Northwest Atlantic Fisheries Organization

Serial No. N1937

NAFO SCR Doc. 91/54

SCIENTIFIC COUNCIL MEETING - JUNE 1991

On Sample Selection when Standardizing Fishing Effort with Multiplicative Model

by

P. S. Gasuikov

Atlantic Scientific Research Institute of Marine Fisheries
and Oceanography (AtlantNIRO), 5, Dm. Donskoy Str.
Kaliningrad 236000, USSR

INTRODUCTION

An analysis of residues has been made for the fishing effort standardization when using multiplicative model in practice (Robson W.E.,1966), (Gavaris S.,1980). In certain program realization of model algorithm special blocks for its accomplishment have been provided. We may refer, as an example, to the complex by S.Gavaris and D.Gascon developed APL language for IBM PC computers.

An analysis of residues has allowed for singling out the anomalous observations for their future removal from the sample and recalculation. However, the data which contain a lot of errors are not always revealed in large values of residues. In addition, visual separation of anomalous observations is of a subjective character. Therefore, a quantitative approach is advisable for this purpose.

BASIS FOR APPROACH TO SAMPLE SELECTION

An approach suggested by Huber (Huber P.,1984) is the basis of paper's methodological part. As it is known, the determination of multiplicative model parameters comes to the definition of parameters of multiplicative regression equation (Robson W.E.,1966) and (Gavaris S., 1980).

Multiplicative model of fishing effort standardization in its general form is described by the following formula

$$U = U_R \cdot P_{1J_1}^{X_{1J_1}} \cdot P_{2J_2}^{X_{2J_2}} \cdot P_{TJ_T}^{X_{TJ_T}} \qquad (1)$$

where U is a standardized value of catch rate;

U_R is a value of catch rate which is characteristic of a definite category combinations selected as a reference point;

P_{ij_i} is a relative power of j_i category in i categories type;

X_{ij} is 1, if u refers to j category in i categories type i;
is 0 in other cases;

T_k is number of categories of different types.

This expression acquires the following form after the transformation

$$\mathrm{LN}\ U = \mathrm{LN}\ U_R + \sum_{i=1} \mathrm{LN}\ P_{ij_i} \cdot X_{ij_i} \qquad (2)$$

where $\ln u_l$ is a dependant variable;
l is an index number in the sample, $l=1,\ldots,M$;
X_{ij} are independent variables.

Model parameters satisfy at the same time the following constraints:

$$\sum_{i=1}^{K_i} \mathrm{LN}\ P_{ij_i} = 0\ ,\ i=1,2,\ldots,T \qquad (3)$$

where J_i is number of categories in i categories type.

A routine scheme of determination for this equation parameters in matrix symbols can be represented as follows (Draper N.R.,Smith H.,1966):

$$\alpha = (X^T \cdot X)^{-1} \cdot X^T \cdot Y \qquad (4)$$

where α is equal to $(U_R,\ldots,P_{ij}\ldots)$ and is parameters vector;
Y is column of a dependent variable $\ln U_1, \ln U_2, \ldots, \ln U_M$;
T is transpose sign,
being estimated calculated values of $\ln U$ by the formula

$$\hat{Y} = X \cdot (X^T \cdot X)^{-1} \cdot X \cdot Y = H \cdot Y\ , \qquad (5)$$

$$H = X \cdot (X^T \cdot X)^{-1} \cdot X. \qquad (6)$$

According to P.Huber terminology (Huber P.,1984), (6) it *is matrix fitting:*

$$Y_i = \sum_k h_{ik} \cdot y_k\ ,$$

$$D^2(y_i) = \sum_k h_{ik} \cdot \sigma^2 = h_i \cdot \sigma^2,$$

wrehe $h_i = h_{ii}$,

σ^2 is variance of residues.

