UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 04-14
Statistics and Econometrics Series 05
Febrary 2004

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

MODEL SELECTION CRITERIA AND QUADRATIC DISCRIMINATION IN ARMA AND SETAR TIME SERIES MODELS

Pedro Galeano and Daniel Peña*

Abstract

We show that analyzing model selection in ARMA time series models as a quadratic discrimination problem provides a unifying approach for deriving model selection criteria. Also this approach suggest a different definition of expected likelihood that the one proposed by Akaike. This approach leads to including a correction term in the criteria which does not modify their large sample performance but can produce significant improvement in the performance of the criteria in small samples. Thus we propose a family of criteria which generalizes the commonly used model selection criteria. These ideas can be extended to self exciting autoregressive models (SETAR) and we generalize the proposed approach for these non linear time series models. A Monte-Carlo study shows that this family improves the finite sample performance of criteria such as AIC, corrected AIC and BIC, for ARMA models, and AIC, corrected AIC, BIC and some cross-validation criteria for SETAR models. In particular, for small and medium sample size the frequency of selecting the true model improves for the consistent criteria and the root mean square error of prediction improves for the efficient criteria. These results are obtained for both linear ARMA models and SETAR models in which we assume that the threshold and the parameters are unknown.

Keywords: Model Selection Criteria; Asymptotic Efficiency; Consistency; Quadratic Discrimination Rule.

* Galeano, Departamento de Estadística, Universidad Carlos III de Madrid, c/Madrid, 126, 28903 Getafe (Madrid), e-mail:pgaleano@est-econ.uc3m.es. Peña, Departamento de Estadística, Universidad Carlos III de Madrid, c/Madrid, 126, 28903 Getafe (Madrid), e-mail:dpena@est-econ.uc3m.es. We acknowledge financial support from BEC2000-0167, MCYT, Spain.

Model selection criteria and quadratic discrimination in ARMA and SETAR time series models

Pedro Galeano and Daniel Peña

Departamento de Estadística, Universidad Carlos III de Madrid, Spain.

Abstract

We show that analyzing model selection in ARMA time series models as a quadratic discrimination problem provides a unifying approach for deriving model selection criteria. Also this approach suggest a different definition of expected likelihood that the one proposed by Akaike. This approach leads to including a correction term in the criteria which does not modify their large sample performance but can produce significant improvement in the performance of the criteria in small samples. Thus we propose a family of criteria which generalizes the commonly used model selection criteria. These ideas can be extended to self exciting autoregressive models (SETAR) and we generalize the proposed approach for these non linear time series models. A Monte-Carlo study shows that this family improves the finite sample performance of criteria such as AIC, corrected AIC and BIC, for ARMA models, and AIC, corrected AIC, BIC and some cross-validation criteria for SETAR models. In particular, for small and medium sample size the frequency of selecting the true model improves for the consistent criteria and the root mean square error of prediction improves for the efficient criteria. These results are obtained for both linear ARMA models and SETAR models in which we assume that the threshold and the parameters are unknown.

KEY WORDS: Model Selection Criteria; Asymptotic Efficiency; Consistency; Quadratic Discrimination Rule.

1 INTRODUCTION

Most of model selection criteria for linear time series can be written as follows:

$$\min_k \left\{ \log \widehat{\sigma}_k^2 + k \times C(T,k) \right\}, \tag{1}$$

where $\widehat{\sigma}_k^2$ is the maximum likelihood estimate of the residual variance, k is the number of estimated parameters for the mean function of the process, T is the sample size and the function $C(T,k)$ converges to 0 when $T \to \infty$. These criteria can be classified into two groups. The first one includes the consistent criteria that, under the assumption that the data come from a finite order autoregressive moving average process, have a probability of obtaining the true order of the model that goes to one when the sample size increases. The Bayesian information criterion, BIC, by Schwarz (1978), where $C(T,k) = \log(T)/T$, and the Hannan and Quinn (1979) criterion, HQC, where $C(T,k) = 2m \log\log(T)/T$ with $m > 1$, are consistent criteria. The second group includes the efficient criteria that asymptotically select the order which produces the least mean square prediction error. The final prediction error criterion, FPE, by Akaike (1969), where $C(T,k) = k^{-1}\log(\frac{T+k}{T-k})$, the Akaike's information criterion, AIC, by Akaike (1973), where $C(T,k) = 2/T$ and the corrected Akaike's information criterion, AICc, by Hurvich and Tsai (1989), where $C(T,k) = \frac{1}{k}\frac{2(k+1)}{T-(k+2)}$, are efficient criteria.

These criteria have been derived from different points of view. The BIC approaches the posterior probabilities of the models. The HQC has been derived to be a consistent criterion such that $C(T,k)$ converges to 0 as fast as possible. The FPE selects the model that minimizes the one step ahead square prediction error. The AIC is an approximately unbiased estimator of the expected Kullback-Leibler information of a fitted model, which can be used as a discrepancy measure between the actual and the fitted model. The AICc is a bias correction form of the AIC that appears to work better in small samples. In this article we consider model selection as a discrimination problem and show that the BIC, AIC and AICc criteria can be derived as approximations to a quadratic discriminant rule. This approach also introduces a correction term that improves the finite sample performance of all these criteria but maintaining their asymptotic properties.

A useful non linear extension of ARMA models are the self-exciting threshold autoregressive (SETAR) models, see Tong (1990). These models can explain interesting features found in real data, such as asymmetric limit cycles, jump phenomena, chaos and so on. A series following a SETAR model is piecewise linear so that the correction considered for linear models can be easily extended to these non linear models. Model selection for SETAR models has been studied in Wong and Li (1998), Kapetanios (2001) and De Gooijer (2001). We analyze the correction term obtained from our approach for SETAR model selection criteria and show that this correction improves the finite sample performance of previous criteria.

The rest of this paper is organized as follows. Section 2 introduces the family of criteria for ARMA time series models. Section 3 discusses the computation of the correction term. Section 4 includes the correction term in SETAR time series models. Section 5 explores the performance of the proposed criteria in a Monte

Carlo experiment and shows that they perform better than the classical criteria for all the models considered.

2 A FAMILY OF MODEL SELECTION CRITERIA BASED ON THE QUADRATIC DISCRIMINANT RULE

The discrimination problem in time series appears as follows. Suppose it is known that a given time series, $x = (x_1, ..., x_T)'$, has been generated by one of the models M_j, $j = 1, ..., j_{\max}$. From the Bayesian point of view we also know the prior probabilities $p(M_j)$. The objective is to select the data generating model given the time series data. In the general case, we assume that the models M_j are Gaussian processes given by $x_t = \mu_{jt} + n_{jt}$, where μ_{jt} are deterministic mean functions and n_{jt} are zero mean ARMA models of the form $\phi_j(B)\, n_{jt} = \theta_j(B)\, a_{jt}$, where $\phi_j(B)$ and $\theta_j(B)$ are polynomials in the lag operator B such that $Bx_t = x_{t-1}$, with no common roots. The models are assumed to be casual and invertible. The series a_{jt} are white noise residuals with variance σ_{ja}^2. The simplest discriminant problem is to assume that the deterministic functions μ_{jt} are different, but the covariance matrices Σ_j are all equal to Σ. This case corresponds to the situation in which all the models have the same ARMA structure. Calling $\mu_j = (\mu_{j1}, ..., \mu_{jT})'$, this is equivalent to consider the hypothesis $M_j : x \in N_T(\mu_j, \Sigma)$, and we have that,

$$p(x \mid M_j) = (2\pi)^{-\frac{T}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu_j)' \Sigma^{-1}(x-\mu_j)\right), \qquad j = 1, \ldots, j_{\max}.$$

Maximizing the likelihood of the data implies minimizing the Mahalanobis distance between the data and the vector of marginal means. The same conclusion is obtained from the Bayesian point of view assuming equal prior probabilities $p(M_j) = 1/j_{\max}$ and maximizing the posterior probability of choosing the right model. A more interesting case appears when the ARMA models are different, that is, $M_j : x \in N_T(\mu_j, \Sigma_j)$, for $j = 1, ..., j_{\max}$, where Σ_j are the covariance matrices of x under each ARMA model n_{jt}. Then, assuming for simplicity $\mu_j = 0$,

$$p(x \mid M_j) = (2\pi)^{-\frac{T}{2}} |\Sigma_j|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu_j)' \Sigma_j^{-1}(x-\mu_j)\right)$$

the standard quadratic classification rule selects the model i if,

$$i = \underset{1 \le j \le j_{\max}}{\arg\max}\, (2\pi)^{-\frac{T}{2}} |\Sigma_j|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu_j)' \Sigma_j^{-1}(x-\mu_j)\right) \tag{2}$$

and the Bayesian rule selects the model i if,

$$i = \underset{1\leq j\leq j_{\max}}{\arg\max} p(M_j)\,(2\pi)^{-\frac{T}{2}}\,|\Sigma_j|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\,(x-\mu_j)'\,\Sigma_j^{-1}\,(x-\mu_j)\right). \tag{3}$$

