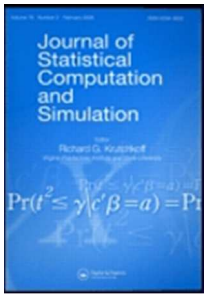

Bootstrap Order Selection for SETAR models

Journal:	*Journal of Statistical Computation and Simulation*
Manuscript ID:	GSCS-2012-0577
Manuscript Type:	Original Paper
Date Submitted by the Author:	08-Dec-2012
Complete List of Authors:	Fenga, Livio; University of California San Diego, Mathematics Politis, Dimitris; University of California San Diego, Mathematics
Areas of Interest:	MODEL SELECTION, TIME SERIES, BOOTSTRAP
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
setar_no_fig.tex	

Bootstrap Order Selection for SETAR models

Livio Fenga*, Dimitris N. Politis†

*University of California San Diego - Department of Mathematics email:lfenga@math.ucsd.edu
†University of California San Diego - Department of Mathematics email:dpolitis@math.ucsd.edu

Abstract

In this paper, we investigate the selecting performances of a bootstrapped version of the Akaike Information Criterion (AIC) for non-linear SETAR-Type data generating processes. Empirical results will be obtained via Monte Carlo simulations. The quality of our method is assessed by comparison with its non-bootstrap counterpart and through a novel procedure based on Artificial Neural Networks.

Keywords: Bootstrap, AIC, SETAR process, order selection, Moving Block Bootstrap, Artificial Neural Networks.

1 Introduction

Nature is nonlinear in general and so are most of its expressions. Non linearity should not therefore be seen as an accident or something happening under certain conditions, but rather as a natural consequence of the complexity of the real world. Despite this complexity, real structures can be brought to a linear framework through *ad hoc* transformations. Such an operation results in an approximation justified by both a theoretical and practical point of view. In fact, the linear set up usually provides with an easier resort to well consolidated theoretical frameworks along with fast and efficient computing algorithms. As a result of that, the problem at hand becomes generally more tractable. This is especially true for the practitioners which, in addition, can benefit from a huge selection of software packages usually featured with a rich portfolio of linear models.

Even though linearity can be considered a limiting case, linear techniques have been widely used to model systems showing strong nonlinearities in many fields of application for example, time series analysis. Many time dependent processes, in fact, show nonlinear dynamics which often makes linear models able to provide only a suboptimal representation of the related Data Generating Process (henceforth DGP). Such a nonlinearity may be generated by facts like non normality, regime dependent behavior, asymmetry or multimodality of the density as well as the presence of aberrant observations, structural breaks, time irreversibility and even sensitivity to initial condition. Especially for "real life" data, nonlinearity can also arise from survey related phenomena such as sampling - non sampling errors, noisy data, methodological changes.

To take in account the particular autocovariance structure of nonlinear processes, different types of models have been proposed. Among them, threshold models are recognized as a powerful tool to address the issue of reproducing the underlying DGP in presence of a multiple regime dependent behavior. In order to cope with such asymmetries, common features of many economic and financial time series, a portfolio of models have been developed; excellent summaries, in this regard, are in [45] and, more recently, in [40]. These models include Markov-switching autoregressive, Threshold autoregressive models, Smooth-transition autoregressive (TAR) [43] and Contemporaneous - threshold STAR [8] models.

One of the simplest and most effective structures in the set of threshold models are the threshold autoregressive (TAR) models. They have the nice property to be linear in different regions of the state space and can be estimated using regression methods. They have been employed in periodic time series, e.g. for business cycle analysis purposes [44], to model the asymmetry in the quarterly US real GNP (see, *inter alia*, [38], [39]), and to deal with cases where limit cycles are present. TAR models assume the transition variable to be exogenous and observable (e.g. a leading indicator). When this variable is expressed by a linear combination of the lagged values of the time series, the model is referred to as Self Exciting Threshold Autoregressive ($SETAR$) model.

For these type of models, the structure determination problem is still an open issue. Therefore, in this paper we present a bootstrap selection criterion designed for the purpose of statistical-based identification of the "best" model in a set of non nested candidate models belonging to the class $SETAR$. As it will be seen, the method relies on a bootstrapped version of the Minimum Akaike Information Criterion Estimate (henceforth MAICE) procedure [1], already successfully tested in

the case of linear ARMA processes [10].

Originally, the bootstrap was designed for *iid* data [9] but in the last 20 years it has seen fruitful application in time series analysis (see e.g. [20] and [35]). In particular, in the present study a Block resample scheme [18] has been employed.

The performances of our method have been evaluated in a comparative way using the outcomes generated by the standard MAICE procedure. An alternative procedure based on neural networks will also be presented.

In section 2, $SETAR$ models and a modified version of Akaike Information Criterion (AIC) for models belonging to this class are presented. Section 3 is devoted to the non-bootstrap estimation method and the selection strategy for $SETAR$ models. The employed bootstrap method and our selection procedure are detailed, respectively in Section 4 and 5. For what regards the empirical evaluation of the proposed procedure, in section 6 a simulation study is presented and the quality of our procedure assessed through a comparison with the outcomes obtained in the non bootstrap world and, in Section 7, by means of a novel approach based on Artificial Neural Networks.

2 Akaike information criterion for SETAR models

Real life stochastic processes, like economic and financial time series, often exhibit non linear patterns such as asymmetric cycles, multimodal jumps as well as multiple regime dependent behaviors and chaos. In such a situation, linear models generally offer poor performances and, as a result of that, the resort to a more complicated non linear structures are in order. In this regard, a very important class of nonlinear models is the Threshold Autoregressive Models (TAR) proposed by Tong (1983).

We decided to focus on model selection for $SETAR$ instead of TAR models because, by design, $SETAR$ models do not require the specification of the exogenous variable. Such a choice, in fact, may introduce strong elements of uncertainty, being related to its determination in a potentially huge set of candidate exogenous variables (several economic indicators are in fact available on a monthly or higher frequency basis). In addition, these variables may be of poor quality or subjected to revisions procedures or filtering techniques (e.g. seasonal adjustment) that may lead to edge effects. Their consistency with the problem at hand and the availability for the time span object of investigation, are also critical issues.

Given data $(y_{-H+1},,y_{-1},y_0,y_1,,y_T)$, a $SETAR$ model is defined as:

$$y_t = \varphi_{j0} + \sum_{k=1}^{p_j} \varphi_{jk} y_{t-k} + \varepsilon_{jt} \qquad \gamma_{j-1} < y_{t-d} \leq \gamma_j \tag{1}$$

where:

φ_{jk} $(k = 1,...p_j)$ are the parameters in state j, $(j = 1,2,...,l)$, ε_{jt} is a 0–mean, variance σ^2 Gaussian White Noise, $-\infty = \gamma_0 < \gamma_1 ... < \gamma_l = +\infty$ are the threshold values and d $(d > 1)$ is the delay parameter.

With T we denote the sample size and with

$$H = max(p_j, d) \qquad j = 1, 2, .., l \tag{2}$$

the number of the initial conditions.

