Quaderni di Statistica
Vol. 4, 2002

A note on estimating autoregressive exponential models

Roberto Baragona
Dipartimento di Sociologia e Comunicazione, Università "La Sapienza"
E-mail roberto.baragona@uniroma1.it
Francesco Battaglia
Dipartimento di Statistica, Probabilità e Stat. Appl., Univ. "La Sapienza"
E-mail francesco.battaglia@uniroma1.it
Domenico Cucina
Dipartimento di Statistica, Probabilità e Stat. Appl., Univ. "La Sapienza"
E-mail c.domenico@email.it

Summary: The exponential autoregressive (EXPAR) models attracted much interest because they are able to account for amplitude-dependent frequency, jump phenomena, and limit cycles. In this paper we examine the estimation method proposed by Haggan and Ozaki, *Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model, Biometrika, 68, 1981, 189-196*. We are trying to improve their grid search procedure by using a genetic algorithm. Further, two entirely different procedures are presented based on indirect inference. The first one implements the calibration step by using the Gauss-Newton algorithm, the second one by using a genetic algorithm. The relative merits of the procedures are investigated by means of a simulation study. This latter shows that implementing the Haggan-Ozaki's method by means of the genetic algorithm performs better than all other procedures. Then, two well-known real data sets are considered, the Canadian lynx data and the sunspot numbers. We attempt to formulate a generalization of the EXPAR model in order to both achieve parsimony in model specification and allow for more flexibility. The estimation procedure makes use of the genetic algorithm. Both parameter estimates and multi-step out-of-sample forecasts are performed for comparison purpose.

Key words: AIC criterion; Genetic algorithms; Heuristic methods; Indirect inference; Least squares.

1. Introduction

The EXPAR models were introduced by Haggan and Ozaki (1981) as an attempt to represent time series that behave as non-linear random vibrations. An EXPAR(p) model may explicitly be written

$$y_t = \{\phi_1 + \pi_1 \exp(-\gamma y_{t-1}^2)\} y_{t-1} + \ldots + \{\phi_p + \pi_p \exp(-\gamma y_{t-1}^2)\} y_{t-p} + e_t, \tag{1}$$

with $\gamma > 0$.

It may be easily recognized that the EXPAR models belong to the class of the state dependent models (Priestley, 1988). If the parameters $(\pi_1, \ldots, \pi_p)$ are all zero, then (1) reduces to a AR(p) model (Box, Jenkins and Reinsel, 1994). Also, note that (1) may be thought of as a smooth threshold model, in the sense that, if $|y_{t-1}|$ is large, then (1) is similar to an autoregressive model with parameters approximately equal to $(\phi_1, \ldots, \phi_p)$, while, if $|y_{t-1}|$ is small, then the autoregressive parameters switch to $(\phi_1 + \pi_1, \ldots, \phi_p + \pi_p)$.

In this article, we examine some procedure for estimating the parameters in Equation (1). The ability of the EXPAR to account for limit cycles depends whether the following conditions on the parameters in Equation (1) be fulfilled (see Priestley, 1988, p. 88):
(I) All the roots of

$$z^p - \phi_1 z^{p-1} \ldots - \phi_p = 0$$

lie inside the unit circle.
(II) Some of the roots of

$$z^p - (\phi_1 + \pi_1) z^{p-1} \ldots - (\phi_p + \pi_p) = 0$$

lie outside the unit circle.

(III)

$$\frac{1-\sum_{j=1}^{p}\phi_j}{\sum_{j=1}^{p}\pi_j} > 1 \text{ or } < 0.$$

Condition (II) means that, for small y_{t-1}, the system tends to "explode" while, for large y_{t-1}, condition (I) implies that the system "damps down" towards zero. Condition (III) aims at excluding the occurrence of unstable singular points.

The paper is organized as follows. In the next Section the Haggan and Ozaki's (1981) estimation procedure is described. In Section 3 alternative methods for estimating model (1) are proposed and described in some detail. In Section 4 a generalization of model (1) is proposed to achieve parsimony as far as the number of parameters is concerned, and to improve the model's flexibility. The goodness-of-fit and the multi-step out-of-sample forecasts are compared for two well-known real time series, the Canadian lynx data and the sunspot numbers series. Conclusions are drawn in Section 5.

