BAYESIAN ESTIMATES IN TIME-SERIES MODELS

J. Andel

June 1985
CP-85-29

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

Within the framework of the Economic Structural Change Program, a cooperative research activity of IIASA and the University of Bonn, FRG, a project is carried out on "Statistical and Econometric Identification of Structural Change"; the project involves studies on the formal aspects of the analysis of structural changes. On the one hand, they include statistical methods to detect non-constancies, such as stability tests, detection criteria, etc., and on the other hand, methods which are suitable for models which incorporate non-constancy of the parameters, such as estimation techniques for time-varying parameters, adaptive methods, etc.

The present paper discusses the application of Bayesian estimation methods in the context of ARMA-models such as the periodic autoregression.

Anatoli Smyshlyaev
Acting Leader
Economic Structural Change Program

BAYESIAN ESTIMATES IN TIME-SERIES MODELS

Jiri Andel

Charles University, Sokolovska 83, 186 00 Prague 8, Czechoslovakia

INTRODUCTION

Autoregressive (AR), moving-average (MA), and autoregressive-moving average (ARMA) models are very popular in time-series analysis. Many problems of estimating their parameters and testing hypotheses are only solved asymptotically. The derivation of asymptotic results is usually not easy. An alternative approach to such problems is the Bayesian approach. It is assumed that the parameters of the models are random variables. There are theorems ensuring that under general assumptions the asymptotic posterior distribution does not depend on the prior distrubiton. As the derivation of the results is usually easier in the Bayesian approach, we can use this procedure particularly for the statistical analysis of more complicated models.

PERIODIC AUTOREGRESSION

In many applications we encounter time series with a seasonal behavior. We may assume that the length of season p is known (e.g., for monthly data exhibiting a one-year seasonality we have p = 12). In the classical approach Box and Jenkins (1970, Ch. 9) recommend to start with the differences

$\nabla_p X_t = X_t - X_{t-p}$ and to apply an ARMA model to the differences of the type $\nabla_1^d \nabla_p^D X_t$. As such differencing is, however, not sufficient for real time series, a nonlinear transformation is sometimes applied before. In the well-known example of the number of airline passengers the log-transformation is applied at the beginning without taking into account its influence on other statistical procedures.

If the length of periodicity p = 2, we can consider the following modification of an AR model:

$$X_{2t+1} = b_{11}X_{2t} + b_{12}X_{2t-1} + Y_{2t+1} ,$$

$$X_{2t+2} = b_{21}X_{2t+1} + b_{22}X_{2t} + Y_{2t+2} ,$$

where $\{Y_t\}$ is a white noise. Simulations show (Andel 1983) that the realization of such a model is very similar to some economic time series. In some cases the model has an explosive behavior. To take account of this behavoir, we put

$$\xi_t = \begin{pmatrix} X_{2t-1} \\ X_{2t} \end{pmatrix} , \qquad \eta_t = \begin{pmatrix} Y_{2t-1} \\ Y_{2t} \end{pmatrix} ,$$

$$A_o = \begin{pmatrix} 1 & 0 \\ -b_{21} & 1 \end{pmatrix} , \qquad A_1 = \begin{pmatrix} -b_{12} & -b_{11} \\ 0 & -b_{22} \end{pmatrix} .$$

Then our model can be expressed in the form of

$$A_o \xi_{t+1} + A_1 \xi_t = \eta_{t+1} .$$

Let z_1, z_2 denote the roots of the polynomial

$$|A_o z + A_1| = z^2 - (b_{11}b_{21} + b_{12} + b_{22})z + b_{12}b_{22} .$$

If $|z_1|<1$, $|z_2|<1$, the process $\{\xi_t\}$ is stationary; otherwise $\{\xi_t\}$ has an explosive behavior.

Our example was a special case of periodic autoregression. We used a model of the second order with p = 2. Generally, the periodic autoregression is given by

$$(\ast) \qquad X_{n+(j-1)p+k} = \sum_{i=1}^{n} b_{ki} X_{n+(j-1)p+k-i} + Y_{n+(j-1)p+k}$$

$$(j = 1,2,\dots \quad ; \quad k = 1,2,\dots,p) \quad .$$

If $\{Y_t\}$ is the usual white noise, we have a model with equal variances. If var $Y_{n+(j-1)p+k} = \sigma_k^2$, where $\sigma_1^2,\dots,$ $\dots,\sigma_p^2$ are not all the same, we have a model with periodic variances. The periodic autoregression is a special case of periodically correlated random sequences, which were introduced by Gladyshev (1961). Jones and Brelsford (1967) expanded $b_{k1},\dots,b_{kn}$ into a Fourier series. Pagano (1978) considered estimators for b_{ki} obtained by modified Yule-Walker equations. He also showed that a periodic AR can be rewritten into a multidimensional AR model. A periodic ARMA model was introduced by Cleveland and Tiao (1979). The problem of the periodic AR is also treated by Troutman (1979). Tiao and Grupe (1980) investigated the errors of misclassification when the periodic structure of an ARMA process was neglected. Newton (1982) shows that a periodic AR can substantially simplify the numerical procedures for estimating parameters in multiple AR models. Andel (1983) presents some results of the Bayesian analysis of the periodic AR model including tests on whether the time series can be described by the classical AR model.

