Catch Rate Standardization Runs for Yellowfin Tuna Caught by Taiwanese Deep Sea Longline Fishery in the Indian Ocean Using Generalized Linear Model and Generalized Linear Mixed Model

Hung-I Liu[1], Shu-Ting Chang[1] and Shui-Kai Chang[2]

[1] Overseas Fisheries Development Council, 19, Sec. 3, Roosevelt Rd, 100 Taipei, Taiwan
[2] Deep Sea Fisheries Division, Fisheries Agency, 2, Chaochow St., 100 Taipei, Taiwan

INTRODUCTION

Yellowfin tuna is one of the most important target species for Taiwanese far seas tuna longline fishery operating in the Indian Ocean. The fishery commenced in the northern and eastern Indian Ocean in the mid 1950s. The catches of this fishery mainly consisted of yellowfin tuna during late 1960s to early 1970s, and then changed to albacore in mid 1970s, and to bigeye tuna since 1980s when super cold freezers were developed and equipped in larger new-built vessels. The catches of yellowfin tuna (Fig. 1) were lower than 20,000 mt before late 1980s and thereafter substantially increased along with the increase of bigeye-targeting activities. In this period, the yellowfin catch has been bumped up to about 80,000 mt in 1993 and around ten years later to about 60,000 mt in 2005. Most of the catches in these two years were caught in the waters off Oman.

In this report, the standardization of CPUE for yellowfin tuna caught by Taiwanese longline fishery in the Indian Ocean is carried out by using generalized liner model (GLM) and generalized liner mixed model (GLMM). Environmental factors have also been included in this study.

MATERIALS AND METHODS

Data sets

Catch and effort data are compiled from logbooks and start from 1967 to 2005 for large deep sea frozen longline fishery. For 1967-1978, only aggregated 5°x5° square monthly data (TASK2) are available, but both original logbooks (LOGBOOK) and aggregated data are available from 1979. In this paper, both original logbooks and aggregated data for Taiwanese longline fishery operated in the Indian Ocean are used to standardize the CPUE of yellowfin tuna. Environment information including the size of area, the sea surface temperature (SST) and mixed layer depth (MLD) are kindly provided by Hiroaki Okamoto of National Research Institute of Far Seas Fisheries of Japan. These data were downloaded from NEAR-GOOS Regional Real Time Data Base of Japan Meteorological Agency and JEDEC (Joint Environmental Data Analysis Center) website of Scripps Institution of Oceanography, respectively. The procedure of this data processing is described in Okamoto et al. (2001).

Statistical Models

Statistical models of GLM and GLMM were used to model the logarithm of the nominal CPUE (defined as the number of fish per 1,000 hooks) in the report. The main effects considered in this analysis are year, season, area (Fig. 2), catch ratio for albacore and bigeye tunas (4 quartile levels of catch composition of albacore or bigeye tunas to the total catches of albacore, bigeye tuna and yellowfin tuna.), SST (five categories separated by 15, 20, 25, and 30°C) and MLD (integer of the MLD/10). In previous reports, there were five areas defined for yellowfin stock assessment. In this study, the Area 1 of previous studies was further divided into new Area 0 and Area 1 to accommodate the fishery feature of Taiwanese longliners in the specific region of new Area 1. Interactions for the main effects are also included into the model.

(1) GLM model: The CPUE is predicted as a linear combination of the explanatory variables. At first, the following form was assumed as a full model.

$$\log(CPUE + c) = \mu + Y + S + A + ALB + BET + SST + MLD + Interactions + \varepsilon$$

where $CPUE$ is the nominal CPUE of yellowfin tuna,

c is the constant value (i.e. 0.1),

μ is the intercept,

Y is the effect of year,

S is the effect of season,

A is the effect of fishing area,

ALB is the effect related to the catch ratio of albacore tuna,

BET is the effect related to the catch ratio of bigeye tuna,

SST is the effect of sea surface temperature,

MLD is the effect of mixed layer depth,

Interactions is the interactions between main effects,

ε is the error term, $\varepsilon \sim N(0, \sigma^2)$.

(2) GLMM model: This model assumes a delta lognormal error distribution for the positive catch rates. The model fits separately the proportion of positive sets assuming a binomial error distribution, and the mean catch rate of positive sets (at least one fish was caught) assuming a lognormal error distribution.

Estimated proportion of successful sets is assumed to be the result of r positive sets of a total n number of sets, and each one is an independent Bernoulli-type realization. The estimated proportion (ρ) is a linear function of fixed effects and interactions, by using logit function as a link between linear factor components and binomial errors. The systematic component is defined as:

$$\log\left(\frac{\rho_{Y,S,A,ALB,BET,SST,MLD}}{1-\rho_{Y,S,A,ALB,BET,SST,MLD}}\right) = \alpha_0 + \alpha_Y + \alpha_S + \alpha_A + \alpha_{ALB} + \alpha_{BET} + \alpha_{SST} + \alpha_{MLD} + \alpha_{Interactions} + \omega$$

with a binomial density:

$$\omega_{Y,S,A,ALB,BET,SST,MLD} \sim Bin(n_{Y,S,A,ALB,BET,SST,MLD}, \rho_{Y,S,A,ALB,BET,SST,MLD})$$

