Iranian Int. J. Sci. ***2****(1) 2001*

A CUSUM Approach to Discrimination of Gaussian Time Series Models

Rahim Chinipardaz
Department of Statistics, University of Shahid Chamran, Ahwaz, Iran

Abstract

Discrimination of time series data can be carried out either in the frequency domain or in the time domain. Applications have been found useful in certain areas such as seismology and medicine. To date most discriminative work has been carried out for linear and stationary time series models. This paper extends the theory by considering the discrimination of state space models which can then also include linear models. A CUSUM approach is adopted to discriminate between two models, although standard CUSUM methods can not be used because of the lack of independence. The CUSUM method allows the discrimination to be viewed as a sequential method where a decision needs to be made as soon as reasonably possible.

Key words*: ARMA, CUSUM, discrimination, state space*

1. Introduction

The problem of the discrimination of time series data has been considered by several authors. Shumway and Unger (1974), Dargahi-Noubary and Laycock (1981), Alagon (1989) and Dargahi-Noubary (1992) use a frequency domain approach to establish the discriminant between two time series models. Gersch (1981), Gersch and Yonemoto (1977), Gersch *et al.*, (1979), Alagon (1986), Chaudhari *et al*., (1991) and Chan, *et al.,* (1996) consider a time domain approach.

In the frequency domain approach, Shumway and Unger (1974) obtained the best linear discriminant when the mean functions of the two models are assumed to be unequal. Dargahi-Noubary and Laycock (1981) and Dargahi-Noubary (1992) look at the spectral ratio for two models when mean functions are assumed equal with the discriminative information residing in the covariance or spectral structures. They suggested using a subset of the frequency bands for discrimination.

In the time domain approach the log-likelihood ratio is usually considered as a criterion for discrimination between two models. For Gaussian models, the log-likelihood ratio is a quadratic form and involves cumbersome matrices. The discriminant function can be expressed in terms of a linear combination of independent chi-squared random variables each with one degree of freedom, and where the coefficients are the eigenvalues of a matrix based on the covariance matrices for the two models. The eigenvalues have to be calculated numerically. Chan, *et al.*, (1996) have obtained an approximate analytic solution for the coefficients for ARMA processes. In a Bayesian approach using predictive discrimination, Geisser (1966) developed a discriminant rule based on the prediction of a random vector, given the training data. This technique was extended by Brolemeling and Son (1987) who derived the marginal posterior mass function of the classification vector for an AR process.

Practical applications of time series discrimination arise in seismology, medicine and other areas. See Shumway (1982, 1988) for a good introduction to the subject. Most results so far are concerned with stationary time series models. This paper investigates the log-likelihood ratio discriminant for state space models, thus enabling some non-stationary models to be encompassed into the discrimination of the time series models area. Also the time domain discrimination of stationary models, e.g. ARMA, can be viewed differently by placing them in state space form.

2. Gaussian state space models

Following the notation of Harvey (1993) state space models are defined by

$\mathbf{y}_t = Z_t ?‡_t + \mathbf{d}_t + ?_t$ (1)

$?‡_t = T_t ?‡_{t-1} + \mathbf{c}_t + R_t ?‡_t$ (2)

where in the measurement equation (1), $\mathbf{y}_t$ is the vector of N observed variables at time t, Z_t is an $N ?Ł m$ matrix, ?Ł is the state vector, $\mathbf{d}_t$ is a vector, and $?_t$^ a vector of serially uncorrelated disturbances. In the transition equation (2), T_t a matrix, $\boldsymbol{c}_t$ a vector, R_t is a matrix and ?ħ a vector of serially uncorrelated disturbances.

Let E(?$_t$€) = 0, Var (?$_t$€) = S_t, E(?$_t$€) = 0, Var (?$_t$€) = Q_t. Let E (?$_t$€) = a_0, Var (?$_0$) = P_0 and assume ?$_t$, ?$_t$ are uncorrelated for all t_1, t_2.

The Kalman filter is used to optimally estimate the state vector, ?'$_t$, using the observed variables up to present time t. Let the estimate of ?$_t$ be a_t, and the mean square error (MSE) matrix of ?ƿ be P_t. Let the suffices $t|t-1$ attached to a vector or matrix give its value at time t, given all information up to time $t-1$.

The prediction equations are

$$\mathbf{a}_{t|t-1} = T_t\,\mathbf{a}_{t-1} + \boldsymbol{c}_t$$
$$P_{t|t-1} = T_t P_{t-1} T_t' + R_t Q_t R_t'.$$

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

$$\hat{\mathbf{y}}_{t|t?1} ? Z_t \mathbf{a}_{t|t?1} ? \mathbf{d}_t\,.$$

The prediction error is

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

which has MSE matrix

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

The updating equations are then

$$\mathbf{a}_t ? \hat{\mathbf{a}}_{t|t?1} ? \hat{P}_{t|t?1} Z_t ? F_t^{?1} (\mathbf{y}_t ? \hat{Z}_t \mathbf{a}_{t|t?1} ? \hat{\mathbf{d}}_t)$$
$$P_t ? P_{t|t?1} ? P_{t|t?1} Z_t ? F_t^{?1} Z_t P_{t|t?1}.$$

For Gaussain models ?$_t$, and ?$_t$ are normally distributed. Let parameters of the models be placed in a vector ?>. Then the log-likelihood based on observations up to present time T can be written

$$L(\boldsymbol{\Psi}) = \frac{-NT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log|F_t| - \frac{1}{2}\sum_{t=1}^{T}\mathbf{v}_t' F_t^{-1}\mathbf{v}_t \qquad (3)$$

See Harvey (1993) for further details.