In this section the rules (2) and (3) are approximated in several ways and corrected versions of AIC, AICc and BIC are obtained from these approximations. For that, we consider the case in which the time series data, $x = (x_1, ..., x_T)'$, has been generated by the class of ARMA Gaussian processes given by:

$$x_t - \phi_1 x_{t-1} - \ldots - \phi_p x_{t-p} = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}, \qquad t = \ldots, -1, 0, 1, \ldots \tag{4}$$

where a_t is a sequence of independent Gaussian distributed random variables with zero mean and variance $\sigma^2_{p,q}$ and we assume that $p \in \{0, ..., p_{\max}\}$ and $q \in \{0, ..., q_{\max}\}$, where $p_{\max}$ and $q_{\max}$ are some fixed upper bounds. We call the ARMA(p, q) model $M_{p,q}$, $\beta_{p,q} = (\phi_{1p}, \ldots, \phi_{pp}, 0, \ldots, 0, \theta_{1q}, \ldots, \theta_{qq}, 0, \ldots, 0)'$ is a $(p_{\max} + q_{\max}) \times 1$ vector of parameters for the $M_{p,q}$ model and we define $\alpha_{p,q} = \left(\beta_{p,q}, \sigma^2_{p,q}\right)$. We denote the parameters of the model that have generated the data as $\alpha_0 = \left(\beta_0, \sigma_0^2\right)$. Thus, the vector of parameters $\alpha_{p,q}$ are the parameters conditioning that the true model is $M_{p,q}$. In this case, let $\widehat{\beta}_{p,q}$ be the maximum likelihood estimate of the vector of parameters $\beta_{p,q}$ and let $\widehat{\sigma}^2_{p,q}$ be the maximum likelihood estimate of the innovations variance. The covariance matrix of x assuming the model $M_{p,q}$ can be written as $\Sigma_T\,(\alpha_{p,q}) = \sigma_a^2 Q_T\,(\beta_{p,q})$, where $Q_T\,(\beta_{p,q})$ is a $T \times T$ matrix depending on the parameters $\beta_{p,q}$. Let $Q_T\,(\beta_{p,q}) = L\,(\beta_{p,q})\,L'\,(\beta_{p,q})$ be the Cholesky decomposition of $Q_T\,(\beta_{p,q})$. We denote, $a\,(\beta_{p,q}) = L\,(\beta_{p,q})^{-1}\,x$ and $S_x\,(\beta_{p,q}) = a\,(\beta_{p,q})'\,a\,(\beta_{p,q})$. We consider the following assumption:

Assumption 1: The models $M_{p,q}$ are casual, invertible and stationary and with polynomials $1 - \phi_1 B - \ldots - \phi_p B^p$ and $1 - \theta_1 B - \ldots - \theta_q B^q$ with no common roots.

2.1 A maximum likelihood approach

From (2), the discriminant rule assigns the data $x = (x_1, \ldots, x_T)'$, to the model $M_{p,q}$ with parameters $\alpha_{p,q}$ that maximizes $p\,(x \mid M_{p,q}) = p\,(x \mid \alpha_{p,q})$. In practice, the parameters are unknown and the first idea is to substitute the unknown parameters $\alpha_{p,q}$ by its maximum likelihood estimates, $\widehat{\alpha}_{p,q}$, but it is well known that this solution will always choose the model with the largest number of parameters. We propose the following way to obtain a suitable approximation of the quadratic rule. We compute the maximum likelihood estimate

$\widehat{\alpha}_{p,q}$ based on x for each possible model and select the one that maximizes,

$$E_{\alpha_0}[\log p(y|\widehat{\alpha}_{p,q})] = \int p(y|\alpha_0) \log p(y|\widehat{\alpha}_{p,q}) dy,$$

that is, the model that maximizes the expectation with respect to future observations generated by the right model. Note that the model which will be selected by this criterion is the one which minimizes the Kullback-Leibler distance between the selected model and the true one. As,

$$E_{\alpha_0}\left[\log \frac{p(y|\alpha_0)}{p(y|\widehat{\alpha}_{p,q})}\right] = \int p(y|\alpha_0) \log \frac{p(y|\alpha_0)}{p(y|\widehat{\alpha}_{p,q})} dy \geq 0$$

and as the integral is always positive, minimizing it implies making $p(y|\widehat{\alpha}_{p,q})$ as close as possible to $p(y|\alpha_0)$. This criterion computes the log-likelihood of each model using the estimates $\widehat{\alpha}_{p,q}$ based on the sample and then compute the expectation with respect to future observations. The model chosen is the one which leads to a larger expected value of this maximized log-likelihood.

Note that this criterion is related to, although different from, the one given by Akaike (1973). He proposed to select the estimation which leads to a larger value of,

$$E_{\widehat{\alpha}}\left[E_y\left[\log p(y|\widehat{\alpha}_{p,q})\right]\right] = \int\int p(y|\alpha_0) \log p(y|\widehat{\alpha}_{p,q}) dy d\widehat{\alpha}_{p,q}$$

where $\widehat{\alpha}$ and y are assumed to be independent. Thus, Akaike computes the expected value with respect to both the distribution of future observations and the distribution of the estimate. Our criterion, is simpler and in practice leads to the same results, as we show next.

Lemma 1 *Under assumption 1, the expectation with respect the true distribution of the data of the logarithm of the probability of x given the parameters is given by:*

1. *if the parameters are evaluated at $\widehat{\beta}_{p,q}$ and $\frac{T}{T-(p+q)}\widehat{\sigma}^2_{p,q}$:*

$$-\frac{T}{2}(\log 2\pi + 1) - \frac{T}{2}\log \widehat{\sigma}^2_{p,q} - \frac{1}{2}\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right| - (p+q) + O_p(1), \tag{5}$$

2. *if the parameters are evaluated at $\widehat{\beta}_{p,q}$ and $\widehat{\sigma}^2_{p,q}$:*

$$-\frac{T}{2}(\log 2\pi + 1) - \frac{T}{2}\log \widehat{\sigma}^2_{p,q} - \frac{1}{2}\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right| - \frac{T(p+q+1)}{(T-p-q-2)} + O_p(1). \tag{6}$$

Proof. Using (2), we have that,

$$E_{\alpha_0}\left[\log p\left(y \mid \widehat{\alpha}_{p,q}\right)\right] = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \widehat{\sigma}^2_{p,q} - \frac{1}{2}\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right| - \frac{1}{2}E_{\alpha_0}\left[\frac{S_y\left(\widehat{\beta}_{p,q}\right)}{\widehat{\sigma}^2_{p,q}}\right], \tag{7}$$

where $S_y\left(\widehat{\beta}_{p,q}\right) = y'Q_T^{-1}\left(\widehat{\beta}_{p,q}\right)y$. Assuming that the model $M_{p,q}$ is the right model, Brockwell and Davies (1991) showed that,

$$E\left[\frac{S_y\left(\widehat{\beta}_{p,q}\right)}{\widehat{\sigma}^2_{p,q}}\right] \simeq \frac{E\left[S_y\left(\widehat{\beta}_{p,q}\right)\right]}{E\left[\widehat{\sigma}^2_{p,q}\right]} = \frac{T\left(T+p+q\right)}{\left(T-p-q-2\right)} + O_p(1), \tag{8}$$

that gives (6).

On the other hand, using that $T\log(1-\frac{(p+q)}{T}) = -\left(p+q\right) + o(1)$, we have that,

$$\begin{aligned} &T\log 2\pi + T\log\frac{T}{T-(p+q)}\widehat{\sigma}^2_{p,q} + \log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right| = \\ &= T\log 2\pi - T\log\left(1-\frac{(p+q)}{T}\right) + T\log\widehat{\sigma}^2_{p,q} + \log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right| = \\ &= T\log 2\pi + T\log\widehat{\sigma}^2_{p,q} + (p+q) + \log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right| + o_p(1). \end{aligned}$$

Moreover, from (8),

$$E\left[\frac{S_x\left(\widehat{\beta}_{p,q}\right)}{\frac{T}{T-(p+q)}\widehat{\sigma}^2_{p,q}}\right] = \frac{\left(T+p+q\right)T}{\frac{T}{T-(p+q)}\left(T-p-q-2\right)} + O_p(1) = \left(T+p+q\right) + O_p(1),$$

which proves (5). ■

The expression (5) leads to the criterion:

$$AIC^*\left(p,q\right) = \log\widehat{\sigma}^2_{p,q} + \frac{2\left(p+q\right)}{T} + \frac{\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right|}{T} = AIC\left(p,q\right) + \frac{\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right|}{T} \tag{9}$$

which is the corrected version of the Akaike criterion. Expression (6) leads to a model selection criterion of the form:

$$AICc^*\left(p,q\right) = \log\widehat{\sigma}^2_{p,q} + \frac{2\left(p+q+1\right)}{T-(p+q-2)} + \frac{\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right|}{T} = AICc\left(p,q\right) + \frac{\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right|}{T} \tag{10}$$

which is the corrected version of the Hurvich and Tsai criterion.

2.2 A Bayesian approach

We analyze the rule in (3) taking into account that this approach requires prior probabilities of the models, $p(M_{p,q})$ and the parameters, $p(\alpha_{p,q}|M_{p,q})$. The Bayesian point of view of maximizing the posterior probability has been extensively considered, see Schwarz (1978), Chow (1981), Haughton (1988) or Raftery, Madigan and Volinsky (1996), but, to the best of our knowledge, it has not been analyzed as a discrimination problem.

