SCRS/1996/063 Col.Vol.Sci.Pap. ICCAT, 46 (2) : 228-232 (1997)

A PROCEDURE FOR USING CATCH-EFFORT INDICES IN BLUEFIN TUNA ASSESSMENTS (REVISED)

Cooke, J.G.[1]

SUMMARY

Based on recent discussions in the SCRS Working Groups, a statistical approach is presented for modelling the mean-variance relationship in catch-per-unit-effort indices. The method enables the variances and covariances of the annual indices to be estimated. Interactions between year and other factors are modelled as random effects. Example applications to western Atlantic northern bluefin tuna data are given.

RÉSUMÉ

Suite aux conclusions des Groupes de travail du SCRS, une approche statistique a été élaborée pour modéliser la relation de la variance moyenne dans les indices de CPUE. Cette méthode permet d'estimer les variances et les co-variances des indices annuels. Les interactions entre l'année et d'autres facteurs sont modélisées en tant qu'effets arbitraires. On présente également des exemples d'application aux données du thon rouge du Nord de l'Atlantique Ouest.

RESUMEN

A partir de discusiones recientes de los Grupos de Trabajo del SCRS, se presenta un enfoque estadístico para la modelización de la relación de la media de varianzas en los índices de Captura por Unidad de Esfuerzo. El método permite estimar las varianzas y covarianzas de los índices anuales. Se modelizan las interacciones entre años y otros factores como efectos aleatorios. Se facilitan ejemplos de aplicación a los datos de atún rojo del Atlántico oeste norte.

[1] Centre for Ecosystem Management Studies, Mooshof, 79297 Winden, Germany.

INTRODUCTION

The question of the appropriate choice of statistical models for the analysis of bluefin tuna catch-effort data has been discussed at recent SCRS meetings (ICCAT, 1995: p.53-543; ICCAT, 1996: section BFT-2-c). In previous years, data had usually been analyzed using a log-normal error model for catch per unit effort (CPUE), but at the 1994 SCRS meeting, the Poisson error model was used for the analysis of several western Atlantic bluefin tuna data sets, following its recommendation in an external review (NRC, 1994). At the 1995 meeting, analyses were presented (Dong and Restrepo, 1996) showing that a negative binomial error distribution provided a better match to the observed mean/variance relationship in the data than a Poisson distribution, for at least one fishery. The SCRS recommended that in future the negative binomial, which includes the Poisson distribution as a special case, should be used in preference to the assumption of a Poisson distribution. In this paper, the negative binomial distribution is used as the basis for a method of analysing catch and effort data that takes account of some further aspects of the distributional properties of catch-effort data.

DESIRABLE FEATURES OF A MODEL FOR CPUE ANALYSIS

An approach to modelling bluefin tuna catch-effort data should preferably possess the following features:

(i) Allow sufficient flexibility in the mean-variance relationship to enable the observed relationship to be approximately fitted;

(ii) Be able to handle small numbers of individual fish caught per trip in some fisheries, with many zero catches;

(iii) Provide a means for computing unique annual abundance indices even in cases where there are significant interactions between year and other factors;

(iv) Provide a means for consistent treatment of the data across different levels of aggregation;

(v) Provide a means for estimating the variances of (and hence relative weights to be applied to) values in annual abundance indices, including covariances between values;

A SUGGESTED APPROACH

An approach which meets the above requirements is to model the distribution of catches either a negative binomial distribution or an overdispersed Poisson distribution, with some covariance between catches in different data cells. The details of the approach, and an explanation of how it addresses each of the above requirements, follow:

(i) Variance-mean relationship
The assumption of a negative binomial distribution for the catch (C), implies that the variance and expectation of the catch are related by:

$$\mathrm{var}(C) = \mu E + k(\mu E)^2 \qquad E(C) = \mu E \tag{1}$$

where μ is the mean catch per unit effort and E is the effort. The parameter k is usually taken to be constant across data cells within an analysis. A slightly more general relationship is proposed here:

$$\mathrm{var}(C) = a\mu E + b\mu^2 E + c\mu^2 E^2 \tag{2}$$

In the applications in this paper we will fix either a or b, leaving only two parameters to estimate. This model for the variance of the catch implies the following mean/variance relationship for CPUE:

$$\mathrm{var}(CPUE) = (a\mu + b\mu^2)/E + c\mu^2 \tag{4}$$

where E is the effort and μ is the expected value of the CPUE. For large enough average catches, the approximate variance of ℓn(CPUE) is:

$$\mathrm{var}(\ell n(CPUE)) = (a/\mu + b)/E + c \tag{5}$$

In terms of the relationship between the variance of ℓn(CPUE) and the level of effort, this is equivalent to the relationship found by Butterworth (1995) to provide a good fit to data from the South African hake fishery. A rationale is that the effort-independent term, c, represents environmental variability while the term in 1/E represents a kind of sampling variability (random encounters with clusters of fish).