3. Discrimination of Gaussian state space models

Suppose an observed time series, $\mathbf{y}_1$, $\mathbf{y}_2$, ... , $\mathbf{y}_T$, 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 (3)

$$DF_T(\mathbf{y}) = \frac{1}{2}\sum_{t=1}^{T}\log\frac{|F_{2,t}|}{|F_{1,t}|} - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{v}_{1,t}' F_{1,t}^{-1}\mathbf{v}_{1,t} - \mathbf{v}_{2,t}' F_{2,t}^{-1}\mathbf{v}_{2,t})$$

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

It 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 situation 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. Mood, *et al.*, (1974)) cannot easily be applied to the problem. This is because the distribution of $DF_T(\mathbf{y})$ will usually be impossible to find analytically and hence the probabilities of misclassification can not 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

$$A_{T+1} = \frac{1}{2}\log\frac{|F_{2,T+1}|}{|F_{1,T+1}|} - \frac{1}{2}\left(\mathbf{v}_{1,T+1}' F_{1,T+1}^{-1}\mathbf{v}_{1,T+1} - \mathbf{v}_{2,T+1}' F_{2,T+1}^{-1}\mathbf{v}_{2,T+1}\right)$$

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_T's. However CUSUM methodology in the area of statistical process control is not appropriate since the A_T's are not independent. In practice a plot of the

CUSUM $DF_T(\mathbf{y})$ may give overwhelming evidence in favor of H_1 or H_2, if it continually increases or decrease, respectively. When the situation is not so clear the percentage of the A_T's 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) is 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_T's.

4. Discrimination of ARMA (p,q) processes

Suppose the models H_1 and H_2 are both taken as ARMA(p,q) processes. The log likelihood ratio discriminant and its approximate distribution has been obtained by Dargahi-Noubary and Laycock (1981) and Dargahi-Noubary (1992) using a frequency domain approach, and by Chan *et al.*, (1996) using a time domain approach, respectively, as discussed in section 1. It may however be more advantageous to express the ARMA processes in state space form. Let the ARMA(p,q) process be

$$\mathbf{y}_t = Z\alpha_t$$

where

$$Z = (1, 0, \dots , 0),$$

and the state equation is

$$\alpha_t = T\,\alpha_{t-1} + R\varepsilon_t$$

where the $m \times m$ matrix T and the $m\times 1$ vector R are given by

$$T = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi_p & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \qquad R = \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_q \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where $m = \max\{p, q+1\}$, and ε_t are uncorrelated with zero mean and variance σ^2.

It can be shown that the prediction error $v_t = y_t - \hat{y}_{t|t-1}$ is normally distributed with zero mean and variance $F_t = ZP_{t|t-1}Z' = [P_{t|t-1}]_{1,1}$ where

$P_{t|t-1}$ is the MSE matrix of $\mathbf{a_t}$ at time *t-1*. The log-likelihood ratio for the two models, H_1 and H_2 becomes

$$DF_T(y) = -\frac{1}{2}\sum_{t=1}^{T}\log\frac{|F_{2,t}|}{|F_{1,t}|} - \frac{1}{2}\sum_{t=1}^{T}\left(\frac{v_{1,t}^2}{F_{1,t}} - \frac{v_{2,t}^2}{F_{2,t}}\right). \qquad (4)$$

It can be shown that the elements of $P_{j,t|t-1}$, j referring to $H_{2,}$ can be obtained by the recursion

$$[P_{j,t|t-1}]_{j,k} = \begin{cases} [P_{j,t-1|t-2}]_{i+1,k+1} - \dfrac{[P_{j,t-1|t-2}]_{i+1,1}\,[P_{j,t-1|t-2}]_{1,k+1}}{[P_{j,t-1|t-2}]_{1,1}} + \sigma_j^2\theta_{j,i}\theta_{j,k} & 1 \le i,k \le m-1 \\ \sigma_j^2\theta_{j,m}\theta_{j,i} & k = m, i \le m-1 \\ \sigma_j^2\theta_{j,m}\theta_{j,k} & i = m, k \le m \\ 0 & \text{otherwise.} \end{cases}$$

If the models are stationary and invertible, after m = {*p*, *q+1*} replications, the value of $F_t = [P_{t|t-1}]_{1,1}$ is equal to σ^2 because

$$[P_{j,t|t-1}]_{i,k} = \begin{cases} \sigma^2 & i = k = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Hence the variance of $v_{j,t}$ will be constant and stable over time. For this case the transition equation is time invariant and the estimate of the initial state vector and variance of the error are the unconditional mean vector and variance matrix of the initial state vector. The mean can be shown to be the zero vector and the covariance matrix

$$\text{vec}(P_0) = (I - T \otimes T)^{-1}\,\text{vec}(\sigma^2 RR')$$

where ⊗ is the Kronecker product and vec (·) denotes that the columns of the matrix are stacked one upon another. See Gardner *et al.,* (1980) or Harvey (1989) for more details.