If one takes into account that $0 \leqslant h_i \leqslant 1$,

$$D^2 (y_i - \hat{y}_i) = (1 - h_i) \cdot \sigma^2.$$

The latter expression may be rewritten as follows:

$$r_i = y_i - \hat{y}_i = (1 - h_i) - \sum_{k \neq i} h_{ik} \cdot y_k .$$

Basing on the (8) formula, it might be concluded that the nearer y_i to the unity the lesser variance of residues.However, this is not the case. P. Huber investigates statistical characteristics of estimates for such a case when $(n+1)$-th element is added to the sample of n elements. Then (Huber P.,1984)

$$D^2 (\hat{y}_{n+1}) = h_{n+1} \cdot \sigma^2 = \frac{x^T \cdot x}{1 + x^T \cdot x} \cdot \sigma^2 .$$

Putting $\hat{\hat{y}} = x^T \cdot \alpha$, calculated value of y_{n+1} element for the sample of n value, one may get

$$D^2 (\hat{y}) = \frac{h_{n+1}}{1 - h_{n+1}} \cdot \sigma^2,$$

which means that $D^2 (\hat{y}) > D^2 (y_{n+1})$, if $h_i > 0.5$. It should simultaneously be taken into account that

$$r_{n+1} = y_{n+1} - \hat{y}_{n+1} = (1 - h_{n+1}) \cdot (y_{n+1} - \hat{y}_{n+1}).$$

As P.Huber writes, "if h_i is close by its value to the unity, a large error in y_i is not necessary revealed in r_i. It may be revealed elsewhere, for instance, in r_k, if h_{ki} appears to be large enough" (Huber P,1984).

Thus, h_i diagonal elements of fitting matrix contain useful information basing on which one may judge on the importance of i-th observation in the model. In particular, "...$h_i \leqslant 0.2$ value looks like a reliable one;the values within 0.2 - 0.5 seem to be speculative ones, and if there exists a possibility of model managing, it is better to avoid values which are greater than 0.5" (Huber P.,1984).

THE DATA USED

Let us apply the above-mentioned results to the solution of catch rate standardization problem when silver hake fishing in NAFO 4WVX Subareas. The initial information for 1977-1989 is given to solve the problem in the paper by Gasiukov (Gasiukov P.,1990). The data for 1990 are given in Table 1. The information contains data on catch and fishing effort specified by countries, vessel types, fishing years, months within a year and some other features, as well (for instance,observational routine).

CALCULATION RESULTS

A number of calculation was done to illustrate the methods suggested. The first one totally corresponds to a routine approach. The results are represented in Tables 2,3 and 4.

Basing on this calculation, a fitting matrix (4) was obtained for the model (2). Diagonal elements are represented in Table 5, being remained the numbering of sample elements in accordance with table 1 in (Gasiukov P., 1990) and table 1 of this paper.

As one may see, the values of H matrix diagonal elements are, as a rule, less than 0.2. A share of such elements is 93.4%. h values for the remainder of elements are within the interval of $0.2 \leq h_i \leq 0.5$ and therefore corresponding to Huber classification refer to speculative values.

Thus, a number of elements are advisable to be removed from the sample to increase the determinative reliability of multiplicative model parameters and then the model parameters are to be redetermined.

The second calculation was made by the sample that there had been removed from the elements for which diagonal elements of fitting matrix appeared to be more than 0.2. In tables 6,7 and 8 relevant results are given.

Residues in the model (2) for the first and second calculations are represented in figures 1 and 2, respectively. As it should be expected, the cloud of points is more solid for the second version. However, the points which are visually classified as "gross" residues were not totally removed with the procedure suggested. In particular, the point marked as 71 is worthy of note. This figure corresponds to elements number in the initial sample. Diagonal element of fitting matrix is 0.187 for the same figure that satisfies the criterion of this element keeping in the second calculation.

In this connection, it may be supposed that the criterion [0.2,0.5] requires greater degree of flexibility. In particular, it may be suggested that visual analysis of residuels be made after the first sample selection, and in the cases similar to the 71 element of sample the interval lower boundary be somewhat changed.