2. The Haggan and Ozaki's estimating procedure

A brief description of the basic procedure proposed by Haggan and Ozaki (1981) (H-O) for estimating (1) follows. It may be considered as a natural benchmark for competitive alternatives because it is quite straightforward and unlike to fail to yield a solution. It does not ensure, however, that the limit cycles conditions be fulfilled.

The algorithm requires that an interval (a, b), $a \geq 0$, be pre-specified for the γ values in (1). This interval is split in M sub-interval, so that a grid of candidate γ values is built. Let $\delta = (b-a)/M$ and $\gamma = a$. Then, for M times, the following steps are performed:

(i) Set $\gamma = \gamma + \delta$

(ii) Estimate ϕ_j and π_j by ordinary least squares regression of y_t on $(y_{t-j}, y_{t-j}\exp(-\gamma y_{t-1}^2), j = 1, \ldots, p)$.

(iii) Compute the Akaike's AIC criterion and repeat step (ii) for $p =$

$1, \ldots, P$, where P is a pre-specified integer greater than 1.
Final estimated parameters are taken that minimize the AIC.

3. Alternative methods for estimating EXPAR models

As we consider the estimation of the parameters of model (1) a difficult task, we took into account two classes of procedures that were devised to handle hard problems, that is the indirect inference and the meta-heuristic (or general heuristic) methods.

3.1 Indirect inference

The indirect inference (Gourieroux, Monfort, Renault, 1993) assumes that some parameters set $\theta = (\theta_1, \ldots, \theta_v)'$ has to be estimated for the model

$$y_t = f(y_{t-1}, y_{t-2}, ...; e_t; \theta), \tag{2}$$

where e_t is a white noise whose probability density function is known. Though model (2) is supposed easy to simulate, usual estimating methods are either not applicable or are likely to fail. Then, an auxiliary model

$$y_t = g(y_{t-1}, y_{t-2}, ...; \eta_t; \beta) \tag{3}$$

is specified. The sequence $\{\eta_t\}$ is a zero-mean white noise. Equation (3) may be preferably, though not necessarily, chosen somewhat close to model (2). The parameters $\beta = (\beta_1, \ldots, \beta_m)'$ in (3) have to be easy to estimate. It must be either $v < m$ (the over-identified case) or $v = m$ (the just-identified case).

The auxiliary model (3) may even be replaced by a sequence of statistics that are related to the model (2). As β statistics, for instance, may be assumed the correlations between y_t^r and $y_{t-\tau}^s$. Several choices of the exponents r and s and of the delay τ may be tried. We experienced, however, that the autocorrelation function from the model (1) often emulates that of an autoregressive model. Further improvements in this direction are possibly advisable if third order moments or other polyspectra related statistics are considered.

Given n observations from a time series, our implementation of the indirect inference method (II-GN in the sequel) consists of the following steps:

(i) The parameters $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_m)'$ in model (3) are estimated.

(ii) Generate, for some suitable integer h, $n \times h$ random numbers $\tilde{\mathbf{e}} = (\tilde{e}_1, \ldots, \tilde{e}_{nh})'$. This generated sequence will not be changed afterwards.

(iii) Let $\theta^{(0)}$ be an initial value for the parameters vector, and simulate $n \times h$ observations $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_{nh})'$ from the model (2) and the white noise $\tilde{\mathbf{e}}$.

(iv) By using the simulated time series $\tilde{\mathbf{y}}$, compute the estimates $\tilde{\beta} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_m)'$ in model (3).

(v) Update the parameters vector $\theta^{(0)}$ to $\theta^{(1)}$ in order to decrease the distance between $\hat{\beta}$ and $\tilde{\beta}$.

(vi) Repeat the steps (iii), (iv) and (v) by using $\theta^{(1)}$ in place of $\theta^{(0)}$ until some convergence criterion is satisfied or a pre-specified maximum number of iterations N is attained.

The step (v) is often called calibration, and the relationship $\tilde{\beta} = \tilde{\beta}(\theta)$ is exploited. Note that there is no need for any explicit functional form, provided that some mild conditions are satisfied. A typical calibration device is, for instance, the well-known Gauss-Newton algorithm for minimizing

$$S(\theta) = (\hat{\beta} - \tilde{\beta}(\theta))' \mathbf{\Omega} (\hat{\beta} - \tilde{\beta}(\theta)), \tag{4}$$

where $\mathbf{\Omega}$ is a suitable definite positive matrix.