For positive observations, which were defined as at least one yellowfin caught, the estimated CPUE rate was assumed to follow a lognormal error distribution (logCPUE) of a linear function of fixed factors and random effect interactions. The systematic component is defined as:

$$\log(\mu_{Y,S,A,ALB,BET,SST,MLD}) = \beta_0 + \beta_Y + \beta_S + \beta_A + \beta_{ALB} + \beta_{BET} + \beta_{SST} + \beta_{MLD} + \beta_{Interactions} + z$$

with a log-normal density:

$$z_{Y,S,A,ALB,BET,SST,MLD} \sim LogNorm(\mu_{Y,S,A,ALB,BET,SST,MLD}, \sigma^2_{Y,S,A,ALB,BET,SST,MLD})$$

Standardization runs

This study has conducted a set of standardization runs using both logbook and Task2 data, by both GLMM and GLM model (Table 1). All runs only keep significant factors ($p<0.01$) in the analysis of CPUE by the effective effort. The calculation was done using GLM, GLIMMIX and MIXED procedure of SAS (Ver. 8. 02). The standardized CPUE were then computed from the least square means (LSMaens) of the estimates of the year effects.

RESULTS AND DISCUSSION

Table 1 shows R^2 of all runs. Basically R^2 of all the runs are similar. The model of all runs explained more than 50% of the variance, and run #8 was the highest, which explained about 68% of the variance. Table 2-9 shows the ANOVA table for the selected model for run #1 to 8. The statistics for the model indicate that the effects of model are highly statistically significant ($p < 0.01$) for eight runs.

Distributions of the standardized residuals for the eight runs are showed in Fig. 3. The distribution of the standardized residuals for runs #1-4 appear to deviate slightly from normal distribution assumption; on the other hand, runs #5-8 appear to meet normal distribution assumption. The normal probability plots in Fig. 4 show slight divergences for tails, however standardized residuals of some of the runs conform to the normal distribution.

The nominal CPUE trend obtained from the eight runs are similar and the trend of run #8 was shown in Fig. 1. The nominal CPUE fluctuated substantially in the 1970's, then remained stable during 1979-1985, fluctuated during 1985-1993, decreased slightly during 1994-2000, and increased again thereafter.

The relative standardized CPUEs obtained from the eight runs are shown in Fig. 5 and 6. Relative values are scaled to the average of estimates. As to cpue trends, for using logbook set by set data, including environmental factor or not does not make much difference when applying the same model, but the trends are quite different between models - GLMM results are flatter than GLM results (runs #1 and #3 vs. runs #2 and #4). But for using monthly aggregated data (TASK2), all the trends look similar (runs #5-8).

Fig. 7 shows relative standardized CPUE trends of run #1 and #8 that with higher R^2 in the logbook data group and TASK2 data group, respectively. The trends of the two series deviated each other in 1980s but are general similar from 1990s onwards.

For comparison, this study has also conducted catch rate standardization applying the same specification of the one done in 2005 (Wang et al., 2005). Option 3 of the Wang et al. (2005) fits GLM model to the TASK2 data sets in main Taiwanese fishing ground of yellowfin tuna (i.e. Area 0, 1, 2 and 5), and CPUEs for albacore and bigeye tunas (categorized as 5 levels) are considered in analysis model. Although there are differences between these two studies in terms of model specifications, trends of the standardized CPUE trends are similar (Fig 7).

In conclusion, the standardized catch rate was stable for a long period in 1980s. The catch rate had a short period of increase in mid-1990s, and then has shown a slowly increasing trend since 1998 to the level in 2005 similar to mid-1990s.

REFERENCES

Okamoto, H., N. Miyabe and T. Matsumoto. 2001. GLM analyses for standardization of Japanese longline cpue for bigeye tuna in the Indian Ocean applying environmental factors. *IOTC Proceedings* 4: 491-522.

Wang, S.P., S.K. Chang and H. Shono. 2005. Standardization of CPUE for yellowfin tuna caught by Taiwanese longline fishery in the Indian ocean using generalized linear model. *IOTC Proceedings* 8:014-023.

Table 1. Standardization runs conducted in this report on Indian Ocean yellowfin tuna CPUE of Taiwanese longline fishery.

Runs	DATA	MODEL	Environmental data	Data period	R^2
1	LOGBOOK	GLM	No	1979~2005	0.560018
2	LOGBOOK	GLMM	No	1979~2005	0.506231
3	LOGBOOK	GLM	SST、MLD	1980~2005	0.551594
4	LOGBOOK	GLMM	SST、MLD	1980~2005	0.507986
5	TASK2	GLM	SST	1968~2005	0.678955
6	TASK2	GLMM	No	1968~2005	0.665907
7	TASK2	GLM	SST、MLD	1980~2005	0.604936
8	TASK2	GLMM	SST	1968~2005	0.681911

Table 2. ANOVA table of the selected model for Run #1.