An empirical apparent variance-mean relationship could conceivably fall outside the range implied by formula (2) (i.e. variance increases less than proportional to μ, or more than proportional to μ^2). This could be caused by various special effects: for example, under-proportionality might reflect catch saturation. Where such effects occur, they have implications for the linearity of the relationship between catch rates and abundance. These go beyond the specific issue of the mean-variance relationship, and would need to be addressed accordingly.

(ii) Discreteness of catches and occurrence of zeros
This requirement is satisfied by either a negative binomial or an overdispersed Poisson distribution for catches.

(iii) Consistency across different levels of disaggregation
Whatever the error model used, GLM analyses usually rely on the assumption that the random error terms in different data cells are mutually independent. This implies that if two cells are combined, the variance of the combined cell is equal to the sum of two individual variances, and hence that the variance of the catch is proportional to the effort. This assumption is inconsistent with a variance functions involving terms in E^2. If the data suggest that there is an E^2 term in the variance function (a non-zero value for c in (2)), then the assumption of independence between data cells must be relaxed to resolve the inconsistency. A suggestion for how to model the covariance between cells is addressed under the following item.

(iv) Treatment of interactions between year and other factors
A common example of this phenomenon in fisheries is a significant year-month or year-area interaction. Due to environmental differences between years, the timing or pattern of fish movements can vary from year to year, generating such interactions. When such interactions are included in a purely fixed-effects GLM, the resulting estimate of an annual CPUE index is no longer unique. The solution proposed here is to treat these interactions as random effects.

Conditional on the values of the random effects, the variance of the catch, C_i in a given cell, i, is presumed to be proportional to the effort, E_i, and related to its conditional expectation, η_i:

$$\mathrm{var}(C_i \mid \eta_i) = a\eta_i E_i + b\eta_i^2 E_i \tag{6}$$

where η_i is itself a random variable, whose value is determined by a product of fixed and random effects:

$$\eta_i = \exp\left(\Sigma_j X_{ij} \beta_j + \Sigma_j \gamma_{j,Y(i,j)} \right) \tag{7}$$

where:
X = data matrix for the fixed-effect covariates
β = vector of fixed effects
Y = data matrix for the random-effect covariates: Y_{ij} is the level in cell i of the jth random-effect covariate
γ = matrix of random effects: for each term j, (single factor or interaction term) that is treated as a random effect, γ_{jk}, (k = 1, ..., n_j) is a set of n_j i.i.d. random variables with zero

expectation and common variance, σ_j^2, where $k = 1, \ldots, n_j$ indexes the n_j levels for term j.

The unconditional expectation and variance of the catch is given by:

$$E(C_i) = \mu_i E_i = E_i \exp(\Sigma_j X_{ij} \beta_j) \quad (8)$$

$$var(C_i) = a\mu_i E_i + b\mu_i^2 E_i + \sigma^2 \mu_i^2 E_i^2 \quad (9)$$

where

$$\sigma^2 = \Sigma_j \sigma_j^2 \quad (10)$$

(The bias corrections arising from the exponential in (7) can be absorbed into the respective coefficients and ignored).

In the case $a = 1$ and $b = 0$, if the conditional distribution of the catch is Poisson, and the distributions of the γ_{jk} are chosen so that their product has a gamma distribution, then the unconditional distribution of the catch is negative binomial. If $a > 1$ or $b > 0$, the conditional distribution is overdispersed with respect to the Poisson, and the unconditional distribution is overdispersed with respect to the negative binomial. In order to be able to make use of standard software, the two cases considered in this paper are: (i) $b = 0$, a free: the conditional distribution of C_i is overdispersed Poisson; and (ii) $a = 1$, b free: the conditional distribution of C_i is negative binomial.