The third calculation was done on having determined 0.185 value as the interval lower boundary. According to this criterion it is necessary to remove from the sample already 16 elements with the following numbers: 1, 7, 24, 30, 43, 44, 52, 71, 73, 77, 87, 123, 132, 143, 156 and 162.

In tables 9,10 and 11 corresponding results obtained after the sample selection of such a criterion are shown. Residuels obtained with these model version are represented in figure 3. In this case cloud of points does not obviously contain the points which could be classified as "gross" residuels.

It is worth noting the method suggested unlike visual selection allows for defining and removing from the sample the points which do not seem to be the anomalous measurements. However, their contribution to the model tuning can substantially affect the results.

Plots of variations in standardized values of catch rate for three versions are represented in figure 4. Not only absolute values are changed when sample selecting, but values of increments, as well.

In some years these variations are rather significant.

Relative increment values are given below.

Year	77	78	79	80	81	82
Sample 1	0.20	-0.22	0.31	-0.14	-1.45	0.51
Sample 2	0.19	-0.23	0.30	-0.19	-1.39	0.51
Sample 3	0.28	-0.15	0.25	-0.18	-1.41	0.50

Year	83	84	85	86	87	88
Sample 1	-0.64	0.13	-0.91	0.01	0.30	-0.49
Sample 2	-0.64	0.13	-1.02	0.08	0.30	-0.44
Sample 3	-0.62	0.13	-0.55	0.03	0.32	-0.44

Variations in increment values between the first and the third version of calculation are especially noted in 1977 and 1985.

Such variations can evidently affect the results of VPA tuning, as well, if standardized values of fishing effort are simultaneously used.

CONCLUSION

It has been suggested to use values of fitting matrix diagonal elements which can be calculated when determining multiplicative model parameters for an objective selection of sample elements to standardize fishing effort.

It is recommended to remove from the sample in accordance with Huber's criterion elements h_i values of which are more than 0.5. The same h_i values allow for singling out sample elements which are of "higher noise". These are to be elements the values of fitting matrix diagonal elements for which are within the interval of 0.2-0.5. Visual analyses of residues can help to ditermination of a lower interval boundary in a more reliable manner.

REFERENCES

1. Robson W.E.1966.Estimation of the relative fishing power of individual ships. ICNAF res. Bull. 3:5-14.
2. Gavaris S. 1980. Use of the multiplicative model to estimate catch rate and effort from commercial data. Can. J. Fish. Aquat. Sci. 37:2272-2275.
3. Draper N.R.,Smith H.1966. Applied regression analysis.John Wiley, New York,p.391.
4. Gasiukov P.S. 1990. Application of multiplicative Model for fising Effort Standardisation in a Spatial Case. NAFO SCR Doc. 90/50:14.

Table 1. Comertial silver hake catch ahd effort for multiplicative model (1990)

nn	Вылов	Усилие	Источн.	Месяц	Год	Район	Реж.	Стр.
156	714	132	2	3	90	460	2	2
157	5546	1708	2	4	90	460	2	2
158	2553	1765	2	5	90	460	2	2
159	1148	[illegible]	2	6	90	460	2	2
160	1704	576	2	7	90	460	2	2
161	51	58	2	8	90	460	2	2
162	374	189	2	3	90	460	2	1
163	20555	7246	2	4	90	460	2	1
164	13467	8306	2	5	90	460	2	1
165	8125	4950	2	6	90	460	2	1
166	3378	1171	2	7	90	460	2	1
167	597	360	2	8	90	460	2	1

Table 2.Statistical characteristics of catch rate standardization for silver hake fishing in the NAFO 4VWX subareas (sample 1)

