E C O N O M I C S B U L L E T I N

Spuriousness of information criteria when selecting the number of breaks in stationary AR(p) process

Mohamed BOUTAHAR
GREQAM and Faculté des Sciences, Université de la Méditerranée, France

Jamel JOUINI
F.S.E.G.N., E.S.S.A.I.T. and L.E.G.I., Université 7 Novembre de Carthage, Tunisie, GREQAM, Université de la Méditerranée, France

Abstract

This note proves analytically and shows by a Monte Carlo analysis the spuriousness that arises by some model selection criteria when selecting the number of breaks in stationary AR(p) process without changes for a regression with mean-shifts. This brings a theoretical support to the Perron's (1997) simulation results which indicate that this phenomenon occurs for an AR(1) process.

Citation: BOUTAHAR, Mohamed and Jamel JOUINI, (2007) "Spuriousness of information criteria when selecting the number of breaks in stationary AR(p) process." *Economics Bulletin*, Vol. 3, No. 38 pp. 1-11
Submitted: April 6, 2007. **Accepted:** August 31, 2007.
URL: http://economicsbulletin.vanderbilt.edu/2007/volume3/EB-07C20014A.pdf

1 Introduction

Finding a theoretical justification for the overestimation was evoked by Bai (1998) who provides a mathematical proof for the phenomenon that when the errors of a linear regression model without any break are integrated of order one there is a tendency to estimate a break date in the middle of the sample. Thus, unlike Bai (1998), our paper is concerned with the case of multiple breaks using some model selection criteria. The fact of overestimating the number of changes when the data-generating process (DGP) is without breaks and the estimated model is with change in mean and change in trend is well illustrated in Nunes, Newbold and Kuan (1996) for a random walk using the Bayesian information criterion. In the same context, Perron (1997) shows by simulations that the conclusions of Nunes, Newbold and Kuan (1996) don't depend on the fact that the DGP is an integrated process of order one, even a stationary AR(1) process leads to an overestimation of the number of changes. Recently, Boutahar and Jouini (2007) provide a mathematical proof and show by simulations that some information criteria tend to detect a spuriously high number of structural changes when the process is trend-stationary without breaks. The important question suggested by their findings is that of distinction between trend-stationary process and random walk when modelling real data series.[1]

Our contribution in this paper consists in generalizing the study considered by Perron (1997) to a stationary AR(p) process for the problem of selecting the number of breaks in the mean of the time series. Indeed, we prove analytically and show by Monte Carlo simulations that some information criteria tend to overestimate the number of shifts.

The remainder of the paper is organized as follows. The second section presents the structural change model and the estimation method. Section 3 defines some model selection criteria. In section 4, we derive the main theoretical results of the paper. Section 5 reports simulation evidence to support the relevance of the theory. Concluding comments are provided in section 6. The proof of the Theorem is given in Appendix A, and the simulation results in Appendix B.

Throughout this paper as a matter of notation, we let "$[\cdot]$" denote integer part, "$\xrightarrow{p}$" convergence in probability, "$\xrightarrow{a.s.}$" convergence almost surely, and "$\xrightarrow{m.q.}$" convergence in quadratic mean.

2 The model and estimators

Consider the following linear regression model of structural break with m changes:

$$y_t = z_t' \delta_j + u_t, \qquad t = T_{j-1} + 1, \ldots, T_j, \tag{1}$$

for $j = 1, \ldots, m+1$, $T_0 = 0$ and $T_{m+1} = T$. y_t is the observed dependent variable, $z_t \in \mathbb{R}^q$ is the vector of covariates, δ_j are the corresponding regression coefficients with $\delta_i \neq \delta_{i+1}$ $(1 \leq i \leq m)$, and u_t is the disturbance. The break dates $(T_1, \ldots, T_m)$ are explicitly treated as unknown and for $i = 1, \ldots, m$, we have $T_i = [\lambda_i T]$ where $0 < \lambda_1 < \cdots < \lambda_m < 1$. Let $\delta = \left(\delta_1', \delta_2', \ldots, \delta_{m+1}'\right)'$.