The random effects are not explicitly observable or estimable, but manifest themselves as covariance between the catches in different cells:

$$V_{ij} = cov(C_i, C_j) = E_i E_j \mu_i \mu_j \Sigma_{k:Y(i,k)=Y(j,k)} \sigma_k^2 \quad (i \neq j) \quad (11)$$

It could conceivably occur that interactions between year and other effects are not fully explained as random effects, because they show a significant trend: after fitting a random-effect term for the interaction, there may remain a significant interaction between time (year as a quantitative variate) and another factor. The choice of course of action in such a case is not merely a statistical issue but is also a matter of interpretation. For example, if there is a significant trend with time in the relative performance of two fleets, this could be interpreted as evidence that one of the fleets has become more efficient: in this case, the interaction term between time and this fleet should be included in the model.

(v) Variances and covariances of values within an annual abundance index
An annual abundance index is generated by including year as a fixed-effect, categorical covariate in the GLM. The variance of the annual abundance index is the sum of a component due to estimation variance and a component due to process variance. It can be calculated as follows (Cooke and Lankester, 1995):

The variance-covariance matrix for the estimation component of the variance does not always exist but its inverse, the information matrix for the year factors, does exist. We use the following notation:

Four variance models were considered for each data set:
I: $a \geq 1, b = 0, c = 0$ (overdispersed Poisson)
II: $a = 1, b \geq 0, c = 0$ (negative binomial)
III $a \geq 1, b = 0, c \geq 0$ (overdispersed Poisson with random effects)
IV: $a = 1, b \geq 0, c \geq 0$ (negative binomial with random effects)

In the case of the large fish, model I co-incided with model II, and model IV co-incided with model II, because the a and b parameter estimates were at the constraints. Hence only six sets of results are listed in Table 2. The estimated annual abundance indices are plotted in Figs 1-6.

DISCUSSION

In each case, the random effects model indicates substantial area-year and area-month-year interactions, but little month-year interaction, hence the month-year interaction term was dropped. The estimated variance of the annual abundance indices in the models without the c parameter are unrealistically small: the assumption of independence between trips is violated. When the c parameter is included, the variances are substantially greater and no obvious outlying points are apparent. Although the results given here are for illustrative purposes only, it is clear that failure to take account of variance components at levels above the individual trip yields annual abundance indices with misleadingly low variances. It appears that the catch effort data series examined here are considerably less informative about trends in abundance than previous analyses have suggested.

REFERENCES

Butterworth, D.S. 1995. A possible alternative approach to generalised linear model analysis of tuna CPUE data. SCRS/95/75.

Dong, Q. and Restrepo, V. 1995. Notes on the Poisson error assumption made to estimate relative abundance of West Atlantic bluefin tuna. ICCAT SCRS/95/88.

Cooke, J.G. and Lankester, K. 1995. Consideration of statistical models for catch-effort indices for use in tuning VPA's. ICCAT SCRS/95/77.

ICCAT. 1995. Report for biennial period, 1994-5. Part I (1994) - Vol. 2. English version. Madrid, 1995.

ICCAT. 1995. Report of the SCRS. BFT-section.

National Research Council. 1994. *An assessment of Atlantic Bluefin Tuna*. National Academy Press, Washington, D.C. 148pp.

Nelder, J.A. and Wedderburn, R.W.M. 1972. Generalized linear models. *J. R. Statist. Soc.* A 135(3):370-84.

A: information matrix for the year factors
B: information matrix for all the fixed-effect parameters
J: identifier matrix for the year factors: $J_{ij} = 1$ if the jth fixed effect parameter in the GLM is the ith year factor, 0 otherwise
K: identifier matrix for the remaining (nuisance) factors: $K_{ij} = 1$ if the jth fixed effect parameter is the ith nuisance parameter, 0 otherwise.
V: variance-covariance matrix of the observations (from formulae 9 and 11)
X: data matrix for the fixed effect parameters: $X_{ij} = 1$ if observation i is affected by the jth parameter.

Then we have:

$$B = X^T V^{-1} X \qquad (12)$$

$$A = J B J^T - J B K^T (K B K^T)^- K B J^T \qquad (13)$$

where $()^-$ denotes pseudo-inverse.

Given a value for the process variance, σ_P^2, of the annual abundance index, the combined information (process plus estimation) matrix of the abundance index is given by:

$$A^* = A (I + \sigma^2_P A)^{-1} \qquad (14)$$

σ^2_P can in principle be estimated from the residual deviance between the abundance index and the relevant component of a fitted stock simulation. A drawback of this approach is that if the stock simulation model is a poor reflection of the true stock dynamics, σ^2_P will be overestimated.

When there are multiple abundance indices, one has the choice of estimating a common value for σ^2_P, or estimating a separate value for each series. If the different indices exhibit different trends, then the use of a common value for σ^2_P will tend to yield a fitted stock simulation that is a compromise between the indices. Fitting a separate σ^2_P for each index will tend to result in a fitted stock simulation that fits the majority of the indices: if there is a 'rogue' index that shows very different trends from the majority of indices, it will receive little weight.