REGRESSION OF MULTIPLICATIVE MODEL

MULTIPLE R............. .744

MULTIPLE R SQUARED..... .553

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	DF	SUMS OF SQUARES	MEAN SQUARES	F-VALUE
INTERCEPT	1	7.689E0001	7.689E0001	
REGRESSION	24	2.614E0001	1.089E0000	7.316
TYPE 1	1	8.182E-001	8.182E-001	5.494
TYPE 2	6	6.010E0000	1.002E0000	6.727
TYPE 3	13	1.450E0001	1.115E0000	7.490
TYPE 4	2	6.042E-001	3.021E-001	2.029
TYPE 5	1	1.261E0000	1.261E0000	8.467
TYPE 6	1	9.544E-001	9.544E-001	6.409
RESIDUALS	142	2.115E0001	1.489E-001	
TOTAL	167	1.242E0002		

Table 3.Coefficients of multiplicative model

(sample 1)

REGRESSION COEFFICIENTS

CATEGORY	CODE	VARIABLE	COEFFICIENT	STD. ERROR	NO. OBS.
1	1	INTERCEPT	0.650	0.197	167
2	5				
3	77				
4	450				
5	2				
6	2				
1	2	1	-0.342	0.146	85
2	3	2	0.670	0.247	3
	4	3	0.234	0.103	24
	6	4	-0.119	0.088	43
	7	5	-0.165	0.093	35
	8	6	-0.350	0.111	21
	9	7	-0.472	0.194	5
3	78	8	-0.228	0.129	26
	79	9	-0.024	0.131	21
	80	10	-0.388	0.165	9
	81	11	-0.256	0.166	9
	82	12	0.649	0.187	7
	83	13	-0.059	0.180	8
	84	14	0.438	0.179	8
	85	15	0.300	0.179	8
	86	16	0.954	0.202	10
	87	17	0.946	0.205	9
	88	18	0.592	0.204	10
	89	19	0.989	0.195	13
	90	20	0.451	0.201	12
4	460	21	-0.166	0.142	121
	470	22	-0.275	0.147	36
5	1	23	0.455	0.156	21
6	1	24	0.216	0.085	127

Table 4.Standardized catch rate values for silver hake (sample 1)

PREDICTED CATCH RATE

STANDARDS USED VARIABLE NUMBERS: 1 5 450 2 1

YEAR	TOTAL CATCH	PROP.	CATCH RATE MEAN	S.E.	EFFORT
77	37095	0.703	2.520	0.449	14719
78	48404	0.879	2.015	0.305	24023
79	51751	0.827	2.463	0.421	21015
80	44525	0.920	1.703	0.341	26148
81	44599	0.833	1.950	0.357	22870
82	60207	0.958	4.786	1.022	12580
83	35837	0.921	2.360	0.493	15185
84	74266	0.967	3.882	0.815	19132
85	75480	0.981	3.379	0.709	22340
86	82689	0.427	6.463	1.523	12794
87	61704	0.926	6.409	1.523	9628
88	74482	0.879	4.501	1.060	16548
89	86729	0.985	6.704	1.523	12937
90	60000	0.974	3.908	0.918	15355

AVERAGE C.V. FOR THE MEAN: .207

Table 5. Diagonal elements of fitting matrix.