THE BAYESIAN APPROACH

If $\theta \in \Omega$ is a random (multidimensional) parameter with a prior density $p(\theta)$, and if a random vector X has a conditional density of $p(x|\theta)$ given θ, then according to the Bayes theorem, the posterior density $p(\theta|x)$ of θ is given by

$$p(\theta|x) = c_x\, p(x|\theta)p(\theta) \quad , \quad \theta \in \Omega \quad ,$$

where c_x is a constant. Usually the modus of the posterior density is taken as an estimator of θ. This is a generalization of the maximum likelihood estimator, which we would get for $p(\theta)$ = const. (Another estimator of θ could be the posterior expectation.)

The main problem with using the Bayesian approach is the choice of the prior density $p(\theta)$. The following three possibilities are most popular:

a. Conjugate prior density. Let $p(x|\theta)$ be given. A system $M_{p(x|\theta)}$ is called a conjugate system, if $p(\theta|x) \in M_{p(x|\theta)}$ holds for every $p(\theta) \in M_{p(x|\theta)}$. Usually we take the minimal system $M_{p(x|\theta)}$ with this property, or the so-called natural system.

Although a conjugate prior density is convenient from the mathematical point of view, there is no logical reason for using it in a given case. Moreover, this procedure does not specify which density from the system $M_{p(x|\theta)}$ should be taken.

b. Uncertainty principle. In the case of "full ignorance" it is recommended to put $p(\theta) = 1$ for $\theta \in \Omega$. This often leads to improper (or vague) densities, e.g. if $\Omega = R_k$ (Euclidean k-dimensional space). The advantage of this method is that the density $p(\theta) = 1$ can asymptotically substitute any other reasonable prior density.

Theorem (DeGroot 1970, § 10.4). Let $\pi_o(\theta|x) = cp(x|\theta)$ and let $\pi(\theta|x) = c'p(x|\theta)p(\theta)$. Let $A \subset \Omega$ be such a set that $m = \inf_{\theta \in A} p(\theta) > 0$. Let $a,b,c > 0$ satisfy

$$\int_A \pi_o(\theta|x)d\theta \geqq 1 - a \quad , \quad \sup_{\theta \in A} p(\theta) \leqq (1+b)m \quad ,$$

$$\sup_{\theta \in \Omega - A} p(\theta) = (1+c)m \quad .$$

Then

$$\int_{\Omega} |\pi(\theta|x) - \pi_o(\theta|x)| d\theta \leqq \varepsilon \quad ,$$

where

$$\varepsilon = \max[(1-a)^{-1}(a+b) \quad , \quad (1+a+b+ac)^{-1}(a+b+ab)] +$$

$$+ (1-a)^{-1}a(1-a+c) \quad .$$

In the typical situation the posterior distribution π_o is the normal one with a variance matrix of the order N^{-1} (where N is the number of observations). Then a and b are nearly zero and ε is very small.

Unfortunately, the uncertainty principle is not logically consistent. For example, let

$$p(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad , \quad 0 < \theta < 1 \quad .$$

If θ is completely unknown, we choose its prior density $p(\theta) = 1$ and compute $p(\theta|x)$ by means of the Bayes theorem. We find that $w = \theta^2$ has a posterior density of $c\ w^{(x-1)/2}(1-w^{1/2})^{n-x}$. However,

$$p(x|w) = \binom{n}{x} w^{x/2} (1-w^{1/2})^{n-x} \quad .$$

If w is considered to be a completely unknown parameter with a constant prior density, then its posterior density is $c' w^{x/2}(1-w^{1/2})^{n-x}$.

c. Jeffrey's principle. To avoid inconsistency it is necessary to choose the prior density proportional to $|J(\theta)|^{1/2}$, where J is the Fisher information matrix.

In practice the prior density is chosen by combining the uncertainty principle and Jeffrey's principle.

SOME RESULTS CONCERNING THE PERIODIC AR

Andel (1983) proved the following assertions:

Theorem. Assume model (*), where $X_1,\dots,X_n$ are given constants and $\{Y_t\}$ are i.i.d. $N(0,\sigma^2)$ variables. If $x_1,\dots,x_N$ is a known realization, put

$$\alpha_k = [\frac{N-n-k}{p}] + 1 \quad \text{(where [ ] is the integer part)},$$

$$q_{ij}^{(k)} = \sum_{h=1}^{\alpha_k} x_{n+k+(h-1)p-i}\, x_{n+k+(h-1)p-j} \quad ,$$

$$Q_k = (q_{ij}^{(k)})_{i,j=1}^{n} \quad , \quad q_k = (q_{01}^{(k)},\dots,q_{0n}^{(k)})' \quad ,$$

$$Q = Q_1+\dots+Q_p \quad ,$$

$$b_k^* = Q_k^{-1}\, q_k \quad , \quad b^* = (b_1^{*'},\dots,b_p^{x'})' \quad ,$$

$$v_k = q_{00}^{(k)} - b_k^{*'}\, q_k \quad , \quad v = v_1+\dots+v_p \quad ,$$

$$H = \mathrm{Diag}\{Q_1,\dots,Q_{p-1}\} - (Q_1,\dots,Q_{p-1})'Q^{-1}(Q_1,\dots,Q_{p-1}) .$$