The $SETAR$ model has parameter

$$\theta = (\phi_{10}, \phi_{11}, ..., \phi_{1p_1}, ..., \phi_{l0}, ..., \phi_{lp_l}; \gamma_1, ..., \gamma_l, d, \sigma_1^2, ..., \sigma_{p_l}^2).$$

We assume (1) to be stationary, ergodic with finite second moments and that the density associated with $(y_1, y_2, ..., y_T)'$ is positive everywhere.

As already pointed out, structure determination for $SETAR$ models is an open, extremely important issue. In fact, the intrinsic model structure, with a potentially high number of parameters involved, makes $SETAR$ models prone at increasing their dimensionality quite easily. In addition, the possibility of distributed lags along with different combinations of threshold delays and threshold values, might mean, for the set of competitive models, to increase their cardinality as well, adding uncertainty to the existing uncertainty. Therefore, selecting the correct model is vital especially with "small to moderate" sample sizes, where the data might not support adequately complex model structures. In this framework, the model builder has to deal with the delicate task to extract the "right" signal using parsimonious structures.

There are several strategies for structure determination in time series analysis. Information criteria (henceforth IC) represent one of the most effective and powerful tools. Among them, Akaike's Information Criterion AIC ([2]), Bayes Information Criterion BIC [42], Hannan-Quinn information criterion HQC [15] have gained widespread acceptance; for an excellent survey see [17] and [27].

In particular, AIC is based on an estimate of the expected relative entropy (the *Kullback - Leibler* divergence) contained in a estimated model, that is the degree of divergence from the "true" theoretical model. In essence, they provide a metric for balancing between goodness of fit and complexity of the model and select an order that minimizes a quantity expressed by the maximized log likelihood plus a penalty term. Being the latter a function of the sample size, many IC have been proposed according to the different penalty functions adopted.

In this study we will focus on a modified version of the AIC criterion conditional upon a specified threshold delay and fixed threshold value, called *pooled-AIC* [45]. Given the complexity of the likelihood function for $SETAR$ models, the exact maximum likelihood is not defined, so the pooled AIC has been derived starting from the conditional maximum likelihood given by:

$$LogL(y \mid \theta) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{j=1}^{k}(T_j \log \sigma_{\varepsilon_j}^2 + \frac{S(\phi_j, \gamma_{j-1}, \gamma_j)}{\sigma_{\varepsilon_j}^2})$$

where p_j and $\hat{\sigma}_{\varepsilon j}^2$ are respectively the lag length and the residual variances for the $j-th$ generic regime; T, ϕ_j , d, γ have already been defined (1) and $S(\phi_j, \gamma_{j-1}, \gamma_j) = \sum_{\gamma_{j-1} \le y_{t-d} < \gamma_j} \varepsilon_{jt}^2$,

$j = 1, ..., l$.

For a generic model, realization of a l-regimes $SETAR$ process, pooled AIC is defined as follows [45]:

$$AIC(p_1, ..., p_l) = \sum_j \left\{ T_j \ln \hat{\sigma}^2_{\varepsilon j} + 2p_j \right\}, \quad j = 1, .., l \tag{3}$$

Being the sum of the AICs computed for each of the l regimes, this "compound statistic" is able to balance fitness and complexity by taking into account the peculiar dynamic structures inherent to each state of the process at hand.

3 Estimation and selection strategy for l-regimes SETAR models

Different methods have been proposed for $SETAR$ parameter estimation - (see [11] for a summary). It is a non trivial task, given the presence of the threshold parameter, which plays the crucial role of allocating the data across the l regimes. In addition, as already pointed out, locating and estimating such a parameter is likely to introduce strong elements of uncertainty in the selection procedure.

The estimation of the parameters in (1), following [45] and [46], requires the definition of l upper bounds for each regime, here denoted by P_j $(j = 1, 2, ..., l)$ as a maximum order a process can reach in a given regime plus one, here denoted by D, for the maximum order allowed for the threshold time delay. The choice of such constants is *a priori* and arbitrary.

Estimation of the threshold values and the threshold delay is usually performed through a *grid-based* search strategy in the (γ x d) space, ($\gamma \in \mathbb{R}$) and $d \in 1, 2, ..., D$, over a set of reasonable threshold values $\gamma \in (\overline{\gamma}, \underline{\gamma})$ for some upper band D.

In this study, the set of allowable threshold values for γ is made up of grid points, $\gamma_i \in \Gamma$, $(i = 1, 2, ...10)$, generated by using equally spaced (in the range 10 - 90 percent) quantiles of the data set under investigation.

For each candidate pair (γ, d), a regression of the type:

$$\boldsymbol{y_j} = \boldsymbol{X}\boldsymbol{\phi_j} + \boldsymbol{\varepsilon_j} \tag{4}$$

is run, where $\boldsymbol{y_j}$ and $\boldsymbol{X}$ are, respectively, the vector and the matrix containing the data for the regime j whereas $\boldsymbol{\phi_j}$ and $\boldsymbol{\varepsilon}$ are the coefficients and the error term for the regime j.

We set the best choice for the pair (γ, d), say (γ°, d°), as the minimizer over the grid points of the residual variance, that is:

$$(\gamma^\circ, d^\circ) = \min_{\gamma, d} \frac{1}{T - H} \boldsymbol{\varepsilon}_j^2$$

where: $\boldsymbol{\varepsilon}_j = (\boldsymbol{\varepsilon}_{j_1}, \boldsymbol{\varepsilon}_{j_2}, ..., \boldsymbol{\varepsilon}_{j_T})'$ and $\{j_1, j_2, ..., j_{T_j}\}$ are the time indices of the subset of the data belonging to the $j - th$ regime and H is as defined in (2).

For a stationary ergodic self-exciting threshold autoregressive model with single threshold parameter, Chan [6] obtained the consistency and limiting distribution of the least-squares estimator for the underlying true parameters. Similar results for the maximum likelihood estimators have been derived by Qian [41].

Given the parameters γ° and d°, the problem for selecting the optimal dimension of a $SETAR$ model is to include the number of lags in each regimes which are relevant in the reconstruction of the underlying DGP. A trade off between model complexity and extraction capabilities of the significant features of the observed DGP is the goal of model selection procedure that works as follows.

Let:

- $\{Y_t : t = -H+1, ..., T\}$ denote the observations generated according to (1),

- $C = C_v : v = 1, 2, ...V$, be the set of competing $SETAR$ models, characterized by the parameters $\{\boldsymbol{\theta}_v \in \boldsymbol{\Theta}_v \subseteq \mathbb{R}^v : v = 1, 2, ..., V\}$, Akaike states that the best model is the one possessing the minimum AIC estimate obtained from comparing the S competing models. This procedure is referred to by Akaike (1974) as MAICE (minimum AIC estimate): it is aimed at extracting, among the candidate models, the order satisfying the following:

$$(\hat{p_1}, \hat{p_2}, ..., \hat{p_l}) = \arg \min_{(p_1, p_2, ..., p_l) \leq P} AIC(p_1, p_2, ..., p_l)$$

Assuming that the bounds for the maximum order of the l regimes are equal, that is $P = P_1 = P_2 = , ..., P_l$, as in the case of our empirical setup presented in the sequel, the exhaustive set of $SETAR$ models is of size $(P_l)^l$.