1	0.192	43	0.206	85	0.168	127	0.144
2	0.111	44	0.215	86	0.120	128	0.148
3	0.107	45	0.160	87	0.256	129	0.147
4	0.107	46	0.149	88	0.123	130	0.139
5	0.100	47	0.139	89	0.134	131	0.162
6	0.109	48	0.143	90	0.134	132	0.194
7	0.243	49	0.161	91	0.152	133	0.150
8	0.125	50	0.149	92	0.122	134	0.144
9	0.114	51	0.183	93	0.135	135	0.139
10	0.108	52	0.193	94	0.127	136	0.138
11	0.113	53	0.171	95	0.147	137	0.129
12	0.162	54	0.170	96	0.124	138	0.127
13	0.155	55	0.171	97	0.142	139	0.145
14	0.145	56	0.169	98	0.157	140	0.152
15	0.102	57	0.159	99	0.140	141	0.137
16	0.097	58	0.157	100	0.135	142	0.140
17	0.099	59	0.158	101	0.138	143	0.412
18	0.103	60	0.158	102	0.147	144	0.122
19	0.132	61	0.157	103	0.161	145	0.114
20	0.112	62	0.157	104	0.178	146	0.111
21	0.101	63	0.181	105	0.176	147	0.117
22	0.097	64	0.158	106	0.181	148	0.125
23	0.112	65	0.157	107	0.165	149	0.122
24	0.203	66	0.157	108	0.159	150	0.118
25	0.107	67	0.181	109	0.157	151	0.128
26	0.096	68	0.134	110	0.162	152	0.141
27	0.098	69	0.184	111	0.174	153	0.126
28	0.105	70	0.139	112	0.158	154	0.117
29	0.125	71	0.187	113	0.155	155	0.134
30	0.269	72	0.156	114	0.160	156	0.371
31	0.121	73	0.262	115	0.174	157	0.136
32	0.102	74	0.115	116	0.158	158	0.129
33	0.099	75	0.096	117	0.155	159	0.126
34	0.160	76	0.104	118	0.160	160	0.131
35	0.144	77	0.188	119	0.165	161	0.148
36	0.144	78	0.157	120	0.153	162	0.359
37	0.141	79	0.095	121	0.138	163	0.144
38	0.155	80	0.173	122	0.132	164	0.132
39	0.147	81	0.098	123	0.291	165	0.130
40	0.141	82	0.161	124	0.164	166	0.130
41	0.150	83	0.094	125	0.150	167	0.145
42	0.164	84	0.108	126	0.148		

Table 6.Statistical characteristics of catch rate standardization for silver hake fishing in the NAFO 4VWX subareas(sample 2)

REGRESSION OF MULTIPLICATIVE MODEL

MULTIPLE R............. .734
MULTIPLE R SQUARED..... .539

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	DF	SUMS OF SQUARES	MEAN SQUARES	F-VALUE
INTERCEPT	1	7.221E0001	7.221E0001	
REGRESSION	22	2.260E0001	1.027E0000	7.064
TYPE 1	1	7.154E-001	7.154E-001	4.920
TYPE 2	4	3.893E0000	9.732E-001	6.693
TYPE 3	13	1.414E0001	1.088E0000	7.480
TYPE 4	2	7.219E-001	3.610E-001	2.482
TYPE 5	1	1.116E0000	1.116E0000	7.677
TYPE 6	1	9.386E-001	9.386E-001	6.455
RESIDUALS	133	1.934E0001	1.454E-001	
TOTAL	156	1.141E0002		

Table 7. Coefficients of multiplicative model
(sample 2)

REGRESSION COEFFICIENTS

CATEGORY	CODE	VARIABLE	COEFFICIENT	STD. ERROR	NO. OBS.
1	1	INTERCEPT	0.711	0.216	156
2	5				
3	77				
4	450				
5	2				
6	2				
1	2	1	-0.325	0.147	79
2	4	2	0.235	0.102	24
	6	3	-0.120	0.087	42
	7	4	-0.160	0.092	34
	8	5	-0.346	0.112	20
3	78	6	-0.218	0.133	25
	79	7	-0.002	0.138	19
	80	8	-0.357	0.165	9
	81	9	-0.183	0.179	7
	82	10	0.686	0.186	7
	83	11	-0.022	0.180	8
	84	12	0.476	0.179	8
	85	13	0.337	0.179	8
	86	14	1.045	0.206	9
	87	15	0.963	0.206	9
	88	16	0.611	0.205	10
	89	17	0.975	0.198	12
	90	18	0.508	0.203	10
4	460	19	-0.268	0.168	113
	470	20	-0.365	0.171	36
5	1	21	0.432	0.156	19
6	1	22	0.219	0.086	118