4.1 Discrimination of two MA(1) processes

As an example consider the discrimination between two MA(1) processes

$$H_j: y_t = \varepsilon_t + \theta_j\varepsilon_{t-1} \qquad (j = 1, 2).$$

For H_j, define the state space vector as

$$\alpha_t = (y_t, \ \theta_j \varepsilon_t)'.$$

Then the measurement equation is

$$y_t = (1, \ 0)\,\alpha_t \qquad t = 1, \ldots, T$$

where

$$\alpha_t = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \alpha_{t-1} + \begin{pmatrix} 1 \\ \theta_j \end{pmatrix} \varepsilon_t$$

The initial state vector is 0 with unconditional covariance matrix

$$P_{1|0} = \begin{pmatrix} 1+\theta_j^2 & \theta_j \\ \theta_j & \theta_j^2 \end{pmatrix}$$

Then

$$[P_{j,t|t-1}]_{i,k} = \begin{cases} \dfrac{\theta_j^{2t}}{1+\theta_j^2+\cdots+\theta_j^{2(t-1)}} & i = k = 1 \\ \theta_j^2 & i = k = 2 \\ 0 & \text{otherwise.} \end{cases}$$

and

$$v_{j,t} = y_t - \frac{\theta_j}{f_{j,t-1}} v_{j,t-1}$$

has a normal distribution with zero mean and variance $f_{j,t}$. After some algebra the log-likelihood ratio for H_1 and H_2 based on observations up to time T is

$$DF_T(y) = \frac{1}{2}\sum_{k=1}^{T} \ln\frac{f_{2,k}}{f_{1,k}} + \frac{1}{2}\sum_{k=1}^{T}\left(\frac{v_{2,k}^2}{f_{2,k}} - \frac{v_{1,k}^2}{f_{1,k}}\right), \qquad (5)$$

where

$$v_{j,t-1} = y_t + \sum_{k=1}^{T-1} \frac{(-\theta_j)^k \, y_{t-k}}{f_{j,t-1} \cdots f_{j,t-k}}.$$

and

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

Equation (5) is the same as that obtained from the classical approach although possibly more difficult to calculate than for this state space form.

5. Simulation Studies

Simulation exercises were carried out to investigate the CUSUM approach. Figure 1 shows the results of one such exercise.

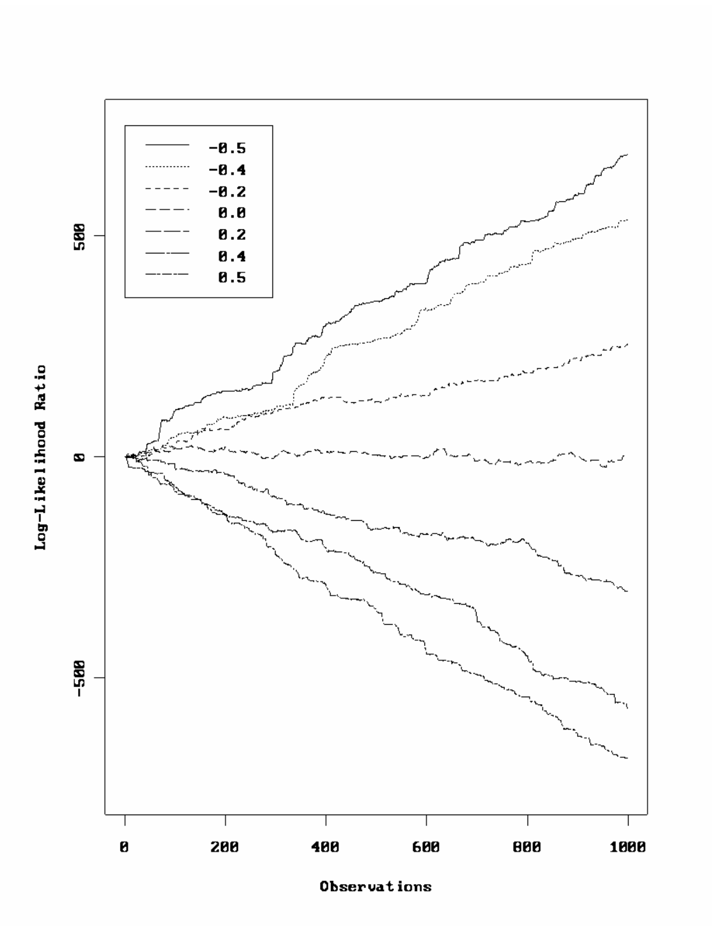


Figure 1 - Log-likelihood ratio for two MA(1) processes(H_1 = -0.5 and H_2 = 0.5) together with different parameter values.

