ECONOMICS BULLETIN

wrong estimation of the true number of shifts in structural break models: Theoretical and numerical evidence

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

Jamel JOUINI
GREQAM, Université de la Méditerranée, France and FSEGN, ESSAIT and LEGI, Université 7 Novembre de Carthage, Tunisie

Abstract

The aim of the paper is to consider the problem of selecting the number of breaks in the mean of a time series. Indeed, we prove analytically and show by a Monte Carlo study that some model selection criteria will tend to choose a spuriously high number of structural breaks when the process is trend-stationary without changes. The important question suggested by our results is that of distinction between trend-stationary process and random walk when modelling real data series.

We are grateful to Anne PEGUIN and Claude DENIAU for their helpful remarks and comments. We remain responsible for all errors and omissions.
Citation: BOUTAHAR, Mohamed and Jamel JOUINI, (2007) "wrong estimation of the true number of shifts in structural break models: Theoretical and numerical evidence." *Economics Bulletin*, Vol. 3, No. 3 pp. 1-10
Submitted: March 10, 2006. **Accepted:** January 10, 2007.
URL: http://economicsbulletin.vanderbilt.edu/2007/volume3/EB-06C20004A.pdf

1 Introduction

The existence of structural change in statistical models acquires a fundamental importance in the literature. In the context of determining the number of breaks using some model selection criteria, Yao (1988), Yao and Au (1989) and Yin (1988) study the detection of mean-shifts using the Bayesian Information Criterion. The most important contribution is that of Bai and Perron (1998) who consider multiple break models estimated by least-squares and propose selection procedures based on a sequence of tests to estimate consistently the number of structural changes. The usefulness of these procedures is illustrated by several works in the literature such as Bai and Perron (2003), and Jouini and Boutahar (2005).

When the data-generating process is a trend-stationary process without any structural change, some model selection criteria have a tendency to estimate the maximum possible number of breaks when we run a regression with mean-shifts. In this paper, we provide a mathematical proof for this phenomenon and compare our theoretical results to those obtained by Nunes, Newbold and Kuan (1996) who show that adopting the Bayesian information criterion when the generating process is a random walk without breaks leads to an estimated number of breaks equal to the maximum allowed. We present simulation evidence to confirm the theory and to produce additional insights. While our paper borrows the idea of proving analytically some findings in the context of estimating the number of breaks from Nunes, Newbold and Kuan (1996), ours is the first to consider multiple criteria. Note that our findings provide useful insights since we can distinguish appropriately between random walk and trend-stationary process when modelling real data series.

The remainder of the paper is organized as follows. The second section presents the model and the estimation method. Section 3 recalls some model selection criteria. In section 4, we provide the main theoretical result of the paper. Section 5 reports simulation experiments to support the theoretical results. 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, "$\Rightarrow$" weak convergence in the space $D\,[0,1]$ under the Skorohod metric (Pollard, 1984), and "$\overset{a.s.}{\to}$" convergence almost surely.

2 The model and estimation method

Consider the following linear regression model of structural change with m breaks:

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

for $j = 1,\dots,m+1$, $T_0 = 0$ 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 $(T_1,\dots,T_m)$ are explicitly treated as unknown and for $i = 1,\dots,m$, we have $T_i = [\lambda_i T]$ where $0 < \lambda_1 < \cdots < \lambda_m < 1$. Let $\delta = \left(\delta_1^{'}, \delta_2^{'}, \dots, \delta_{m+1}^{'}\right)^{'}$.

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

δ_j 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 $T_i - T_{i-1} \geq h$.[1]. Thus, the break point 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 experiments, 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 model selection criteria

To detect the number of breaks, Yao (1988) suggests the use of the Bayesian Information Criterion defined as

$$BIC(m) = \ln\left(S_T\left(\hat{T}_1, \ldots, \hat{T}_m\right)/T\right) + p^* \ln(T)/T, \tag{2}$$

where $p^* = (m+1)q + m$ is the number of unknown parameters. The author shows that, for the change in mean model, $\hat{m}$ is a consistent estimator of m^0, the true number of breaks, provided $m^0 \leq M$ with M a fixed upper bound for m and the error term of the model is normally distributed. Another criterion proposed by Yao and Au (1989) is defined as

$$YIC(m) = \ln\left(S_T\left(\hat{T}_1, \ldots, \hat{T}_m\right)/T\right) + mC_T/T, \tag{3}$$

where $C_T = 0.368T^{0.7}$.[2] Liu, Wu and Zidek (1997) propose a modified Schwarz' criterion that takes the form

$$MIC(m) = \ln\left(S_T\left(\hat{T}_1, \ldots, \hat{T}_m\right)/(T - p^*)\right) + 0.299p^* [\ln(T)]^{2.1}/T. \tag{4}$$