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Model	292	446998.0197	1530.8151	2311.68	<.0001
Error	530325	351186.6334	0.6622		
Corrected Total	530617	798184.653			

R-Square	Coeff Var	Root MSE	lnbetcpue Mean
0.560018	144.7603	0.813763	0.562145

Source	Df	Type III SS	Mean Square	F-value	P-value
Y	26	1925.92168	74.07391	111.86	<.0001
S	3	476.82606	158.94202	240.02	<.0001
A	5	18435.87254	3687.17451	5567.98	<.0001
ALB	3	33114.36406	11038.12135	16668.60	<.0001
BET	3	94998.15481	31666.05160	47818.70	<.0001
Y*S	78	5136.46114	65.85207	99.44	<.0001
Y*ALB	78	5251.05769	67.32125	101.66	<.0001
Y*BET	78	12033.38040	154.27411	232.97	<.0001
S*ALB	9	1973.26144	219.25127	331.09	<.0001
S*BET	9	1631.72799	181.30311	273.78	<.0001

Table 3. ANOVA table of the selected model for Run #2.

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Model	166	287095.909	1729.4934	2505.72	<.0001
Error	405709	280028.1574	6902		
Corrected Total	405875	567124.0664			

R-Square	Coeff Var	Root MSE	lnbetcpue Mean
0.506231	103.0122	0.830794	0.806501

Source	DF	Type III SS	Mean Square	F-value	P-value
Y	26	1786.6277	68.7164	99.56	<.0001
A	5	2388.3537	477.6707	692.06	<.0001
ALB	3	29209.0920	9736.3640	14106.2	<.0001
BET	3	133710.2293	44570.0764	64573.8	<.0001
Y*A	129	7346.3744	56.9486	82.51	<.0001

Table 4. ANOVA table of the selected model for Run #3.

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Model	307	392899.8089	1279.8039	1967.19	<.0001
Error	490949	319399.1239	0.6506		
Corrected Total	491256	712298.9329			

R-Square	Coeff Var	Root MSE	Lnbetcpue Mean
0.551594	126.2971	0.806582	0.638639

Source	Df	Type III SS	Mean Square	F-value	P-value
Y	25	1603.65749	64.1463	98.60	<.0001
S	3	254.33858	84.7795	130.31	<.0001
A	5	10877.96210	2175.5924	3344.11	<.0001
ALB	3	24412.71899	8137.5730	12508.30	<.0001
BET	3	91286.22909	30428.7430	46772.10	<.0001
SST	3	3126.38794	1042.1293	1601.86	<.0001
MLD	22	210.26758	9.5576	14.69	<.0001
Y*S	75	4307.02964	57.4270	88.27	<.0001
Y*ALB	75	3886.03646	51.8138	79.64	<.0001
Y*BET	75	9488.28795	126.5105	194.46	<.0001
S*ALB	9	688.64939	76.5166	117.61	<.0001
S*BET	9	1381.42924	153.4921	235.93	<.0001

Table 5. ANOVA table of the selected model for Run #4.

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Model	185	276633.8208	1495.318	2163.1	<.0001
Error	387590	267935.4208	0.6913		
Corrected Total	387775	544569.2416			

R-Square	Coeff Var	Root MSE	lnbetcpue Mean
0.507986	101.3431	0.831436	0.820417

Source	DF	Type III SS	Mean Square	F-value	P-value
Y	25	1704.3876	68.1755	98.62	<.0001
A	5	1517.1907	303.4381	438.95	<.0001
ALB	3	26310.0725	8770.0242	12686.5	<.0001
BET	3	128627.7628	42875.9209	62023.40	<.0001
SST	3	342.1514	114.0505	164.98	<.0001
MLD	22	549.9156	24.9962	36.16	<.0001
Y*A	124	6104.1775	49.2272	71.21	<.0001

Table 6. ANOVA table of the selected model for Run #5.

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Model	166	17252.57787	103.93119	261.54	<.0001
Error	20529	8157.90879	0.39738		
Corrected Total	20695	25410.48665			

R-Square	Coeff Var	Root MSE	lnbetcpue Mean
0.678955	75.09636	0.630384	0.839434

Source	DF	Type III SS	Mean Square	F-value	P-value
Y	37	1353.56579	36.58285	92.06	<.0001
S	3	23.36008	7.78669	19.59	<.0001
A	5	269.46688	53.89337	135.62	<.0001
ALB	3	2504.47349	834.82449	2100.80	<.0001
BET	3	3462.10839	1154.03613	2904.08	<.0001
SST	4	156.51348	39.12837	88.46	<.0001
Y*S	111	178.97355	1.61237	4.06	<.0001

Table 7. ANOVA table of the selected model for Run #6.

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Model	229	24345.57407	106.31255	164.55	<.0001
Error	18905	12214.43988	0.6461		
Corrected Total	19134	36560.01395			

R-Square	Coeff Var	Root MSE	lnbetcpue Mean
0.665907	121.0806	0.803801	0.663856

Source	DF	Type III SS	Mean Square	F-value	P-value
Y	37	556.677498	15.045338	23.29	<.0001
A	5	452.386604	90.477321	140.04	<.0001
ALB	3	3654.236057	1218.078686	1885.29	<.0001
BET	3	4193.154115	1397.718038	2163.33	<.0001
Y*A	101	932.518495	5.152036	7.97	<.0001

Table 8. ANOVA table of the selected model for Run #7.

