A POSSIBLE ALTERNATIVE APPROACH FOR GENERALISED LINEAR MODEL ANALYSIS OF TUNA CPUE DATA

SCRS/1995/075

Col.Vol.Sci.Pap. ICCAT, 45 (2) : 123-124 (1996)

Butterworth, D.S.

Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7700, South Africa

SUMMARY

Key aspects of an approach currently being implemented for the South African hake fishery are briefly described and motivated. Normality is assumed for the residual error distribution, with adjustment for heteroscedasticity by application of a weighting function which is determined empirically from the relationship between residual variance and effort. The approach may merit consideration for application to CPUE data for tuna fisheries.

RESUMÉ

Le présent document décrit et commente brièvement les aspects fondamentaux d'une approche qui est actuellement appliquée dans la pêcherie de merlu d'Afrique du Sud. On suppose une distribution normale de l'erreur résiduelle, avec un ajustement pour l'hétéroscédasticité en appliquant une fonction pondérale qui est déterminée de façon empirique à partir du rapport entre la variance résiduelle et l'effort. L'application de cette approche aux données de CPUE des pêcheries thonières pourrait utilement être envisagée.

RESUMEN

Se describen y justifican los principales aspectos de un enfoque que actualmente se aplica a la pesquería sudafricana de la merluza. En la distribución del error residual se supone normalidad, con ajustes respecto a la heteroseraticidad aplicando una función de ponderación que se determina de forma empírica partiendo de la relación entre la varianza residual y el esfuerzo. Este enfoque merecería ser tenido en cuenta para su aplicación a los datos de CPUE de las pesquerías de túnidos.

THE APPROACH

An approach currently being implemented in South Africa in the analysis of CPUE data from the demersal trawl fishery for hake (the country's most important fishery) may merit inclusion among those considered for the analysis of CPUE for tuna at ICCAT. Essentially, the approach involves:

i) modelling $\ell n\ (\delta + \text{CPUE})$ as a sum of linear effects;

ii) assuming that residuals are normally distributed, but with variance related to effort, so that the regression is weighted accordingly;

iii) choosing δ to give the most normal-like distribution of residuals (usually by means of selecting for zero skewness);

iv) taking abundance to be indexed by the model-predicted CPUE, rather than the estimated (exponentiated) year factor alone (in the absence of year interactions).

THE RATIONALE

The primary rationale for the implementation of this approach in South Africa was to be able to maintain utilisation of normal error model packages, both in the interests of simplicity, and because the real overall error structure results from a combination of different processes which no single "pure" error model is likely to capture anyway. The idea is that essentially what is required is homoscedasticity for minimum variance estimation, so rather impose that empirically. Thus, with the hake fishery for example, a plot of residual variance for $\ell n\ (\delta + \text{CPUE})$ against effort (E) has hyperbolic shape, and is reasonably fitted by the form $(\alpha + \beta/E)$, which is what might be expected from the combination of environmental variation with constant CV together with Poisson-like sampling variability. The regression fit then simply includes a weight given by the inverse of this form. (One should, of course, iteratively reweight in such an exercise - to get overall MLE's for α and β - though in the case of the hake data, convergence is virtually immediate.) The potential advantage of such an approach for tuna species such as bluefin, say, would be the provision of an objective approach to the model error question, which would avoid a debate of virtual irresolvables such as whether environmental or sampling contributions to the variance are likely to dominate.

For some tuna fisheries, it could be that discreteness effects cause problems with this approach for a highly disaggregated analysis with low catch numbers. If so, one could revert to the basic Poisson(-like) model approach for which the catch number C is modelled with E as an offset, but again include weighting (assuming the computer packages can incorporate that) based on a plot of residual variance against expected C.

The prescription for δ in iii) above is intended primarily to move away from the arbitrariness of other specifications. Normality is more pertinent to confidence interval estimation (not of particular pertinence for the problem at hand), rather than central trend estimates, but it does give some feeling of security, being suggestive of estimation that is likely to be more robust.

The final point iv) above concerns the fact that if CPUE is assumed to provide an index of abundance (I^*), then strictly:

$$I^* \quad \alpha \quad \exp\text{ (year effect + other effects) - }\delta$$

rather than simply:

$$I^* \quad \alpha \quad \exp\text{ (year effect)}$$

i.e. the other effects do not then cancel out. What is being done for the hake analyses is to evaluate I^* having chosen plausible standard values for the variables associated with these other effects (vessel size, month, etc. as relevant), based on the average or maximum for the fishery as pertinent. This does not have much effect on the results of the hake analyses, but does seem logically essential if one is to work with the $\ell n\ (\delta + \text{CPUE})$ approach.