We remark that these criteria have not the same penalty term. Indeed, that of the criterion MIC is heavier than those of the criteria BIC and YIC. As we will show in the next section, this affects the results for all the processes. The estimated number of break dates $\hat{m}$ is obtained by minimizing the above-mentioned criteria given an upper bound M for m.

Perron (1997) carried out Monte Carlo simulations to study the behavior of the above-mentioned information criteria in the context of selecting the number of break dates in the trend function of a series in the presence of serial correlation. The criteria perform reasonably well when the errors are not correlated but overestimate the number of changes when serial correlation is present. When the errors are not correlated but a lagged dependent variable is present, the criterion BIC performs badly when the autoregressive coefficient is large. On the other hand, the criterion MIC performs better under the null hypothesis of structural stability but underestimates the number of breaks when some are present. His results show that the conclusions of Nunes, Newbold and Kuan (1996)

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

[2] Note that this sequence is proposed by Liu, Wu and Zidek (1997).

don't depend on the fact that the data-generating process is a random walk; even an AR(1) process with a correlation degree smaller than one leads to an overestimation of the number of breaks.

4 A poor identification of the number of breaks

Finding a theoretical explanation for the overestimation was evoked by Bai (1998) who gives a mathematical proof for the phenomenon that when the errors of a linear regression model without any break are $I(1)$, 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.

We define the sum of squared residuals

$$\begin{aligned} S_T\left(\hat{T}_1,\ldots,\hat{T}_m\right) &= \min_{(T_1,\ldots,T_m)}\left[\sum_{t=1}^{T} y_t^2\right. \\ &\left.-\sum_{i=1}^{m+1}\left(\sum_{t=T_{i-1}+1}^{T_i} y_t z_t'\right)\left(\sum_{t=T_{i-1}+1}^{T_i} z_t z_t'\right)^{-1}\left(\sum_{t=T_{i-1}+1}^{T_i} z_t y_t\right)\right]. \end{aligned} \tag{5}$$

Nunes, Newbold and Kuan (1996) show that for a random walk without changes, and when estimating mean-shift (i.e. $z_t = 1$) and trend-shift (i.e. $z_t = (1,t)'$) models, the criteria overestimate the number of breaks. This conclusion is supported by the simulation results reported in Table 1.

Our contribution in this paper consists in providing a mathematical proof for the problem of overestimation that appears when estimating a mean-shift model using data generated by a trend-stationary process without any break:

$$y_t = a + bt + u_t, \qquad 1 \le t \le T, \tag{6}$$

where $a \in \mathbb{R}$, $b \in \mathbb{R}$ and $u_t \sim i.i.d.\ N\left(0,\sigma_u^2\right)$. The following Theorem indicates that the estimation of a model with M breaks leads to the selection of a spurious number of changes (namely the maximum permitted number) by the information criteria.

Theorem. *Suppose that the data are generated according to the model (6) with $b \neq 0$ and that we estimate a model with change in mean, i.e. $z_t = 1$. Then, we have for $T \to \infty$*

1.

$$\frac{S_T\left(\hat{T}_1,\ldots,\hat{T}_m\right)}{T^3} \overset{a.s.}{\to} C(m), \tag{7}$$

where

$$C(m) = b^2\left(\frac{1}{3} - \frac{1}{4}\left(1 + l(m)\right)\right), \tag{8}$$

and

$$l(m) = \max_{0<\lambda_1<\cdots<\lambda_m<1} \sum_{i=1}^{m+1} \lambda_i \lambda_{i-1}\left(\lambda_i - \lambda_{i-1}\right). \tag{9}$$

2. *The function $l(m)$ is such that $l(1) = 1/4$, $l(2) = 8/27$, and for any $m \geq 3$*

$$l(m) = \frac{\left\{\frac{2\sqrt{2}}{3}m^3 + 2\left(2-\sqrt{2}\right)m^2 + \left(5\sqrt{2} + \frac{\sqrt{2}}{3} - 8\right)m + 6 - 4\sqrt{2}\right\}}{\left(2+\sqrt{2}\left(m-1\right)\right)^3}.$$

Therefore $l(m) < 1/3$ for all m and hence $C(m) > 0$.

