{2}\log\frac{(1-\theta_1^2)(1-\theta_1^{2j})}{(1-\theta_2^2)(1-\theta_2^{2j})} + \frac{1}{2}\left\{\frac{\sum_{m=0}^{j-1}(-\theta_2)^m(1-\theta_2^{2(j-m)})y_{j-m}}{\sum_{m=0}^{j-1}(-\theta_1)^m(1-\theta_1^{2(j-m)})y_{j-m}}\right\}^2. \quad (3)$$

4.1 Comparison between Two Approaches

The classical method, led to ([8]) and the new method given in ([9]) and ([10]) were compared with the simulation. One thousand time series each with the length 200 was simulated and allocated to H_1 or H_2 according to the two methods. Results are given in Table 1. As can be seen both methods work well with the classical method slightly superior.

Table 1: Misclassification table for 1000 time series of length 200 for MA(1) processes

θ_1	θ_2	I	II
-0.2	0.2	1.4	0.1
0.2	0.4	6.5	2.0
-0.5	0.5	0.0	0.0
0.4	0.7	1.4	0.1
-0.3	-0.5	5.1	2.2
0.1	0.8	0.0	0.0

I: Percentage misclassified with the state space method

II: Percentage misclassified with the classical method

5. Adopting Bilinear Models in SSM

Consider a univariate time series $\{y_t\}$, $t = 1, 2, \ldots, T$ generated by a bilinear model, BL (p, q, k, r),

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \eta_{t-j} + \sum_{i=1}^{k} \sum_{j=1}^{r} \beta_{i,j} y_{t-i} \eta_{t-j} + \eta_t, \tag{1}$$

where $\{\eta_t\}$ is a white noise process and $\{\phi_i\}$, $\{\theta_j\}$ and $\{\beta_{ij}\}$ are real valued constants. General form of a bilinear model has been studied by many authors, including Granger and Anderson ([16]), Pham ([26]), Subba Rao and Gaber ([34]), Liu and Brockwell ([23]) and Grahn ([15]). These models have successfully been used for real data as economics and control theory ([24]).

The discrimination between two bilinear models can be motivated with these two reasons. Firstly, as noted by some authors (e.g. [34]), these models may give a reasonably good approximation to seismological data. Theses data are most interested in discrimination when an unknown observation has to be allocated to one of the known categories, namely explosion or earthquake. Secondly, these models have advantages that they include ARMA models as special case. Unfortunately, so far the discrimination between two bilinear models has not been considered by authors. The main reason is that the discrimination based on log-likelihood ratio leads to a very complicated equation and distribution of the discriminant function is unknown and subsequently misclassification rate cannot be obtained theoretically. In this section the method is adopted to a bilinear model. Note that in this section BL(1, 1, 1, 1) models will be considered. But it can be extended to more complex models of BL(p, q, k, r).

Consider the BL(1, 1, 1, 1) model

$$y_t = \phi y_{t-1} + \theta\eta_{t-1} + \beta\eta_{t-1}y_{t-1} + \eta_t$$

where $\eta_t \sim N(0, \sigma^2)$ and ϕ, θ and β are unknown parameters. This model can be put into the state space form

$$y_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \boldsymbol{\alpha}_t,$$

$$\boldsymbol{\alpha}_t = \begin{bmatrix} y_t \\ \epsilon_t \end{bmatrix} = \begin{bmatrix} \phi & \theta + \beta y_{t-1} \\ 0 & 0 \end{bmatrix}, \quad R_t = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The model is not stationary, nor time invariant. The initial distribution of $\mathbf{a}_0$ will be in terms of a diffuse or non-informative prior, i.e. $P_0 = \kappa I$, where κ is a positive scalar with the diffuse prior being observed as $\kappa \to \infty$ (see [14] and [17]). Now

$$P_{1|0} = G_1 P_0 G_1' + R_1 Q_1 R_1' = \begin{bmatrix} \kappa[\theta^2 + (\theta + \beta y_0)^2] + \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{bmatrix},$$

and

$$F_1 = Z_1 P_{1|0} Z' = [P_{1|0}]_{1,1} = \kappa\left[\phi^2 + (\theta + \beta y_0)^2\right] + \sigma^2,$$

which becomes infinite as $\kappa \to \infty$ and hence the variance of the first error prediction is infinite. Combine (4) and (6) to obtain the recursion formula for the error covariance matrix which is

$$P_{t+1|t} = G_{t+1}\left(P_{t|t-1} - P_{t|t-1}Z_t' F_t^{-1} Z_t P_{t|t-1}\right) G_{t+1}' + R_{t+1}Q_{t+1}R_{t+1}' \quad , t = 1, \ldots, T.$$