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Model	139	9542.62821	68.652	157.14	<.0001
Error	14265	6231.97282	0.43687		
Corrected Total	14404	15774.60104			

R-Square	Coeff Var	Root MSE	lnbetcpue Mean
0.604936	105.2125	0.660963	0.628217

Source	DF	Type III SS	Mean Square	F-value	P-value
Y	25	263.279445	10.531178	24.11	<.0001
S	3	20.398414	6.799471	15.56	<.0001
A	5	336.240376	67.248075	153.93	<.0001
ALB	3	1729.634438	576.544813	1319.71	<.0001
BET	3	2717.603844	905.867948	2073.53	<.0001
SST	3	5.604123	1.868041	4.28	0.0050
MLD	22	58.611304	2.664150	6.10	<.0001
Y*S	75	120.634973	1.608466	3.68	<.0001

Table 9. ANOVA table of the selected model for Run #8.

Source	DF	Sum of Squares	Mean Square	F-value	P-value
Model	233	24810.49275	106.4828	172.8	<.0001
Error	18781	11573.30055	0.61622		
Corrected Total	19014	36383.7933			

R-Square	Coeff Var	Root MSE	lnbetcpue Mean
0.681911	118.8255	0.784999	0.660632

Source	DF	Type III SS	Mean Square	F-value	P-value
Y	37	512.559211	13.852952	22.48	<.0001
A	5	191.523042	38.304608	62.16	<.0001
ALB	3	3198.702223	1066.234070	1730.27	<.0001
BET	3	4114.110048	1371.370020	2225.44	<.0001
SST	4	577.866357	144.466589	234.44	<.0001
Y*A	181	878.292415	4.852444	7.87	<.0001

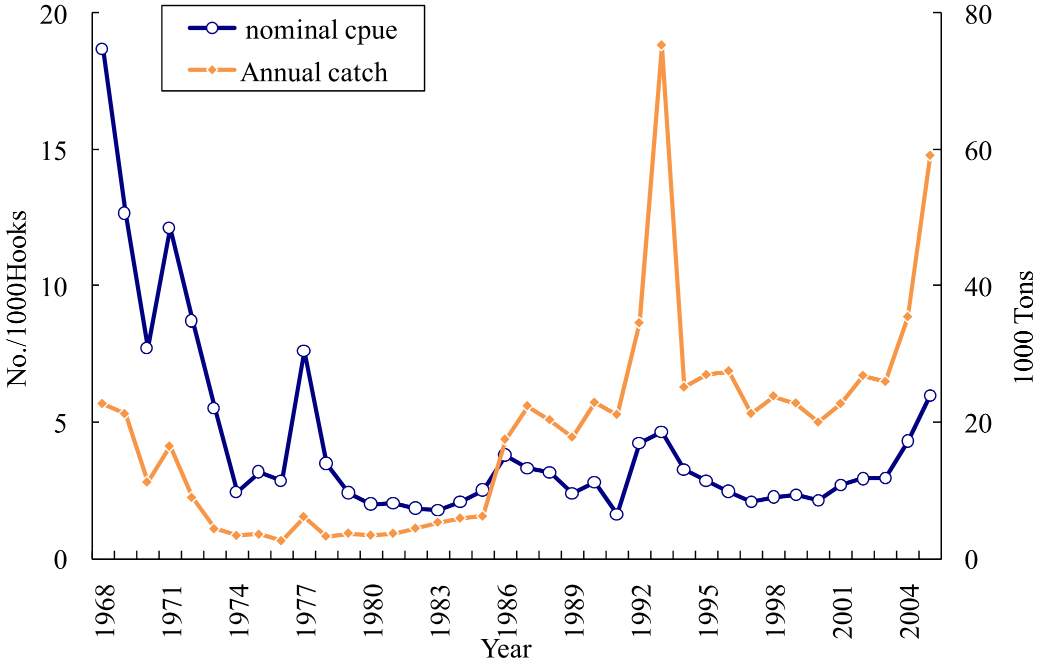


Fig. 1. Trends of annual nominal catches and nominal CPUE of yellowfin tuna caught by Taiwanese deep sea frozen longline fishery in the Indian Ocean.