The models for H_1 and H_2 were $\theta_1 = -0.5$, $\theta_2 = 0.5$ respectively. A time series of total length 1000, generated according to H_1 gave the discriminative CUSUM, the solid line in the figure (The upper CUSUM). A time series generated according to H_2 gave the lower CUSUM in the figure. As these CUSUMs progress they give more and more evidence as leaning to the appropriate model. Additional time series were generated with ?»taking various values between -0.5 and 0.5, and although neither model is the correct one, evidence as to one or the other can be seen from the various CUSUMs that reside between the initial two.

Figure 2 shows the log likelihood ratio for H_1 and Models H_2 for a time series with length 1000. Each plot in this figure shows the log likelihood ratio for two different models with parameters written in brackets. Observations in this simulation were taken from H_1. As can be seen from this figure the log likelihood ratio increase with time. Comparison between plots shows that the log likelihood ratio increases more if the difference between the two models is large.

This simulation was repeated with observation taken from H_2. Figure 3 shows the log likelihood ratio against time. As can be seen the log likelihood ratio decreases if observations are taken from H_2. Again this figure shows that the log likelihood ratio will decrease more if the difference between two models is large.

Figure 4 and 5 show simulations for two AR(2) processes and two ARMA(1,1) processes, respectively. Similar results also occur.

Another simulation exercise was carried out to assess the discrimination of some ARMA processes. Models were chosen for H_1 and H_2 respectively. Then one thousand time series each of 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 1. 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 and others show that the method works well.

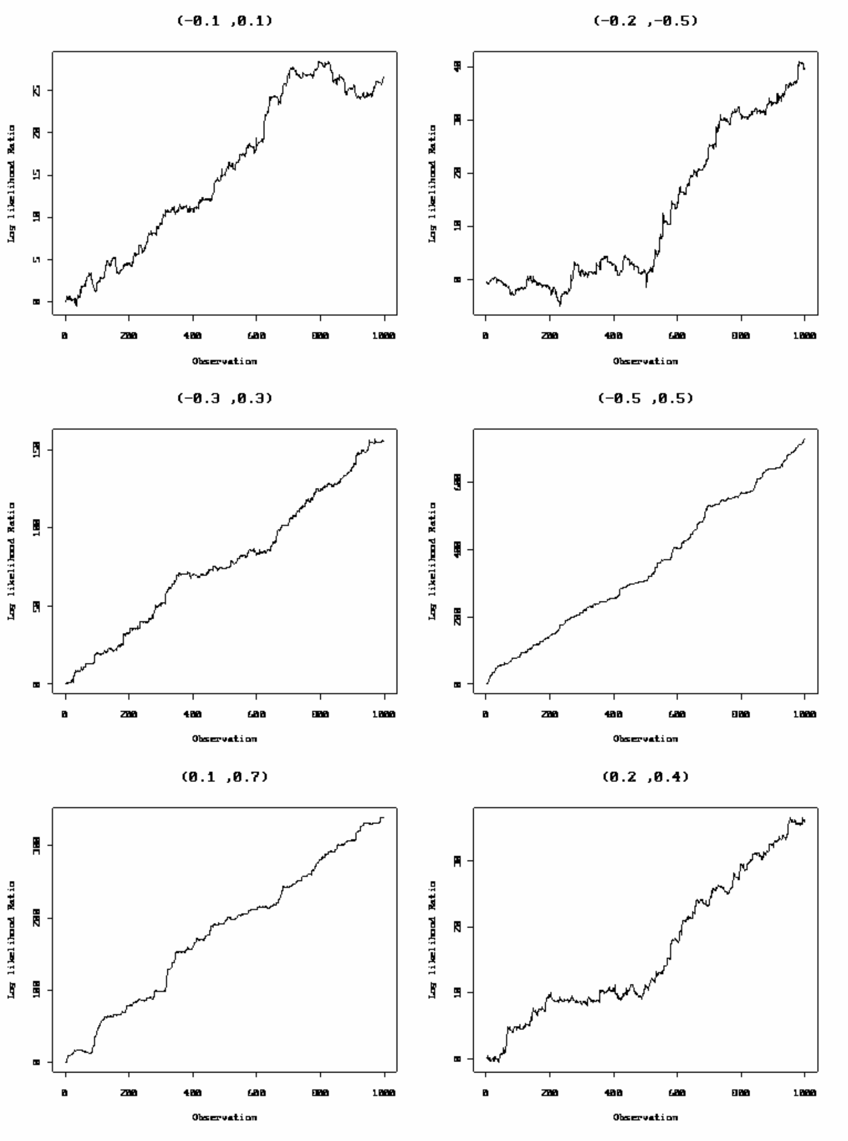


Figure 2 - Discrimination (L.L.R) between two MA(1) models, (H_1 and H_2). Observations from H_1.