3.2 The genetic algorithm

The meta-heuristic methods (see, for instance, Osman and Kelly, 1996) may be used as well to cope with hard estimation problems. Their use is advisable when the solution space is quite large, and estimation is to be done by maximizing a complicated function that is likely to possess several local maxima. All algorithms in this class typically incorporate some stochastic devices. Meta-heuristic methods may be considered, for instance, the tabu search, the simulated annealing and the genetic algorithms. These methods often yield very similar solutions, and we con-

fine our discussion only to genetic algorithms (see Mitchell, 1996, for an introduction), mainly because they offer a natural way to display several candidate solutions simultaneously and allow explicit interaction amongst them.

A brief account of the genetic algorithm for parameters model estimation will be given along the guidelines provided by Chatterjee and Laudato (1997). Note that any real parameter x will be approximated by a binary string c according to the formula

$$x = a + c(b - a)/(2^{\ell} - 1), \tag{5}$$

where x is assumed to belong to the interval (a, b) and ℓ is the pre-specified length of the binary string c. So, v parameters are represented by a sequence of v binary strings of length ℓ each. The steps of the genetic algorithm may be summarized as follows:

(i) A positive integer s is chosen that represents the size of the population. Unlike the common statistical meaning, in the context of the genetic algorithm the population is a subset of the generally huge set of all the elements of the space of the solutions. The initial population is randomly generated to include s candidate solutions. These solutions are called chromosomes.

(ii) For each chromosome the objective function is computed. The aim of the procedure is to optimize the objective function. This latter is called fitness function.

(iii) The chromosomes are paired, and, in each couple, the one which possesses the larger fitness function is copied into the other with probability p_s. This probability is the selection pressure, with obvious meaning, and has to be pre-specified. This method for producing the next generation is called tournament selection.

(iv) The chromosomes provided by step (iii) may exchange some of their bits by means of the crossover operator. A probability p_c is pre-specified, so that $sp_c/2$ pairs are randomly chosen. For each pair, a cutting point k, say, is randomly chosen in the interval $(1, \ell - 1)$. The bits from $k+1$ to ℓ of the first chromosome in the couple replace the corresponding ones in the second chromosome, and vice versa.

(v) For each chromosome in the population inversion may take place with pre-specified probability p_i as follows. Two cutting points are randomly chosen in $(1, \ell)$, k_1 and k_2, say, where $k_1 \leq k_2$. The bits between the two cutting points are taken in reverse order.

(vi) The last operator is mutation, and may occur with pre-specified probability p_m per bit. The bit changes from 1 to 0 or vice versa.

(vii) The steps from (ii) through (vi) are repeated, independently for each model parameter, until some pre-specified criterion is met or the maximum number of generations N is attained.

3.3 Implementing the indirect inference by a genetic algorithm

We embed the genetic algorithm into the indirect inference method, to replace the Gauss-Newton algorithm in the calibration step. Let us denote by II-GA our implementation of this method. The procedure steps (i) and (ii) are left unchanged as explained in Section 3.1. On the other hand, step (iii) has to be modified because a genetic algorithm simultaneously handles the s vectors that form the current population.

Note that the chromosomes encode the EXPAR model (1) parameters only, which has the role of model (2). So, let $(\theta_1, \ldots, \theta_v)'$=$(\phi_1, \ldots, \phi_p, \pi_1, \ldots, \pi_p, \gamma)'$. We have $v = 2p + 1$. To produce the initial population, a suitable interval has to be pre-specified for each scalar parameter. Then, s vectors $(\theta_1^{(0)}, \ldots, \theta_s^{(0)})$ are randomly generated by using the binary encoding (5). The auxiliary parameters in (3) are needed to compute the fitness function, and do not appear explicitly in the procedure. The fitness function may be written, for the i-th chromosome in the j-th generation,

$$f(\theta_i^{(j)}) = \exp(-\sum_{k=1}^{m}\{\hat{\beta}_k - \tilde{\beta}_k(\theta_i^{(j)})\}^2, \qquad i = 1, \ldots, s, \qquad j = 1, \ldots, N. \tag{6}$$

Maximizing the fitness function (6) is equivalent to minimizing (4), where we assume $\Omega = I$, the identity matrix.