FITTING THE MODEL

The fixed effects part of the model was fitted using the Newton-Raphson iteration on the log-likelihood, which is equivalent to an iterative weighted least-squares fit (Nelder and Wedderburn, 1972). The random effects part was fitted using the method of steepest descent. The combined model was fitted by alternating successive iterations of the fixed and random fitting procedures. Typically rather more iterations were required when random effects were included.

APPLICATION TO THE US BLUEFIN ROD & REEL CATCH-EFFORT DATA

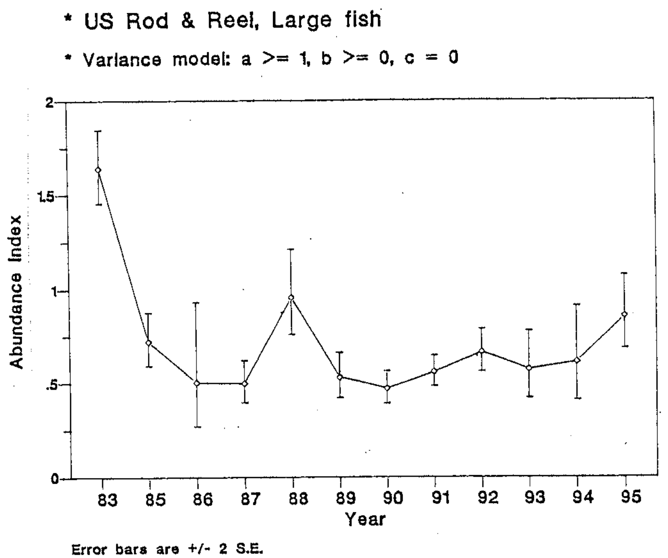
To illustrate the approach, it was applied to the US Rod & Reel bluefin catch-effort data for small and large fish respectively. Table 1 lists the factors included in the regression analysis.

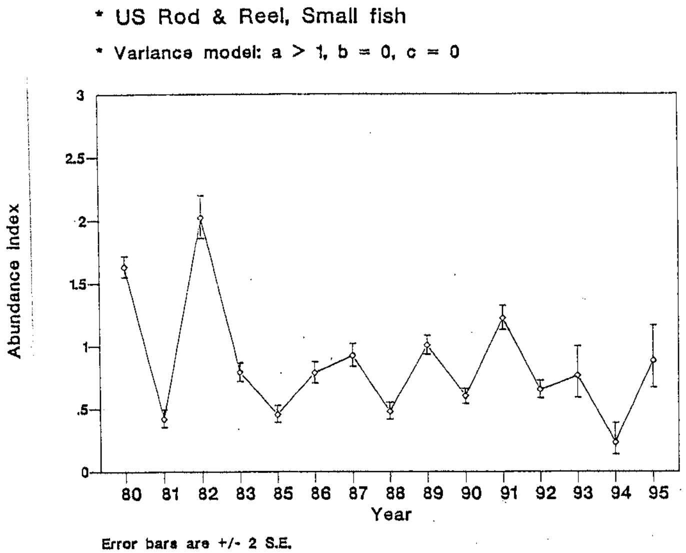
Table 1. Factors included in regression analysis

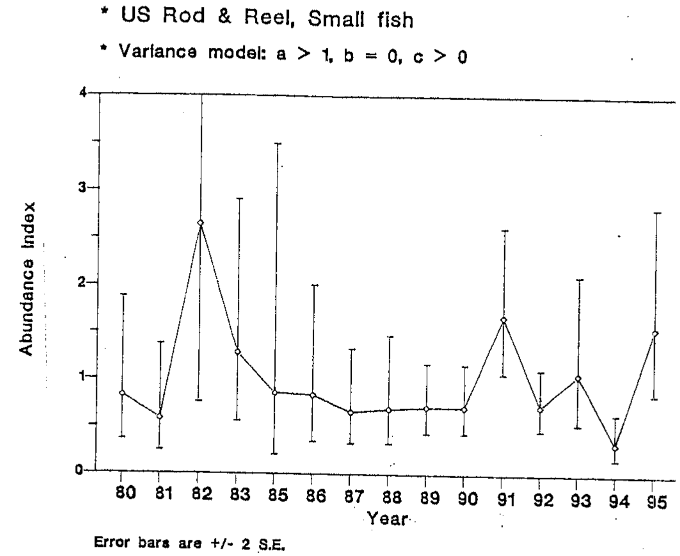
Effort measure (offset):	Line-hours	
Unit of data:	Fishing trip	
Dependent variable:	Catch (numbers of fish)	
Assumed distribution:	Overdispersed Poisson or negative binomial	
Fixed-effect factors:	*Large fish*	*Small fish*
Year	1983, 1985-95	1980-83, 1985-95
Month	Jul-Sep (3)	Jun-Oct (5)
Phone vs dockside interview	(2 levels)	
Boatype: private or charter	(2 levels)	
Target:	bluefin or other	bluefin only
Random effects (variance models with c > 0 only)		
Area.Year		
Area.Month.Year		