After some algebra

$$[P_{t+1|1}]_{j,k} = \begin{cases} \sigma^2 \left\{1 + (\theta + \beta y_{t-1})^2 \left(1 - \frac{\sigma^2}{[P_{t|t-1}]_{1,1}}\right)\right\} & j = k = 1 \\ 0 & \textit{otherwise.} \end{cases}$$

The recursion for the variance of the error prediction is

$$\begin{aligned} f_t &= Z_t P_{t|t-1} Z_t' + S_t = \left[P_{t|t-1}\right]_{1,1} \\ &= \sigma^2 \left\{1 + (\theta + \beta y_{t-1})^2 \left(1 - \frac{\sigma^2}{f_{t-1}}\right)\right\} \quad t = 1, \ldots, T, \end{aligned}$$

since $\mathrm{Var}(\eta_t) = s_t = 0$.

Now because $f_1 \to \infty$ as $\kappa \to \infty$, f_2 can be taken as

$$\sigma^2 \left\{1 + (\theta + \beta y_{t-1})^2\right\},$$

and since T will be finite, subsequent values of f_t can be found starting with f_2. Let the estimate of the state vector $\boldsymbol{\alpha}_t$ at time $t-1$ be $\mathbf{a}_{t|t-1}$. The recursion for $\mathbf{a}_{t|t-1}$ is given by

$$\mathbf{a}_{t+1|t} = (G_{t+1} - K_t Z_t)\, \mathbf{a}_{t|t-1} + K_t y_t,$$

where

$$K_t = G_{t+1} P_{t|t-1} Z_t' F_t^{-1}.$$

After some algebra

$$\mathbf{a}_{t+1|t} = \begin{bmatrix} (\theta + \beta y_t) \left\{ \frac{\sigma^2}{f_t} \left(y_t - \mathbf{a}_{t|t-1}(1) \right) \right\} + \phi y_t \\ 0 \end{bmatrix}, \quad t = 1, \ldots, T.$$

where $\mathbf{a}(m)$ means the mth element of vector $\mathbf{a}$. As $f_1 \to \infty$, $a_{2|1} \to y_1$ as $\kappa \to \infty$. The error prediction is therefore obtained as

$$\begin{aligned} v_{t+1} &= y_{t+1} - Z_{t+1} \mathbf{a}_{t+1|t} \\ &= \begin{bmatrix} y_2 - \phi y_1 & t = 1 \\ y_{t+1} - (\theta + \beta y_t) \left\{ \frac{\sigma^2}{f_t} \left(y_t - \mathbf{a}_{t|t-1}(1) \right) \right\} + \phi y_t, & t = 2, \ldots, T-1 \end{bmatrix} \end{aligned}$$

The error prediction at step k, v_k, has a normal distribution with zero mean and variance f_t. Thus the log-likelihood function for the kth step is

$$L_k(\mathbf{y}) = L(y_2, \ldots, y_k) = -\frac{1}{2} \left[\sum_{t=2}^{k} \log f_t + \sum_{t=2}^{k} \frac{v_j^2}{f_t} \right] - \frac{k}{2} \log 2\pi, \quad k = 2, \ldots, T.$$

6. Discrimination of Bilinear Models

The log-likelihood ratio for $H_1 : \boldsymbol{\psi} = (\theta_1, \phi_1, \beta_1)$ and $H_2 : \boldsymbol{\psi} = (\theta_2, \phi_2, \beta_2)$ reduces to

$$DF_k(\mathbf{y}) = \frac{1}{2} \sum_{t=2}^{k} \log \frac{f_{2,t}}{f_{1,t}} + \frac{1}{2} \sum_{t=2}^{k} \left\{ \frac{v_{2,t}^2}{f_{2,t}} - \frac{v_{1,t}^2}{f_{1,t}} \right\} \quad k = 2, \ldots, T.$$

where

$$f_{i,t} = \begin{cases} \sigma^2 \left\{1 + (\theta_i + \beta_i y_{t-1})^2\right\} & t = 2 \\ \sigma^2 \left\{1 + (\theta_i + \beta_i y_{t-1})^2 \left(1 - \frac{\sigma^2}{f_{i,t-1}}\right)\right\} & t = 3, \ldots, T \end{cases}$$

and

$$v_{i,t} = \begin{cases} y_2 - \phi_i y_1 & t = 2 \\ y_t - (\theta_i + \beta_i y_{t-1}) \left\{\frac{\sigma^2}{f_{i,t-1}} \left(y_{t-1} - \mathbf{a}_{i,t-1|t-2}(1)\right)\right\} & \\ \qquad +\phi_i y_{t-1}, & t = 3, \ldots, T, \end{cases}$$

for $i = 1, 2$ and $\mathbf{a}_{i,t|t-1}(1)$ is given in (3).

Then as before the log-likelihood ratio can be used sequentially to discriminate between H_1 and H_2.