Lemma 2 *Under assumption 1, the logarithm of the probability of x given the parameters is given by:*

$$\log p(x|M_{p,q}) = \frac{1}{2}(p+q+1-T)\log(2\pi) - \frac{1}{2}(p+q+1)\log T - \frac{T}{2}\log\widehat{\sigma}^2_{p,q} \tag{11}$$
$$-\frac{1}{2}\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right| - \frac{1}{2}T + \log p(\widehat{\alpha}_{p,q}|M_{p,q}) + O_p(1).$$

Proof. Let,

$$h(\alpha_{p,q}) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_T(\alpha_{p,q})| - \frac{1}{2}x'\Sigma_T(\alpha_{p,q})^{-1}x + \log p(\alpha_{p,q}|M_{p,q}),$$

then, applying the Laplace's method, see Tierney and Kadane (1986), we obtain,

$$p(x|M_{p,q}) \approx (2\pi)^{\frac{p+q+1-T}{2}}|H(\widehat{\alpha}_{p,q})|^{\frac{1}{2}}|\Sigma_T(\widehat{\alpha}_{p,q})|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}x'\Sigma_T(\widehat{\alpha}_{p,q})^{-1}x\right)p(\widehat{\alpha}_{p,q}|M_{p,q}),$$

where $\widehat{\alpha}_{p,q}$ is the maximum likelihood estimate of $\alpha_{p,q}$ and H is minus the inverse Hessian of h evaluated at $\widehat{\alpha}_{p,q}$. Then,

$$\log p(x|M_{p,q}) \approx \frac{p+q+1-T}{2}\log(2\pi) + \frac{1}{2}\log|H(\widehat{\alpha}_{p,q})| - \frac{1}{2}\log|\Sigma_T(\widehat{\alpha}_{p,q})| - \frac{1}{2}x'\Sigma_T(\widehat{\alpha}_{p,q})^{-1}x + \log p(\widehat{\alpha}_{p,q}|M_{p,q}).$$

Raftery, Madigan and Volinsky (1996) proved that $\log|H(\widehat{\alpha}_{p,q})| = -(p+q+1)\log T + O_p(1)$ because $H(\widehat{\alpha}_{p,q})$ is asymptotically equal to the inverse of the observed information matrix, which in turn is asymp-

totically equal to T times a constant matrix. Then,

$$\log p(x|M_{p,q}) = \frac{1}{2}(p+q+1-T)\log(2\pi) - \frac{1}{2}(p+q+1)\log T - \frac{1}{2}\log|\Sigma_T(\widehat{\alpha}_{p,q})|$$
$$-\frac{1}{2}x'\Sigma_T(\widehat{\alpha}_{p,q})^{-1}x + \log p(\widehat{\alpha}_{p,q}|M_{p,q}) + O(1) = \frac{1}{2}(p+q+1-T)\log(2\pi) - \frac{1}{2}(p+q+1)\log T$$
$$-\frac{T}{2}\log\widehat{\sigma}^2_{p,q} - \frac{1}{2}\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right| - \frac{1}{2}T + \log p(\widehat{\alpha}_{p,q}|M_{p,q}) + O_p(1),$$

which proves the stated result. ■

Therefore, dividing by T in (11), taking the same prior probabilities for all the parameters and ignoring some constant terms leads to the corrected BIC selection criterion,

$$BIC^*(p,q) = \log\widehat{\sigma}^2_{p,q} + \frac{\log(T)(p+q+1)}{T} + \frac{\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right|}{T} = BIC(p,q) + \frac{\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right|}{T}. \tag{12}$$

All the criteria obtained can be written in a compact way as members of the family of criteria,

$$\min_{(p,q)}\left\{\log\widehat{\sigma}^2_{p,q} + (p+q)\times C(T,p+q+1) + \frac{\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right|}{T}\right\}, \tag{13}$$

where the term $\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right|$ is computed as we will show in section 3. Note that the properties of efficiency and consistency of the criteria in (13) are not affected by the correction term. For that, first we state the following Theorem whose proof is in the appendix, which shows that the criteria (9) and (10) are efficient.

Theorem 3 *Under the assumptions: (A1)* $\{x_t\}$ *is generated by a stationary process* $x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \ldots = a_t$, $t = \ldots, -1, 0, 1, \ldots$ *where* a_t *is a sequence of independent Gaussian distributed random variables with zero mean and variance* σ_a^2 *and* $\sum_{j=1}^{\infty}|\phi_j| < \infty$*; (A2) The sequence* $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots$*, is nonzero for* $|B| \leq 1$*; (A3)* $p_{\max}$ *is a sequence of positive integers such that* $p_{\max} \to \infty$ *and* $p_{\max}/\sqrt{T} \to 0$ *as* $T \to \infty$*; (A4)* $\{x_t\}$ *is not degenerate to a finite order autoregressive process, the AIC* and AICc* are efficient.*

Thus, AIC* and AICc* are similar to AIC and AICc for large samples, but includes the correction factor $\log\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right|/T$, that we will show is important in finite samples. For the criterion in (12), the consistency property is preserved due to Theorem 3 in Hannan (1980, p. 1073). Therefore, the criterion (12) is a consistent criterion.

Remark: The main contribution of this section is to view the model selection problem as a kind of discrimination analysis and present an unified approach of criteria proposed in the literature from different points of view. As far as we known, the connection between this two decision problems has not been previously made. The technical details in both maximum likelihood and Bayesian points of view are included for completeness. The inclusion of the correction term was analyzed in Hannan (1980) and concluded that it can be omitted because it tends to zero with T. We will show in the simulation experiment its importance in finite sample series. Hannan (1980) omitted this term to derive the consistency properties of the BIC. Theorem 3 shows that the results in Hannan (1980) and Shibata (1980) are not affected by the inclusion of this term.

3 ANALYSIS OF THE CORRECTION TERM

A key point in the previous discussion is obtaining a suitable expression for $\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right|$. We note that the determinant depends on the maximum likelihood estimates under the hypothesis of an ARMA(p,q) model. Leeuw (1994) provides an expression for $|Q_T(\beta_{p,q})|$ in closed form which only depends on the parameters $\beta_{p,q}$ and only requires the computation of two determinants of order $m = \max(p,q)$. The expression can be written as:

$$|Q_T(\beta_{p,q})| = \frac{\left|(R'R - SS') + (RV - US)' H_1' H_1 (RV - US)\right|}{|R'R - SS'|}, \tag{14}$$

where R, S, U and V are $m \times m$ matrices given by:

$$R = \begin{cases} 0 & i < j \\ 1 & i = j \\ -\phi_{i-j} & i > j \end{cases} \qquad U = \begin{cases} 0 & i < j \\ 1 & i = j \\ -\theta_{i-j} & i > j \end{cases}$$

$$S = \begin{cases} -\phi_{m+(i-j)} & i \leq j \\ 0 & i > j \end{cases} \qquad V = \begin{cases} -\theta_{m+(i-j)} & i \leq j \\ 0 & i > j \end{cases}$$

and the matrix H_1 is the $T \times m$ matrix consisting of the first m columns of the adjoint of the $T \times T$ matrix H, which has the same structure as U with 0 outside the first m inferior diagonals. The ϕ_i elements of the previous matrices are 0 if $m > p$ and the θ_i are 0 if $m > q$. Therefore, we can obtain $\left|Q_T\left(\widehat{\beta}_{p,q}\right)\right|$ plugging in the estimates $\widehat{\beta}_{p,q}$ in the formulas for the matrices R, S, U, V and H.

Let us analyze the term $|Q_T(\beta_{p,q})|$. As $Q_T(\beta_{p,q}) = \left(\sigma_x^2/\sigma_a^2\right) R_T(\alpha_{p,q})$, where σ_x^2 is the variance of x_t and

$R_T(\alpha_{p,q})$ is the correlation matrix of order T, we have $|Q_T(\beta_{p,q})| = \left(\sigma_x^2/\sigma_a^2\right)^T |R_T(\alpha_{p,q})|$. Durbin (1960) and Ramsey (1974) show that:

$$\sigma_x^2/\sigma_a^2 = \prod_{i=1}^{\infty} \left(1 - \phi_{ii}^2(\beta_{p,q})\right)^{-1}, \qquad |R_T(\alpha_{p,q})| = \prod_{i=1}^{T-1} \left(1 - \phi_{ii}^2(\beta_{p,q})\right)^{T-i}$$

respectively, where $\phi_{ii}(\beta_{p,q})$ are the partial autocorrelations of the process. It is important to note that these partial autocorrelations are obtained under the assumption of an ARMA(p,q) model. Then,

$$|Q_T(\beta_{p,q})| = \frac{\prod_{i=1}^{T-1} \left(1 - \phi_{ii}^2(\beta_{p,q})\right)^{T-i}}{\prod_{i=1}^{\infty} \left(1 - \phi_{ii}^2(\beta_{p,q})\right)^{T}} = \prod_{i=1}^{T-1} \left(1 - \phi_{ii}^2(\beta_{p,q})\right)^{-i} \prod_{i=T}^{\infty} \left(1 - \phi_{ii}^2(\beta_{p,q})\right)^{-T}. \tag{15}$$