Table 8.Standardized catch rate values for silver hake (sample 2)

PREDICTED CATCH RATE

STANDARDS USED VARIABLE NUMBERS: 1 5 450 2 1

YEAR	TOTAL CATCH	PROP.	CATCH RATE MEAN	CATCH RATE S.E.	EFFORT
77	37095	0.684	2.672	0.527	13881
78	48404	0.877	2.163	0.358	22380
79	51751	0.813	2.665	0.531	19416
80	44525	0.920	1.861	0.404	23919
81	44599	0.822	2.209	0.506	20191
82	60207	0.958	5.270	1.205	11425
83	35837	0.921	2.599	0.584	13789
84	74266	0.967	4.273	0.964	17379
85	75480	0.981	3.720	0.839	20292
86	82689	0.423	7.499	1.893	11027
87	61704	0.926	6.912	1.744	8928
88	74482	0.879	4.862	1.220	15319
89	86729	0.978	7.007	1.715	12377
90	60000	0.955	4.383	1.104	13688

AVERAGE C.V. FOR THE MEAN: .226

Table 9.Statistical characteristics of catch rate standardization for silver hake fishing in the NAFO 4VWX subareas(sample 3)

REGRESSION OF MULTIPLICATIVE MODEL

MULTIPLE R............ .737
MULTIPLE R SQUARED..... .544

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	DF	SUMS OF SQUARES	MEAN SQUARES	F-VALUE
INTERCEPT	1	7.325E0001	7.325E0001	
REGRESSION	22	1.912E0001	8.692E-001	6.929
TYPE 1	1	2.172E-002	2.172E-002	0.173
TYPE 2	4	3.030E0000	7.575E-001	6.039
TYPE 3	13	1.136E0001	8.741E-001	6.968
TYPE 4	2	7.417E-001	3.708E-001	2.956
TYPE 5	1	1.343E-001	1.343E-001	1.070
TYPE 6	1	9.259E-001	9.259E-001	7.381
RESIDUALS	128	1.606E0001	1.254E-001	
TOTAL	151	1.084E0002		

Table 10.Coefficients of multiplicative model

(sample 3)

REGRESSION COEFFICIENTS

CATEGORY	CODE	VARIABLE	COEFFICIENT	STD. ERROR	NO. OBS.
1	1	INTERCEPT	0.863	0.221	151
2	5				
3	77				
4	450				
5	2				
6	2				
1	2	1	-0.064	0.153	76
2	4	2	0.228	0.097	23
	6	3	-0.132	0.082	42
	7	4	-0.179	0.086	34
	8	5	-0.241	0.109	17
3	78	6	-0.336	0.132	24
	79	7	-0.187	0.135	19
	80	8	-0.473	0.157	9
	81	9	-0.305	0.170	7
	82	10	0.577	0.187	6
	83	11	-0.108	0.170	8
	84	12	0.376	0.170	8
	85	13	0.237	0.170	8
	86	14	0.683	0.208	9
	87	15	0.651	0.214	8
	88	16	0.262	0.207	10
	89	17	0.623	0.201	12
	90	18	0.137	0.206	10
4	460	19	-0.324	0.170	111
	470	20	-0.412	0.174	34
5	1	21	0.166	0.160	19
6	1	22	0.219	0.081	113

Table 11.Standardized catch rate values for silver hake (sample 3)

PREDICTED CATCH RATE

STANDARDS USED VARIABLE NUMBERS 1 5 450 2 1

YEAR	TOTAL CATCH	PROP.	CATCH RATE MEAN	CATCH RATE S.E.	EFFORT
77	37095	0.683	3.077	0.630	12055
78	48404	0.877	2.217	0.363	21832
79	51751	0.813	2.556	0.511	20250
80	44525	0.920	1.914	0.405	23258
81	44599	0.822	2.260	0.503	19738
82	60207	0.922	5.448	1.254	11052
83	35837	0.921	2.753	0.604	13017
84	74266	0.967	4.465	0.981	16635
85	75480	0.981	3.886	0.854	19424
86	82689	0.423	6.015	1.535	13747
87	61704	0.922	5.822	1.505	10599
88	74482	0.879	3.950	1.003	18855
89	86729	0.978	5.678	1.415	15274
90	60000	0.955	3.486	0.887	17212