4 The employed bootstrap scheme

The most popular bootstrap methods for dependent data are block ([13], [21] and [33]), local [28], wild ([23], [22], [25], [26], [3]) and Markov ([29], [32]) bootstrap and sieve (see [12] and references cited therein). They all are nonparametric procedures, except for the last which is semiparametric[1].

Block Bootstrap is the framework where the bootstrap method employed in this paper stems from. By design, it is very close to the bootstrap for *iid* observations; in fact, in both the approaches, observations are drawn with replacement. In the block bootstrap however, instead of resampling from a collection of n individual and independent observations, contiguous sequences of data points of length m are resampled. The bootstrap resample is obtained by sampling T/m, (T denoting the sample size) blocks randomly with replacement and with equal probability for each block to be drawn. These blocks are then put together to a time series of length T.

Initially, this method has been studied for the case of a set of non overlapping blocks of fixed length m: $\{X_j : j = 1, \ldots, m\}, \{X_{m+j} : j = 1, \ldots, m\}, \ldots,$. An important extension of the method,

[1]For an overview on the bootstrap for time dependent data, see [4], [16], [35]. In particular, [16] deals with the higher order performances of the different schemes.

the Moving Block Bootstrap, was independently proposed by Künsch [18] and Liu and Singh [21]. This procedure considers overlapping blocks, that is when the observations $\{X_{k-1+j} : j = 1, \ldots, m\}$ belong to the j-th block (the performances of these two approaches have investigated in [14] and [19] in terms of the estimation error). For both the procedures, the crucial point is to choose the blocklength so that the dynamic structure of neighbored observations *within* the blocks is preserved and the data belonging to different blocks are separated far enough in time to show a vanishing autocorrelation structure ([47]). In more details, in order to achieve consistency ([14]) for the generic blocklength $m = m_t$ the following must be true: $m \to \infty$ and $T^{-1}m \to 0$ as $T \to \infty$; that is the blocksize m tends to infinity but at a rate slower than the sample size T.

By design, the bootstrapped time series has a nonstationary (conditional) distribution. To achieve stationarity, Politis and Romano [33] proposed a modified version of this scheme, the Stationary Bootstrap. Here the index t of the first observation in the block is drawn from the discrete uniform distribution on $1, .., t$, whereas m is sampled from the geometric distribution such that: $\boldsymbol{P}(m = g) = (1 - p)^{g-1}p; \quad g = 1, 2, ...$ and $p \in (0, 1)$. Finally, in order to deal with the boundaries effects between neighbored blocks, *ad hoc* procedures (tapering) have been studied ([30], [31]). The scheme adopted in this paper is the Stationary Bootstrap (henceforth SB). We undertake this procedure by selecting the optimal block length according to the methodology proposed in [36] and [37].

5 Definition of βAIC and the proposed βMAICE procedure

Let $P(\theta, G)$ be the $SETAR$(l, p_1, ..., p_l) probability process generating the observations $\{y_t\}_{t\in T}$. Via SB we generate pseudo data y^* by replacing G with the empirical distribution function G^* in the probability mechanism. B pseudo series are then generated from $P(\theta, G^*)$ and, by recomputing AIC for each of these replications, a bootstrap version of the criterion, called βAIC, is obtained, that is:

$$\beta AIC(p_1^*, ..., p_l^*) = \sum_j \left\{ T_j \ln {\sigma^*}^2_{\varepsilon j} + 2p_j^* \right\} \quad j = 1, .., l \tag{5}$$

Our selection procedure is iterative and consists in fitting $(\gamma \times d)$ regressions and increasing order $SETAR$ models (up to a predefined order P_j, $j = 1, 2, ..., l$) to the original time series and its bootstrap replications.

The method is computer intensive especially because of the fact that it explores all the $P = \prod_{j=1}^{l}$ $SETAR$ possibilities for each bootstrap replication. Now our procedure is illustrated step-by-step.

1. Let $\{y_t\}_{t\in T}$ be a real-valued stationary process of the form (1);
2. $\{y_t\}_{t\in T}$ is divided into overlapping blocks of m consecutive observations: $B_i = (y_i, y_{i+1}, ..., y_{(i+m-1)})$ where $i = 1, ...T - m + 1$ are the starting points of the new blocks, drawn iid with uniform distribution on the set $1, 2, ..., N$, and m's are drawn iid from a geometric distribution. The

different $m's$ are determined through the Politis and Romano's procedure, based to the definition of spectral estimation via the flat-top lag-window[2] [37];

3. the generic b, $b = 1, 2, ..., B$, bootstrap pseudo-series $y_1^*, y_2^*, ...,$ is obtained by joining together $k \approx \frac{T}{m} \in N$ blocks;

4. define a maximum order for both P_j, $j = 1, 2, ..., l$, the maximum lag lengths allowed for the l regimes j and D, the maximum time delay, so that they are likely to encompass the true model;

5. for each bootstrap replication b, the selected threshold value and threshold delay parameters, that is the pair (γ^*, d^*), is chosen as follows:

   (a) grid points $\gamma_i \in \Gamma$ $(i = 1, 2, ...10)$, are generated by using equally spaced (in the range 10 - 90 percent) quantiles of y^*. The set of possible choice for $d \in D$ is also selected;

   (b) for each candidate pair (γ, d) the following regression is run:

$$\boldsymbol{y_j^*} = \boldsymbol{X^*}\boldsymbol{\phi_j^*} + \boldsymbol{\varepsilon_j^*}, j = 1, 2, ..., l$$

   where $\boldsymbol{y_j^*}$ and $\boldsymbol{X^*}$ are, respectively, the vector and the matrix containing the bootstrapped data for the regime j whereas $\boldsymbol{\phi_j^*}$ and $\boldsymbol{\varepsilon^*}$ are the bootstrap coefficients and the bootstrap error term for the regime j.

   (c) the selected pair (γ^*, d^*) is the minimizer of the residual variance, that is:

$$(\gamma^*, d^*) = \min_{\gamma, d} \frac{1}{T - H} \boldsymbol{\varepsilon^*}_j^2$$

   where T is the sample size, $\boldsymbol{\varepsilon^*}_j = (\boldsymbol{\varepsilon^*}_{j_1}, \boldsymbol{\varepsilon^*}_{j_2}, ..., \boldsymbol{\varepsilon^*}_{j_T})'$, $\{j_1, j_2, ..., j_{T_j}\}$ are the time indices of the subset of the data belonging to the $j - th$ regime and H has been defined in (2);

   (d) given (γ^*, d^*), an exhaustive set C of size $P = \prod_{j=1}^{l}$ of tentative $SETAR(l, p_1, ..., p_l)$ models is fitted recursively up to the maximum order (step 4) chosen, so that βAIC is computed for each candidate model. That is $\beta AIC(c)$, $c = 1, 2, ...C$;

   (e) The optimal model order, say $(p_1^*, ..., p_l^*)$, is selected by minimizing βAIC, over the models in the set C, that is:

$$(p_1^*, ..., p_l^*) = \arg \min_{(p_1, ..., p_l) \leq P} \beta AIC(p_1, ..., p_l)$$

[2]The expected block size is a function of the following three parameters: $\sum_{\xi=-\infty}^{\infty} |\xi| R(\xi)$, $\sum_{\xi=-\infty}^{\infty} R(\xi) = g(0)$, $\int_{-\pi}^{\pi} (1 + cos(w)) g^2(w)\, dw$, where with $R(\xi)$ the autocovariance function, estimated at the lag $\xi = 1, 2, ...\Xi$, is denoted. The estimation of these parameters involves the choice of the spectral bandwidth, say, Ψ. Politis and Romano [34] suggest considering the smallest integer, say $\hat{\psi}$, after which the correlogram shows correlations not significantly different from 0, i.e. $R(\xi) \simeq 0$ for $\xi > 0$. After determining $\hat{\psi}$ the recommendation is to take $\Psi = 2\hat{\psi}$.