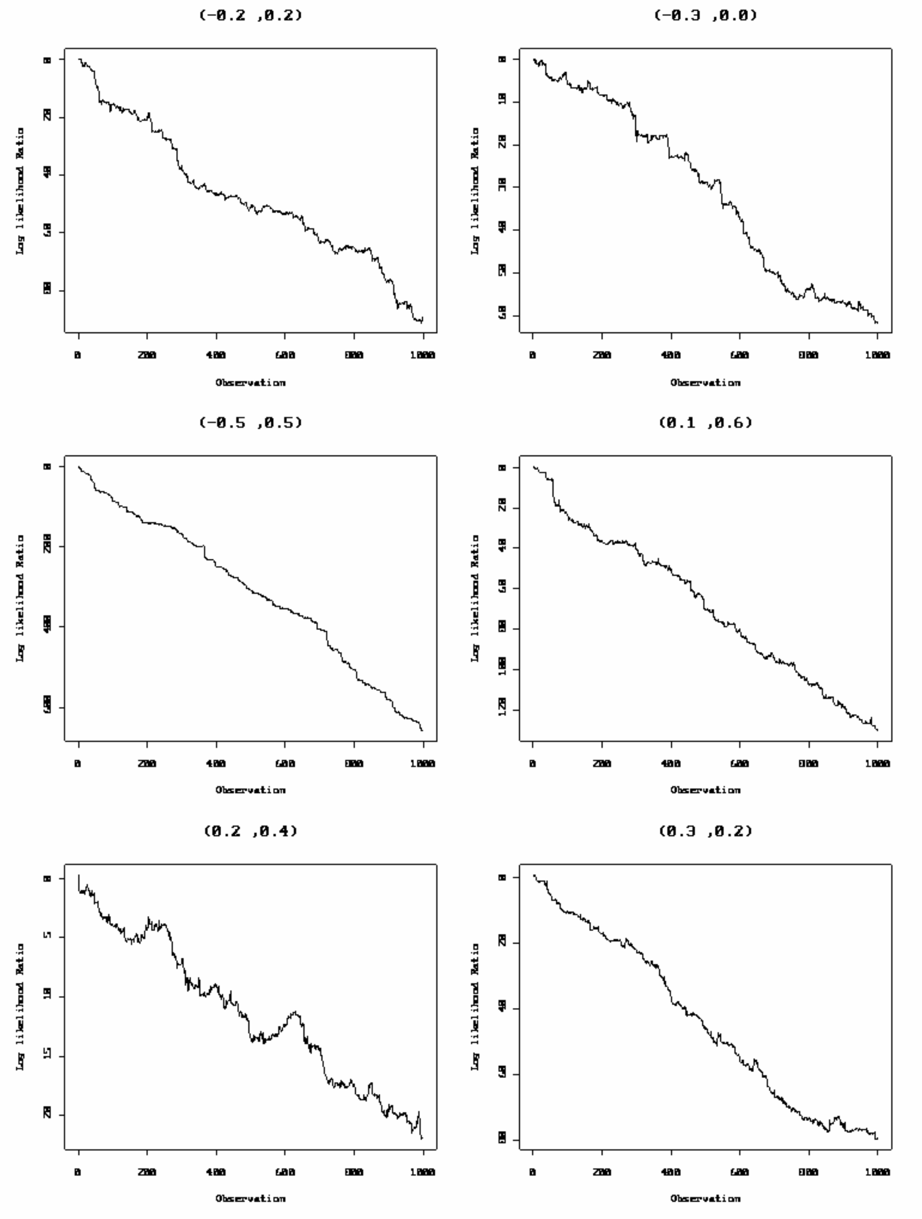


Figure 3 - Discrimination (L.L.R) between two MA(1) models, (H_1 and H_2). Observations from H_2.