The steps (iii) and (iv) in Section 3.1 specify how to compute the estimates $\tilde{\beta}$ in (6). Then, the steps from (ii) to (vii) in Section 3.2 follow. Recall that, unlike the Gauss-Newton algorithm, the genetic algorithm

evolves simultaneously s candidate solutions. Each one follows its own path, and the best found through all N generations is assumed as the one which provides us with the model (1) parameter estimates.

3.4 Implementing the Haggan-Ozaki's method by a genetic algorithm

We propose another estimation method that uses the genetic algorithm to find a suitable γ value to insert in the H-O procedure. Let HO-GA denote this procedure. In fact, once some $\gamma > 0$ is stated, and for fixed p, the estimates of the remaining parameters $(\phi_1, \ldots, \phi_p,\ \pi_1, \ldots, \pi_p)'$ in (1) are uniquely determined by ordinary least squares. So, we argue that a genetic algorithm may be designed which efficiently replaces the grid search procedure. This means that a larger set of tentative solutions may be explored in a relatively short time. If, for instance, we took the interval $(0, 10)$ such that $0 < \gamma \leq 10$, then a 10000 points grid would offer the values $0.001, 0.002, \ldots, 10$ as candidate solutions. On the other hand, encoding γ as in (5), with $\ell = 16$, say, defines the set $0.000153, 0.000305, \ldots, 10$ (numbers are approximated to 6 decimal places). Unlike the grid search, not all candidate solutions are examined. Theorems on genetic algorithms, however, make us confident that some optimal, or sub-optimal, solution will be discovered in a reasonable amount of time. The outcome from model (1) is heavily influenced by the exponential functions, that are known likely amplify even slight incorrect approximations.

The algorithm develops along the following steps:

(i) Choose a suitable large integer ℓ and an interval (a, b) where the γ values have to be searched for. Let N denote the maximum number of generations.

(ii) Randomly generate s binary strings each ℓ bits long. This is the initial population.

(iii) Decode each binary string according to (5). This provides a set of s values for γ.

(iv) For each given γ, compute the fitness function. This requires several steps to be performed.

$(iv-1)$ Regress y_t on $y_{t-1},\ldots,y_{t-p}$, $e^{-\gamma y_{t-1}^2}y_{t-1},\ldots,e^{-\gamma y_{t-1}^2}y_{t-p}$, $t=p+1,\ldots,n$, and retain the ordered estimated parameters $\hat{\phi}_1,\ldots,\hat{\phi}_p$, $\hat{\pi}_1,\ldots,\hat{\pi}_p$.

$(iv-2)$ Insert such parameters, along with γ, in (1), and compute the residuals $\hat{e}_t$, $t=p+1,\ldots,n$.

$(iv-3)$ Compute the residual variance $\hat{\sigma}_e^2=\frac{1}{n}\sum_{t=p+1}^{n}\hat{e}_t^2$.

$(iv-4)$ Set the fitness function equal to $\exp(-\hat{\sigma}_e^2)$.

(v) Perform steps from (iii) to (vi) as described in Section 3.2. The current population is obtained.

(vi) The steps from (iii) through (v) above are repeated until the maximum number of generations N is attained. Alternatively, the procedure may be stopped according whether some pre-specified criterion is fulfilled.

3.5 A simulation experiment

A simulation experiment allowed the four estimation methods to be compared. For 100 times a sequence of 5500 normally independently distributed random variates with zero mean and variance 0.001 were generated using the algorithms AS183 (Wichmann and Hill, 1982) and AS241 (Wichura, 1988). These variates were used to compute 5500 observations from the model (1) with $p=2$ and parameters $\phi_1=1.95$, $\phi_2=-0.96$, $\pi_1=0.23$, $\pi_2=-0.24$, and $\gamma=1$. These values were taken from Haggan and Ozaki (1981) and fulfill the conditions (I), (II) and (III). Initial values were set to zero, and the first 2250 observations were discarded to obtain 1000 valid observations. The remaining 2250 observations were set apart to compare the out-of-sample one-step-ahead forecasts. We held $p=2$ fixed.