6. out of the group of the B winning models (one for each bootstrap replication) the winner one, say c_0, is identified on the basis of its relative frequency. That is:
$\#[\beta AIC(c_{0,b}) < \beta AIC(c_{j,b})] \qquad j = 1...C-1 \quad , \quad b = 1, 2, ...B.$

In practice, the procedure picks the model winning the greatest number of time in the set of the bootstrap replications.

In steps 2 and 3 our procedure coincides with SB which is in turn theoretically justified by the non linear dynamics generated by $SETAR$ models. Step 4 shows that, unfortunately, our procedure does not solve the problem related to the arbitrary choice of the maximum $SETAR$ lag order in each regime nor the maximum threshold time delay. The remaining steps summarize our procedure.

6 Simulation study

Even though several tests for $SETAR$-Type nonlinearity have been developed ([24], [5]), in our simulation study we always assume the data to be a realization of a $SETAR$ process.

In order to evaluate the small-sample performance of the method presented in Section 5, two different Monte Carlo studies[3] have been carried out: the first set of simulations (Section 6.1) are aimed at comparing the performances of AIC and βAIC on the basis of a set of competitive models belonging to the class $SETAR$. In the second study (Section 7), the set of candidate models has been broadened in order to include a portfolio of artificial neural networks (henceforth $ANNs$).

After the illustration of each experimental design and simulation, a discussion of the results will follow.

6.1 Experimental design and simulations

For both the studies, 1000 time series have been generated from six different DGPs (detailed in Table 1) from a $SETAR$ model as formalized in (1), with $\varepsilon_{jt} \sim N(0, \sigma_j^2)$ and $j = 1, 2$. For all the DGPs, the true value of the delay parameter d, always supposed unknown, has been set to 1, whereas the maximum order investigated is 2. Parameters d and γ are jointly estimated according to a grid search strategy as explained in Section 5. We also assume the number of regimes l known and set to 2 and, without loss of generality, the bounds for the maximum order of both the regimes to be equal, that is $P = P_1 = P_2$. The sample size T takes the value 100, 200, 500.

Throughout the whole experiment, for each of the simulated time series, $B = 125$ bootstrap replications have been generated according to the method presented in Section 4 and, for each of them, a sample of size $T + 1000$ observations is also generated. The first 1000 observations has then been discharged to take into account the effect of the initial conditions (the starting value has been always set to 0).

[3]All the simulations have been implemented using the software R (8.1 version) and performed by employing hardware resources of the University of California, San Diego. In particular, we made use of the computer EULER (maintained by the Mathematical Department) and the supercomputer IBM–TERAGRID.

Table 1: The employed $SETAR$ DGPs

Parameters	DGP 1	DGP 2	DGP 3	DGP 4	DGP 5	DGP 6
d	1	1	1	1	1	1
γ	.1	.1	.1	.1	.4	.1
ϕ_{10}	.2	-.1	-.1	-.1	.7	-.2
ϕ_{11}	.2	-.2	-.4	-.4	-.7	.8
ϕ_{12}	.1	.2	-.3	-.4	-.7	-.7
ϕ_{13}			-.2	-.2	-.5	.5
ϕ_{20}	.4	.3	.4	.4	.3	.2
ϕ_{21}	.2	-.3	-.3	-.5	-.8	.8
ϕ_{22}	-.1	.2	-.4	-.4	-.6	-.8
ϕ_{23}			.3	-.3	-.5	.6
σ_1^2	.9	.9	.9	.8	.8	.9
σ_2^2	.6	.6	.6	.6	.6	.8

Finally, the block selection algorithm via flat–top window employed in this study is fully automatic and has been developed by Dr. Andrew Patton (London School of Economics) [4].

As it can be seen by inspecting Table 1, the absolute values of the coefficients in DGPs 1 and 2 are decreasing regime-wise and small. Such a choice has been made in order to assess the behavior of our selection procedure in the case of weak threshold autoregressive processes.

The proposed procedure will be evaluated by employing the following measures of performances:

a) frequency $f(p_1, p_2)$ of selection of the true order (p_1, p_2);
b) frequency $f(d)$ of selection of the true delay parameter d;
c) estimation of the bias of the threshold parameter $E(\hat{\gamma} - \gamma)$;
d) estimation of the standard error of $\hat{\gamma}$, $\sigma(\hat{\gamma})$.

These results are presented in Table 2. By examining this table, it is possible to assess the effectiveness of our method, especially for small and moderate sample sizes.

In fact, the frequency of selection of the true model shows that $\beta MAICE$ procedure performs consistently better than the traditional $MAICE$ procedure. Averaging over all the DGPs and sample sizes considered, the percentage gain[5] upon the standard procedure is around 22%, however, if we restrict our attention to $T = 100$, this percentage rises up to 34.4%. The reason of that, lies in the proportionally decreasing gains showed by βAIC upon the standard procedure for increasing sample sizes. In fact, for T=500, smaller discrepancies (in average around 12.4%) has been recorded in the frequency of selection of the correct model. Such a behavior is consistent with our findings on the asymptotic equivalence of AIC and its bootstrap counterpart for $ARMA$ models [10] and,

[4]The Matlab code is available from his website http://fmg.lse.ac.uk/patton/ code.html

[5]The percentage gain is computed as following: $100 * [f^*(p_1, p_2) - f(p_1, p_2)]/f(p_1, p_2)]$, where, with $f^*(p_1, p_2)$, the frequency of selection of the correct model achieved in the bootstrap world is denoted.

Table 2: Performances of AIC and βAIC (absolute values

DGP	Selection Criterion	T	$f(p_1, p_2)$	$f(d)$	$bias(\hat{\gamma})$	$\sigma(\hat{\gamma})$
1	AIC	100	151	270	.71	.75
		200	228	337	.66	.77
		500	272	367	.62	.70
	βAIC	100	215	286	.69	.71
		200	259	341	.62	.64
		500	286	371	.61	.62
2	AIC	100	201	407	.65	.70
		200	289	530	.63	.63
		500	364	597	.60	.62
	β AIC	100	304	500	.62	.69
		200	397	623	.58	.55
		500	433	689	.56	.41
3	AIC	100	426	537	.50	.62
		200	527	651	.47	.58
		500	610	746	.43	.47
	βAIC	100	569	769	.40	.53
		200	634	846	.37	.50
		500	662	911	.25	.34
4	AIC	100	477	702	.51	.59
		200	565	793	.49	.58
		500	621	891	.46	.33
	βAIC	100	629	885	.35	.48
		200	683	997	.31	.43
		500	716	1000	.21	.24
5	AIC	100	530	712	.39	.26
		200	624	895	.34	.22
		500	687	951	.29	.19
	βAIC	100	694	901	.30	.23
		200	755	1000	.29	.21
		500	789	1000	.12	.10
6	AIC	100	553	783	.26	.27
		200	651	902	.21	.24
		500	764	948	.18	.11
	βAIC	100	732	972	.11	.19
		200	816	1000	.10	.15
		500	845	1000	.07	.05

on the other hand, would suggest a particularly fruitful employment of our method when small or moderate sample sizes are available, that is the case of many practical applications in the econometric field.