Consequently, the criteria (13) can be written as follows:

$$\min_{(p,q)} \left\{ \log \widehat{\sigma}_{p,q}^2 + (p+q) \times C(T, p+q) - \sum_{i=1}^{T-1} \frac{i}{T} \log\left(1 - \phi_{ii}^2\left(\widehat{\beta}_{p,q}\right)\right) - \sum_{i=T}^{\infty} \log\left(1 - \phi_{ii}^2\left(\widehat{\beta}_{p,q}\right)\right) \right\}.$$

To understand further the corrected criteria, let us obtain the difference AIC$^*(p+1,q)$ −AIC$^*(p,q)$ which is given by

$$AIC^*(p+1,q) - AIC^*(p,q) = \log \frac{\widehat{\sigma}_{p+1,q}^2}{\widehat{\sigma}_{p,q}^2} + \frac{2}{T} - \sum_{i=1}^{T-1} \frac{i}{T} \log\left(1 - \phi_{ii}^2\left(\widehat{\beta}_{p+1,q}\right)\right) -$$
$$- \sum_{i=T}^{\infty} \log\left(1 - \phi_{ii}^2\left(\widehat{\beta}_{p+1,q}\right)\right) + \sum_{i=1}^{T-1} \frac{i}{T} \log\left(1 - \phi_{ii}^2\left(\widehat{\beta}_{p,q}\right)\right) + \sum_{i=T}^{\infty} \log\left(1 - \phi_{ii}^2\left(\widehat{\beta}_{p,q}\right)\right),$$

and this can be approached by

$$\frac{\widehat{\sigma}_{p+1,q}^2 - \widehat{\sigma}_{p,q}^2}{\widehat{\sigma}_{p,q}^2} + \frac{2}{T} + \sum_{i=1}^{T-1} \frac{i}{T} \left(\phi_{ii}^2\left(\widehat{\beta}_{p+1,q}\right) - \phi_{ii}^2\left(\widehat{\beta}_{p,q}\right)\right) + \sum_{i=T}^{\infty} \left(\phi_{ii}^2\left(\widehat{\beta}_{p+1,q}\right) - \phi_{ii}^2\left(\widehat{\beta}_{p,q}\right)\right).$$

The first term measures the relative change between the variances, $\widehat{\sigma}_{p+1,q}^2$ and $\widehat{\sigma}_{p,q}^2$. The second term is a penalization for the inclusion of one additional parameter. The third and the fourth terms measure the discrepancy between all the partial autocorrelation coefficients under both hypothesis, ARMA(p,q) and ARMA$(p+1,q)$, with weights that increase with the lag. Therefore, AIC$^*(p+1,q)$ <AIC$^*(p,q)$ if either: (a) $\widehat{\sigma}_{p+1,q}^2$ is significantly smaller than $\widehat{\sigma}_{p,q}^2$, or (b) the weighted sum of the partial autocorrelation coefficients computed under the ARMA(p,q) model is greater than the corresponding sum under the ARMA$(p+1,q)$

model. The same interpretation applies to BIC* and AICc*, and the only difference is the penalization term for including one additional parameter.

In the case of autoregressive fitting, where the criteria (13) takes the form:

$$\min_p \left\{ \log \widehat{\sigma}_p^2 + p \times C(T,p) + \frac{\log \left| Q_T \left(\widehat{\beta}_p \right) \right|}{T} \right\}, \tag{16}$$

and we note that then (14) is reduced to:

$$|Q_T(\beta_p)| = \frac{1}{|R'R - SS'|} \tag{17}$$

and using that $\sigma_x^2/\sigma_a^2 = \prod_{i=1}^{p} \left(1 - \phi_{ii}^2(\beta_p)\right)^{-1}$ and taking into account that $\phi_{ii}^2(\beta_p) = 0$ if $i > p$,

$$|Q_T(\beta_p)| = \frac{\prod_{i=1}^{T-1} \left(1 - \phi_{ii}^2(\beta_p)\right)^{T-i}}{\prod_{i=1}^{p} \left(1 - \phi_{ii}^2(\beta_p)\right)^{T}} = \prod_{i=1}^{p} \left(1 - \phi_{ii}^2(\beta_p)\right)^{-i}. \tag{18}$$

Therefore, for AR(p) models, the criteria (16) can be written as follows:

$$\min_p \left\{ \log \widehat{\sigma}_p^2 + p \times C(T,p) - \sum_{i=1}^{p} \frac{i}{T} \log \left(1 - \phi_{ii}^2 \left(\widehat{\beta}_p \right)\right) \right\},$$

and the difference AIC$^*(p+1)$ −AIC$^*(p)$ can be approached by

$$\frac{\widehat{\sigma}_{p+1}^2 - \widehat{\sigma}_p^2}{\widehat{\sigma}_p^2} + \frac{2}{T} + \sum_{i=1}^{p} \frac{i}{T} \left(\phi_{ii}^2 \left(\widehat{\beta}_{p+1} \right) - \phi_{ii}^2 \left(\widehat{\beta}_p \right) \right) + \frac{p+1}{T} \phi_{p+1,p+1}^2 \left(\widehat{\beta}_{p+1} \right).$$

As in the case of ARMA models, the first term measures the relative change between the variances, $\widehat{\sigma}_p^2$ and $\widehat{\sigma}_{p+1}^2$, the second term is the penalization for including one additional parameter and the third term measures the discrepancy between the first p partial autocorrelations under both hypothesis, AR(p) and AR($p+1$). Finally, the last term measures the significance of the $p+1$ autocorrelation coefficient.

4 MODEL SELECTION IN SETAR MODELS

One of the most often used nonlinear time series model is the self-exciting threshold autoregressive (SETAR) model. A time series data, $x = (x_1, ..., x_T)'$ has been generated by the class of SETAR processes if it follows

the model:

$$x_t = \phi_{j0} + \sum_{i=1}^{p_j} \phi_{ji} x_{t-i} + a_{jt}, \qquad \text{if } r_{j-1} \leq x_{t-d} < r_j \tag{19}$$

where we assume that a_{jt}, $j = 1, \ldots, k$, is a white noise series with zero mean and finite variances $\sigma^2_{a_j}$, $d \in \{0, ..., d_{\max}\}$ is a positive integer and $-\infty = r_0 < r_1 < \ldots < r_{k-1} < r_k = \infty$ are the thresholds. We also assume that $p_j \in \left\{0, ..., p_j^{\max}\right\}$ where $p_j^{\max}$ are some fixed upper bounds. We call $M_{p_1,\ldots,p_k,d}$ the SETAR$(p_1, \ldots, p_k, d)$ model.

Exact maximum likelihood estimates of the parameters of the model (19) are not considered because the thresholds $r_1, \ldots, r_{k-1}$ are not continuous. Therefore, the parameters of the model (19) are estimated by conditional likelihood. We assume the following assumption:

Assumption 2: The models $M_{p_1,\ldots,p_k,d}$ are stationary and ergodic with finite second moments and the stationary distribution of $x = (x_1, ..., x_T)'$ admits a density that is positive everywhere.

Chan (1993) shows that under assumption 2, the conditional least squares estimators of the parameters of a stationary ergodic threshold autoregressive model are strongly consistent. Let $\phi_j = \left(\phi_{j0}, \ldots, \phi_{jp_j}\right)'$, $j = 1, \ldots, k$, be the vector of parameters. The sums of squares function is:

$$\begin{aligned} S_x\left(\phi_1, \ldots, \phi_k, r_0, \ldots, r_k\right) &= S_x\left(\phi_1, r_0, r_1\right) + \cdots + S_x\left(\phi_k, r_{k-1}, r_k\right) = \\ &= \sum_{r_0 < x_{t-d} \leq r_1} a_t^2 + \cdots + \sum_{r_{k-1} \leq x_{t-d} < r_k} a_t^2, \end{aligned} \tag{20}$$

and the conditional least squares of $(\phi_1, \ldots, \phi_k, r_1, \ldots, r_{k-1})$ are the values that minimize (20), which we denote by $\widehat{\phi}_1, \ldots, \widehat{\phi}_k$ and $\widehat{r}_1, \ldots, \widehat{r}_{k-1}$. The residual variances are defined as,

$$\widehat{\sigma}_1^2 = \frac{S_x\left(\widehat{\phi}_1, r_0, \widehat{r}_1\right)}{T_1}, \qquad \widehat{\sigma}_j^2 = \frac{S_x\left(\widehat{\phi}_j, \widehat{r}_{j-1}, \widehat{r}_j\right)}{T_j}, \; j = 2, \ldots, k-1, \qquad \widehat{\sigma}_k^2 = \frac{S_x\left(\widehat{\phi}_k, \widehat{r}_{k-1}, r_k\right)}{T_k}$$

where T_j are the number of observations in each regime for the estimates $\widehat{r}_1, \ldots, \widehat{r}_{k-1}$.