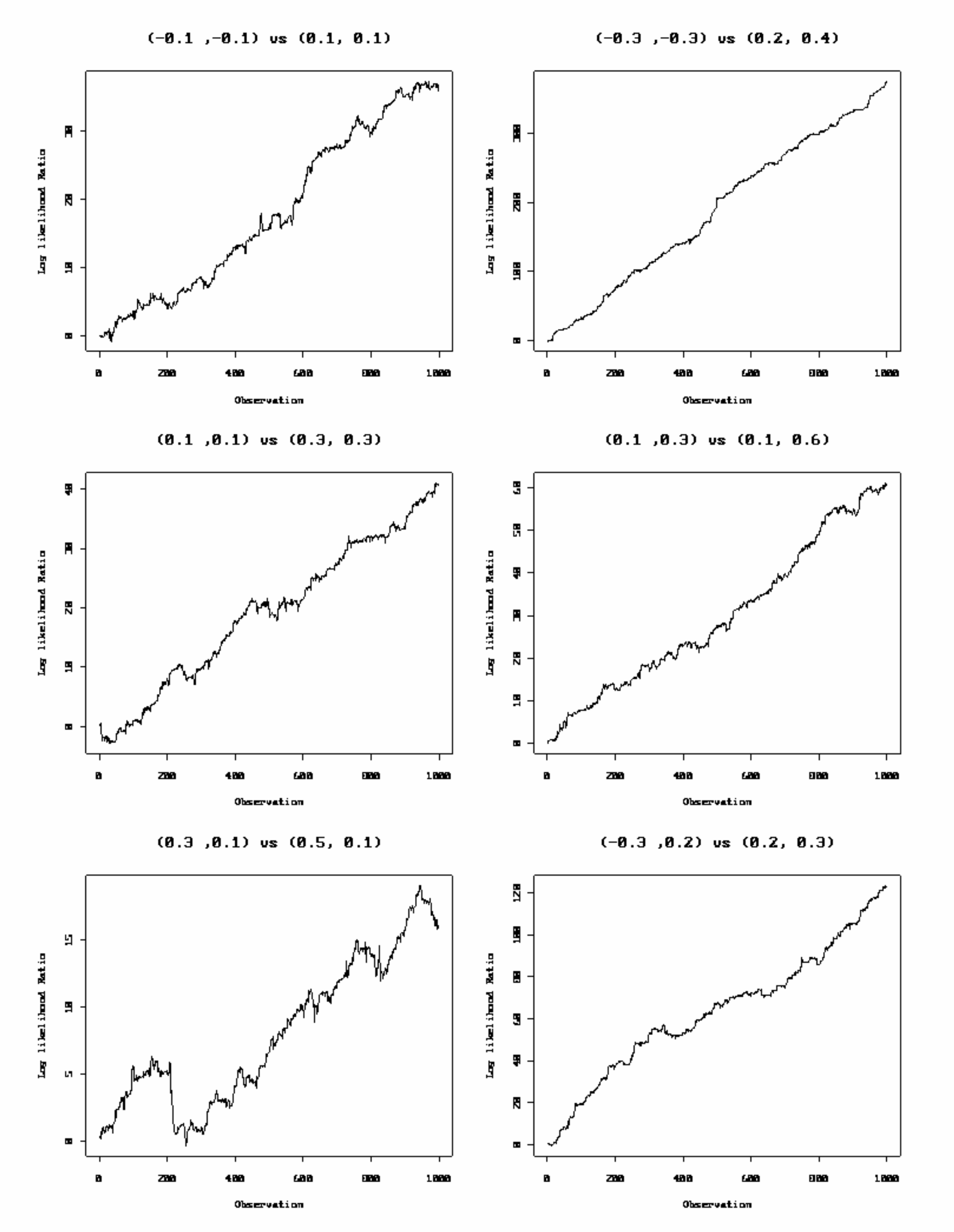
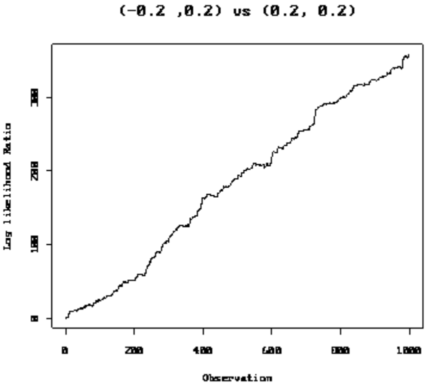
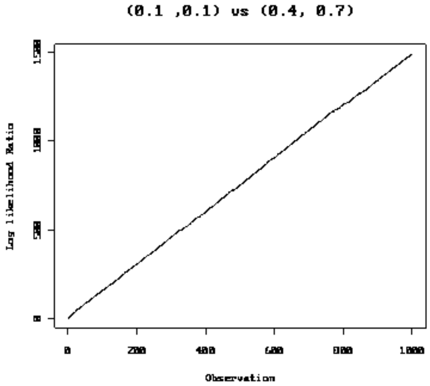
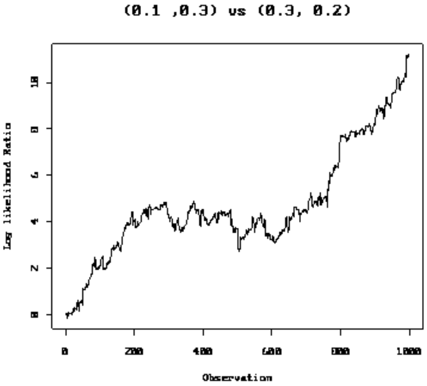
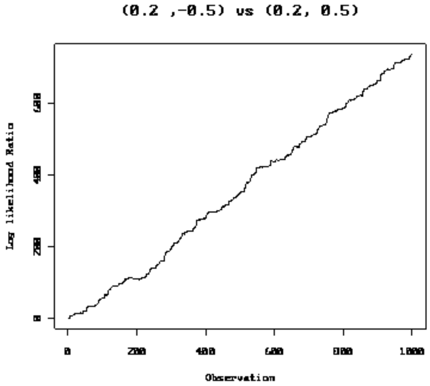


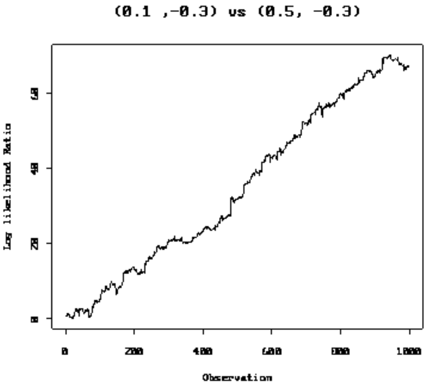
Figure 4 - Discrimination (L.L.R) between two AR(2) models, (H_1 and H_2). Observations from H_1.