For the H-O method, the interval for the γ parameter was $(0,2)$ and a grid of 10000 points was used.

In the II-GN method, we used $h=10$. The auxiliary model was chosen a just-identified AR(5). Lower and upper bounds for the parameters to vary through the procedure were assumed. The intervals were chosen $(-2,2)$, $(-1,1)$, $(-3,3)$, $(-3,3)$, and $(0,2)$ respectively. The maximum

number of iterations was set equal to 1000, but relaxation of the step in the direction of minus the gradient was allowed 100 times at most, with a factor 0.5. The algorithm stops anyway either if the sum of squares of the differences between the auxiliary parameters estimates $\hat{\beta} - \tilde{\beta}$ or the sum of the squared differences between two consecutive sets of parameters θ falls below a tolerance value ϵ, say, set to 0.000001. The iterations stop as well either if the gradient norm becomes less than ϵ or no improvement of the residual sum of squares is found greater than ϵ. The initial values to start the Gauss-Newton algorithm were chosen as follows. The initial value of the γ parameter was chosen uniformly randomly in the interval $(0, 2)$. The remaining initial parameters were computed by least squares. If the conditions (I), (II) and (III) are not fulfilled, then a new random γ is generated.

For the II-GA method, we assumed again $h = 10$ and the just-identified AR(5) auxiliary model. Also, the same intervals for the parameters were chosen. The genetic algorithm parameters were pre-specified $s = 50$, $p_s = 1$, $p_c = 0.6$ and $p_m = 0.001$. The selection pressure p_s was chosen equal to its maximum value to accelerate the convergence. We did not use the inversion operator, that is we set $p_i = 0$. In fact, in the present context the meaning of a chromosome is locus-dependent. We found that the inversion operator caused the rate of convergence of the algorithm to decrease considerably. A string of length $\ell = 16$ bits was assumed as a chromosome. The initial population was formed by randomly selecting the γ parameter and computing the remaining ones by least squares. The chromosomes that did not fulfill the conditions (I), (II) and (III) were discarded. The algorithm was allowed to complete 1000 generations. We recorded, however, the iteration where the best solution was attained.

The same genetic algorithm parameters were chosen for the HO-GA algorithm. Note that the search for the γ parameter was done amongst $2^{16} - 1 = 65535$ numbers, because zero was excluded. We recall that for the H-O algorithm a grid of 10000 numbers was used, and computations were performed for each one. The genetic algorithm did not examine all the 65535 candidate solutions, but explored the solutions space efficiently enough to yield the best γ value with high precision. On the average over 100 replications, the objective function was computed 1364 times.

For comparison, note that the objective function was computed, on the average over 100 replications, 77 times by the II-GN, 5477 times by the II-GA and obviously 10000 times by the H-O algorithm.

For all of the four methods, computations were done on the mean-deleted data. Moreover, no parameters set was accepted unless the conditions (I), (II) and (III) in Section 2 were fulfilled. In Table 1 are reported the average estimated parameters and their standard errors enclosed in parentheses.

Table 1. Average estimates for 100 replications from an EXPAR(2), 1000 observations

parameter	ϕ_1	ϕ_2	π_1	π_2	γ^2			
true value	1.95	-0.96	0.23	-0.24	1.0			
algorithm						d^2	$\hat{\sigma}^2$	mse
H-O	1.83	$-.84$	$.40$	$-.41$	1.41	72.77	34	35
	$(.27)$	$(.27)$	$(.27)$	$(.27)$	$(.43)$			
II-GN	1.84	$-.85$	$.33$	$-.37$	$.94$	45.60	43	43
	$(.09)$	$(.09)$	$(.14)$	$(.13)$	$(.59)$			
II-GA	1.86	$-.87$	$.36$	$-.42$	1.20	29.77	50	50
	$(.07)$	$(.07)$	$(.11)$	$(.19)$	$(.36)$			
HO-GA	1.86	$-.87$	$.37$	$-.38$	1.45	44.40	36	37
	$(.07)$	$(.07)$	$(.11)$	$(.11)$	$(.39)$			