Little attention has been paid to model selection in SETAR models. Wong and Li (1998) derive the AICc for these models for the case $k = 2$ and propose a procedure for selecting and estimating a SETAR model and compare via a simulation study three model selection criteria, AIC, AICc and BIC, which for k regimens

are given by:

$$
\begin{aligned}
BIC\left(p_1,\ldots,p_k\right) &= \sum_{j=1}^{k}\left\{T_j\log\widehat{\sigma}_j^2+\log T_j\left(p_j+1\right)\right\}\\
AIC\left(p_1,\ldots,p_k\right) &= \sum_{j=1}^{k}\left\{T_j\log\widehat{\sigma}_j^2+2\left(p_j+2\right)\right\}\\
AICc\left(p_1,\ldots,p_k\right) &= \sum_{j=1}^{k}\left\{T_j\log\widehat{\sigma}_j^2+\frac{T_j\left(T_j+p_j+1\right)}{T_j-p_j-3}\right\}.
\end{aligned}
\tag{21}
$$

The procedure proposed by Wong and Li (1998) when $k=2$ and $r_1=r$ and d are unknown works as follows: (a) Fix the maximum autoregressive and delay orders $\{p_1^{\max},p_2^{\max},d^{\max}\}$; (b) Assume $r\in[l,u]\subset R$, where l is the $0.25\times100\%$ percentile and u is the $0.75\times100\%$ percentile of x_t; (c) Let $x_{(1)},\ldots,x_{(T)}$ be the order statistics of x_t; (d) Let $I_r=\{[0.25T],\ldots,[0.75T]\}$. Set $r=x_{(i)}$, $i\in I_r$; (e) Calculate,

$$
\min\left\{C\left(p_1,p_2,d,x_{(i)}\right):p_1\in\{1,\ldots,p_1^{\max}\},p_2\in\{1,\ldots,p_2^{\max}\},d\in\{1,\ldots,d^{\max}\},x_{(i)}\in I_r\right\},
$$

where $C\left(p_1,p_2,d,x_{(i)}\right)$ is the model selection criteria used. The autoregressive orders $\left(p_1,p_2,d,x_{(i)}\right)$, the delay parameter, d and the estimated threshold are the ones that minimize $C\left(p_1,p_2,d,x_{(i)}\right)$.

This procedure gives the autoregressive orders and the delay parameter selected by the criterion and the estimated threshold. Wong and Li (1998) carried out a Monte Carlo experiment for different models and sample sizes for the criteria in (21), and conclude that the AICc is the preferable criterion for small sample sizes but BIC is preferable for medium and large sample sizes.

De Gooijer (2001) proposes a procedure for selecting and estimating the parameters of a SETAR model for three cross-validation criteria. The first four steps of the procedure proposed for the first criterion that we denote by C_1 are as in Wong and Li (1998). The rest of the procedure works as follows: (e) Omit one observation of the series and with the remaining data set obtain conditional least squares estimates of the parameters of the corresponding model, which we denote by $\widehat{\phi}_j^t$, predict the omitted observation and obtain the predictive residual, $a_t\left(\widehat{\phi}_j^t,\widehat{r}_{j-1},\widehat{r}_j\right)$; (f) Repeat the previous step for all the observations, and the final model is the one that minimizes the C_1 criterion defined as follows:

$$
C_1\left(p_1,\ldots,p_k\right)=T\log\left(T^{-1}\sum_{t=1}^{T}a_t^2\left(\widehat{\phi}_j^t,\widehat{r}_{j-1},\widehat{r}_j\right)\right).
\tag{22}
$$

Stoica et al (1986) proved that for a given model,

$$C_1(p_1,\ldots,p_k) = AIC(p_1,\ldots,p_k) + O(T^{-\frac{1}{2}}). \tag{23}$$

Therefore, based on the definition of AICc and (23), De Gooijer (2001) define the following criterion:

$$Cc(p_1,\ldots,p_k) = C_1(p_1,\ldots,p_k) + \sum_{j=1}^{k}\left\{\frac{T_j(T_j+p_j+1)}{T_j-p_j-3}\right\}. \tag{24}$$

Moreover, De Gooijer (2001) proposes a generalization of a model selection criterion introduced by McQuarrie et al (1997) for linear models. This criterion is not efficient or consistent but it has a good performance in finite samples. This criterion for SETAR models is,

$$AICu(p_1,\ldots,p_k) = AICc(p_1,\ldots,p_k) + \sum_{j=1}^{k} T_j \log\left\{\frac{T_j}{T_j-p_j-2}\right\}, \tag{25}$$

and the cross validation criteria proposed by De Gooijer (2001) has the form:

$$Cu(p_1,\ldots,p_k) = Cc(p_1,\ldots,p_k) + \sum_{j=1}^{k} T_j \log\left\{\frac{T_j}{T_j-p_j-2}\right\}. \tag{26}$$

As SETAR models are piecewise linear we can include the correction term in (17) in each regime. Therefore, we will explore the performance of the Wong and Li (1998) procedure but modifying the criteria BIC, AIC, AICc and AICu by:

$$C^*(p_1,\ldots,p_k) = C(p_1,\ldots,p_k) + \sum_{j=1}^{k}\log\left|Q_{T_j}\left(\widehat{\phi}_j\right)\right|, \tag{27}$$

where $C(p_1,\ldots,p_k)$ represents by the BIC, AIC, AICc and AICu in (21) and (25) respectively. In order to compute the correction term in each regime we first estimate the parameters of the model by conditional likelihood and then obtain the correction term in each regime as in (17).

In the same way, the cross-validation criteria proposed by De Gooijer (2001) can be modified as:

$$C_1^*(p_1,\ldots,p_k) = C_1(p_1,\ldots,p_k) + \sum_{j=1}^{k}\log\left|Q_{T_j}\left(\widehat{\phi}_j\right)\right|, \tag{28}$$

where $C(p_1,\ldots,p_k)$ represents $C_1(p_1,\ldots,p_k)$, $Cc(p_1,\ldots,p_k)$ and $Cu(p_1,\ldots,p_k)$. In this case, the pro-

cedure in De Gooijer (2001) is modified by adding the correction term in the last step by estimating by conditional least squares with all the observations in the series. Then, the final model selected is the one that minimizes one of the criteria in (28). The correction term is computed as in the previous case.

5 MONTE CARLO EXPERIMENTS

5.1 Simulations for ARMA models

To evaluate the performance of the proposed criteria for ARMA models and different sample sizes, 1000 realizations were generated from the following six models, (M1) $x_t = 0.9x_{t-1} + a_t$, (M2) $x_t = 1.4x_{t-1} - 0.7x_{t-2} + a_t$, (M3) $x_t = 0.4x_{t-1} - 0.8x_{t-2} + 0.6x_{t-3} + a_t$, (M4) $x_t = a_t + 0.8a_{t-1}$, (M5) $x_t = 0.8x_{t-1} + a_t + 0.7a_{t-1}$ and (M6) $x_t = 1.4x_{t-1} - 0.7x_{t-2} + a_t + 0.8a_{t-1}$, where a_t are independent identically distributed standard normal. These models have been chosen to represent different situations. The first three represents some common AR structures: a strong AR(1) dependency (M1), an AR(2) with two complex roots (M2), and an AR(3) model with a real factor and two complex ones (M3). The second three models include ARMA models which require long AR approximations with real roots (M4 and M5) and mixtures of real and complex roots (M6). The first three models are used to show the performance of the corrected criteria in finite autoregressive order models and the last three in ARMA processes. Based on the previous sections, we compare the performance of criteria of the form:

$$\min_{p,q} \left\{ \log \widehat{\sigma}^2_{p,q} + (p+q) \times C(T, p+q) \right\}$$

and,

$$\min_{p,q} \left\{ \log \widehat{\sigma}^2_{p,q} + (p+q) \times C(T, p+q) + \frac{\log \left| Q_T \left(\widehat{\beta}_{p,q} \right) \right|}{T} \right\}.$$

In all cases, 1000 series were generated from each model M1 to M6 with sample sizes $T = 31$, 51 and 101. For autoregressive processes, where $q = 0$, we fit each model to the first $T-1$ observations of each series by maximum likelihood estimation and in each model, we obtain $\left| Q_T \left(\widehat{\beta}_p \right) \right|$ as in (17). We fix $p_{\max} = 15$, so that 16 models are fitted for each series. For ARMA processes we fit the whole set of $(p_{\max}+1) \times (q_{\max}+1)$ ARMA orders to the first $T-1$ observation where $p_{\max} = 4$ and $q_{\max} = 4$ by maximum likelihood estimation, so that 25 models are fitted for each series. In each model, we obtain $\left| Q_T \left(\widehat{\beta}_{p,q} \right) \right|$ by (14). In both cases, with the model chosen by each criteria and the fitted parameters, we obtain the prediction in the least mean

Table 1: Frequency of times of correct selection and root mean square prediction errors for the models selected by each criterion. If C is a given criterion, C* is the corrected one by including the correction term proposed in this article