(-0.2 ,0.2) vs (0.2, 0.2)
Log likelihood Ratio
Observation


(0.1 ,0.1) vs (0.4, 0.7)
Log likelihood Ratio
Observation


(0.1 ,0.3) vs (0.3, 0.2)
Log likelihood Ratio
Observation


(0.2 ,-0.5) vs (0.2, 0.5)
Log likelihood Ratio
Observation


(0.1 ,-0.3) vs (0.5, -0.3)
Log likelihood Ratio
Observation

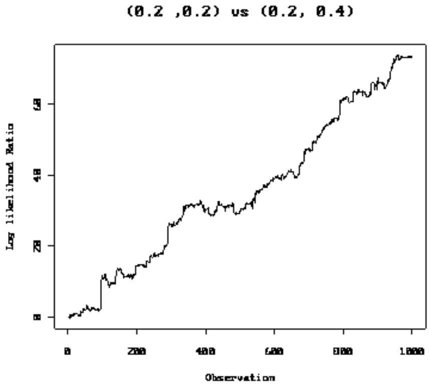

(0.2 ,0.2) vs (0.2, 0.4)
Log likelihood Ratio
Observation

Figure 5 - Discrimination (L.L.R) between two ARMA(1,1) models, (H_1 and H_2). Observations from H_1.

Table 1 - The number of correct allocations and wrong allocations For 1000 time series of length 500

Models

H_1: $x_t - 0.1x_{t-1} - 0.1x_{t-2}$ = ?$_t$°

H_2: $x_t + 0.1x_{t-1} + 0.1x_{t-2}$ = ?$_t$L

Series generated from	No. Allocated to H_1	H_2	No. final allocated	Av. No. of points until decision
H_1	912	70	12	143
H_2	79	890	31	155

Models

H_1: $x_t - 0.2x_{t-1} - 0.2x_{t-2}$ = ?$_t$e

H_2: $x_t + 0.2x_{t-1} + 0.2x_{t-2}$ = ?$_t$

Series generated from	No. Allocated to H_1	H_2	No. final allocated	Av. No. of points until decision
H_1	994	5	1	86
H_2	10	989	1	90

Models

H_1: $x_t + 0.1x_{t-1} + 0.1x_{t-2}$ = ?$_t$7

H_2: $x_t + 0.2x_{t-1} + 0.2x_{t-2}$ = ?$_t$ρ

Series generated From	No. Allocated to H_1	H_2	No. final allocated	Av. No. of points until decision
H_1	792	122	86	163
H_2	197	721	82	165

Models

H_1: $x_t - 0.3x_{t-1} - 0.3x_{t-2} = ?_t$
H_2: $x_t + 0.3x_{t-1} + 0.3x_{t-2} = ?_t$

Series generated from	No. Allocated to H_1	H_2	No. final allocated	Av. No. of points until decision
H_1	1000	0	0	61
H_2	3		0	65
	997			

6. Comparison between state space and classical methods

The last section is devoted to the comparison between the state space method for discrimination between ARMA models and those found in Chan et al., termed classical discrimination.

As in Chan *et al.*, discrimination between two MA(1) models, H_1 and H_2

$$H_1 : y_t = ?_1 ?_{t-1} + ?_t, \quad ?_t ? N(0, ?^2)$$
$$H_2 : y_t = ?_2 ?_{t-1} + ?_t \quad ?_t ? N(0, ?^2)$$

leads to a discriminant

$$d(Z) = (?_1 - ?_2)\sum_{j=1}^{T} \frac{\left(?_1 + ?_2 + 2\cos\frac{j?}{T+1}\right) z_j^2}{\left(1 + ?_1^2 + 2?_1 \cos\frac{j?}{T+1}\right)\left(1 + ?_2^2 + 2?_2 \cos\frac{j?}{T+1}\right)}, \tag{6}$$

where $$z_j = \left(\frac{2}{T+1}\right)\sum_{k=1}^{T} x_k \sin\frac{jk?}{T+1}.$$

Then Z is allocated to H_1 if d is larger than

$$\sum_{j=1}^{T} \log \frac{\left(1 + ?_2^2 + 2?_2 \cos\frac{j?}{T+1}\right)}{\left(1 + ?_1^2 + 2?_1 \cos\frac{j?}{T+1}\right)}$$

and to H_2 otherwise. This model can be considered as a state space model shown in section (4.1).

Setting $d(y) = \sum_{j=1}^{T} A_j$, where

$$A_j = \frac{1}{2} log \frac{(1-?_1^2)(1-?_1^{2j})}{(1-?_2^2)(1-?_2^{2j})} + \frac{1}{2}\left\{ \frac{\sum_{m=0}^{j-1}(-?_2)^m (1-?_2^{2(j-m)}) y_{j-m}}{\sum_{m=0}^{j-1}(-?_1)^m (1-?_1^{2(j-m)}) y_{j-m}} \right\}^2$$

then y is allocated to H_1 if the percentage A'_ts is greater than 50%.