Further comparison indexes are (d^2), that is the sum (multiplied by 100) of the squared differences between the true parameters (ϕ, π) and their respective estimates, the average estimated residual variance $(\hat{\sigma}^2)$, multiplied by 10000, and the average mean square errors of the one-step-ahead forecasts (mse) on out-of-sample 2250 observations, multiplied by 10000 as well. The least d^2 is yielded by the II-GA algorithm, while the least $\hat{\sigma}^2$ and mse are obtained by using the H-O algorithm. Nevertheless, the II-GA yields the largest $\hat{\sigma}^2$ and mse, and the H-O the largest standard errors of the estimates. So, from Table 1 we may see that the overall performance of the algorithm HO-GA is better than the other procedures. In fact, all figures concerned with the standard errors of the estimates,

and the indexes d^2, $\hat{\sigma}^2$ and mse, are quite small. Also, the genetic algorithm seems able to improve both the H-O and the II-GA methods with respect to the standard errors of the estimates and the d^2 index. Then, we may note that the less both bias and variance of the estimated γ, the better is the overall performance as far as the remaining parameters are concerned. For instance, the average parameter estimates from the II-GN algorithm are closer to the true values than that yielded by the H-O algorithm. Nonetheless, the variance of the γ parameter II-GN estimates is large enough to make the statistics used for comparison to be worse than that computed by the H-O algorithm. So, the accuracy of the estimates of the parameters of the EXPAR model depends considerably on the quality of the estimate of γ.

4. Application to real time series and EXPAR generalization

The EXPAR models were applied by many authors for modelling the well-known Canadian lynx data set (see, for instance, Tong, 1990, Chapter 7, p. 357-418). Haggan and Ozaki (1981) proposed an EXPAR(11) model. Computations were performed on the mean-deleted $\log_{10}$ data. We used their estimated parameters for computing the multi-step forecasts from 1921 to 1934, that is the time origin was assumed the year 1920, and lead times were $1, 2, \ldots, 14$. Then the HO-GA procedure was used to provide the parameter estimates and the multi-step forecasts were computed with the same time origin and lead times as before. In the Figure 1, the forecasts for the years 1921-1934 are reported, obtained from the Haggan and Ozaki's parameters set and from the HO-GA procedure. The forecasts are similar, but those obtained by using the latter method are generally closer to the observed values. In fact, the mean square error of the forecasts is 0.0597, while that computed for the former method is 0.0705.

Figure 1. Forecasts of the Canadian lynx data (vertical axis) for the years 1921-1934 (horizontal axis). The solid line represents the original transformed data, the dotted line the forecasts computed by using the parameters given in the Haggan and Ozaki's paper, and the dashed line our forecasts from the HO-GA procedure.

The use of genetic algorithms allows also the estimation of more elaborated models, where many parameters have to be obtained by searching methods. We suggest that model (1) may be generalized by allowing the γ parameters to be different in each term. Such generalization may be written

$$y_t = \{\phi_1 + \pi_1 \exp(-\gamma_1 y_{t-1}^2)\} y_{t-1} + \ldots + \{\phi_p + \pi_p \exp(-\gamma_p y_{t-1}^2)\} y_{t-p} + e_t. \tag{7}$$

In order to exhibit limit cycle behavior, the same conditions (I) and (II) concerned with model (1) have to hold true for the model (7). The condition (III), however, has been derived by assuming a single γ, so that it is no longer applicable for our model (7). Nonetheless, a sufficient condition for model (7) to have no unstable singular point may be stated as

(III')

$$1 - \sum_{j=1}^{p} \phi_j > \sum_{j=1}^{p} |\pi_j| \text{ or } < p\min(0, \pi^*),$$

where π^*=min($\pi_1, \ldots, \pi_p$). The proof may be readily developed along the same guidelines provided by Haggan and Ozaki (1981, p. 191). Only we have to observe that the simplified formula (1.7) displayed therein does not hold in our case, as we allow for multiple γ's. So, in deriving condition (III'), we have to consider the absolute value of the coefficients π, in the first inequality, and their least value in the second one.

We argue that, using model (7), some parsimony may be gained as far as the number of parameters is concerned. Furthermore, model (7) is likely to allow for more flexibility and adaptation to the data. Estimating model (7) is easily done by using the HO-GA method, while a grid search, in such multi-dimensional parameter context, would be much more heavy.