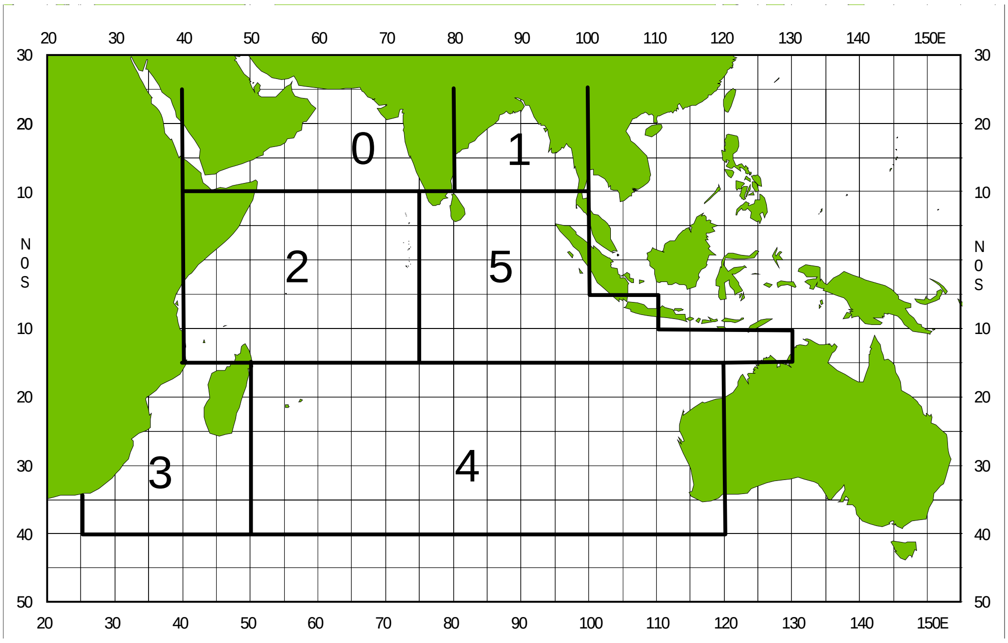
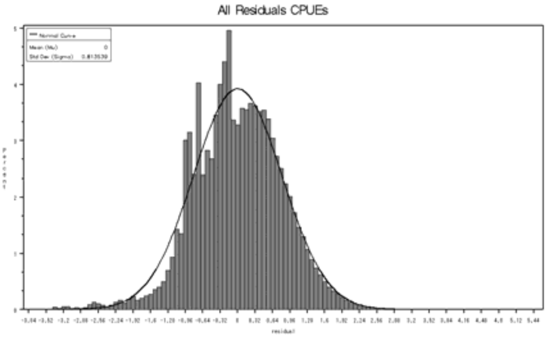
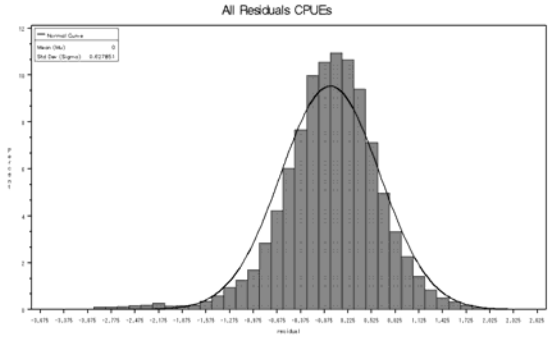


Fig. 2. Area stratification used fro the standardization of CPUE for yellowfin tuna in the Indian Ocean.