Let the prior density of a,b be a^{-1} for $a > 0$. Then $(npv)^{-1}(N-n-np) \sum_{k=1}^{p} (b_k-b_k^*)'Q_k(b_k-b_k^*)$ has a posterior $F_{np,N-n-np}$ distribution, and $[n(p-1)v]^{-1}[N-n(p+1)]\Delta'H\Delta$ has a posterior $F_{n(p-1),N-n(p+1)}$ distribution, where

$$\Delta = (\Delta_1',\dots,\Delta_{p-1}')' \quad , \quad \Delta_k = b_k - b_p - (b_k^*-b_p^*) \quad .$$

Similar assertions are also derived for the model with periodic variances.

GENERALIZATIONS

Model (*) can be generalized to a model with exogenous and endogenous variables. The computations are slightly easier when we write the absolute term (μ_k) separately. The model reads

$$X_{n+(j-1)p+k} = \mu_k + \sum_{i=1}^{n} b_{ki}\, X_{n+(j-1)p+k-1} + \\ + \sum_{s=1}^{S} \sum_{r=o}^{m_s} a_{ksr}\, \psi_{s,n+(j-1)p+k-r} + \\ + Y_{n+(j-1)p+k} \quad , \qquad j = 1,2,\ldots; \\ k = 1,2,\ldots p.$$

Quite analogously we can get a multiple model for the case that X_t are random vectors.

PROBLEMS OF STRUCTURAL CHANGES

For reasons of simplicity, we only discuss here the main ideas in the field of classical AR models. Their generalization to a periodic AR model is then obvious.

Let X_t be created by the model

$$X_t = \sum_{k=1}^{n} a_k^{(t)}\, X_{t-k} + Y_t \quad ,$$

where Y_t are independent variables, $Y_t \sim N(0,\sigma_t^2)$. We assume that in some unknown moment, $t = \theta$, there can be a change of parameters:

$$a_k^{(t)} = a_k^o \quad (k = 1,\ldots,n) \quad , \quad \sigma_t^2 = \sigma_o^2 \quad \text{for } t < \theta,$$

$$a_k^{(t)} = a_k^1 \quad (k = 1,\ldots,n) \quad , \quad \sigma_t^2 = \sigma_\theta^2 \quad \text{for } t \geqq \theta.$$

Some methods for the detection of point θ are described by Segen and Sanderson (1980) and by Basseville and Benveniste (1983). Their procedures are based on the cumulative sum (CUSUM) technique. Simulations show, however, that the CUSUM method is in many cases not sensitive enough.

Another promising model for detecting structural changes can be the model with exogenous and endogenous variables, e.g.

$$X_t = \mu + b_1 X_{t-1} + \dots + b_n X_{t-n} + a\,\psi_t + Y_t \quad ,$$

where

$$\psi_t = 0 \quad \text{for } t < \theta \quad ,$$

$$\psi_t = 1 \quad \text{for } t \geq \theta \quad .$$

The problem is to estimate parameters μ, $b_1,\dots,b_n$, a, θ, and σ^2 = var Y_t. The change occurs in time $t = \theta$, mainly influencing the mean value of the process. This formulation reminds of the piece-wise linear regression with unknown break-points, where the solution is often based on the maximum likelihood method.

Both cases mentioned here can be generalized to more complicated models. Further research in this field is necessary, because the problem is interesting and important. It is clear that the theoretical results should be complemented by available computer programs.

REFERENCES

Andel, J. (1983). Statistical analysis of periodic autoregression. Appl. Mat. 28:364-385.

Basseville, M., and A. Benveniste (1983). Sequential detection of abrupt changes in spectral characteristics of digital signals. IEEE Trans. of Inf. Th. IT-29:709-724.

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

Cleveland, W.P., and G.C. Tiao (1979). Modeling seasonal time series. Rev. Economie Appliquee 32:107-129.

DeGroot, M.H. (1970). Optimal Statistical Decisions. New York: Mc Graw-Hill.

Gladyshev, E.G. (1961). Periodically correlated random sequences. Soviet Math. 2:385-388.

Jones, R.H., and W.M. Brelsford (1967). Time series with periodic structure. Biometrika 54:403-408.

Newton, H.J. (1982). Using periodic autoregression for multiple spectral estimation. Technometrics 24:109-116.

Pagano, M. (1978). On periodic and mulitple autoregression. Ann. Statist. 6:1310-1317.

Segen, J., and A.C. Sanderson (1980). Detecting change in a time series. IEEE Trans. on Inf. Th. IT-26:249-255.

Tiao, C.G., and M.R. Grupe (1980). Hidden periodic autoregressive-moving average models in time series data. Biometrika 67:365-373.

Troutman, B.M. (1979). Some results in periodic autoregression. Biometrika 66:219-228.