Proof: See Appendix A.

Thus, from the Theorem we conclude that $S_T\left(\hat{T}_1, \ldots, \hat{T}_m\right) = O_p\left(T^3\right)$, which implies that for any fixed m, only the first term in the criteria matters asymptotically since the penalty term goes to 0 as the sample size T increases. Consequently, the criteria select the maximum permitted number of breaks M since $S_T\left(\hat{T}_1, \ldots, \hat{T}_m\right)$ is monotonically decreasing in m. In particular, the criteria BIC and YIC will be minimized at $m = M$ with percentages higher than those of the criterion MIC. These two criteria suggest that a trend-stationary process is a series generated by a stationary process with M structural breaks. This conclusion is supported by the simulation results considered in Table 2 for the mean-shift model.

The fact that from Nunes, Newbold and Kuan (1996) $S_T\left(\hat{T}_1, \ldots, \hat{T}_m\right) = O_p\left(T^2\right)$ implies that the magnitude of the sum of squared residuals is higher for the trend-stationary process than the random walk. Consequently, the proportion of selecting the upper bound as estimate of the number of breaks is higher for the former than the latter for all the criteria (see Tables 1 and 2 for the case of mean-shifts).

5 Monte Carlo analysis

Based on Monte Carlo simulations we attempt to confirm the above-mentioned theoretical results. h and M take value 5, the sample size is fixed at $T = 150$ and the disturbances $\{u_t\}$ are independent and identically distributed standard normal. All the reported simulation results are based on 1000 replications.

Experiment 1. We consider an $I(1)$ process without any structural change as a data-generating process. The corresponding results are reported in Table 1. The results of the mean-shift and trend-shift models are very similar for the criteria BIC and YIC. The estimator of the number of break dates obtained by the criterion MIC has some distribution on the set $\{0, 1, \ldots, M\}$ where the frequency of selecting the upper bound M is the highest. The criteria BIC and YIC select the maximum possible number of breaks on the overwhelming majority of occasions. These criteria suggest that the random walk is a series generated by a stationary process with M break dates. This bias towards the overestimation of the number of breaks is less severe for the criterion MIC especially in the case of a change in trend model. Hence, a heavy penalty reduces the tendency of choosing the upper bound as estimate of the number of changes.

Experiment 2. This experiment considers a series generated according to a trend-stationary process. Our simulation experiment is carried out with $a = 2.0$ and $b = 0.1$. The results are provided in Table 2. We first consider the case of change in mean. All the criteria perform badly in the sense that they select the maximum permitted number of breaks $m = 5$ (100% of the time for the criteria BIC and YIC and nearly 100% of the time for the criterion MIC). As for the random walk, the criteria BIC and YIC have higher tendency to overestimate the number of changes than the criterion MIC. As the theory predicts, we remark that the frequency of selecting 5 break points is higher for the trend-stationary process than the random walk (see Tables 1 and 2 for the case of mean-shifts). The results show that the conclusions of Nunes, Newbold and Kuan (1996) don't depend on the fact that the data-generating process is an $I(1)$ process; even a trend-stationary process leads to an overestimation of the number of breaks.

For the trend-shift model, the information criteria correctly select the true number of breaks, namely 100% of the time for the criterion MIC, 98% for the BIC and 91% for the YIC.

6 Conclusion

This paper has discussed the problem of selecting the number of breaks using some model selection criteria. We have remarked that for some data-generating processes without any break, the estimation of a model with structural change results in the appearance of a spurious number of breaks. Our results are simply rigorous proofs of this fact. The tendency towards the overestimation of the number of changes is less severe when a heavy penalty is used. Our results, together with those of Nunes, Newbold and Kuan (1996), suggest that choosing a random walk or a trend-stationary process to model real data series can also be done based on the study of structural change models.