The estimation method, proposed by Bai and Perron (1998), is based on the ordinary least-squares (OLS) principle. The method first consists in estimating the regression coefficients δ_j

[1] Note that unlike Boutahar and Jouini (2007), this study concerns the case of stationary processes.

by minimizing the sum of squared residuals $\sum_{i=1}^{m+1}\sum_{t=T_{i-1}+1}^{T_i}(y_t - z_t'\delta_i)^2$. Once the estimate $\hat{\delta}(T_1,\ldots,T_m)$ is obtained, we substitute it in the objective function and denote the resulting sum of squared residuals as $S_T(T_1,\ldots,T_m)$. The estimated break dates $\left(\hat{T}_1,\ldots,\hat{T}_m\right)$ are then determined by minimizing $S_T(T_1,\ldots,T_m)$ over all partitions $(T_1,\ldots,T_m)$ such that $T_i - T_{i-1} \geq h$.[2] Thus, the break date estimators are global minimizers of the objective function. Finally, the estimated regression coefficients are such that $\hat{\delta} = \hat{\delta}\left(\hat{T}_1,\ldots,\hat{T}_m\right)$. In our Monte Carlo study, we use the efficient algorithm developed in Bai and Perron (2003), based on the principle of dynamic programming, to estimate the unknown parameters.

3 The information criteria

A common procedure to select the dimension of a model is to consider an information criterion. In this context, Yao (1988) uses the Bayesian information criterion defined as

$$BIC(m) = \ln\left(S_T\left(\hat{T}_1,\ldots,\hat{T}_m\right)/T\right) + p^*\ln(T)/T,$$

where $p^* = (m+1)q + m$ is the number of unknown parameters. He shows that the estimator of the number of breaks $\hat{m}$ is consistent (at least for normal sequence of random variables with mean-shifts) for m^0, the true number of breaks, provided $m^0 \leq M$ with M a known upper bound for m. Another criterion proposed by Yao and Au (1989) is given by

$$YIC(m) = \ln\left(S_T\left(\hat{T}_1,\ldots,\hat{T}_m\right)/T\right) + mC_T/T,$$

where $\{C_T\}$ is any sequence satisfying $C_T T^{-2/n} \to \infty$ and $C_T/T \to 0$ as $T \to \infty$ for some positive integer n. In our simulation experiments, we use the sequence $C_T = 0.368T^{0.7}$ proposed by Liu, Wu and Zidek (1997).

Note that these information criteria cannot directly take into account the presence of serial correlation in the errors. The estimated number of break dates $\hat{m}$ is determined by minimizing the above-mentioned criteria given M a fixed upper bound for m. The usefulness of these criteria is illustrated by Jouini and Boutahar (2005) to explore the empirical evidence of the instability by uncovering structural breaks in some U.S. time series. To that effect, they pursue a methodology composed of different steps and propose a modelling strategy to implement it. Their results indicate that the time series relations have been altered by various important facts and international economic events such as the two Oil-Price Shocks and changes in the International Monetary System.

4 Spuriousness of the criteria

We consider an AR(p) process without breaks which is commonly expressed as

[2] Note that h is the minimal number of observations in each segment ($h \geq q$, not depending on T). Bai and Perron (2003) suggest that if tests for structural breaks are required, then h must be set to $[\varepsilon T]$ for some arbitrary small positive number ε.

$$A(B)\, y_t = u_t, \qquad 1 \leq t \leq T, \tag{2}$$

where B is a backward shift operator such that $B^n y_t = y_{t-n}$ for $n \in \mathbb{N}$, and $A(B) = 1 - \sum_{j=1}^{p} a_j B^j$ denotes the autoregressive polynomial of finite order p. We provide, in the next Theorem, a description of the results relating to this process under the condition that the following assumptions are satisfied.

