The Philippine Statistician, 1995-1996
Vol. 44-45, Nos. 1-8, pp. 27-32

Parameter Changes in Autoregressive Processes: A Bayesian Approach

Arnulfo P. Supe[1]

ABSTRACT

The posterior distribution of the break point in autoregressive processes which undergo structural change is derived. The noninformative prior density is used. Simulation study is done for the changing AR(1) and the AR(2) processes.

KEY WORDS: Structural change, Break point, Prior distribution, Posterior distribution

When modeling time-series data using the Box-Jenkins approach, parameters are assumed not to vary with time. While this may be sufficient in most practical applications, there are instances that model parameters also change after some specific time points. For example, evidences have shown that economic time series models before and after the 1983 Ninoy Aquino assassination exhibited marked differences.

In this paper, the autoregressive process will be examined for possible shifts in parameters. The Bayesian approach will be employed in identifying the time point when the parameters of an autoregressive process are said to undergo some changes. The noninformative prior distribution will be used. Broemeling and Tsurumi (1987) gives an extensive discussion of structural changes in linear and time series models. Salazar (1982) analyzed structural change in some time series models. Their works, which use the Bayesian approach, employ the Normal-Gamma prior distribution. The noninformative prior will be used here and the derived results will be tested by simulation.

1. THE MODEL

Let (1,2, ..., ν, ν+1, ... n) be discrete time points. The autoregressive model to be considered is

$$Y_t = \begin{cases} \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_r Y_{t-r} + \varepsilon_t & ; \quad t = 1,2,\cdots, \nu \\ \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_r Y_{t-r} + \varepsilon_t & ; \quad t = \nu+1,\cdots,n \end{cases} \tag{1.1}$$

where $1 \le \nu \le n$ and unknown ; $r < \nu$

$\alpha_i \neq \beta_i$ for at least one i

$Y_0, Y_{-1}, \ldots, Y_{1-r}$ are known constants

ε_t ; $t = 1,2, \ldots, n$ are independently distributed normal random variables with means 0 and variance σ^2.

When $\nu = n$, this is the usual autoregressive process model with no parameter shift but if $1 \le \nu \le n-1$, then structural change is said to occur and ν is called the break point. The

[1] MSU-Iligan Institute of Technology, Iligan City

analysis will consist of deriving the posterior distribution of ν. We choose as our estimate of the break point, the value of ν which attains the maximum posterior probability.

2. POSTERIOR ANALYSIS

In this section, the posterior density, $\pi(\nu|X)$ of the break point ν given the sample data X, will be derived using Bayes' Formula

$$\pi(\nu|X) = \frac{h(\nu)f(X|\nu)}{m(X)} \qquad \text{where} \quad m(X) = \int\cdots\int f(X|\nu)h(\nu)d\nu .$$

Since m(X) does not involve ν , the above expression will be rewritten as

$$\pi(\nu|X) \propto h(\nu)f(X|\nu)$$

where the symbol " $\propto$ " means " is proportional to".

$$\text{Let} \quad Y_1 = Y_1(\nu) = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_\nu \end{bmatrix} \qquad Y_2 = Y_2(\nu) = \begin{bmatrix} Y_{\nu+1} \\ Y_{\nu+2} \\ \vdots \\ Y_n \end{bmatrix}$$

$$Z_1 = Z_1(\nu) = \begin{bmatrix} Y_0 & Y_{-1} & \cdots & Y_{1-r} \\ Y_1 & Y_0 & \cdots & Y_{2-r} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{\nu-1} & Y_{\nu-2} & \cdots & Y_{\nu-r} \end{bmatrix} \qquad Z_2 = Z_2(\nu) = \begin{bmatrix} Y_\nu & Y_{\nu-1} & \cdots & Y_{\nu-r} \\ Y_{\nu+1} & Y_\nu & \cdots & Y_{\nu+1-r} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{n-1} & Y_{n-2} & \cdots & Y_{n-r} \end{bmatrix}$$

$$Y = Y(\nu) = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \qquad Z = Z(\nu) = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}$$

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_r \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{bmatrix} \qquad \Phi = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