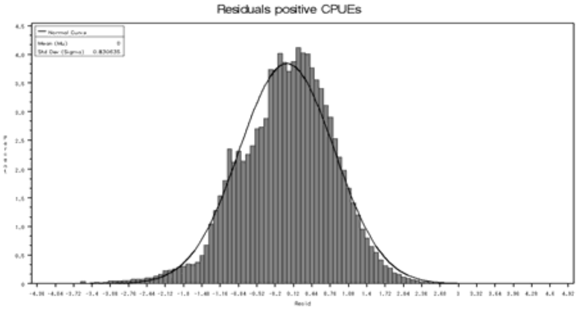
Run #1



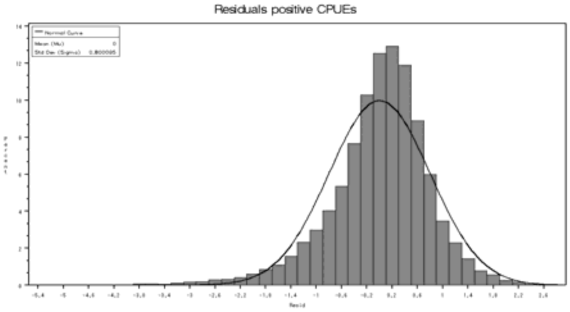
Run #5



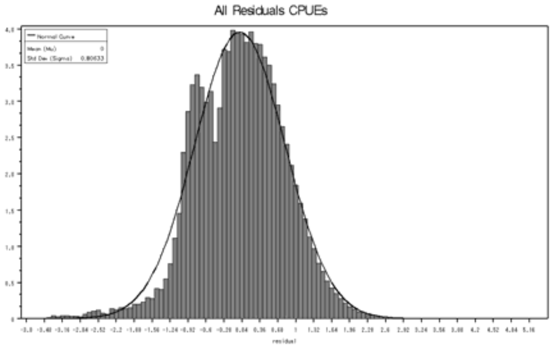
Run #2



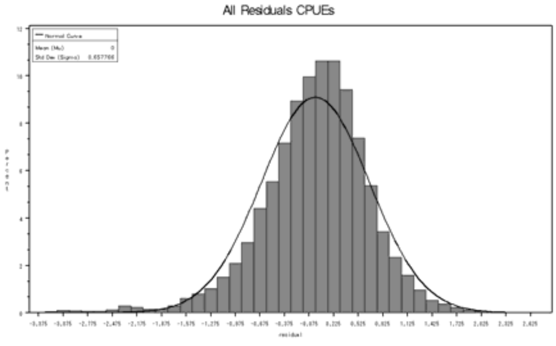
Run #6



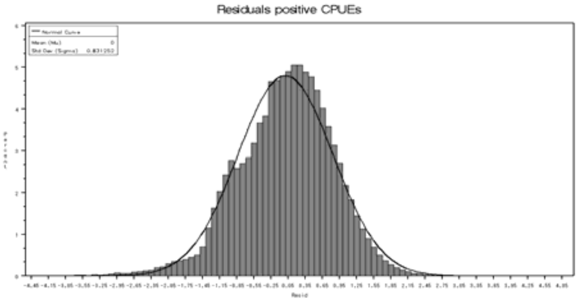
Run #3



Run #7



Run #4



Run #8

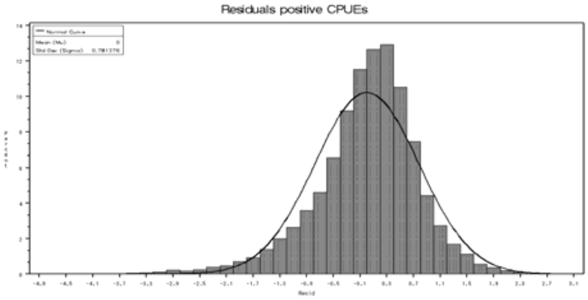
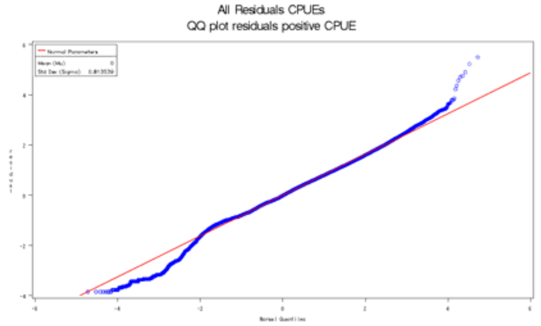
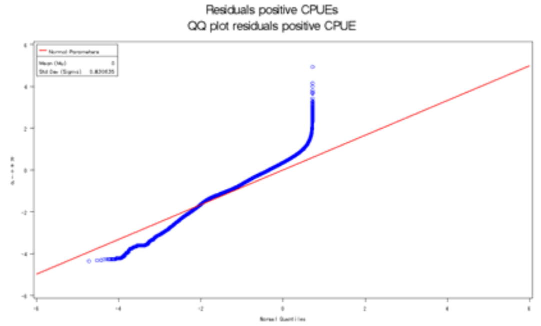


Fig. 3. Distributions of the standardized residuals for the standardization models fitted to the catch and effort data.