AVERAGE C.V. FOR THE MEAN: .226

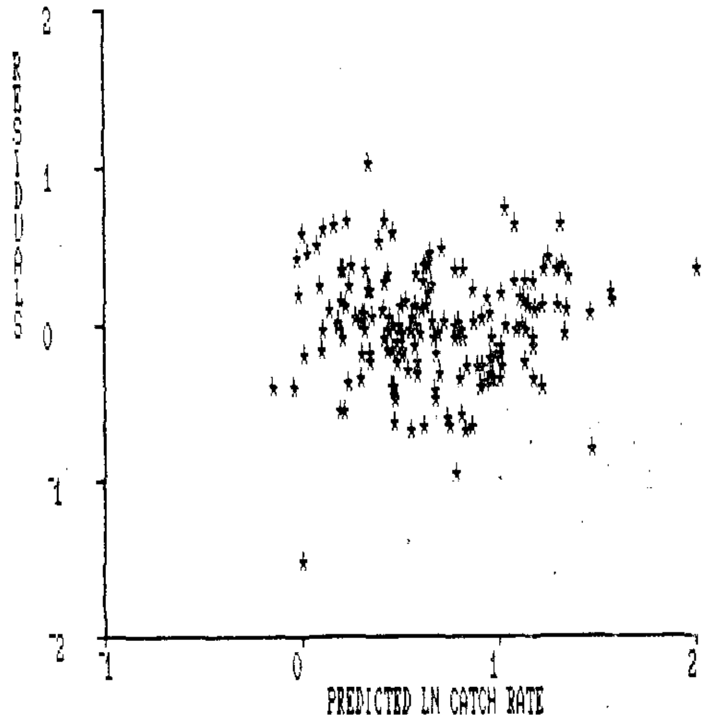

Figure 1. Residual plot from silver hake standardized catch rate analysys (sample 1).

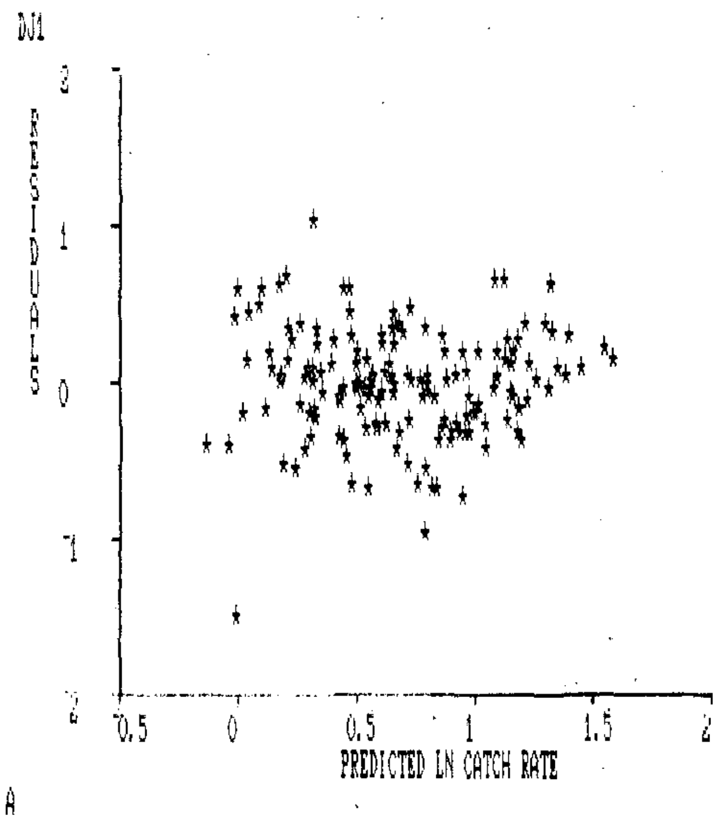

Figure 2. Residual plot from silver hake standardized catch rate analysys (sample 2).