Our method seems to increase AIC's selection performances, especially when processes show a strong structure, as in the case of DGPs 5 and 6. Averaging over these processes, our procedure picks the correct model 71.3% of the time versus 54.2% of the standard procedure for T=100 whereas for T=500 these percentages rise up to 81.7% and 72.6% respectively. The percentage gain of our procedure is around 31.7% and 12.6%, for $T = 100$ and $T = 500$ respectively. A reason for such results is likely to be correlated to the quantity of information conveyed by the original process captured by each of the block extracted by the bootstrap procedure. When blocks "dense of information" are joined together, they are likely to produce a good representation of the original process so that the selection procedure is more likely to determine the correct model.

For "milder" structures the results are still good: considering all the sample sizes of the DGPs 3 and 4, our procedure chooses the correct model in average 59.9% of the time versus the 45.2% of its non bootstrap counterpart. Conversely, for weakly structured processes (DGPs 1 and 2), both the methods offer poor performances; however, regarding DGP 2, βMAICE procedure is still doing better, even thought the overall performances seem to be not satisfactory: for example, with $T = 100$, our method picks the right model 30.4% of the time whereas the standard procedure find the correct model only 20.1% of the time.

Our method breaks down (as well as its non bootstrap counterpart) in the case of DGPs of the type 1. Even though improvements can be noticed upon the MAICE procedure, the overall performances turn out to be unsatisfactory for both the approaches. Even with the largest sample size considered, the frequency of selection is about 28.6% (27.2% for the MAICE procedure).

Always in Table 2, the frequency of selection of the correct delay d detected by both the procedures is presented. As far as the considered DGPs are concerned, it seems that the βAIC outperforms the standard AIC, except for DGP 1, where no substantial improvements can be noticed in favor of our procedure. The performances of both the methods now seem to be less affected by the chosen parametrization. They may be considered somehow satisfactory also in the case of DGP 2. However, for this process our method produces better results: the frequency of selection of the right delay is 50.0%, for T=100 and 68.9% for T=500 whereas the standard procedure makes the correct choice 40.7% and 59.7% of the time respectively. Such a result seems to depend heavily on the sample size. In fact, if we exclude DGP 1, averaging over the remaining ones, our selection criterion finds the right delay 92% and 80.5% of the time versus 82.7% and 62.8% of the MAICE procedure, for T=500 and T=100, respectively.

We now assess the quality of our method for the delimitation of the two regimes, by focusing on the estimation of the threshold parameter γ. Heavy distortions in the estimation of this parameter, associated with high standard errors, are often introduced by both the procedures. Averaging over all the sample sizes, weak processes of the type of $DGP1$ and $DGP2$ show an average bias of 0.61 for the βMAICE procedure and 0.65 for the MAICE procedure whereas the average standard error is 0.60 and 0.70 respectively. For the remaining DGPs the average reduction in terms of bias and standard error achieved by our method is more substantial: considering all the sample size, the

average bias now is 0.24 (βMAICE) versus 0.38 when the MAICE procedure is applied, whereas the standard error is 0.29 and 0.37 respectively. Even though a systematic error in the estimation of the parameter γ is noticeable over all the DGPs, it seems to decrease with increasing sample size for both the methods, with a proportion in favor of our procedure. In fact, averaging over all the DGPs, the bias ranges from 0.41, for T=100, through 0.30 for T=500 when our procedure is applied, versus 0.50 and 0.43 respectively in the case of MAICE procedure. Also, the standard error shows a similar behavior: in fact, considering all the DGPs, for T=100 its average value is 0.47 for βMAICE procedure and 0.53 in the case of the standard procedure. For T=500 these values reduce to 0.29 and 0.40 respectively.

Finally, we briefly discuss the distribution of the orders chosen by the two procedures. In particular, we present in Figure **??** [FIGURE 1] this distribution in the case of DGP 3, for two different sample sizes: T=100 and T=500. Due to the non nested structure of the $SETAR$ model, the graphs show on the X-axis the tentative models according to the codification shown in Table 4 (under this convention the true model is the number 11). By comparing the two graphs, noticeable discrepancies can be observed between the two distributions along with a tendency to overestimation in the case of the small sample size. Such discrepancies appear to be less pronounced for T=500 and may be an indication of an asymptotic equivalence between the AIC for $SETAR$ models and its bootstrapped counterpart, already verified in the ARMA case [10].

FIGURE 1
Distribution of the orders chosen by βMAICE and $MAICE$ procedures. DGP 3, T=100 (left graph) and T=500 (right graph).

7 Artificial neural network approach for the assessment of the robustness of the proposed method

We now propose a novel approach aimed at comparing the performances of both the presented procedure and the standard one. As we have seen in Section 5, such a comparison may be performed by fitting a set, say A, of non nested competition models on the artificial process and by evaluating them on the basis of a predetermined decision rule.

In this approach, a subset $a \subset C$ of models *not* belonging to the class of the true, artificial one, is created. By including the subset a in C, we define a new set of candidate models, say $C' = a + C$, where C is the subset of models belonging to the class of the original process ($SETAR$). The subset a is aimed at perturbing the results obtained when the Monte Carlo experiment is carried out employing only models belonging to the set C, that is when $a \equiv \varnothing$. Basically, what we want is to provoke a deterioration of the selection performances, so that the resulting loss can be taken as an indicator of the reliability of the methods under test. In practice, we assess their robustness versus an *"ad hoc"* additional set of candidate models.

The a-induced loss is determined not only by the new winners belonging to the subset a, but it also depends on the increased level of uncertainty as a result of the bigger number of competitive models involved in the selection process.

7.1 Choice of the models belonging to the subset a

The choice of the type of models belonging to the subset a, depends greatly on the specificity of the problem at hand. In general, however, such models should embody two key features: good explicative capabilities and flexibility. With regard to the first characteristic, in order to produce significant selection results, these models by construction should be flexible enough to capture efficiently the dynamics of the original process. The second one, refers to their ability of being easy to adjust, in order to yield a reasonable "contrast" with the performances given by the set C of candidate models. In other words, by allowing an easy and effective way to tune them up, such models can be designed and refined to capture a "balanced" proportion of winning models along their different specifications. In fact, we are not here interested in finding the best model, conversely we want to build up a more informative experiment in which the set a plays the ancillary role of a mere tool.

For the sake of our empirical study, the employment of non-parametric models seems to be an appropriate choice. In particular, we use artificial neural networks of the type "feed forward". Given their great flexibility, ANNs are considered a powerful computational tool for nonlinear problems in many fields, therefore they can be successfully employed to handle model selection problems with a variety of DGPs, including $SETAR$ processes. In addition, ANNs can be easily controlled in order not to overshadow set C performances neither to give not significant results (i.e. a frequency of selection close to 0).