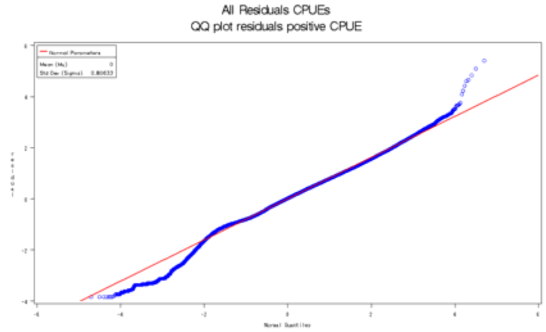
Run #1



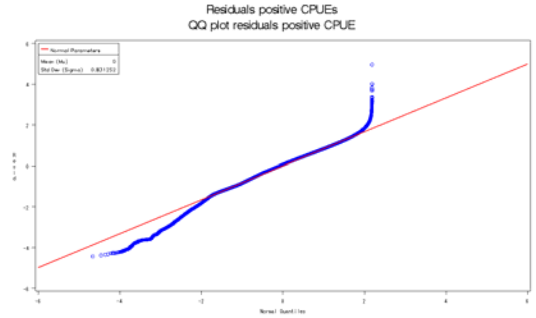
Run #2



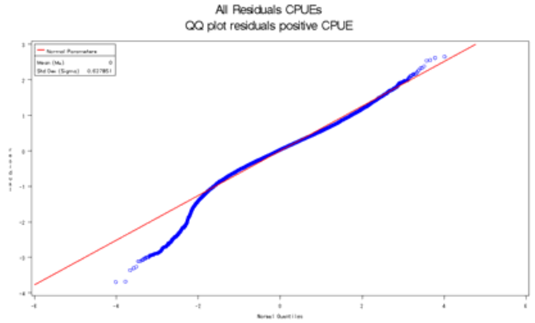
Run #3



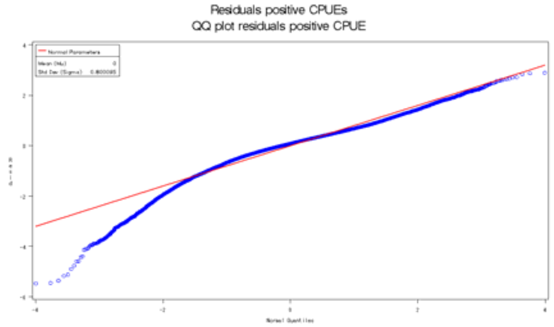
Run #4



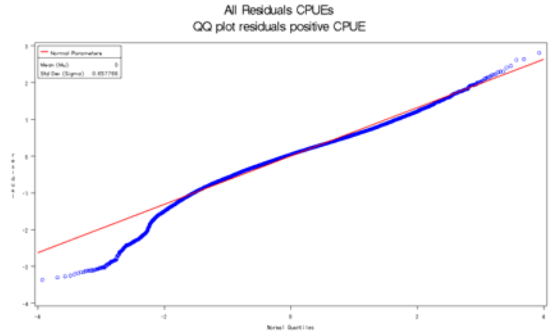
Run #5



Run #6



Run #7



Run #8

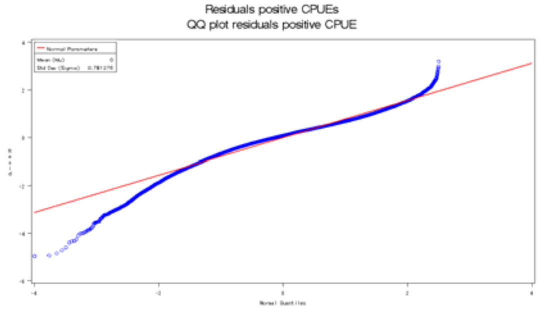


Fig. 4. The normal probability plots for the standardization models fitted to the catch and effort data.

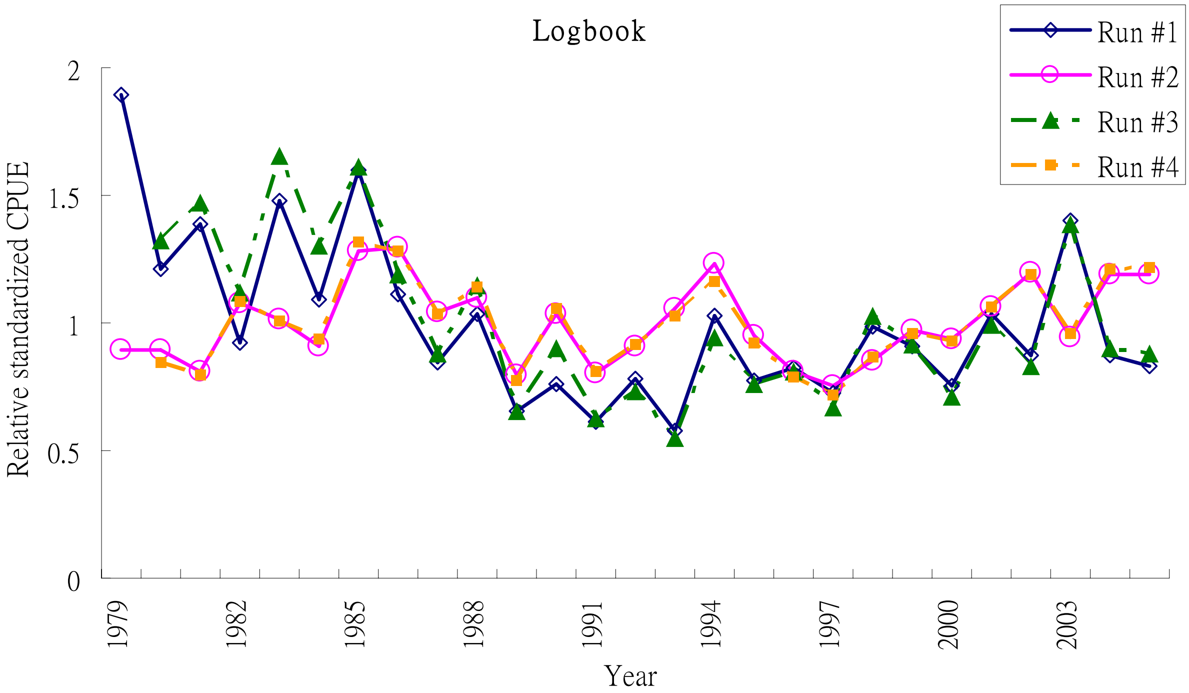


Fig. 5. Trends of relative standardized CPUE of yellowfin tuna caught by Taiwanese longline fishery in the Indian Ocean from run #1 to 4 on logbook data.