Only the first 100 observations were used for parameter estimation. The residual variance, that is the mean square error of the within-sample one-step-ahead forecasts, was computed. Also, we computed the AIC criterion to take the number of parameters in each model into account. Note that some caution is needed when considering the AIC in the present context, because here the likelihood is only approximately proportional to the residual variance. Further, the model is not linear in the parameters γ_j, and they have a limited range of dependence on the data. Then, we calculated the out-of-sample forecasts for lead times at $1, 2, \ldots, 14$ steps ahead. Such forecasts were compared with the observations from 1921 through 1934. The mean square error was computed on these 14 forecasts. We compared the models EXPAR(2), EXPAR(6) and EXPAR(11), with a single γ and with 2, 6 and 11 γ's respectively. By using the HO-GA method, we obtained the results listed in Table 2. As expected, by allowing for more γ's makes the residual variance to decrease. According to the AIC criterion, the best model is the EXPAR(11) with a single γ, and the EXPAR(11) with more γ's is the second best. The mean square forecasts error, however, does not exhibit this behavior, and best forecasts are yielded by the EXPAR(2) model with 2 γ's, though the EXPAR(2) model with one γ is only slightly worse. This circumstance seems to suggest that the adherence of the model to the sample data does not en-

sure more accurate multi-step forecasts. Note that we considered out-of-sample forecasts, so that the parameter estimates did not take into account the last 14 observations. These latter are better predicted by the EXPAR(2) model, in spite of the fact that the least residual variance is yielded by the EXPAR(11) model with 11 γ's. The former model forecasting ability is possibly due to the fact that it describes well the overall behavior of the time series.

Table 2. Comparison amongst some EXPAR(p) models with 1 and p γ's for the Canadian lynx data

model	residual variance	AIC criterion	multi-step forecasts mse
EXPAR(2) one γ	0.0498	-289.97	0.0437
EXPAR(2) 2 γ's	0.0479	-291.86	0.0419
EXPAR(6) one γ	0.0440	-286.36	0.1085
EXPAR(6) 6 γ's	0.0404	-284.89	0.0911
EXPAR(11) one γ	0.0296	-306.00	0.0917
EXPAR(11) 11 γ's	0.0267	-296.31	0.0818

The sunspot numbers (see Tong, 1990, Chapter 7, p. 419-429) were investigated as well by using several models and estimation methods. We made out computations on the mean-deleted transformed data $2\{(1 + y_t)^{1/2} - 1\}$ as suggested in Tong (1990, p. 420). We considered the AR(9) model reported by Tong (1990, p. 423), and the self-exciting threshold autoregressive SETAR$(2; 11, 3)$ model proposed by Ghaddar and Tong (1981, p. 247). Then, we took into account the EXPAR(2), the EXPAR(6) and the EXPAR(9) models with one γ and with 2, 6 and 9 γ's respectively. For estimating the parameters of each model we used the observations from 1700 to 1979, while the observations from 1980 to 1995 were reserved for the multi-step forecasts. The results are displayed in Table 3. Models are compared by means of the residual variance, the AIC criterion and the mean square forecasts error. Time origins are 1979, 1984, 1987 and lead times $1, 2 \ldots, 8$. We consider as well a wider time span with time origin 1979 and lead times $1, 2 \ldots, 13$.

Table 3. Multi-step forecasts mean square error for the sunspot numbers: comparison amongst AR(9), SETAR$(2;11,3)$ and some EXPAR(p) models with 1 and p γ's

model	residual variance	AIC criterion	mse 80-87	mse 85-92	mse 88-95	mse 80-92
AR(9)	4.05	409.64	3.60	16.5	9.01	16.19
SETAR $(2;11,3)$	3.73	402.59	1.82	33.51	17.34	22.27
EXPAR(2)						
one γ	4.90	454.99	7.08	65.28	31.39	32.97
2 γ's	4.83	452.96	3.77	85.33	29.32	38.46
EXPAR(6)						
one γ	4.47	445.27	7.64	54.74	19.46	21.11
6 γ's	4.34	447.00	11.85	42.01	20.62	21.89
EXPAR(9)						
one γ	3.66	401.29	4.99	20.43	8.21	13.02
9 γ's	3.57	410.32	2.62	16.34	10.65	10.27