$T = 30$													
M	BIC	BIC*	AIC	AIC*	AICc	AICc*	M	BIC	BIC*	AIC	AIC*	AICc	AICc*
1	793	851	469	612	667	749	1	1.28	1.16	1.58	1.35	1.17	1.12
2	763	829	454	662	694	782	2	1.13	1.08	1.41	1.26	1.03	0.99
3	566	587	422	578	596	623	3	1.22	1.10	1.39	1.23	1.07	1.03
4	495	595	187	309	360	460	4	1.03	1.01	1.12	1.05	1.07	1.02
5	302	423	101	185	224	352	5	1.09	1.05	1.10	1.02	1.09	1.01
6	314	510	111	270	262	474	6	1.04	0.99	1.07	1.03	1.04	1.03
$T = 50$													
M	BIC	BIC*	AIC	AIC*	AICc	AICc*	M	BIC	BIC*	AIC	AIC*	AICc	AICc*
1	879	902	538	641	675	727	1	1.06	1.06	1.10	1.10	1.09	1.09
2	878	914	553	677	695	771	2	1.06	1.06	1.10	1.08	1.08	1.07
3	802	819	554	681	690	745	3	1.11	1.11	1.15	1.13	1.12	1.11
4	655	728	246	314	326	399	4	1.09	1.09	1.24	1.16	1.17	1.14
5	554	628	226	283	314	397	5	1.13	1.13	1.18	1.16	1.17	1.14
6	598	725	265	384	385	512	6	1.03	1.00	1.06	1.01	1.01	1.00
$T = 100$													
M	BIC	BIC*	AIC	AIC*	AICc	AICc*	M	BIC	BIC*	AIC	AIC*	AICc	AICc*
1	931	940	584	629	633	671	1	1.17	1.17	1.17	1.16	1.16	1.16
2	906	917	589	655	663	704	2	0.99	0.98	0.94	0.94	0.93	0.90
3	944	954	622	702	685	758	3	1.09	1.09	1.12	1.10	1.09	1.10
4	833	852	378	418	441	458	4	1.14	1.14	1.23	1.23	1.23	1.22
5	826	843	418	435	463	493	5	0.91	0.89	0.92	0.92	0.92	0.92
6	857	888	500	537	564	607	6	1.02	1.01	1.03	1.03	1.04	1.03

square error sense for the last observation and we compare it with the true value. In this way, we obtain the prediction error for this observation and the chosen model. As BIC and BIC* are consistent criteria, that is, they aim at choosing the right order, and AIC, AICc, AIC* and AICc* are efficient criteria, that is, they aim at choosing the best predictor model, we analyze the consistency and the efficiency properties for all the criteria in small and medium samples.

It is important to note that the fitting of the $(p_{\max}+1)\times(q_{\max}+1)$ ARMA models can lead to serious estimation problems. See Hannan and Rissanen (1982), Hannan and Kavalieris (1984), Poskitt (1987), Pukkila et al (1990) or Pötscher (1990), among others. These authors have proposed algorithms to estimate the true orders (p,q) of an ARMA process by means of the consistent estimation of m=max(p,q). Knowing m avoids the estimation of overparametrized ARMA processes that leads to some inconsistency problems. Here, for simplicity, we consider the whole search over all the candidate models. When a estimated singular covariance matrix is found in the Monte-Carlo experiment, it is rejected from the comparison with other

Table 2: Mean and Standard Error of the model order chosen by the criteria compared for M1, M2 and M3

$T = 30$						
M	BIC	BIC*	AIC	AIC*	AICc	AICc*
1	1.50 (1.54)	1.23 (0.73)	3.93 (4.26)	2.04 (1.94)	1.71 (1.41)	1.41 (0.94)
2	2.57 (1.71)	2.19 (0.68)	4.86 (3.99)	2.83 (1.69)	2.60 (1.28)	2.31 (0.81)
3	3.23 (1.69)	2.80 (0.74)	5.12 (3.62)	3.30 (1.42)	3.19 (1.15)	2.89 (0.76)
$T = 50$						
M	BIC	BIC*	AIC	AIC*	AICc	AICc*
1	1.20 (0.70)	1.13 (0.48)	3.27 (3.69)	2.13 (2.23)	1.87 (1.81)	1.56 (1.24)
2	2.18 (0.66)	2.10 (0.43)	4.01 (3.39)	2.90 (1.87)	2.78 (1.67)	2.44 (1.05)
3	3.15 (0.77)	3.04 (0.56)	4.53 (2.74)	3.64 (1.52)	3.56 (1.32)	3.31 (0.92)
$T = 100$						
M	BIC	BIC*	AIC	AIC*	AICc	AICc*
1	1.09 (0.40)	1.07 (0.33)	2.66 (3.00)	2.20 (2.33)	2.13 (2.15)	1.86 (1.72)
2	2.12 (0.44)	2.10 (0.36)	3.69 (2.90)	3.13 (2.18)	3.04 (2.01)	2.76 (1.63)
3	3.07 (0.36)	3.05 (0.31)	4.38 (2.59)	3.86 (1.89)	3.89 (1.87)	3.56 (1.32)

models.

We will compare the 3 criteria with their corrected versions in the six models, so that 54 comparisons are made. The frequencies over 1000 where the criteria chose the right order for the six models are shown in columns 2 to 7 in Table 1. It can be seen that for small sample size, $T = 30$, the improvement in the number of times in which the correct model is selected can be as large as 143% (see AIC and AIC* in M6) and for $T = 100$ as large as 13% (see AIC and AIC* in M3). On the other hand, columns 9 to 14 in Table 1 show the root mean square prediction error estimated for each criteria in all the sample sizes. For $T = 30$, the corrected criteria improves the forecast performance of the original ones. For $T = 50$ and $T = 100$, the root mean square prediction errors of the corrected criteria are also smaller in most of the cases, but the differences between the corrected and the original criteria are small, especially for $T = 100$.

Table 2 presents the mean and the standard error or the orders chosen by each criteria for the autoregressive fitting for the first three models. For all the sample sizes considered, the corrected criteria performs better than the original versions. Note that the inclusion of the term $\left|Q_T\left(\widehat{\beta}_p\right)\right|$ is very effective in reducing the standard error of the order taken by the criteria, especially for the case of $T = 30$, where the standard deviation is reduced by more than 50% by introducing the correction term.

Finally, we carried out a last experiment to analyze the forecasting performance of AR approximations to

Table 3: Root Mean Square Prediction Error for M4

T	BIC	BIC*	AIC	AIC*	AICc	AICc*
30	1.33	1.23	1.51	1.36	1.13	1.11
50	1.15	1.14	1.17	1.14	1.12	1.13
100	1.09	1.10	1.09	1.08	1.09	1.09

ARMA models. We generate 1000 series from the model M4 for the sample sizes $T = 31$, 51 and 101. We fit an autoregressive model by maximum likelihood with $p_{\max} = 15$ for the first $T-1$ observations, and obtain the prediction error for the last observation for the model chosen by each criterion. With the 1000 series we estimate the root mean square prediction error. Table 3 shows the results. For $T = 30$, in all the cases, the corrected criteria outperform the original criteria. For the sample sizes $T = 50$ and $T = 100$, the corrected criteria perform better than the original ones in the cases of BIC and AIC, but the differences between the original and the corrected criteria are quite small. We conclude that the correction term improves the small and medium sample performance of all the corrected criteria.

5.2 Simulations for SETAR models

To evaluate the performance of the proposed criteria for SETAR models and different sample sizes, 1000 realizations were generated from the following two stationary SETAR models,

$$(M7)\begin{cases} x_t = -0.8x_{t-1} + a_{1t}, & x_{t-1} \leq 0 \\ x_t = -0.2x_{t-1} + a_{2t}, & x_{t-1} > 0 \end{cases} \qquad (M8)\begin{cases} x_t = 0.5x_{t-1} + a_{1t}, & x_{t-1} \leq 0 \\ x_t = -0.5x_{t-1} + a_{2t}, & x_{t-1} > 0 \end{cases}$$

where $a_{jt} \sim N(0,1)$, $j = 1, 2$. Based on section 4, we compare the performance of the criteria in (21), (25) with respect to the criteria in (27) and the criteria in (22), (24) and (26) with respect to the criteria in (28). In all cases, 1000 series were generated from models M7 and M8 with sample sizes $T = 31$, 51 and 101. We proceed as in Wong and Li (1998) and De Gooijer (2001) by using a grid to estimate the threshold parameter r. We fit each model to the first $T-1$ observations of each series by conditional likelihood in each model, we obtain the correction term in (17) in each regime. We first assume that the delay parameter is known and fix $p_1^{\max} = p_2^{\max} = 5$ for $T = 31$, 51 and 101, so that taking into account that the number of possible values of the threshold parameter is $(T-1)/2$, we compare 375, 625 and 1250 models respectively. In every case, we consider the following measures of the performance of the model selection criteria: (a) the frequency detection of the correct order $(p_1, p_2) = (1, 1)$, (b) the root mean square error of estimation of

Table 4: Frequency of times of correct selection, root mean square errors of the threshold parameter and root mean square prediction errors assuming that d is known