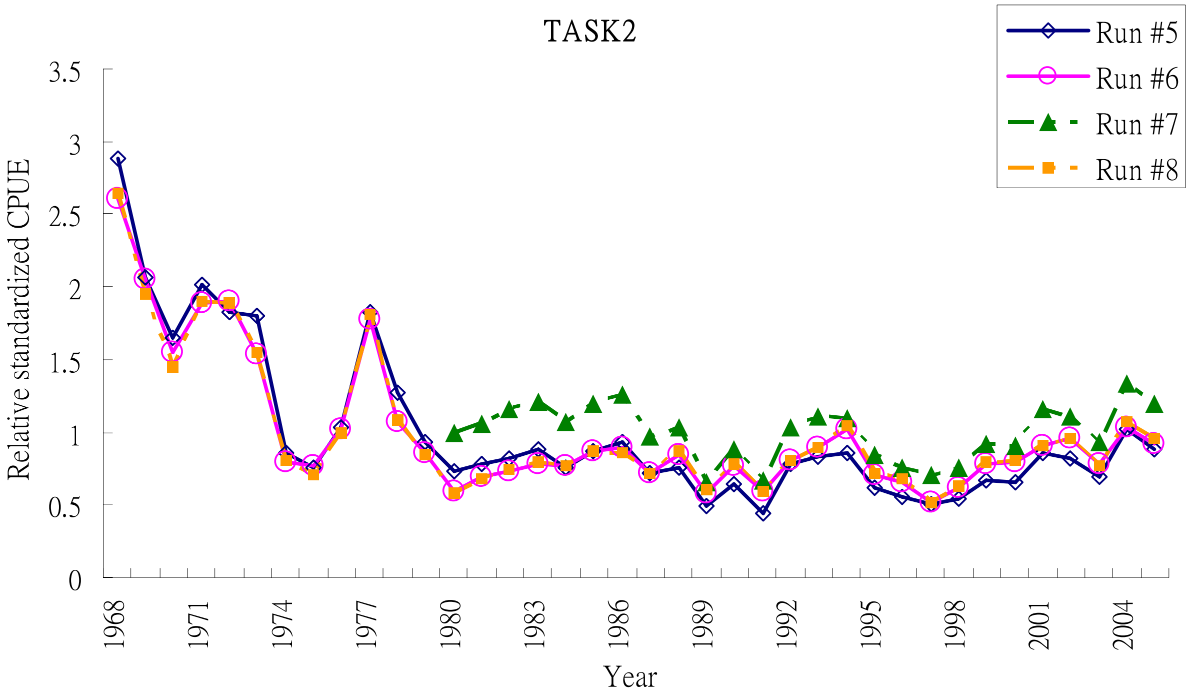


Fig. 6. Trends of relative standardized CPUE of yellowfin tuna caught by Taiwanese longline fishery in the Indian Ocean from run #5 to 8 on aggregated TASK2 data.

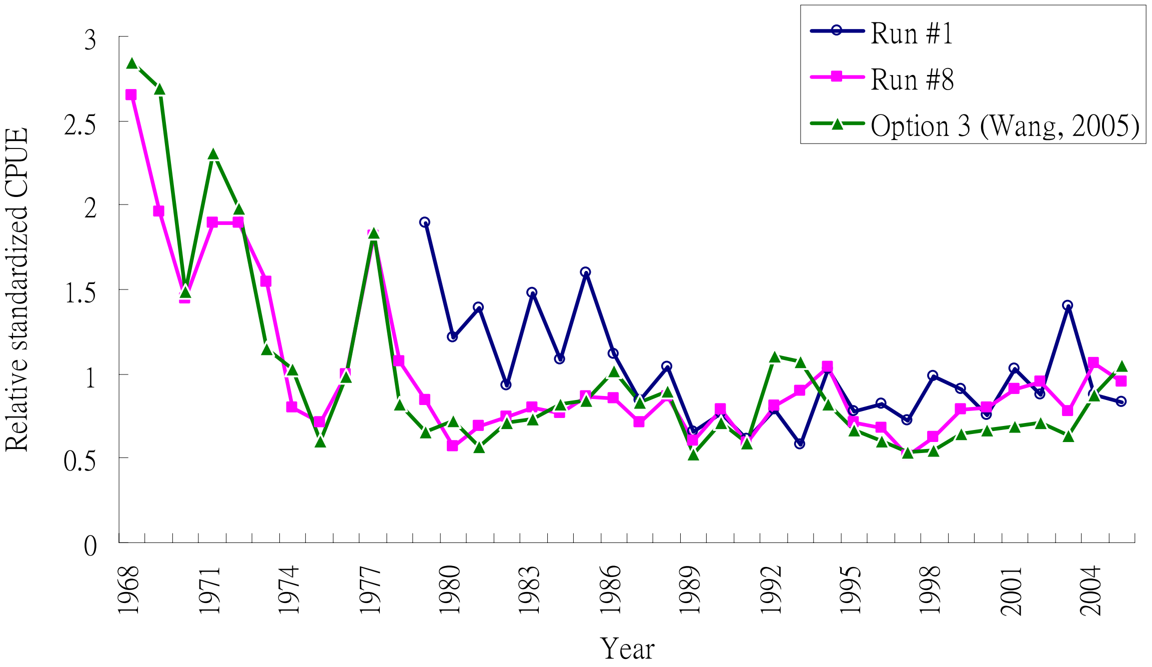


Fig. 7. Trends of relative standardized CPUE from run #1 and #8 that with higher R^2 in the logbook data group and TASK2 data group, respectively. Additional standardization result using the same specifications as the option-3 run of Wang et al (2005) was also shown in the figure.