Model (1.1) can be rewritten as

$$\begin{aligned}
\varepsilon_1 &= Y_1 - \alpha_1 Y_0 - \ldots - \alpha_r Y_{1-r} \\
\varepsilon_2 &= Y_2 - \alpha_1 Y_1 - \ldots - \alpha_r Y_{2-r} \\
&\vdots \qquad\qquad \vdots \\
\varepsilon_\nu &= Y_\nu - \alpha_1 Y_{\nu-1} - \ldots - \alpha_r Y_{\nu-r} \\
\varepsilon_{\nu+1} &= Y_{\nu+1} - \beta_1 Y_\nu - \ldots - \beta_r Y_{\nu+1-r} \\
&\vdots \qquad\qquad \vdots \\
\varepsilon_n &= Y_n - \beta_1 Y_{n-1} - \ldots - \beta_r Y_{n-r}
\end{aligned}$$

The conditional likelihood of (Φ,δ,ν), given the sample observations $(\boldsymbol{Z},\boldsymbol{Y})$ is

$$L(\Phi,\delta,\nu|\,(\boldsymbol{Z},\boldsymbol{Y})) \propto \delta^{\frac{n}{2}} exp\left\{-\frac{\delta}{2}\left(\sum_{t=1}^{\nu}(Y_t-\alpha_1 Y_{t-1}-\ \ldots\ -\alpha_r Y_{t-r})^2 + \sum_{t=\nu+1}^{n}(Y_t-\beta_1 Y_{t-1}-\ \ldots\ -\beta_r Y_{t-r})^2\right)\right\}$$

or in matrix form

$$L(\Phi,\delta,\nu|(\boldsymbol{Z},\boldsymbol{Y})) \propto \begin{cases} \delta^{\frac{n}{2}} \exp\left(-\frac{\delta}{2}\left[(\mathbf{Y}(\nu)-\mathbf{Z}(\nu)\Phi)'(\mathbf{Y}(\nu)-\mathbf{Z}(\nu)\Phi)\right]\right);\ 1\le\nu\le n-1 \\ \delta^{\frac{n}{2}} \exp\left(-\frac{\delta}{2}\left[(\mathbf{Y}_1(n)-\mathbf{Z}_1(n)\Phi)'(\mathbf{Y}_1(n)-\mathbf{Z}_1(n)\Phi)\right]\right);\ \nu=n \end{cases} \tag{2.1}$$

To derive the posterior density of ν, a prior density for the parameters (Φ,δ,ν) must be specified. There are a number of ways by which the prior density can be determined and Berger (1985), gives a number of methods. The use of prior density is one of the most controversial aspects of Bayesian analysis. One of the criticisms is the issue of whether different priors give different results. Bayesians argue that this seldom occurs and in cases that this happens, then an important prior information that affects the data really exists and hence the need for a more detailed Bayesian analysis.

Jeffreys' noninformative prior will be used here because it is the least controversial choice and it has been shown that many classical estimation results correspond to the Bayesian results when the noninformative prior is used. It also has an important feature that it is not affected by a restriction on the parameter space. Use of this prior is also, in a sense, robust because no subjective prior beliefs are assumed.

The posterior density of ν will now be derived for $1 \le \nu \le n$-1. The steps involved when $\nu = n$ is similar, and will not be shown. Let the joint prior of (Φ,δ,ν) be given by Jeffreys' noninformative prior

$$\rho(\Phi,\delta,\nu) \propto \frac{1}{\sigma^2} \qquad \text{and let} \quad \delta = \frac{1}{\sigma^2} \tag{2.2}$$

Combining (2.1) and (2.2) by Bayes Theorem, we get

$$\pi(\Phi,\delta,\nu|(\mathbf{Z},\mathbf{Y})) \propto \delta^{\frac{n-2}{2}} \exp\left\{-\frac{\delta}{2}\left((\alpha-\alpha^*)'(\mathbf{Z}_1'\mathbf{Z}_1)(\alpha-\alpha^*) + (\beta-\beta^*)'(\mathbf{Z}_2'\mathbf{Z}_2)(\beta-\beta^*) + g(\nu)\right)\right\}$$

where

$$\alpha^* = (Z_1'Z_1)^{-1}(Z_1'Y_1)$$

$$\beta^* = (Z_2'Z_2)^{-1}(Z_2'Y_2)$$

$$g(\nu) = Y'Y - (Y'Z)(Z'Z)^{-1}(Z'Y)$$

Integrating out α and β,

$$\pi(\delta,\nu|(Z,Y)) \propto \delta^{\frac{n-2-2r}{2}} \exp\left\{-\frac{\delta}{2}[g(\nu)]\right\} \cdot |[Z'Z]|^{-\frac{1}{2}}$$