6.1 Simulation Study

A simulation exercise was carried out to assess the discrimination of some bilinear processes using CUSUM approach. Models were chosen for H_1 and H_2 respectively. One thousand time series, which with the length 500 were simulated from each of the models. Every series was then allocated to H_1 or H_2 but with the possibility of no decision. A 95% decision level was chosen as discussed in section 3. Results are shown in Table 2. The various models chosen are shown together with the number of correct and incorrect allocations. Also shown is the average number of points needed before a decision was made. These results show that the method works well.

Table 2: The number of correct allocations and wrong allocations for 1000 time series of length 500

Models: $H_1: \ y_t + 0.1y_{t-1} + 0.1\epsilon_{t-1} + 0.1y_{t-1}\epsilon_{t-1} = \epsilon_t$
$H_2: \ y_t + 0.2y_{t-1} + 0.2\epsilon_{t-1} + 0.2y_{t-1}\epsilon_{t-1} = \epsilon_t$

Series generated from	No. allocated to H_1	H_2	No final allocation	Av. No. of points until decision
H_1	704	71	225	120
H_2	96	640	264	131.95

Models: $H_1: y_t - 0.2y_{t-1} - 0.2\epsilon_{t-1} - 0.2y_{t-1}\epsilon_{t-1} = \epsilon_t$
$H_2: y_t + 0.2y_{t-1} + 0.2\epsilon_{t-1} + 0.2y_{t-1}\epsilon_{t-1} = \epsilon_t$

Series generated from	No. allocated to H_1	H_2	No final allocation	Av. No. of points until decision
H_1	989		4 7	78.6
H_2	3	987	10	80.27

Models: $H_1: y_t + 0.2y_{t-1} + 0.2\epsilon_{t-1} + 0.2y_{t-1}\epsilon_{t-1} = \epsilon_t$
$H_2: y_t + 0.4y_{t-1} + 0.4\epsilon_{t-1} + 0.4y_{t-1}\epsilon_{t-1} = \epsilon_t$

Series generated from	No. allocated to H_1	H_2	No final allocation	Av. No. of points until decision
H_1	923	22	55	116.68
H_2	25	914	61	108.43

Models: $H_1: y_t - 0.5y_{t-1} + 0.2\epsilon_{t-1} + 0.1y_{t-1}\epsilon_{t-1} = \epsilon_t$
$H_2: y_t + 0.5y_{t-1} + 0.2\epsilon_{t-1} + 0.1y_{t-1}\epsilon_{t-1} = \epsilon_t$

Series generated from	No. allocated to H_1	H_2	No final allocation	Av. No. of points until decision
H_1	995	3	2	69.76
H_2	4	994	2	61.24

Models: $H_1 : y_t + 0.1y_{t-1} - 0.4\epsilon_{t-1} + 0.2y_{t-1}\epsilon_{t-1} = \epsilon_t$
$H_2 : y_t + 0.1y_{t-1} + 0.4\epsilon_{t-1} + 0.2y_{t-1}\epsilon_{t-1} = \epsilon_t$

Series generated from	No. allocated to H_1	H_2	No final allocation	Av. No. of points until decision
H_1	994	5	1	77
H_2	5	988	7	77.33

Models: $H_1 : y_t + 0.1y_{t-1} + 0.1\epsilon_{t-1} - 0.5y_{t-1}\epsilon_{t-1} = \epsilon_t$
$H_2 : y_t + 0.1y_{t-1} + 0.1\epsilon_{t-1} + 0.5y_{t-1}\epsilon_{t-1} = \epsilon_t$

Series generated from	No. allocated to H_1	H_2	No final allocation	Av. No. of points until decision
H_1	993	2	5	75.39
H_2	0	991	9	73.47

References

[1] H. Akaike, Covariance matrix computation of the state variable of a stationary Gaussian Process, *Ann. Inst. Statist. Math. Part B*, 30 (1978), 499-504.

[2] J. Alagon, Spectral Discrimination for Two Group of Time Series, *J. Tim. Ser. Anal.*, 10(3) (1989), 203-214.

[3] T. W. Anderson and R. R. Bahudur, Classification into two multivariate normal distribution with different covariance matrix,*Ann. Math. Stat.,* 33 (1989), 420-431.

[4] W. Bell and S. Hillmer, Initializing the Kalman Filter for nonstationary time series models, *J. Tim. Ser. Anal.,* 12 (4) (1989), 283-300.

[5] R. R. Blandford, Discrimination of Earthquakes and Explosions, AFTAC-TR-93-044 HQ. Air force Technical Application Center, *Patrick Air Force Base*, F1 (1993).

[6] P. Burridge and K. F. Wallis, Calculating the variance of seasonally adjusted series, *J. Am. Statist. Assoc.,* 80 (1985), 541-552.

