Northwest Atlantic Fisheries Organization

Serial No. N1843 NAFO SCR Doc. 90/108

SCIENTIFIC COUNCIL MEETING - SEPTEMBER 1990

Bootstrap Estimation of the Confidence Intervals of Stock and TAC Assessments with the Use of Dynamic Surplus Production Models

by

Z. I. Kizner

All-Union Research Institute of Marine Fisheries and Oceanography (VNIRO)
17 V. Krasnoselskaya, Moscow B-140, 107140, USSR

SPECIAL SESSION ON MANAGEMENT UNDER UNCERTAINTIES
RELATED TO BIOLOGY AND ASSESSMENTS

ABSTRACT

Having dot commercial fishery data and using a dynamic production model it is possible to evaluate estimates of stock abundance for every year of the period of intensive fishery (including the last one) and TACs for a few years ahead. But being dependent upon the initial fishery data, these estimates are to be regarded random ones, and therefore it is necessary to provide them by confidence intervals. Taking into account nonlinearity of the models under consideration and the fact that usually data series are sufficiently short, the only real way to get the confidence intervals is the residual (conditional) bootstrap. The corresponding procedure is described and discussed. Two hypotheses (the error is an additive or multiplicative one) and two kinds of bootstrap technique (parametric and nonparametric bootstrap) are compared.

1. INTRODUCTION

When dealing with the problem of stock and TAC assessment one has to operate sometimes with commercial fishery data which do not reflect the age structure of the exploitable population. In such a case a surplus production model can serve as a mathematical instrument of the investigation.

The initial data which are used for model fitting, especially, catch per unit effort series contain errors of different nature. Those errors can be regarded random ones. It implies that any dynamic production model for TAC forecasting should be constructed as an **observation error model** (using the terminology of Walters, 1986). Therefore the fitting procedure for a nonlinear model can not be a simple regression but must contain a certain iterative procedure providing gradual tuning of the model to best describing real stock dynamics. In such a case, any direct analytical estimation of confidence intervals of the model parameters, the stock and TAC assessments (which are random values too) can be carried out very rarely, and the bootstrap technique occurs the only real way to get the corresponding confidence intervals.

As an example, the dynamic production model with the control through fishing effort suggested by the present author (Kizner, 1989) is considered below. The model may be shown to be stable (Kizner, 1990), therefore its confidence intervals are rather narrow as compared to other versions of the model (e.g. the model with the control trough catch). The model itself is described in the next section, while the fitting procedure is presented in the section 3. The bootstrap procedure is described in the section 4 were two approaches are compared.

2. THE MODEL

Two equations expressing balance of the stock biomass and the proportion between the biomass and *cpue*:

$$B_{t+1} = B_t + G(B_t) - C_t, \tag{1}$$

$$v_t = qB_t, \tag{2}$$

will serve as a basis of the following constructions.

Here

B_t - biomass at start of the year t,

v_t - *cpue* at start of the year t,

C_t - catch in the year t,

$G(\ )$ - production function: $G(B_t) = rB_t(1-B_t/K)$ and $G(B_t) = rB_t(1-lnB_t/lnK)$ according to Schaefer and Fox respectively,

q, r, K - positive constants: q - catchability coefficient, K - carrying capacity.

Here and below in this section we operate only with 'model' (estimated) variables (except C_t and f_t).

Substitution of (2) into (1) and replacing C_t in (1) by $f_t(v_t + v_{t+1})/2$. reduces the system to one equation with respect to *cpue*:

$$v_{t+1} = v_t + qG(v_t/q) - qf_t(v_t + v_{t+1})/2,$$

which gives:

$$v_{t+1} = \frac{(1 - qf_t/2)v_t + qG(v_t/q)}{1 + qf_t/2}. \tag{3}$$

Here f_t is the fishing effort in the year t. For example, when G is the Schaefer function, the governing equation is:

$$v_{t+1} = v_t \frac{1 - qf_t/2 + r(1-v_t/qK)}{1 + qf_t/2}. \tag{4}$$

The TAC forecasts are calculated in this model as

$$TAC_{n+l} = f_{0.1}(v_{n+l} + v_{n+l+1})/2,$$

where every new *cpue* value is related to the previous one through the relationship analogous to (3). E.g., for the Schaefer-type model

$$v_{n+k+1} = v_{n+k} \frac{1-qf_{0.1}/2 + r(1-v_{n+k}/qK)}{1+qf_{0.1}/2},$$

for $k = 1,\dots,l$ (v_{n+1} is determined by (4), $f_{0.1}$ is a given control action and is determined by the model parameters).