Integrating out δ,

$$\pi_1(\nu|(Z,Y)) \propto |Z'Z|^{-\frac{1}{2}} \cdot [g(\nu)]^{-(\frac{n-2r}{2})}$$

This gives us the posterior distribution of ν below, from which most of the inferences will be based.

$$\therefore \pi_1(\nu|(Z,Y)) \propto \begin{cases} |[Z(\nu)'Z(\nu)]|^{-\frac{1}{2}}[g(\nu)]^{-(\frac{n-2r}{2})} & ; \ 1 \le \nu \le n-1 \\ |[Z(n)'Z(n)]|^{-\frac{1}{2}}[g(n)]^{-(\frac{n-r}{2})} & ; \quad \nu = n \end{cases} \tag{2.3}$$

where $g(\nu) = Y(\nu)'Y(\nu) - [Y(\nu)'Z(\nu)][Z(\nu)'Z(\nu)]^{-1}[Z(\nu)'Y(\nu)]$

$$g(n) = Y_1(n)'Y_1(n) - [Y_1(n)'Z_1(n)][Z_1(n)'Z_1(n)]^{-1}[Z_1(n)'Y_1(n)]$$

3. SIMULATION STUDY

To illustrate the effectiveness of the method derived in the previous section, sets of normally distributed random variables were generated using SAS. These values were used as the error terms in creating 20 data files for each of the parameter combinations used in Tables 3.1 and 3.2. A computer program to compute the posterior probabilities of the break point was developed by the author using Turbo Pascal 6.0. All computations were done using extended precision.

Summaries of the results are given in Tables 3.1 and 3.2 for the AR(1) and AR(2) respectively. The label ' *HPP at* ν ' means that the highest posterior probability is attained at the simulated break point ν. The column heading ' *HPP near* ν ' refers to the number of times that the break point ν simulated by the appropriate model is contained in the interval HPP_t plus or minus 5% of the sample size n, where HPP_t is the time point where the Highest Posterior Probability occurs. As an example, the third row of Table 3.1 indicates a sample size of 50, a break point of 30 and '*HPP near* ν' of 15. This means that 15 out of 20 data files have the simulated break point ν(=30) contained in the interval (HPP_t - 1, HPP_t + 1) . Here 5% of n(=20) is 1. For sample sizes 15, 25, 30, 50, and 75, the value 5% of the sample size was rounded off to 1, 2, 2, 3, and 4 respectively. The column heading 'HPP near ν' is similar

to finding an HPD(Highest Posterior Density) Credible Set which is the Bayesian analogue of the Confidence Set used in Classical Statistics.

Table 3.1
Summary of Simulation Results for the Model

$$Y_t = \begin{cases} \alpha_1 Y_{t-1} + \varepsilon_t \; ; \; t = 1,2, \ldots, \nu \\ \beta_1 Y_{t-1} + \varepsilon_t \; ; \; t = \nu+1, \ldots, n \end{cases}$$

Parameter Values	Sample Size	Break Point	HPP at ν	HPP near ν	Percentage HPP near ν
$\alpha_1 = 0.4$ $\beta_1 = 0.6$	50	30	0	1	5%
$\alpha_1 = 0.3$ $\beta_1 = 0.9$	50	30	2	10	50%
$\alpha_1 = 0.4$ $\beta_1 = 1.0$	50	30	5	15	75%
$\alpha_1 = 1.0$ $\beta_1 = 1.1$	50	30	30	30	100%
$\alpha_1 = -0.5$ $\beta_1 = +0.5$	50	30	8	14	70%
$\alpha_1 = -0.5$ $\beta_1 = +0.5$	75	40	8	14	70%
$\alpha_1 = -0.5$ $\beta_1 = +0.5$	100	50	5	14	70%