As it is well known, the building process of an ANN is not a standard problem, so that it is usually difficult to assess the optimality of a network architecture for a given system. However, for the instrumental role played by ANNs in our framework, the problem of choosing the optimal number of neurons and hidden layers is not critical, in the sense that we focused our attention mainly on making the outcomes of the set a in line with the purposes of our study and on keeping the computational time reasonably low.

As we have already mentioned, feed forward is the architecture used in this second experiment. It is one of the most common neural networks, in which neurons in one layer are connected to neurons in the next layer successively. Each neuron expresses a value in the hidden and output layer which is in turn determined by a activation function that transfers a weighted summation of the value of each neuron in the previous layer. This network has been proved [7] to be able of fitting any nonlinear function $v = f(u)$ to an arbitrary degree of precision.

The general formulation of this architecture is the following:

$$v_k = \sum_{j=1}^{N_2} w_{jk2} \xi \left(\sum_{i=1}^{N_1} w_{ij1} u_i + \lambda_j \right).$$

Here $\xi(x)$ is the activation function (we used a sigmoid function of the type $\xi(x) = \frac{1}{1+e^{-x}}$), N_1 and N_2 are the number of inputs $(u_i, i = 1, 2, ..., N_1)$, and hidden layer neurons respectively and w_{ijk} are the connection weights from the neuron i in layer k to node j in layer $k+1$. The parameter λ is the neural bias, implemented here using an additional weight associated with a constant unitary signal.

7.2 The employed selection criterion for artificial neural network

As a selection criterion for the best architecture network, we employed the following version of AIC:

$$AIC^{ANN} = TLog(\hat{\sigma}^2_{\varepsilon}) + m \tag{6}$$

where T is the data length, $\hat{\sigma}^2_{\varepsilon}$ is the residual variance and m is the number of the network parameters. As usual, the AIC is applied on a predetermined candidate network space and the architecture achieving the least AIC value is the winning one. The bootstrap version of (6) that is:

$$AIC^{*^{ANN}} = TLog(\hat{\sigma}^2)^* + m^* \tag{7}$$

is straightforward as well as the extension of βMAICE procedure to ANNs. In practice when performed on the set a, βMAICE procedure is carried out replacing (5) with (7), whereas the standard selection strategy (Section 3) is performed using (3) in place of (6).

The list of the candidate networks is reported in Table 5. Here, in order to create a *continuum* with Table 2, the first network is labeled by the number 33 the second by 34 and so on.

We focus our attention on a fully connected one–hidden layer network, with 2 and 4 neurons, and 1 output neuron. In input, we employed configuration with 1 up to 4 neurons (one for each time delay). As it is possible to see, the number of ANNs has been kept relatively low (in total there are 8 ANNs), in order not to introduce too much uncertainty in the selection process and for computational time reasons.

7.3 Empirical results

The same Monte Carlo experiment presented in Section 6.1 is this time carried out on the redefined set of candidate models C'. To diminish uncertainty, the same set of time series as the previous experiment has been employed. For all the ANNs the number of iterations has been fixed to 500.

In the sequel, we compare the two procedures in terms of their *percentage loss* of correctly identified time series when the new set C' is evaluated. In practice, we consider the following two ratios:

$-100 * \frac{f^{C'}(p_1,p_2)-f(p_1,p_2)}{f(p_1,p_2)}$ and $-100 * \frac{(f^{C'}(p_1,p_2))^*-f^*(p_1,p_2)}{f^*(p_1,p_2)}$, for $MAICE$ and β-$MAICE$ procedure respectively, where $f^{C'}(p_1,p_2)$ and $(f^{C'}(p_1,p_2))^*$ are the frequency of selection of the correct model achieved under C' in the non bootstrap and bootstrap world respectively, and $f(p_1,p_2)$ and $f^*(p_1,p_2)$ are as already defined (Section 6.1).

In Table 3, the percentage loss for each DGP and sample size, is reported. By inspecting this table, it is possible to see the strong impact of the ANNs on the $SETAR$ performances: averaging over all the DGPs and sample sizes the standard MAICE procedure achieves a correct selection of the process 30.5% of the time whereas our procedure does 40.6 %. In the previous experiment it was 47.4% and 57.9% respectively.

FIGURE 2

Distribution of the orders chosen by βMAICE and $MAICE$ procedures in the case of C' competition set. DGP 3, T=100 (left graph) and T=500 (right graph).

The general tendency is a more remarkable loss with increasing sample size: in fact, for T=500, all the DGPs show an average percentage loss equal to 44.3% and 37.0% for MAICE and βMAICE procedure respectively whereas for T=100 those percentages drop to 32.4 and 28.7. This could be an evidence of ANNs capabilities in exploiting an increasing information set more efficiently than the $SETAR$ models.

In the case of weak structures, ANNs seem to have the strongest impact: in fact, the percentage of selection of the right model, averaging over all the sample size, for DGP 1 is only 11.6% when the MAICE procedure is applied and 13.9% for the βMAICE procedure. The percentage loss in average is 45.3% (MAICE) and 44.6 (βMAICE). Therefore, only slight improvements are produced by our procedure. In case of the DGP 2 an average reduction of 38.3% is recorded, over all the sample size, for our procedure which therefore seems to do better than the standard one (loss equals to 44.8%).

The effectiveness of our method in counteracting the impact of the ANNs, is more substantial considering the other DGPs. By averaging the frequency of selection of the true model over all the sample size of both DGPs 3 and 4, we can see that our procedure show less considerable reductions: 32.3% versus 38.7% of the standard procedure. Considering DGPs 5 and 6 such figures become, respectively, 22.8 % and 28.0 %, that confirms the milder impact of ANNs in the case of more structured processes.

In the light of this second experiment, our approach proves to be more effective in counteracting the impact of the ANNs than the standard one, especially with small sample sizes. Both the approaches seem to be more robust in case of more structured process: this fact is likely to be

Table 3: Performances of AIC and βAIC (absolute values) in the case of the competition set C'.