M	$T=30$	BIC	BIC*	AIC	AIC*	AICc	AICc*	AICu	AICu*	C_1	C_1^*	Cc	Cc*	Cu	Cu*
7	(p_1,p_2)	306	377	254	331	800	856	903	923	363	474	818	854	895	923
7	$RMSE$	0.62	0.62	0.60	0.60	0.41	0.40	0.40	0.39	0.55	0.54	0.40	0.40	0.40	0.39
7	$RMSPE$	1.41	1.37	1.38	1.32	1.13	1.13	1.12	1.12	1.22	1.22	1.14	1.12	1.12	1.11
8	(p_1,p_2)	306	361	243	310	779	825	876	913	378	450	786	840	890	921
8	$RMSE$	0.84	0.84	0.82	0.82	0.65	0.66	0.65	0.65	0.73	0.74	0.64	0.65	0.63	0.63
8	$RMSPE$	1.64	1.63	1.64	1.64	1.16	1.15	1.13	1.12	1.31	1.31	1.14	1.13	1.12	1.12
M	$T=50$	BIC	BIC*	AIC	AIC*	AICc	AICc*	AICu	AICu*	C_1	C_1^*	Cc	Cc*	Cu	Cu*
7	(p_1,p_2)	420	512	198	298	629	686	789	827	286	395	634	688	796	836
7	$RMSE$	0.58	0.56	0.55	0.54	0.47	0.47	0.47	0.47	0.53	0.52	0.45	0.45	0.45	0.44
7	$RMSPE$	1.15	1.14	1.19	1.19	1.06	1.06	1.06	1.06	1.15	1.15	1.07	1.07	1.07	1.06
8	(p_1,p_2)	412	507	216	307	633	704	802	846	313	406	659	724	820	855
8	$RMSE$	0.74	0.74	0.73	0.74	0.63	0.63	0.63	0.64	0.70	0.70	0.63	0.65	0.64	0.64
8	$RMSPE$	1.19	1.19	1.24	1.23	1.15	1.15	1.13	1.12	1.22	1.22	1.15	1.14	1.12	1.12
M	$T=100$	BIC	BIC*	AIC	AIC*	AICc	AICc*	AICu	AICu*	C_1	C_1^*	Cc	Cc*	Cu	Cu*
7	(p_1,p_2)	743	796	359	431	537	591	753	789	358	439	522	589	765	799
7	$RMSE$	0.54	0.52	0.52	0.51	0.50	0.48	0.50	0.47	0.51	0.49	0.49	0.47	0.48	0.46
7	$MSPE$	1.06	1.06	1.08	1.06	1.07	1.06	1.06	1.06	1.08	1.08	1.07	1.07	1.06	1.06
8	(p_1,p_2)	771	819	365	446	545	600	773	816	398	461	542	611	773	805
8	$RMSE$	0.59	0.59	0.58	0.59	0.57	0.58	0.57	0.57	0.58	0.58	0.56	0.57	0.56	0.57
8	$RMSPE$	1.06	1.05	1.07	1.07	1.07	1.07	1.06	1.05	1.07	1.07	1.07	1.06	1.06	1.06

the threshold parameter and (c) the root mean square prediction error for the last observation based on the model chosen by each criteria, the fitted parameters and the true value. The results are in Table 4. It can be seen that for small sample size, $T = 30$, the improvement in the number of times in which the correct model is selected can be as large as 30.5 % (see C_1 and C_1^* in M7), for $T = 50$ as large as 50.5 % (see AIC and AIC* in M7) and for $T = 100$ as large as 22.6 % (see AIC and AIC* in M7). We note that the AICu, AICu*, Cu and Cu* have larger frequency detection for $T = 30$ but the frequency detection decreases when the sample size increases.

On the other hand, the root mean square error of estimation (RMSE) of the threshold parameter are very close for the original and corrected criteria, whereas the root mean square prediction error (RMSPE) is usually smaller for the corrected criteria.

Now, we assume that the delay is unknown and fix $p_1^{\max} = p_2^{\max} = 5$ and $d^{\max} = 4$ for $T = 31$, 51 and 101, so that taking into account that the number of possible values of the threshold parameter is $(T-1)/2$, we compare the 1500, 2500 and 5000 models respectively. In every case, we consider the following measures of the performance of the model selection criteria: (a) the frequency detection of the correct order

Table 5: Frequency of times of correct selection, root mean square errors of the threshold parameter and root mean square prediction errors assuming that d is unknown

M	$T = 30$	BIC	BIC*	AIC	AIC*	AICc	AICc*	AICu	AICu*	C_1	C_1^*	Cc	Cc*	Cu	Cu*
7	(p_1, p_2)	205	303	163	248	739	808	862	897	292	403	765	824	863	913
7	d	540	559	553	571	582	616	589	617	540	540	573	578	573	575
7	(p_1, p_2, d)	135	187	117	165	463	517	527	566	187	240	469	499	517	536
7	$RMSE$	0.65	0.64	0.63	0.62	0.43	0.42	0.42	0.41	0.55	0.55	0.41	0.41	0.41	0.41
7	$RMSPE$	1.62	1.61	1.64	1.62	1.18	1.17	1.17	1.17	1.29	1.28	1.19	1.19	1.18	1.18
8	(p_1, p_2)	184	249	147	221	733	802	846	879	292	390	754	809	860	885
8	d	569	555	562	568	601	590	608	598	611	599	649	632	653	633
8	(p_1, p_2, d)	119	146	98	141	461	485	527	538	203	254	512	526	577	570
8	$RMSE$	0.90	0.90	0.88	0.88	0.68	0.69	0.67	0.67	0.78	0.78	0.67	0.68	0.66	0.67
8	$RMSPE$	1.90	1.86	1.90	1.86	1.27	1.25	1.24	1.22	1.52	1.51	1.28	1.27	1.25	1.24
M	$T = 50$	BIC	BIC*	AIC	AIC*	AICc	AICc*	AICu	AICu*	C_1	C_1^*	Cc	Cc*	Cu	Cu*
7	(p_1, p_2)	223	340	85	168	529	624	744	800	187	292	554	645	746	812
7	d	335	337	337	320	364	365	369	371	384	380	394	403	406	419
7	(p_1, p_2, d)	99	130	40	63	225	250	296	318	107	149	252	288	326	360
7	$RMSE$	0.63	0.61	0.61	0.60	0.48	0.47	0.47	0.47	0.53	0.53	0.46	0.46	0.45	0.45
7	$RMSPE$	1.25	1.22	1.30	1.26	1.17	1.16	1.14	1.14	1.24	1.23	1.18	1.18	1.16	1.15
8	(p_1, p_2)	247	325	105	183	570	655	773	824	217	293	594	668	784	834
8	d	419	421	400	423	483	474	500	483	480	478	528	521	544	529
8	(p_1, p_2, d)	146	182	66	115	325	347	419	427	143	178	359	380	455	460
8	$RMSE$	0.84	0.84	0.82	0.82	0.68	0.68	0.68	0.68	0.72	0.72	0.66	0.66	0.65	0.66
8	$RMSPE$	1.44	1.44	1.39	1.35	1.20	1.20	1.17	1.16	1.31	1.29	1.20	1.19	1.18	1.18
M	$T = 100$	BIC	BIC*	AIC	AIC*	AICc	AICc*	AICu	AICu*	C_1	C_1^*	Cc	Cc*	Cu	Cu*
7	(p_1, p_2)	652	747	221	330	388	480	667	773	235	316	401	481	662	773
7	d	491	542	489	527	504	542	522	527	552	567	567	572	582	582
7	(p_1, p_2, d)	351	421	135	192	210	286	376	436	195	251	301	351	436	481
7	$RMSE$	0.53	0.52	0.52	0.50	0.49	0.46	0.49	0.47	0.49	0.49	0.48	0.47	0.49	0.46
7	$MSPE$	1.04	1.03	1.05	1.05	1.06	1.06	1.04	1.02	1.06	1.06	1.04	1.02	1.04	1.03
8	(p_1, p_2)	808	888	371	466	547	632	838	863	421	532	602	662	798	863
8	d	732	727	667	667	677	692	717	712	747	773	773	788	788	793
8	(p_1, p_2, d)	632	662	291	371	431	486	632	637	376	456	517	562	657	697
8	$RMSE$	0.58	0.59	0.60	0.59	0.56	0.57	0.56	0.56	0.56	0.56	0.54	0.57	0.55	0.56
8	$RMSPE$	1.04	1.03	1.05	1.05	1.04	1.04	1.04	1.03	1.06	1.06	1.07	1.05	1.05	1.05

$(p_1, p_2) = (1, 1)$, (b) the frequency detection of selecting the correct delay parameter $d = 1$, (c) the frequency detection of the correct order and delay parameter, (d) the root mean squared error of estimation of the threshold parameter and (e) the root mean square prediction error for the last observation based on the model chosen by each criteria, the fitted parameters and the true value. The results are given in Table 5. It can be seen that for small sample size, $T = 30$, the improvement in the number of times in which the correct orders (p_1, p_2, d)=$(1, 1, 1)$ are selected can be as large as 43.8 % (see AIC and AIC* in M8), for $T = 50$ as large as 74.2 % (see AIC and AIC* in M8) and for $T = 100$ as large as 42.2 % (see AIC and AIC* in M7). As in the case in which d is assumed known, the AICu, AICu*, Cu and Cu* have the larger frequency detection for the true autoregression orders and delay parameters for $T = 30$ but the frequency detection decreases when the sample size increases. We note that sometimes the corrected criteria have a shorter frequency detection of the delay parameter but this is not a drawback for them because in this simulation the interest is in detecting the true autoregressive orders and delay parameter and not only the delay parameter.