ν = break point number of datasets = 20 HPP = Highest Posterior Probability

For the AR(1) model, the change cannot be detected for small changes in the parameter. This is seen in Table 3.1 above, where only a 5% success is recorded when the parameter changes from 0.4 to 0.6. However, the capture of the break point improves dramatically as the numerical change in the parameter increases. This improvement can be noted in Table 3.1 where the success rate is 50% for $\alpha_1 = 0.3$ and $\beta_1 = 0.9$, and 70% for $\alpha_1 = -0.5$ and $\beta_1 = +0.5$ for sample sizes 50, 75 and 100. When the process changes from being stationary($\alpha_1 = 0.4$) to nonstationary($\beta_1 = 1.0$) , 75% of HPP is near the simulated break point. Finally the break point is captured 100% of the time for even a small change($\Delta = 0.1$) of the parameter when the process starts as being nonstationary.

The AR(2) process enjoys a much better success rate than the AR(1) model. One reason that can be forwarded is that more parameters are subjected to change for AR(2) than for AR(1). The summary is given in Table 3.2 below. For parameter combinations $\alpha_1 = 0.4$, $\beta_1 = 0.7$, $\alpha_2 = 0.7$, $\beta_2 = 0.6$ the percentage of HPP near the simulated break point ranges from 85% to 95% for sample sizes 50, 75 and 100 and when the break point is near the center. The higher percentage is attained at a sample size of 100. This means that an improvement is made when the sample size is increased. When the variance is allowed to change, the success rate remains the same although the actual probabilities for the simulated break point improves. The break point is captured 95% of the time when it occurs near the beginning ($\nu = 5$) of the series. When the break point is near the end($\nu = 45$), 95% of the highest posterior probability hovers near 50, indicating no change occurred. In most cases where the graph(not shown) clearly indicates a break point, the highest posterior probability also occurs at that particular break point.

Table 3.2
Summary of Simulation Results for the Model

$$Y_t = \begin{cases} \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t \; ; \; t = 1,2, \ldots, \nu \\ \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \varepsilon_t \; ; \; t = \nu+1, \ldots, n \end{cases}$$

Parameters	Sample Size	Break Point	HPP at ν	HPP near ν	Percentage HPP near ν
$\alpha_1 = 0.4 \quad \beta_1 = 0.3$ $\alpha_2 = 0.7 \quad \beta_2 = 0.6$	50	30	7	17	85%
$\alpha_1 = 0.4 \quad \beta_1 = 0.3$ $\alpha_2 = 0.7 \quad \beta_2 = 0.6$	75	40	4	17	85%
$\alpha_1 = 0.4 \quad \beta_1 = 0.3$ $\alpha_2 = 0.7 \quad \beta_2 = 0.6$	100	50	4	17	85%
$\alpha_1 = 0.4 \quad \beta_1 = 0.3$ $\alpha_2 = 0.7 \quad \beta_2 = 0.6$	50	5	19	19	95%
$\alpha_1 = 0.4 \quad \beta_1 = 0.3$ $\alpha_2 = 0.7 \quad \beta_2 = 0.6$	50	45	0	19	95%
$\alpha_1 = 0.4 \quad \beta_1 = 0.3$ $\beta_1 = 0.7 \quad \beta_2 = 0.6 \quad \sigma_1 \neq \sigma_2$	50	30	8	17	85%
$\alpha_1 = -0.6 \quad \beta_1 = 0.5$ $\alpha_2 = -0.4 \quad \beta_2 = 0.7$	50	30	30	30	100%

ν = break point number of datasets = 20 HPP = Highest Posterior Probability

REFERENCES

BERGER, J. O. (1985), Statistical Decision Theory and Bayseisan Analysis, pp. 74-113.

BROEMELING and TSURUMI (1987), Econometrics and Structural Change, pp. 161-174.

SALAZAR, D. H. (1982), Structural Changes in Time Series Models, *Journal of Econometrics*, pp. 147-163.