As before, the highest order EXPAR model with a single γ exhibits the least AIC value. Such finding, however, is not in agreement with the forecasting ability of the models. In fact, the best forecasts not always are obtained by using models that have the least AIC. The SETAR$(2;11,3)$ model, for instance, provides the best multi-step forecasts for the years 1980-1987. The results change, however, if different time intervals are considered. Thus, the least mean square forecasts error is observed for the EXPAR(9) with 9 γ's in 1985-1992, for the EXPAR(9) with a single γ in 1988-1995. In the wider time span 1980-1992, the EXPAR(9) with 9 γ's is able to produce the best multi-step forecasts. The cyclical behavior of this time series is changing over time, and our models may describe it better in certain years than others. It seems that the EXPAR(9) model with 9 γ's almost always yields the most accurate forecasting performance.

5. Conclusions

We investigated four methods for estimating an EXPAR model and compared their performances on the basis of a simulation study. The method proposed by Haggan and Ozaki (1981) was used as a benchmark to evaluate our three proposal methods:

(1) An algorithm based on the indirect inference by using the Gauss-Newton algorithm for calibration.

(2) Another indirect inference based algorithm where the calibration is done by using a genetic algorithm.

(3) The Haggan and Ozaki's procedure where the search for the γ parameter was performed by a genetic algorithm.

The genetic algorithm was found able to improve both the indirect inference method and the Haggan and Ozaki's procedure. Implementing this latter with the genetic algorithm also provided the best overall performance amongst all of the four methods, and reduced the computational effort. Such improvement was observed as well as far as forecasting the Canadian lynx data by an EXPAR(11) model is concerned.

Finally, we argued that using more than a single parameter γ in the EXPAR model specification could allow for improved forecasts. For estimating the parameters of these models, it is straightforward to generalize the HO-GA procedure. The applications to the Canadian lynx data and to the sunspot numbers show that some decrease of the residual variance may be gained. Further, the mean square forecasting errors were found smaller for the EXPAR models with multiple γ's. The model order does not need be chosen too large, because even low order EXPAR models were often able to provide accurate multi-step forecasts.

Acknowledgments: Financial support from MIUR, Italy, "Stochastic models and simulation methods for dependent data," 2000, and from CNR, Italy, "Non linear models and new computational methods in time series analysis and forecasting," grant C003399, 2000, are gratefully acknowledged. We downloaded the Canadian lynx data from http://www-personal.buseco.monash.edu.au/ ˜ hyndman/TSDL (Time Series Data Library maintained by Rob Hyndman and Muham-

mad Akram). The sunspot numbers were taken from http://sidc.oma.be (RWC Belgium World Data Center for the Sunspot Index). The authors are grateful to an anonymous reviewer for helpful comments.

References

Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (1994) *Time Series Analysis: Forecasting and Control*, Prentice Hall, Englewood Cliffs.

Chatterjee, S., Laudato, M. (1997) Genetic algorithms in statistics: procedures and applications, *Communications in Statistics - Theory and Methods*, 26(4), 1617-1630.

Ghaddar, D. K., Tong, H. (1981) Data transformation and self-exciting threshold autoregression, *Applied Statistics*, 30, 238-248.

Gourieroux, C., Monfort, A., Renault, E. (1993) Indirect inference, *Journal of Applied Econometrics*, 8, S85-S118.

Haggan V., Ozaki T. (1981) Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model, *Biometrika*, 68, 189-196.

Mitchell, M. (1996) *An Introduction to Genetic Algorithms*, The MIT Press.

Osman, H. I., Kelly, J. P. (Eds.) (1996) *Meta-Heuristics: Theory and Applications*, Kluwer Academic Publishers, Boston.

Priestley, M. B. (1988) *Non-linear and Non-stationary Time Series Analysis*, Academic Press.

Tong, H. (1990) *Non-linear time series. A Dynamical System approach*, Oxford Science Publications, Clarendon Press, Oxford.

Wichmann, B. A., Hill, I. D. (1982) An efficient and portable pseudo-random number generator, *Applied Statistics*, 31, 188-190; correction, 33 (1984), 123.

Wichura, M. J. (1988) The percentage points of the normal distribution, *Applied Statistics*, 37, 477-484.