The stock biomass estimates can be obtained using the equation (2).

3. FITTING PROCEDURE

First the initial (start) 'model' *cpue* must be evaluated as

$$v_2 = (cpue_1^{obs} + cpue_2^{obs})/2,$$

where the actual (observed) *cpue*s are provided by the marc '*obs*'

Then the first approximations of the model parameters q, r, K, must be given (the values of the parameters of the corresponding 'process error' models can be taken) and the first approximations of the estimated v_i ($i = 3, \dots, n+1$) must be evaluated through (3) (from (4), for th case of the Schaefer surplus production function).

Every next approximation of the estimates of the series $\{v_i\}$ and of the set of the model parameters must be found in the course of the iterative procedure of minimizing the functional

$$\sum_{i=2}^{n}[(v_i+v_{i+1})/2 - cpue_i^{obs}]^2 \qquad (5)$$

if the error is supposed to be additive or

$$\sum_{i=2}^{n}[\ln((v_i+v_{i+1})/2) - \ln cpue_i^{obs}]^2 \qquad (6)$$

if the error is supposed to be multiplicative.

On the output of the procedure described one has the final estimates of q, r, K, as well as v_i for $i = 3, \dots, n+1$.

4. BOOTSTRAP ESTIMATION OF THE CONFIDENCE INTERVALS

The initial data series $\{spue_i^{obs}\}$ is in fact only a sample (and usually rather short one) from any set of possible *spue* values (parent population). If a number of such samples were available, we could repeat the whole computational procedure over and over again to obtain a lot of estimates of the model parameters, of TACs and biomass values, and then using conventional statistical methods to evaluate corresponding confidence intervals. The residual (conditional) bootstrap (Efron, 1982) which is based on this very idea is actually a kind of Monte-Carlo approach to evaluating the statistical characteristics of the above mentioned estimates by means of simulation of artificial input data statistically similar to the initial data series.

Starting from the maximum likelihood principle, the residual bootstrap regards the residuals $\varepsilon_t = cpue_t^{obs} - v_t$ as being a sample representing the random component in the input data, when the error is an additive one. Supposing all ε_t equally distributed, we can take any permutation of the residuals and add every term of the new sequence to corresponding 'model' (estimated) *cpue* values to obtain a new data series (a replication) similar to but different from the initial one. In the case of the multiplicative error, the ratios $\delta_t = cpue_t^{obs}/v_t$ should be rearranged, and then every 'model' *cpue* value must be multipled by corresponding δ_t to get a new artificial data series.

The described approach is called nonparametric bootstrap.

A modification of this method, called parametric bootstrap, can be obtained by the use of a generator of pseudorandom numbers distributed just as the residuals are. Since the method of least squares is used when fitting the model, it is only natural to use a generator of normal (if the error in the initial data is supposed to be additive) or lognormal (in the case of the multiplicative error) numbers; minimization of the functionals (5) and (6) should be performed in the first and the second case correspondingly.

A comparison of different variants of the described bootstrap technique was carried out with the use of a computer program made by the author in co-operation with V.Babajan and M. Matushansky (1989). It was found that for an approximately 25-year series it is sufficient to produce about 200-250 replications to get accurate estimates of the confidence intervals. Another result is that for the present model the hypothesis of the multiplicative nature of the error and, consequently, minimization of the functional (6) and utilization of the generator of lognormal numbers are preferable.

The results of application of this approach to analysis of the Cape hake fishery (in the ICSEAF Division 1.5) can serve as illustration of the parametric bootstrap estimation of confidence intervals of the model parameters, TAC forecasts and Current stock size (see fig. and table).