Appendix A: Proof of the Theorem

Proof of Part 1. We have for $z_t = 1$

$$S_T\left(\hat{T}_1, \ldots, \hat{T}_m\right) = \min_{(T_1,\ldots,T_m)} \left\{ \sum_{t=1}^{T} (a + bt + u_t)^2 - \sum_{i=1}^{m+1} \frac{\left(\sum_{t=T_{i-1}+1}^{T_i} (a + bt + u_t)\right)^2}{T_i - T_{i-1}} \right\}. \tag{10}$$

We have $\left(1/T^3\right) \sum_{t=1}^{T} u_t^2 \overset{a.s.}{\to} 0$ and $\left(1/T^3\right) \sum_{t=1}^{T} u_t \overset{a.s.}{\to} 0$ since by the strong law of large numbers $(1/T) \sum_{t=1}^{T} u_t^2 \overset{a.s.}{\to} E\left(u_1^2\right) = \sigma_u^2$ and $(1/T) \sum_{t=1}^{T} u_t \overset{a.s.}{\to} E\left(u_1\right) = 0$. $\sum_{t=1}^{T} tu_t$ is a martingale transform, then by applying the Lemma 1 of Lai and Wei (1982), we obtain

$$\left| \sum_{t=1}^{T} tu_t \right|^2 = O\left(\sum_{t=1}^{T} t^2 \ln \left(\sum_{t=1}^{T} t^2 \right) \right) = O\left(T^3 \ln(T)\right), \qquad \text{almost surely,}$$

consequently

$$\left| \frac{1}{T^3} \sum_{t=1}^{T} tu_t \right| = O\left(\frac{(\ln(T))^{1/2}}{T^{3/2}} \right), \qquad \text{almost surely,}$$

and then

$$\frac{1}{T^3} \sum_{t=1}^{T} tu_t \overset{a.s.}{\to} 0.$$

Therefore

$$\sum_{t=1}^{T} (a + bt + u_t)^2 = \frac{b^2}{3} T^3 + o\left(T^3\right), \qquad \text{almost surely.} \tag{11}$$

Using the same arguments as above and since $T_i = [\lambda_i T]$, we have with probability one

$$\sum_{i=1}^{m+1} \frac{\left(\sum_{t=T_{i-1}+1}^{T_i} (a + bt + u_t)\right)^2}{T_i - T_{i-1}} = \frac{b^2}{4} T^3 \sum_{i=1}^{m+1} (\lambda_i + \lambda_{i-1})^2 (\lambda_i - \lambda_{i-1}) + o\left(T^3\right). \tag{12}$$

From (10)–(12), we deduce that

$$S_T\left(\hat{T}_1, \ldots, \hat{T}_m\right) = \min_{(\lambda_1,\ldots,\lambda_m)} \left\{ \frac{b^2}{3} T^3 - \frac{b^2}{4} T^3 \sum_{i=1}^{m+1} (\lambda_i + \lambda_{i-1})^2 (\lambda_i - \lambda_{i-1}) + o\left(T^3\right) \right\}.$$

Then

$$\frac{S_T\left(\hat{T}_1,\ldots,\hat{T}_m\right)}{T^3} \xrightarrow{a.s.} b^2\left(\frac{1}{3}-\frac{1}{4}\max_{0<\lambda_1<\cdots<\lambda_m<1}\sum_{i=1}^{m+1}\left(\lambda_i+\lambda_{i-1}\right)^2\left(\lambda_i-\lambda_{i-1}\right)\right)$$
$$= b^2\left(\frac{1}{3}-\frac{1}{4}\left(1+\max_{0<\lambda_1<\cdots<\lambda_m<1}\sum_{i=1}^{m+1}\lambda_i\lambda_{i-1}\left(\lambda_i-\lambda_{i-1}\right)\right)\right).$$

This proves the first part of the Theorem.