Assumption 1. *We assume that the roots of the polynomial $A(z) = 1 - \sum_{j=1}^{p} a_j z^j$ are strictly outside the unit disk, i.e. the process y_t is stationary.*

Assumption 2. *Let $F_t = \sigma(u_i : i \leq t)$ be the sigma-field generated by the past history of $\{u_t\}$. We assume that $\{u_t\}$ is a martingale difference sequence with respect to the sigma-field F_t such that*

$$\begin{aligned} E\left(u_t^2 | F_{t-1}\right) &= \sigma_u^2, \qquad \textit{almost surely}, \\ \sup E\left(|u_t|^{2+\alpha} | F_{t-1}\right) &< \infty, \qquad \textit{almost surely}, \end{aligned}$$

for some $\alpha > 0$.

We define the sum of squared residuals

$$\begin{aligned} S_T\left(\hat{T}_1, \ldots, \hat{T}_m\right) &= \min_{(T_1, \ldots, T_m)} \left[\sum_{t=1}^{T} y_t^2 \right. \\ & \left. - \sum_{i=1}^{m+1} \left(\sum_{t=T_{i-1}+1}^{T_i} y_t z_t' \right) \left(\sum_{t=T_{i-1}+1}^{T_i} z_t z_t' \right)^{-1} \left(\sum_{t=T_{i-1}+1}^{T_i} z_t y_t \right) \right]. \end{aligned} \tag{3}$$

We can now state the following results.

Theorem. *Suppose that the data are generated according to the process (2) and that we estimate a model with change in mean, i.e. $z_t = 1$. Then, under Assumptions 1 and 2 we have for $T \rightarrow \infty$*

1.

$$\frac{S_T\left(\hat{T}_1, \ldots, \hat{T}_m\right)}{T} \xrightarrow{p} L(a_1, \ldots, a_p), \tag{4}$$

where

$$L(a_1, \ldots, a_p) = e_1' \sum_{j=0}^{\infty} D^j K D^{j\prime} e_1, \tag{5}$$

with $e_1 = (1, 0, \ldots, 0)'$ a p-vector, $K = diag\left(\sigma_u^2, 0, \ldots, 0\right)$, and

$$D=\begin{pmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix} \text{ the companion matrix of the polynomial } A(z). \tag{6}$$

2. *The limit in (5) is such that*

$$L(a_1,\ldots,a_p)>0. \tag{7}$$

Proof: See Appendix A.

Thus, from the Theorem we can deduce that $S_T\left(\hat{T}_1,\ldots,\hat{T}_m\right)/T=O_p\left(L\left(a_1,\ldots,a_p\right)\right)$. The following remarks illustrate two specific cases of the process (2) to clarify the results of the Theorem.

Remark 1. If $p=1$, then $L(a_1)=\sigma_u^2/\left(1-a_1^2\right)$ and thus $\lim_{|a_1|\to 1}L(a_1)=\infty$. This implies that when the coefficient a_1 approaches -1 or 1, $L(a_1)$ goes to infinity and consequently the term $S_T\left(\hat{T}_1,\ldots,\hat{T}_m\right)/T$ explodes.

Remark 2. Let $p=2$. We assume without loss of generality that

$$A(z)=1-a_1z-a_2z^2=(1-\alpha_1z)(1-\alpha_2z).$$

Since the model is assumed to be stationary, we have $|\alpha_1|<1$ and $|\alpha_2|<1$. We then obtain

$$L(a_1,a_2)=\frac{\sigma_u^2(1+\alpha_1\alpha_2)}{\left(1-\alpha_1^2\right)\left(1-\alpha_2^2\right)(1-\alpha_1\alpha_2)},$$

which goes to infinity if at least one of the following conditions is satisfied, *i)* $|\alpha_1|\to 1$; *ii)* $|\alpha_2|\to 1$; and *iii)* $\alpha_1\alpha_2\to 1$.[3] Hence, as the roots of $A(z)$ approach the boundary of the stationarity region, the limit $L(a_1,a_2)$ becomes large and then so becomes the term $S_T\left(\hat{T}_1,\ldots,\hat{T}_m\right)/T$. These results may be generalized for an AR(p) process for any $p>2$, and we then obtain the same conclusions as for $p=2$.