Regarding the RMSE and the RMSPE, the results are similar to the case in which d is assumed known.

A APPENDIX

Proof of Theorem 3. Shibata (1980) considers order selection criteria of the form:

$$S_T^o(p) = (T - p_{\max} + \delta_T(p) + 2p)\,\widehat{\sigma}_p^2.$$

The order chosen for the selection criteria $S_T^o(p)$ is efficient if $\delta_T(p)$ verifies the conditions imposed in Theorem 4.2 of Shibata:

1. $\operatorname*{p\,lim}_{T\to\infty}\left(\max_{1\le p\le p_{\max}} \frac{|\delta_T(p)|}{T-p_{\max}}\right) = 0,$
2. $\operatorname*{p\,lim}_{T\to\infty}\left(\max_{1\le p\le p_{\max}} \frac{|\delta_T(p)-\delta_T(p_T^*)|}{(T-p_{\max})L_T(p)}\right) = 0,$

where plim denotes limit in probability, $L_T(p)$, is the following function,

$$L_T(p) = \frac{p\sigma_a^2}{T - p_{\max}} + \sum_{i=p+1}^{\infty}\sum_{j=p+1}^{\infty}\phi_i\phi_j\Sigma_{ij}$$

where $\Sigma_{ij} = Cov\left(x_t, x_{t-|i-j|}\right)$ and p_T^* is a sequence of positive integers with $1 \le p_T^* \le p_{\max}$ which attain the minimum of $L_T(p)$ for each T (see Shibata, 1980, p.154). The AIC can be written in terms of $S_T^o(p)$

taking $\delta_T(p) = \delta_T^{AIC}(p) = T\exp\left(\frac{2p}{T}\right) - (T - p_{\max}) - 2p$. Shibata (1980) has shown that this term verifies the two conditions, and this gives the asymptotic efficiency of AIC. We can write AIC* in terms of $S_T^o(p)$ taking $\delta_T(p) = \delta_T^{AIC^*}(p) = T\exp\left(\frac{2p}{T}\right)\left(\log\left|Q_T\left(\widehat{\beta}_p\right)\right|\right)^{\frac{1}{T}} - (T - p_{\max}) - 2p$. Therefore,

$$\delta_T^{AIC^*}(p) = \delta_T^{AIC}(p) - T\exp\left(\frac{2p}{T}\right)\left(1 - \left(\log\left|Q_T\left(\widehat{\beta}_p\right)\right|\right)^{\frac{1}{T}}\right).$$

We show that $\delta_T^{AIC^*}(p)$ verifies both conditions. First we write,

$$\frac{\left|\delta_T^{AIC^*}(p)\right|}{T - p_{\max}} = \left|\frac{\exp\left(\frac{2p}{T}\right)\left(\log\left|Q_T\left(\widehat{\beta}_p\right)\right|\right)^{\frac{1}{T}}}{1 - \frac{p_{\max}}{T}} - \frac{\frac{2p}{T}}{1 - \frac{p_{\max}}{T}} - 1\right|. \tag{29}$$

Hannan (1973) shows that $\left(\log\left|Q_T(\gamma)\right|\right)^{\frac{1}{T}} \to 1$, for every γ belonging to the parametric space, and consequently, $\left(\log\left|Q_T\left(\widehat{\beta}_p\right)\right|\right)^{\frac{1}{T}} \to 1$ and the limit when $T \to \infty$ of the maximum of the values (29) in the set $1 \le p \le p_{\max}$ is 0. This proves the first condition.

For the second condition, we write the following decomposition,

$$\frac{\left|\delta_T^{AIC^*}(p) - \delta_T^{AIC^*}(p_T^*)\right|}{(T - p_{\max})L_T(p)} \le \frac{\left|\delta_T^{AIC}(p) - \delta_T^{AIC}(p_T^*)\right|}{(T - p_{\max})L_T(p)} +$$
$$+\frac{\left|T\exp\left(\frac{2p_T^*}{T}\right)\left(1 - \left(\log\left|Q_T\left(\widehat{\beta}_{p_T^*}\right)\right|\right)^{\frac{1}{T}}\right) - T\exp\left(\frac{2p}{T}\right)\left(1 - \left(\log\left|Q_T\left(\widehat{\beta}_p\right)\right|\right)^{\frac{1}{T}}\right)\right|}{(T - p_{\max})L_T(p)}.$$

Shibata (1980) showed that the first term tends to 0 implying that AIC is efficient. For the second expression, for any p such that $1 \le p \le p_{\max}$ including p_T^*, it can be shown that,

$$\lim_{T\to\infty} T\exp\left(\frac{2p}{T}\right)\left(1 - \left(\log\left|Q_T\left(\widehat{\beta}_p\right)\right|\right)^{\frac{1}{T}}\right) = -\log\left(-\sum_{i=1}^{p} i\log\left(1 - \phi_{ii}^2(\beta_p)\right)\right) < \infty.$$

As this limit is bounded for every p and $(T - p_{\max})L_T(p) \to \infty$ when $T \to \infty$, for every $1 \le p \le p_{\max}$, the second expression also tends to 0. Then, $\delta_T^{AIC^*}(p)$ verifies the second condition. Therefore, AIC* is efficient. As AICc* is asymptotically equivalent to AIC*, AICc* is also efficient. ■

REFERENCES

Akaike, H. (1969) Fitting autoregressive models for prediction. *Ann. Inst. Stat. Math.* **21**, 243-247.

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. Proceedings

of the 2nd International Symposium on Information Theory (Akademiai Kiadó, Budapest), 267-281.

Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods.* Springer-Verlag, New York, Inc.

Chan, K. S. (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21**, 520-533.

Chow, G. C. (1981) A comparison of the information and posterior probability criteria for model selection. *J. Econometrics* **16**, 21-33.

De Gooijer, J. G. (2001) Cross-validation criteria for SETAR model selection. *J. Time Ser. Anal.,* **22**, 267-281.

Durbin, J. (1960) The fitting of Time Series Models. *Rev. Inst. Int. Stat.* **28**, 3, 233-244.

Haughton, D. M. A. (1988) On the choice of a model to fit data from an exponential family. *Ann. Statist.* **16**, 342-355.

Hannan, E. J. (1973) The asymptotic theory of linear time-series models. *J. Appl. Prob.* **10**, 130-145.

Hannan, E. J. (1980) Estimation of the order of an ARMA process. *Ann. Statist.* **8**, 1071-1081.

Hannan, E. J. and Kavalieris, L. (1984) A method for autoregressive-moving average estimation. *Biometrika,* **72**, 273-280.

Hannan, E. J. and Quinn, B. G. (1979) The determination of the order of an autoregression. *J. R. Stat. Soc. B,* **41**, 190-195.

Hannan, E. J. and Rissanen, J. (1982) Recursive estimation of mixed autoregressive-moving average order. *Biometrika,* **69**, 81-94.

Hurvich, C. M. and Tsai, C. (1989) Regression and time series model selection in small samples. *Biometrika,* **76**, 297-307.

Kapetanios, G. (2001) Model selection in threshold models. *J. Time Ser. Anal.,* **22**, 733-754.

Leeuw, J. van der (1994) The covariance matrix of ARMA errors in closed form. *J. Econometrics,* **63**, 397-405.

McQuarrie, A., Shumway, R. and Tsai, C. L. (1997) The model selection criterion AICu. *Stat. Probab. Lett.* **34,** 285-292.

Poskitt, D. S. (1987) A modified Hannan-Rissanen strategy for mixed autoregressive-moving average order determination. *Biometrika,* **74**, 781-790.

Pötscher, B. M. (1990) Estimation of autoregressive moving-average order given an infinite number of models and approximation of spectral densities. *J. Time Ser. Anal.,* **11**, 165-179.

Pukkila T., Koreisha, S. and Kallinen, A. (1990) The identification of ARMA models. *Biometrika,* **77**, 537-548.

Raftery, A. E., Madigan, D. and Volinsky, C. T. (1996) Accounting for model uncertainty in survival analysis improves predictive. In Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M., eds., *Bayesian Statistics 5,* 323-349, Oxford University Press, Oxford.

Ramsey, F.L. (1974) Characterization of the partial autocorrelation function. *Ann. Statist.* **2**, 1296-1301.

Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Statist.* **6**, 461-464.

Shibata, R. (1980) Asymptotically efficient selection of the order of the model for estimating parameters of a linear process *Ann. Statist.* **8**, 147-164.

Stoica, P., Eykhoff, P., Janssen, P. and Söderström, T. (1986) Model-structure selection by cross-validation. *Int. J. Control* **43**, 1841-1878.

Tierney, L. and Kadane, J. B. (1986) Accurate approximations for posterior moments and marginal densities. *J. Amer. Stat. Assoc.* **81**, 82-86.

Tong, J. (1990) *Non-linear Time Series: A Dynamical System Approach.* Oxford: Oxford University Press.

Wong, C. S. and Li, W. K. (1998) A note on the corrected Akaike information criterion for the threshold autoregressive models. *J. Time Ser. Anal.,* **19**, 113-124.