DGP	Selection Criterion	T	$f^{C'}(p_1, p_2)$	% loss
1	AIC	100	91	39.7
		200	127	44.3
		500	131	51.8
	βAIC	100	132	38.6
		200	145	44.0
		500	140	51.0
2	AIC	100	117	41.8
		200	176	39.1
		500	169	53.6
	βAIC	100	211	30.6
		200	250	37.0
		500	228	47.3
3	AIC	100	252	40.8
		200	337	36.1
		500	319	47.7
	βAIC	100	361	36.6
		200	419	33.9
		500	390	41.1
4	AIC	100	335	29.8
		200	364	35.6
		500	359	42.2
	βAIC	100	464	26.2
		200	527	22.8
		500	478	33.2
5	AIC	100	421	20.6
		200	447	28.4
		500	436	36.5
	βAIC	100	558	19.6
		200	576	23.7
		500	601	23.5
6	AIC	100	433	21.7
		200	477	26.7
		500	504	34.0
	βAIC	100	580	20.8
		200	623	23.7
		500	630	25.4

Table 4: Codification of the candidate models

Delay	orders	$p_2 = 1$	$p_2 = 2$	$p_2 = 3$	$p_2 = 4$
d=1	$p_1 = 1$	*1* (1,1)	*2* (1,2)	*3* (1,3)	*4* (1,4)
	$p_1 = 2$	*5* (2,1)	*6* (2,2)	*7* (2,3)	*8* (2,4)
	$p_1 = 3$	*9* (3,1)	*10* (3,2)	*11* (3,3)	*12* (3,4)
	$p_1 = 4$	*13* (4,1)	*14* (4,2)	*15* (4,3)	*16* (4,4)
d=2	$p_1 = 1$	*17* (1,1)	*18* (1,2)	*19* (1,3)	*20* (1,4)
	$p_1 = 2$	*21* (2,1)	*22* (2,2)	*23* (2,3)	*24* (2,4)
	$p_1 = 3$	*25* (3,1)	*26* (3,2)	*27* (3,3)	*28* (3,4)
	$p_1 = 4$	*29* (4,1)	*30* (4,2)	*31* (4,3)	*32* (4,4)

Table 5: Codification of the employed ANNs

ANN Models				
Neurons	*33* (1,2,1)	*34* (2,2,1)	*35* (3,2,1)	*36* (4,2,1)
	37 (1,4,1)	*38* (2,4,1)	*39* (3,4,1)	*40* (4,4,1)

correlated with the smaller amount of uncertainty associated in the selection process of such models. However, our method exploits more efficiently the information conveyed in a well defined process and, as a result of that, the impact of the bootstrapped ANNs is less effective.

Finally, in Figure **??** [FIGURE 2] the distribution of the percentage of winning models chosen by both the procedures is presented for the DGP 3. The considered sample sizes are T=100, 500. As in the previous case, a codification is needed and it is as reported in Table 4 and 5. By inspecting both the graphs, it is clear that strong competitors are the ANNs which show, in the input layer, 3 and 4 neurons - that is the networks labeled by number *35*, *39*, and *40*. This fact is consistent with the structure of the DGP at hand, which shows two autoregressions of order 3. In particular, the network with the label *39* performs consistently better than $SETAR$ models in both the procedures and for both the sample sizes.

References

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control*, AC–19, 716–723.

[2] Akaike, H. (1973). Information theory and an extension of maximum likelihood principle. *Second International Symposium on Information Theory*, Akademia Kiado, 267–281

[3] Akaike, H. (1993). Bootstrap and wild bootstrap for high–dimensional linear models. *Ann. Statist.*, 21, 255–285.

[4] Bhlmann, P. (2002). Bootstraps for time series. *Statist. Science*, 17, 52–72.

[5] Chan, KS and Tong, H. (1990). On likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society*, B, 52, 496–476.

[6] Chan, K. S. (1993). Consistency aad limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics*, 21, 520–533.

[7] Cybenko, G. (1989). Approximation by Superimposition of a Sigmoidal Function. *Math. Control Signal & Sys.*, 2, 303.

[8] Dueker, M. J., Sola, M. and Spagnolo, F. (2007). Contemporaneous threshold autoregressive models: estimation, testing and forecasting. *Journal of Econometrics*, 141, 517–547.

[9] Efron, B. (1979). Bootstrap methods: Another look at jackknife . *Ann. Statist.*, 7, 1–26.

[10] Fenga, L. and Politis, D. Bootstrap ARMA order selection. *Journal of Statistical Simulation and Computation*, Volume 81, Issue 7, July 2011, pages 799-814.

[11] Ferrara, L. and Guegan, D. (2003). Detection of business cycles using SETAR models. *Preprint MORA–IDHE*, 12, Ecole Normale Superieure, Cachan, France.

[12] Franke, J. and Kreiss, J.P. (1992). Bootstrapping Stationary Autoregressive Moving - Average Models. *J. Time Series Analysis*, 13, 297–317.

[13] Hall, P. (1985). Resampling a coverage process. *Stoch. Proc. Appl.*, 19. 259–269.

[14] Hall, P., Horowitz, J. L. and Jing, B. Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, 3, 561– 574

[15] Hannan J. (1980). The estimation of the order of an arma process. *Ann. Statist.*, 8, 1071–1081

[16] Hrdle, W., Horowitz, J.L. and Kreiss, J.P. (2003). Bootstrap methods for time series. *International Statist. Review*, 71, 435–459.

[17] Konishi, S. and Kitagawa G. (2008). *Information Criteria and Statistical Modeling*. Springer–Verlag, New York.

[18] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17, 1217–1241.

[19] Lahiri, S.N.(1999). Theoretical comparisons of block bootstrap methods. *Annals of Statistics*, 27, 386–404.

[20] Lahiri, S.N. (2003). *Resampling Methods for Dependent data*. Springer, New York.

[21] Liu, R.Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Lepage, R. and Billard, L. (eds), Exploring the Limits of the Bootstrap*, 225–248, Wiley, New York.

[22] Liu, R.Y. and Singh, K. (1992). Efficiency and robustness in resampling. *Ann. Statist.*, 20, 370–384.

[23] Liu, R.Y. (1988). Bootstrap procedures under some non i.i.d. models. *Ann. Statist.*, 16, 1696–1708.

[24] Lukkonen, R., Saikkonen, P., and Terisvirta, T. (1988). Testing Linearity Against Smooth Transition Autoregressive Models. *Biometrika*, 75, 491–499

[25] Mammen, E. (1992). *Bootstrap, wild bootstrap, and asymptotic normality*. Springer Lecture Notes in Statistics 77, Springer, Heidelberg, Berlin.

[26] Mammen, E. (1992). *When does bootstrap work? Asymptotic results and simulations*. Springer Lecture Notes in Statistics 77, Springer, Heidelberg, Berlin.

[27] Miller, A. J. (1990). *Subset selection in regression*. London: Chapman and Hall.

[28] Paparodtis, E. and Politis, D.N. (2000). The local bootstrap for kernel estimators under general dependence conditions. *Ann. Inst. Statist. Math.*, 52. 139–159.

[29] Paparoditis, E. and Politis, D.N. (2001). A Markovian local resampling scheme for nonparametric estimators in time series analysis. *Econometric Theory*, 17, 540–566.

[30] Paparoditis, E. and Politis, D.N. (2001). Tapered block bootstrap. *Biometrika*, 88, 1105–1119

[31] Paparoditis, E. and Politis, D.N. (2002). The tapered block bootstrap for general statistics from stationary sequences. *The Econometrics Journal*, 5, 131–148.

[32] Paparoditis, E. and Politis, D.N. (2002). The local bootstrap for Markov processes. J. Statist. Planning Inference . *J. Statist. Planning Inference*, 108, 301–328.

[33] Politis, D.N. and Romano, J.P. (1994). The Stationary Bootstrap. *J. Amer. Statist. Assoc.*, 89, 1303–1313

[34] Politis, D.N. and Romano, J.P. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Series Anal.* 16:67103.

[35] Politis, D.N. (2003). The Impact of Bootstrap Methods on Time Series Analysis. *Statist. Science*, 18, 219–230.