As a consequence of these two remarks, for any fixed m, only the first term in the criteria matters asymptotically since the second term goes to 0 as the sample size T increases. Thus, some breaks will spuriously be inferred by the information criteria since the sum of squared residuals $S_T\left(\hat{T}_1,\ldots,\hat{T}_m\right)$ is monotonically decreasing in m. The results of the Remark 1 then constitute a theoretical support to the simulation results of Perron (1997).

[3] Note that the condition $\alpha_1\alpha_2=1$ implies that the roots of $A(z)$ do not lie in the stationarity region of the AR(1) process (e.g., Box and Jenkins, 1976, page 58).

5 Monte Carlo design

We report some Monte Carlo experiments to support the relevance of the theory. To that effect, we set $h = 5$, $1 \leq M \leq 5$, the sample size is fixed at $T = 150$ and the disturbances $\{u_t\}$ are independent and identically distributed standard normal. The simulation results are based on 1000 replications. We simulate series according to AR(1) ($a_1 = 0.8$) and AR(2) ($\alpha_1 = 0.7$ and $\alpha_2 = 0.6$) processes and we run regressions with change in mean ($z_t = 1$). The results given in Table 1 indicate that all the criteria perform badly since they overestimate the number of structural changes, which implies that the considered DGP are series generated by stationary processes with M breaks. We now set $a_1 = 0.9$ and $\alpha_2 = 0.8$, and all the other parameters are kept constant. From the results provided in Table 2, we observe that the overestimation of the number of breaks becomes more severe when we increase the autoregressive parameters, which confirms the theoretical results of the two Remarks. The obtained simulation results show that the conclusions of Perron (1997) don't depend on the fact that the DGP is an AR(1) process; even a stationary autoregressive process of order higher than one leads to an overestimation of the number of breaks.

We have carried out other Monte Carlo experiments with the same characteristics of the considered data-generating processes, but with the modification that here we estimate a model with trend-shifts $\left(z_t = (1, t)'\right)$. The obtained results (not reported but available upon request) provide the same conclusions as for the case of mean-shifts. A future investigation then consists in finding a theoretical explanation for the spuriousness that arises by the information criteria when the DGP is a stationary AR(p) process without breaks and when we run a regression with trend-shifts. This should be possible at the expense of a more complicated mathematical treatment.

6 Conclusion

This paper has discussed the problem of selecting the number of structural changes using some standard information criteria for variables generated by stationary autoregressive processes without any break. We have observed that the estimation of a model with mean-shifts implies the detection of a spurious number of structural changes. Our findings are then rigorous proofs of this fact. Our study is then justified by our aim to provide a mathematical support to the spuriousness that arises by the model selection criteria when choosing the number of breaks.

Appendix A: Proof of the Theorem

Proof of Part 1. Let $\phi_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})'$, then from (2)

$$\phi_t = D\phi_{t-1} + \varepsilon_t,$$

where D is given by (6) and $\varepsilon_t = (u_t, 0, \ldots, 0)'$. From Theorem 1 of Lai and Wei (1983), we obtain

$$\frac{1}{T}\sum_{t=1}^{T} \phi_t \phi_t' \overset{a.s.}{\to} F = \sum_{j=0}^{\infty} D^j K D^{j\prime}, \qquad as\ T \to \infty,$$

where K is defined in the Theorem. As a consequence and since $y_t = e_1'\phi_t$, we have

$$\frac{1}{T}\sum_{t=1}^{T} y_t^2 \overset{a.s.}{\to} e_1' F e_1, \tag{8}$$

where e_1 is a p-vector given in the Theorem. Applying Theorem 3.1.1 of Brockwell and Davis (1987) and since the model (2) is causal:

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}.$$

From Brockwell and Davis (1987) (Remark 2 page 212), we deduce that

$$T\text{var}\left(\frac{1}{T}\sum_{t=1}^{T} y_t\right) \to \sigma_u^2 \left(\sum_{j=0}^{\infty} \psi_j\right)^2, \qquad as\ T \to \infty,$$

consequently $\text{var}\Big((1/T)\sum_{t=1}^{T} y_t\Big) \to 0$, and hence $(1/T)\sum_{t=1}^{T} y_t \overset{m.q.}{\to} E(y_1) = 0$, which implies that

$$\frac{1}{T}\sum_{t=1}^{T} y_t \overset{p}{\to} 0. \tag{9}$$

For $i = 2, \ldots, m+1$,

$$\frac{1}{T_i - T_{i-1}} \sum_{t=T_{i-1}+1}^{T_i} y_t = \frac{T_i}{T_i - T_{i-1}} \frac{1}{T_i} \sum_{t=1}^{T_i} y_t - \frac{T_{i-1}}{T_i - T_{i-1}} \frac{1}{T_{i-1}} \sum_{t=1}^{T_{i-1}} y_t.$$

Then

$$\frac{1}{T_i - T_{i-1}} \sum_{t=T_{i-1}+1}^{T_i} y_t \overset{p}{\to} \frac{\lambda_i}{\lambda_i - \lambda_{i-1}} E(y_1) - \frac{\lambda_{i-1}}{\lambda_i - \lambda_{i-1}} E(y_1) = E(y_1) = 0.$$

Note that $(1/T_1)\sum_{t=1}^{T_1} y_t \overset{p}{\to} 0$. Hence for $i = 1, \ldots, m+1$,

$$\frac{1}{T_i - T_{i-1}} \sum_{t=T_{i-1}+1}^{T_i} y_t \overset{p}{\to} 0. \tag{10}$$

Using the same arguments as above, we obtain $(1/T)\sum_{t=T_{i-1}+1}^{T_i} y_t \xrightarrow{p} 0$, which implies together with (10) that

$$\frac{1}{T}\sum_{i=1}^{m+1}\frac{\left(\sum_{t=T_{i-1}+1}^{T_i} y_t\right)^2}{T_i - T_{i-1}} \xrightarrow{p} 0. \tag{11}$$

Note that for $z_t = 1$,

$$\frac{S_T\left(\hat{T}_1,\ldots,\hat{T}_m\right)}{T} = \min_{(T_1,\ldots,T_m)}\left\{\frac{1}{T}\sum_{t=1}^{T} y_t^2 - \frac{1}{T}\sum_{i=1}^{m+1}\frac{\left(\sum_{t=T_{i-1}+1}^{T_i} y_t\right)^2}{T_i - T_{i-1}}\right\}.$$

From (8) and (11), we obtain

$$\frac{S_T\left(\hat{T}_1,\ldots,\hat{T}_m\right)}{T} \xrightarrow{p} L\left(a_1,\ldots,a_p\right) = e_1'\sum_{j=0}^{\infty} D^j K D^{j\prime} e_1.$$

This proves the first part of the Theorem.

Proof of Part 2. To prove that $L\left(a_1,\ldots,a_p\right) > 0$, we apply the Proposition 3.1 of Boutahar (1991) to show that the matrix F is a positive definite matrix and hence $L\left(a_1,\ldots,a_p\right) = e_1' F e_1 > 0$ since $e_1 \neq 0$. This proves the second part of the Theorem.