Proof of Part 2. We have

$$l(m) = \max_{0<\lambda_1<\cdots<\lambda_m<1}\sum_{i=1}^{m+1}\lambda_i\lambda_{i-1}\left(\lambda_i-\lambda_{i-1}\right) = \max_{0<\lambda_1<\cdots<\lambda_m<1} L\left(\lambda_1,\ldots,\lambda_m\right).$$

We easily show that $l(1) = 1/4 < 1/3$ and $l(2) = 8/27 < 1/3$. For $m \geq 3$, we have the following relations: $\lambda_2 = 2\lambda_1$, $\lambda_3 = \left(2+\sqrt{2}\right)\lambda_1, \ldots, \lambda_m = \left(2+(m-2)\sqrt{2}\right)\lambda_1$. It follows that $\lambda_i - \lambda_{i-1} = \sqrt{2}\lambda_1$ for $i = 3,\ldots,m$. Note that we have

$$\frac{\partial L}{\partial \lambda_m} = \lambda_{m-1}\left(\lambda_m-\lambda_{m-1}\right)+\lambda_m\lambda_{m-1}+1-2\lambda_m = 0.$$

Substituting λ_{m-1} and λ_m, we obtain

$$\lambda_1 = \frac{1}{2+(m-1)\sqrt{2}} \text{ or } \lambda_1 = \frac{1}{2+(m-3)\sqrt{2}}$$

The second solution leads to $\lambda_{m-1} = 1$, it is then not considered. We have

$$\begin{aligned} l(m) &= \lambda_2\lambda_1\left(\lambda_2-\lambda_1\right)+\sum_{i=3}^{m}\lambda_i\lambda_{i-1}\left(\lambda_i-\lambda_{i-1}\right)+\lambda_m\left(1-\lambda_m\right) \\ &= 2\lambda_1^3+\sqrt{2}\lambda_1^3\sum_{i=3}^{m}\left(2+(i-2)\sqrt{2}\right)\left(2+(i-3)\sqrt{2}\right) \\ &\quad +\left(2+(m-2)\sqrt{2}\right)\lambda_1\left(1-\left(2+(m-2)\sqrt{2}\right)\lambda_1\right). \end{aligned}$$

Substituting λ_1, we obtain

$$l(m) = \left\{\frac{2\sqrt{2}}{3}m^3+2\left(2-\sqrt{2}\right)m^2+\left(5\sqrt{2}+\frac{\sqrt{2}}{3}-8\right)m+6-4\sqrt{2}\right\}/\left(2+\sqrt{2}(m-1)\right)^3,$$

and then

$$l(m) \in \left[\frac{12\sqrt{2}+18}{\left(2+2\sqrt{2}\right)^3},\frac{1}{3}\right[, \qquad \forall\, m \geq 3.$$

It follows that $l(m) < 1/3$ for all m, which implies that $C(m) > 0$ for $b \neq 0$. This proves the second part of the Theorem.

Appendix B: Simulation Results

Table 1. Percentage of breaks selected
by the information criteria

	Mean-shifts			Trend-shifts		
$\hat{m}$	BIC	YIC	MIC	BIC	YIC	MIC
0	0.0	0.0	0.0	0.0	0.0	0.1
1	0.0	0.0	0.2	0.0	0.0	1.3
2	0.0	0.0	1.2	0.0	0.0	7.3
3	0.0	0.1	3.8	0.1	0.0	18.4
4	0.8	1.6	13.4	1.1	0.3	26.6
5	99.2	98.3	81.4	98.8	99.7	46.3

Table 2. Percentage of breaks selected
by the information criteria

	Mean-shifts			Trend-shifts		
$\hat{m}$	BIC	YIC	MIC	BIC	YIC	MIC
0	0.0	0.0	0.0	98.1	90.5	100.0
1	0.0	0.0	0.0	1.7	7.3	0.0
2	0.0	0.0	0.0	0.2	1.9	0.0
3	0.0	0.0	0.0	0.0	0.3	0.0
4	0.0	0.1	7.5	0.0	0.0	0.0
5	100.0	99.9	92.5	0.0	0.0	0.0

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

[5] Lai, T.L., and C.Z. Wei (1982) "Least squares estimates in stochastic regression models with application to identification and control in dynamic system", *The Annals of Statistics* **10**, 154-166.

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

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

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

[9] Pollard, D. (1984) *Convergence of Stochastic Processes*, Springer-Verlag: New York.

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

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

[12] Yin, Y.Q. (1988) "Detection of the number, locations and magnitudes of jumps", *Communications in Statistics-Stochastic Models* **4**, 445-455.