[7] H. T. Chan, R. Chinipardaz, and T. F. Cox, Discrimination of AR , MA and ARMA time series models, *Comm. Stat.- Theory and Methods,* 26 (1996), 1247-1260.

[8] G. Chaudhuri, J. D. Borwanker, and P. R. K. Rao, Bhattacharyya Distance-Based Linear Discriminant Function for Stationary Time Series, *Comm. Stat.- Theory and Methods,* 20(7) (1991), 2195-2205.

[9] R. Chinipardaz and T. F. Cox, Nonparametric Discrimination of Time series, *Metrika,* 59 (2004), 13-20.

[10] R. Dahlhaus, Fitting time series models to nonstationary processes, *Annals of Statistics,* 25 (1997), 1-37.

[11] R. Dahlhaus, On the Kullback-Leibler Information Divergence of Locally Stationary Processes, *Stochastic Processes and Their Applications,* 62 (1996), 139-168.

[12] G. R. Dargahi-Noubary, Discrimination between gaussian time series Based on Their Spectral Differences, *Comm. Stat.- Theory and Methods,* 21(9) (1992), 2439-2458.

[13] G. R. Dargahi-Noubary, *Time series Applications to Seismology (1999).* NOVA Science Publishers, Commack.

[14] G. R. Dargahi-Noubary and P. J. Laycock, Spectral Ratio Discriminants and Information Theory, *J. Tim. Ser. Anal.,* 2 (1981), 71-86.

[15] J. Fan and Q. Yao, *Nonlinear Time Series; Nonparametrics and Parametrics Methods (2003).* Springer, New York.

[16] T. Grahn, A Conditional Least Square Approach to Bilinear Time Series Estimation, *Tech. Report, (1994),* University of Heidelberg.

[17] C. V. J. Granger and P. Anderson, *An Introduction to Bilinear Time Series Models, (1984),* Vandenhock and Ruprecht, Gottingen.

[18] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter (1989),* Cambridge University Press, New York.

[19] A. C. Harvey, *Time Series Models, (1993),* Harvester, Wheatshaf.

[20] A. C. Harvey and G. P. A. Phillips, Maximum likelihood estimation of regression models with autoregressive-moving average disturbances, *Biometrika,* 66 (1979), 49-58.

[21] R. H. Jones, Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics,* 22 (1981), 389-395.

[22] G. Kakizawa, R. H. Shumway, and M. Taniguchi, Discrimination and clustering for multivariate time series, *J. Am. Statist. Assoc.,* 93(441) (1998), 328-340.

[23] G. Kitagawa, A nonstationary time series model and its fitting by a recursive filter, *J. Tim. Ser. Anal.,* 2 (1981), 103-116.

[24] J. Liu and P. J. Brockwell, On the general bilinear time series model, *J. App. Probab.* 25 (1988), 553-564

[25] R. R. Mohler and R. R. Bilinear, *Control Processes*, Academic Press, (1973), New York.

[26] A. F. Mood, F. A. Graybill, and D. C. Boes, *Introduction to Theory of Statistics,* (1974), McGraw-Hill, New York.

[27] D. T. Pham, Bilinear Markovian representation and bilinear models, *Stochastic Processes and Their Applications,* 23, (1985), 291-300.

[28] M. B. Priestley, Evolutionary spectra and non-stationary processes, *Journal of the Royal Society, B,* 28 (1965), 228-240.

[29] M. B. Priestley,*Non-Linear and Non-Stationary Time Series Analysis* (1991), Academic Press, London.

[30] K. Sakiyama M. and R. Taniguchi, Discriminant Analysis for Locally Stationary Processes, *Journal of Multivariate Analysis.* 90 (2004), 282-300.

[31] R. H. Shumway, *App. Stat. Tim. Ser. Anal.* (2004), Prentice Hall, New Jersey.

[32] R. H. Shumway, Discriminant Analysis for Time Series. *Handbook of Statistics,*2 (2004), 1-46, eds. P. R. Krishnaiah and L. N. Kanal Amsterdam: North-Holland.

[33] R. H. Shumway, Time-Frequency Evolutionary Clustering and Discriminant Analysis, *Statistics and Probability Letters,* 63 (3) (2003), 307-314.

[34] R. H. Shumway and D. S. Stoffer, *Time series Analysis and Its Applications,* (2000), Springer-Verlag, New York.

[35] T. Subba Rao and M. M. Gaber, *An Introduction to Bispectral Analysis and Bilinear Time Series,* (1984). Springer-Verlag, New York.

Rahim Chinipardaz
Department of Mathematics
College of Mathematics and Computer Science
Shahid Chamran University
Ahvaz, Iran
E-mail: chinipardaz_r@scu.ac.ir