REFERENCE

1. **Babajan, V.K., Kizner, Z.I. and M.V.Matushansky.** Three versions of a dynamic production model (brief comments on BKM computer program). - ICSEAF/SAC/89/S.P./44, 1989.
2. **Efron, B.** The Jackknife, the Bootstrap and Other Resampling Plans. - Society for Industrial and Applied mathematics, Philadelphia, 1982.
3. **Kizner, Z.I.** On the stability of dynamic surplus production models used in ICSEAF. Resume. ICSEAF/89/M.P./12, 1989.

4. **Kizner, Z.I.** On the stability of dynamic surplus production models used in ICSEAF. ICSEAF/SAC/90/S.P./3, 1990.

5. **Walters, C.J.** Adaptive Managment of Renewable Resources. McMillan, New York, 1986.

Table. Mane results of the computations (numbers in parentheses are the coefficients of variation, i.e. the ratios of standard errors to corresponding estimates).

```
Estimates and coefficients of variation

Model parameters
  q =      0.001412 (0.2691)
  K =   1221.294312 (0.1758)
  r =      0.554281 (0.2098)
Parameters of the equilibrum CPUE vs effort relationship
  a =      1.724625 (0.1095)   b =     0.004394 (0.1876)
f MSY strategy
  MSY         169.234924 (0.0494)
  E msy       196.257065 (0.0906)
  B msy       610.647156 (0.1758)       At start of the current year
  CPUE msy      0.862313 (0.1095)        Bt           665.309737 (0.2314)
f 0.1 strategy                           Bt/K           0.544758 (0.0945)
 E 0.1       176.631348 (0.0906)         Bt/B msy       1.089516 (0.0945)
 B 0.1       671.711853 (0.1758)
 CPUE 0.1      0.948544 (0.1095)        RY=G(Bt)     165.804106 (0.0521)

Fitting statistics
  SS =   0.330712   Residual mean =  0.005056    S.D.E. =  0.125385
```