[36] Politis, D. N. and White, H. (2004). Automatic block–length selection for the dependent bootstrap. *Econometric Reviews*, 23, 53–70.

[37] Patton,A., Politis, D. N. and White, H. (2009). CORRECTION TO "Automatic block–length selection for the dependent bootstrap". *Econometric Reviews*, 28, 372–375.

[38] Potter, S.M. (1995). A nonlinear approach to US GNP. *Journal of Applied Econometrics*, 10, 109–125.

[39] Proietti, T. (1998). Characterizing Asymmetries in Business Cycles Using Smooth–Transition Structural Time–Series Models. *Studies in Nonlinear Dynamics Econometrics*, 3, Article 2.

[40] Psaradakis Z., Sola M., Spagnolo F., Spagnolo N. (2009) Selecting Nonlinear Time Series Models Using Information Criteria. *Journal of Time Series Analysis*, 30(4): 369-394.

[41] Qian, L. (1998). On maximum likelihood estimators for a threshold autoregression. *J. Statist. Plann. Inference*, 75, 21–46.

[42] Schwarz, G. (1981). Estimating the dimension of a model. *Ann. Statist.*, 6, 461–464

[43] Teräsvirta, T. (1994). Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association*, 89, 208-218.

[44] Tiao, G.C. and Tsay,R.S. (1994). Some advances in non–linear and adaptive modelling in time series. *Journal of Forecasting*, 13, AC–19, 109–140.

[45] Tong, H. (1990). *Non–Linear Time Series: A Dynamical System Approach.* Oxford University Press.

[46] Tsay, R. S. (1989). Testing and Modeling Threshold Autoregressive Processes. *Journal of American Statistical Association*, 84, 231–240.

[47] Vogel, R.M., and Shallcross, A.L. (1996). The moving blocks bootstrap versus parametric time series models. *Water Resources Research*, 32, 1875–1882.

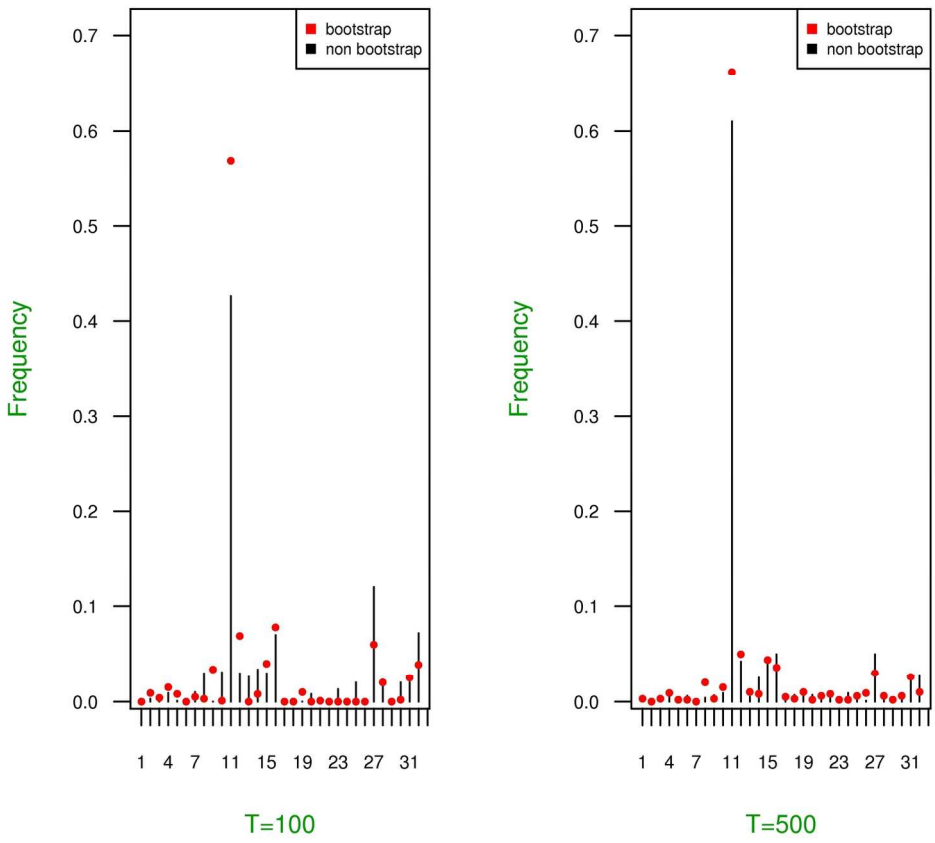

Figure 1
156x156mm (300 x 300 DPI)

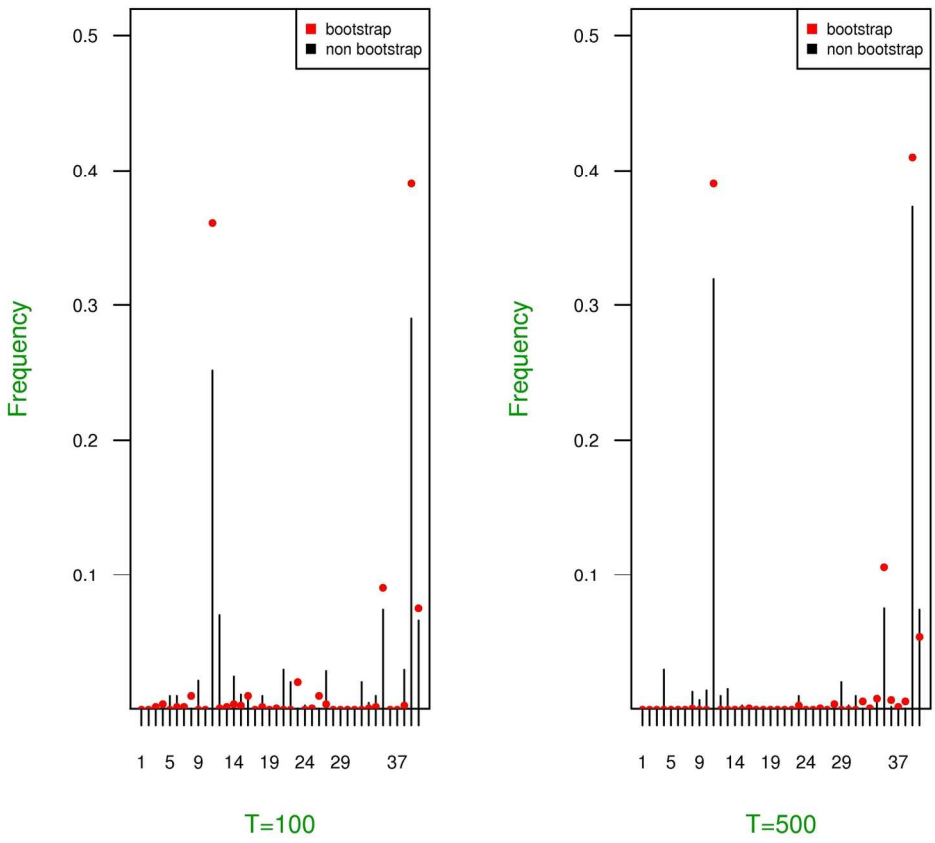

Figure 2
156x156mm (300 x 300 DPI)