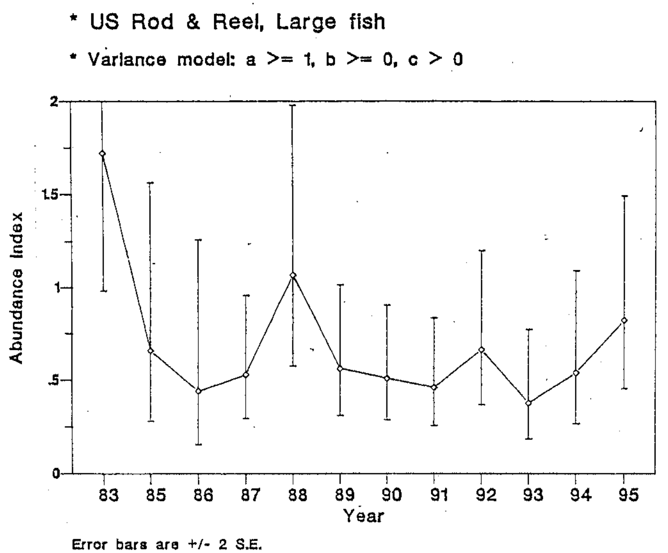
TABLE 2. Estimates of variance parameters, random effects, and year factors

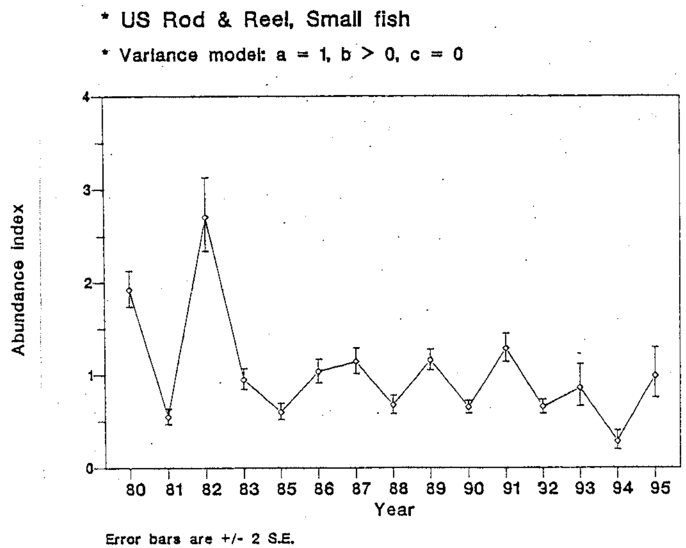
Fish	Large	Large	Small	Small	Small	Small
Variance model	a ≥ 1 b ≥ 0 c = 0	a ≥ 1 b ≥ 0 c ≥ 0	a ≥ 1 b = 0 c = 0	a = 1 b ≥ 0 c = 0	a ≥ 1 b = 0 c ≥ 0	a = 1 b ≥ 0 c ≥ 0
Variance parameters a	1.000	1.000	5.450	1	4.590	1
b	0.000	0.000	0	69.5	0	55.3
c	0	0.309	0	0	0.707	0.899
Random effects (s.d. of log effect)						
Area.Year		0.459			0.583	0.642
Area.Month.Year		0.314			0.606	0.698
Abundance index (log year factors)						
1980			0.489 (0.025)	0.654 (0.051)	-0.196 (0.422)	-0.090 (0.476)
1981			-0.871 (0.086)	-0.604 (0.077)	-0.537 (0.434)	-0.399 (0.487)
1982			0.705 (0.042)	0.994 (0.073)	0.971 (0.638)	1.132 (0.722)
1983	0.494 (0.060)	0.542 (0.286)	-0.233 (0.048)	-0.049 (0.060)	0.245 (0.419)	0.437 (0.482)
1984						
1985	-0.326 (0.099)	-0.417 (0.440)	-0.788 (0.075)	-0.515 (0.073)	-0.160 (0.720)	0.085 (0.797)
1986	-0.689 (0.316)	-0.824 (0.539)	-0.238 (0.055)	0.032 (0.061)	-0.184 (0.449)	-0.020 (0.500)
1987	-0.698 (0.114)	-0.635 (0.302)	-0.077 (0.049)	0.134 (0.060)	-0.416 (0.357)	-0.602 (0.383)
1988	-0.042 (0.119)	0.065 (0.314)	-0.736 (0.072)	-0.391 (0.075)	-0.367 (0.384)	-0.285 (0.422)
1989	-0.638 (0.115)	-0.580 (0.303)	0.006 (0.037)	0.147 (0.048)	-0.336 (0.252)	-0.338 (0.287)
1990	-0.759 (0.092)	-0.677 (0.294)	-0.511 (0.051)	-0.427 (0.053)	-0.340 (0.251)	-0.323 (0.290)
1991	-0.584 (0.074)	-0.776 (0.303)	0.200 (0.040)	0.252 (0.060)	0.514 (0.228)	0.688 (0.281)
1992	-0.409 (0.087)	-0.411 (0.302)	-0.426 (0.055)	-0.420 (0.058)	-0.327 (0.220)	-0.228 (0.263)
1993	-0.559 (0.157)	-0.978 (0.368)	-0.267 (0.135)	-0.149 (0.132)	0.065 (0.349)	0.039 (0.384)
1994	-0.490 (0.203)	-0.619 (0.360)	-1.469 (0.263)	-1.268 (0.182)	-1.096 (0.340)	-1.105 (0.330)
1995	-0.153 (0.115)	-0.196 (0.304)	-0.123 (0.141)	-0.007 (0.138)	0.443 (0.305)	0.506 (0.367)