Appendix B: Simulation Results

Table 1. Percentage of breaks selected by the information criteria

		AR(1) process		AR(2) process	
M	$\hat{m}$	BIC	YIC	BIC	YIC
1	0	6.9	12.6	5.1	9.4
	1	93.1	87.4	94.9	90.6
2	0	0.2	1.1	0.0	0.2
	1	3.4	6.6	1.2	33.0
	2	96.4	92.3	98.8	96.5
3	0	0.2	0.6	0.0	0.2
	1	0.1	1.5	0.0	0.2
	2	6.6	13.6	2.8	6.5
	3	93.1	84.3	97.2	93.1
4	0	0.1	0.5	0.0	0.1
	1	0.0	1.0	0.0	0.1
	2	0.8	3.9	0.1	0.5
	3	7.3	13.9	2.7	6.4
	4	91.8	80.7	97.2	92.9
5	0	0.0	0.4	0.0	0.0
	1	0.0	0.7	0.0	0.1
	2	0.3	2.8	0.0	0.2
	3	1.0	5.9	0.0	0.7
	4	11.6	18.1	3.4	8.9
	5	87.1	72.1	96.6	90.1

Table 2. Percentage of breaks selected by the information criteria

		AR(1) process		AR(2) process	
M	$\hat{m}$	BIC	YIC	BIC	YIC
1	0	1.4	2.6	1.6	2.7
	1	98.6	97.4	98.4	97.3
2	0	0.0	0.0	0.0	0.0
	1	0.8	2.4	0.6	1.4
	2	99.2	97.6	99.4	98.6
3	0	0.0	0.0	0.0	0.0
	1	0.0	0.1	0.0	0.0
	2	1.4	3.8	0.7	2.1
	3	98.6	96.1	99.3	97.9
4	0	0.0	0.0	0.0	0.0
	1	0.0	0.1	0.0	0.0
	2	0.1	1.0	0.0	0.0
	3	2.4	7.0	0.9	2.2
	4	97.5	91.9	99.1	97.8
5	0	0.0	0.0	0.0	0.0
	1	0.0	0.1	0.0	0.0
	2	0.1	0.9	0.0	0.0
	3	0.1	0.9	0.0	0.3
	4	3.8	8.4	1.0	2.7
	5	96.0	89.7	99.0	97.0

References

[1] Bai, J. (1998) "A note on spurious break", *Econometric Theory* **14**, 663-669.

[2] Bai, J., and P. Perron (1998) "Estimating and testing linear models with multiple structural changes", *Econometrica* **66**, 47-78.

[3] Bai, J., and P. Perron (2003) "Computation and analysis of multiple structural change models", *Journal of Applied Econometrics* **18**, 1-22.

[4] Boutahar, M. (1991) "Asymptotic distribution of least squares estimates in unstable ARX (p, s) models", *Stochastics and Stochastics Reports* **37**, 105-126.

[5] Boutahar, M., and J. Jouini (2007) "Wrong estimation of the true number of shifts in structural break models: theoretical and numerical evidence", *Economics Bulletin* **3**, 1-10.

[6] Box, G.E.P., and G.M. Jenkins (1976) *Time Series Analysis: Forecasting and Control*, Holden-Day.

[7] Brockwell, P.J., and R.A. Davis (1987) *Time Series: Theory and Methods*, Springer.

[8] Jouini, J., and M. Boutahar (2005) "Evidence on structural changes in US time series", *Economic Modelling* **22**, 391-422.

[9] Lai, T.L., and C.Z. Wei (1983) "Asymptotic properties of general autoregressive models and strong consistency of least squares estimates of their parameters", *Journal of Multivariate Analysis* **13**, 1-13.

[10] Liu, J., Wu, S., and J.V. Zidek (1997) "On segmented multivariate regressions", *Statistica Sinica* **7**, 497-525.

[11] Nunes, L.C., Newbold, P., and C.-M. Kuan (1996) "Spurious number of breaks", *Economics Letters* **50**, 175-178.

[12] Perron, P. (1997) "L'estimation de modèles avec changements structurels multiples", *Actualité Économique* **73**, 457-505.

[13] Yao, Y-C. (1988) "Estimating the number of change-points via Schwarz' criterion", *Statistics and Probability Letters* **6**, 181-189.

[14] Yao, Y-C. and S.T. Au (1989) "Least squares estimation of a step function", *Sankhyā* **51**, 370-381.