A simulation exercise was carried out to compare the above two methods of discrimination. A time series of total length 200 was generated from H_2 and then allocated to H_1 or H_2 according to the two methods. This was repeated 1000 times. For the state space method allocation was considered using all of observations. Results are given in the Table 2. As can be seen both state space discrimination and classical discrimination work well but with the classical method slightly superior.

Table 2: Misclassification table for 1000 time series of length 200 for MA(1) processes

I: Percentage misclassified with the state space method
II: Percentage misclassified with the classical method
III: Misclassified with state space and (6)

Models:

$H_1 : \mathbf{y}_t = -0.2?_{t-1} + ?_t$

$H_2 : \mathbf{y}_t = 0.2?_{t-1} + ?_t$

	I	II	III
Misclassified	1.4	0.1	0.1

Models:

$H_1 : \mathbf{y}_t = 0.2?_{t-1} + ?_t$

$H_2 : \mathbf{y}_t = 0.4?_{t-1} + ?_t$

	I	II	III
Misclassified	6.5	2.0	4.8

Models:

$H_1 : \mathbf{y}_t = -0.5?_{t-1} + ?_t$

$H_2 : y_t = 0.5?_{t-1} + ?_t$

	I	II	III
Misclassified	0.0	0.0	0.0

Models:

$H_1 : \mathbf{y}_t = 0.4?_{t-1} + ?_t$

$H_2 : \mathbf{y}_t = 0.7?_{t-1} + \varepsilon_t$

	I	II	III
Misclassified	1.4	0.1	0.2

Models:

$H_1 : \mathbf{y}_t = -0.3?_{t-1} + ?_t$

$H_2 : \mathbf{y}_t = -0.5?_{t-1} + ?_t$

	I	II	III
Misclassified	5.1	2.2	5.6

Models:

$H_1 : \mathbf{y}_t = 0.1?_{t-1} + ?_t$

$H_2 : \mathbf{y}_t = 0.8?_{t-1} + ?_t$

	I	II	III
Misclassified	0.0	0.0	0.0

References

Alagon, J. (1989) *Spectral Discrimination for Two Group of Time Series*, Journal of Time Series Analysis, **10**(3), 203-214.

Alagon, J., (1986) *Discriminant Analysis for Time Series*, Ph.D thesis, University of Oxford.

Broemeling, L. D. and Son, M. S. (1987). *The Classification Problem with Autoregressive Processes*, Communicatins in Statistics: Theory and Methods, **16**, 927-37.

Chan, H. T., Chinipardaz, R. and Cox, T. F. (1996) *Discrimination of AR,, MA and ARMA time series models*, Communications in Statistics: Theory and Methods, **26**, 1247-1260.

Chaudhuri, G., Borwanker, J. D. and Rao, P. R. K. (1991). *Bhattacharyya Distance-Based Linear Discriminant Function for Stationary Time Series*, Communications in Statistics: Theory and Methods, **20**(7), 2195-2205.

Dargahi-Noubari, G.R. (1992) *Discrimination Between Gaussian Time Series Based on Their Spectral Differences*, Communications in Statistics: Theory and Methods, **21**(9), 2439-2458.

Dargahi-Noubari, G.R. and Laycock, P.J. (1981) *Spectral Ratio Discriminants and Information Theory,* Journal of Time Series Analysis, **2**, 71-86.

Gardner, G., Harvey, A.C. and Phillips, G.D.A. (1980) *An Algorithm for Exact Maximum Likelihood Estimation of Autoregressive-Moving Average Model by Means of Kalman Filtering Algorithm,* Applied Statistics, **29**, 311-322.

Geisser, S. (1966) *Predictive Discrimination*, Multivariate Analysis, 149-163, ed. Krishnaiah, P.R., Academic press, New York

Gersch, W. and Yonemoto, J. (1977) *Parametric time series models for multivariate EEG analysis,* Comp. Biomed. Res., **10**, 113-125.

Gersch, W.(1981) *Nearest Neighbor Rule Classification of Stationary and Non-Stationary Time Series,* In: Applied Time Series Analysis II, (Ed.by D. Findley), PP. 221-270, Academic Press, New York.

Gersch, W., Martinelli, F. and Yonemoto, J. (1979) *Automatic classification of EEG's: Kullback Leibler-nearest neighbour rule*, Science, **205**, 193-195.

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

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

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

Shumway, R. H. (1988) *Applied Statistical Time Series Analysis*, Prentice Hall, New Jersey.

Shumway, R.H. (1982) *Discriminant Analysis for Time Series*, Handbook of Statistics, 2, 1-46, (Eds. P. R. Krishnaiah and L. N. Kanal), PP. 1- 46, Amsterdam: North-Holland.

Shumway, R.H. and Unger, A.N.(1974) *Linear Discriminant Functions for Stationary Time Series*, JASA, **72**, 881-885.