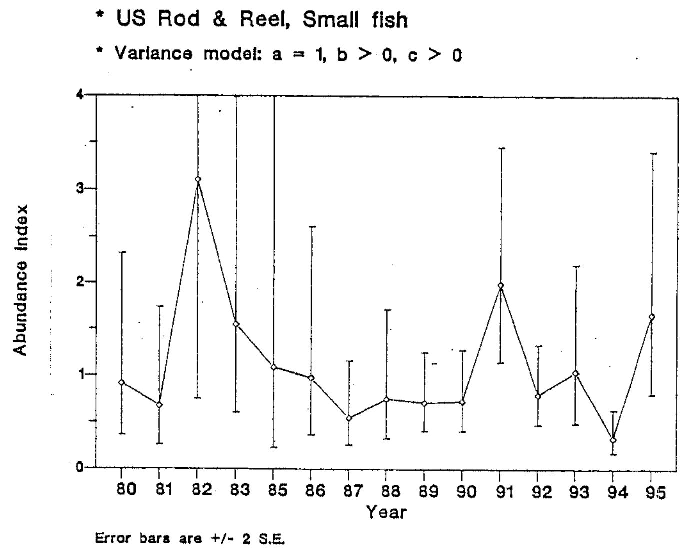

* US Rod & Reel, Large fish
* Variance model: a >= 1, b >= 0, c = 0
Abundance Index
Year
83 85 86 87 88 89 90 91 92 93 94 95
Error bars are +/- 2 S.E.


* US Rod & Reel, Small fish
* Variance model: a > 1, b = 0, c = 0
Abundance Index
Year
80 81 82 83 85 86 87 88 89 90 91 92 93 94 95
Error bars are +/- 2 S.E.


* US Rod & Reel, Small fish
* Variance model: a > 1, b = 0, c > 0
Abundance Index
Year
80 81 82 83 85 86 87 88 89 90 91 92 93 94 95
Error bars are +/- 2 S.E.


* US Rod & Reel, Large fish
* Variance model: a >= 1, b >= 0, c > 0
Abundance Index
Year
83 85 86 87 88 89 90 91 92 93 94 95
Error bars are +/- 2 S.E.


* US Rod & Reel, Small fish
* Variance model: a = 1, b > 0, c = 0
Abundance Index
Year
80 81 82 83 85 86 87 88 89 90 91 92 93 94 95
Error bars are +/- 2 S.E.


* US Rod & Reel, Small fish
* Variance model: a = 1, b > 0, c > 0
Abundance Index
Year
80 81 82 83 85 86 87 88 89 90 91 92 93 94 95
Error bars are +/- 2 S.E.