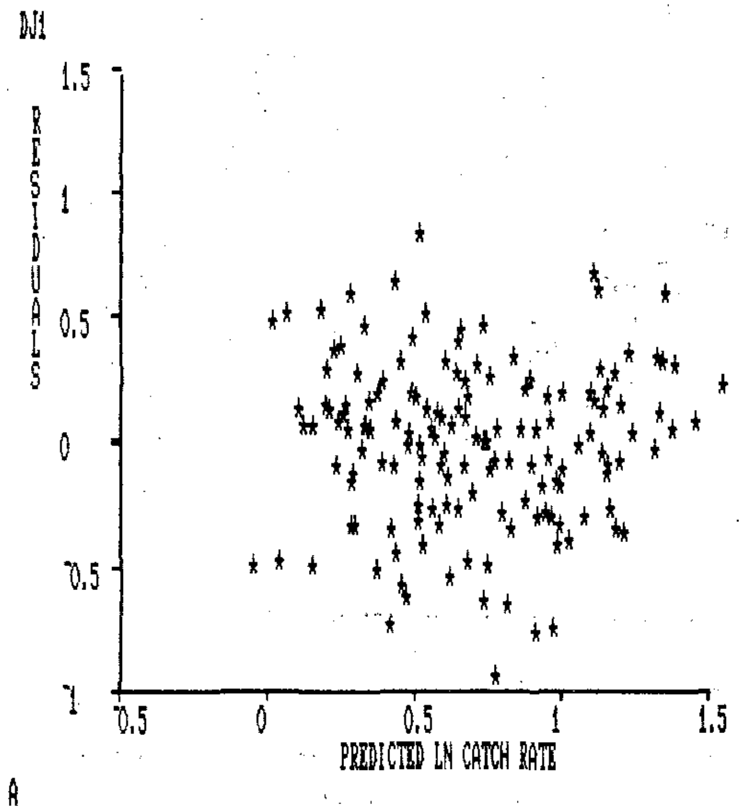

Figure 3. Residual plot from silver hake standardized catch rate analysys (sample 3).

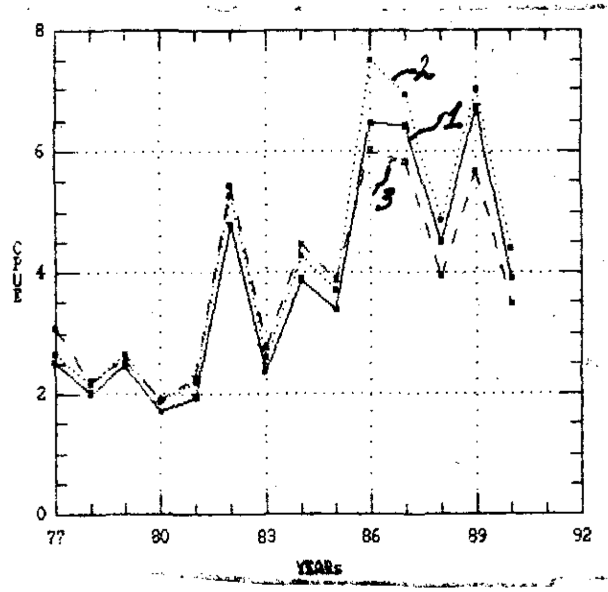

Figure 4. Silver hake standardized catch rate.
(1 - sample 1)
(2 - sample 2)
(3 - sample 3)