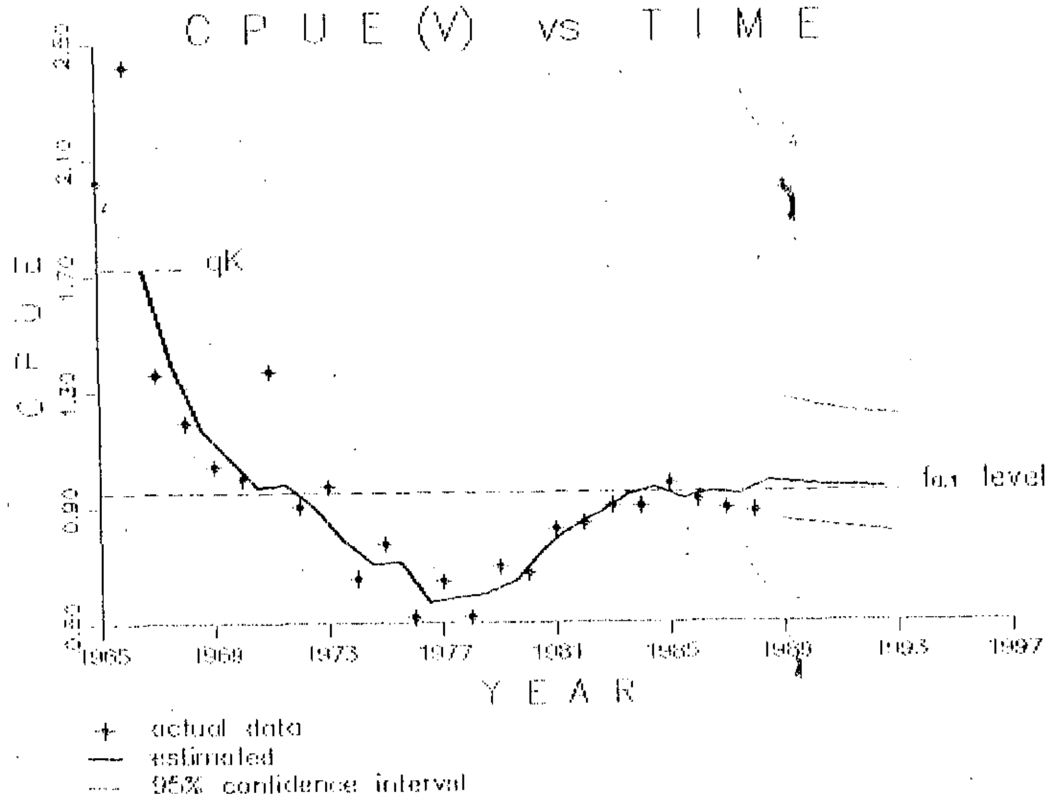

Fig. 1. Actual data, estimated *cpue* and 95% confidence intervals for forecasted TACs for Cape hake in ICSEAF Division 1.5.

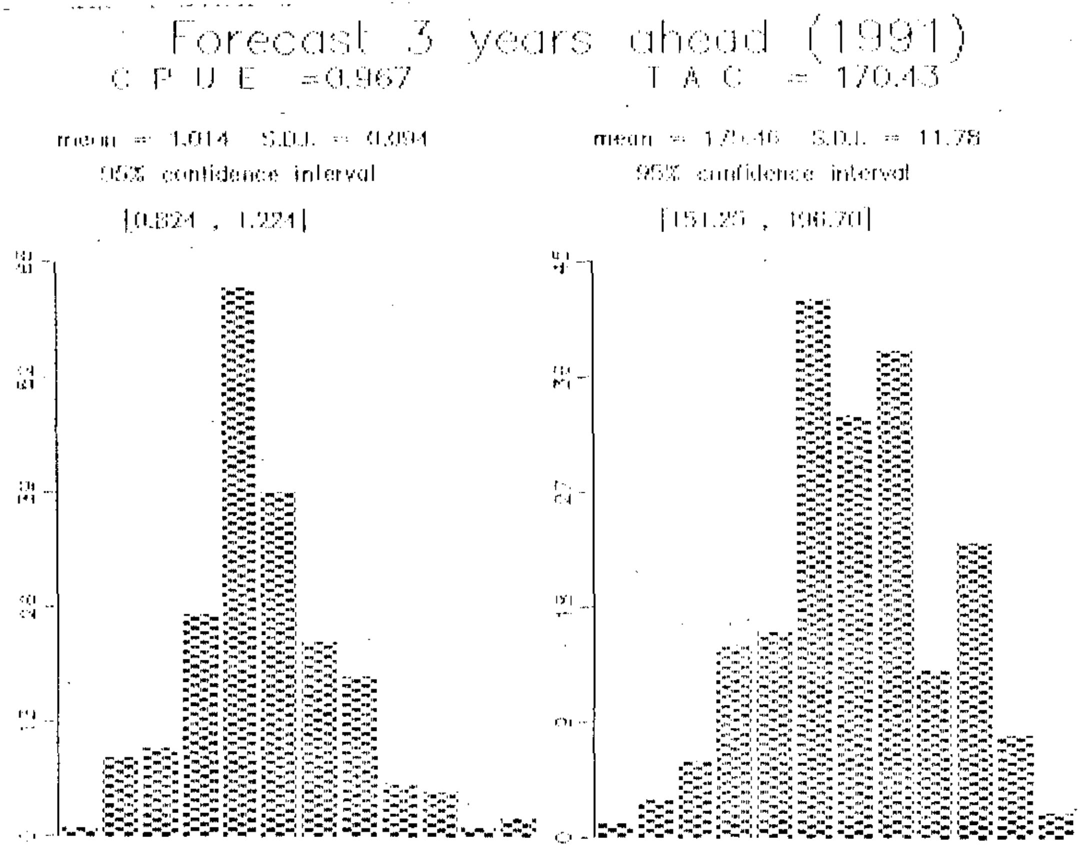

Fig. 2. Hystograms of forecasted *cpues* and TACs 3 years ahead for 200 bootstrap replications (the data are shown in fig.1).