SCIENTIFIC COMMITTEE
SIXTH REGULAR SESSION

10-19 August 2010
Nuku'alofa, Tonga

Confidence interval estimation of CPUE year trend in delta-type two- step model

WCPFC-SC6-2010/ME-WP-03

Hiroshi SHONO [1]

[1] National Research Institute of Far Seas Fisheries, Fisheries Research Agency, 5-7-1, Orido, Shimizu-ku, Shizuoka-shi, Shizuoka-ken 424-8633, Japan

Confidence interval estimation of CPUE year trend in delta-type two-step model

HIROSHI SHONO*

National Research Institute of Far Seas Fisheries, Fisheries Research Agency, Shizuoka 424-8633, Japan

ABSTRACT: A procedure is suggested for estimation of the approximate confidence intervals of the extracted catch per unit effort (CPUE) year trend in the delta-type two-step model used for CPUE standardization with a lot of zero-catch data. This method is a simple way to combine the Taylor expansion and delta method and is suitable for practical use. This model was applied to the catch and effort data with more than 80% zero catch for silky shark in the North Pacific Ocean caught by Japanese training vessels. As a result, realistic values of the 95% confidence interval of CPUE year trend are obtained. A method for left–right unsymmetrical interval estimation based on the asymptotic normality of the natural logarithm of CPUE is also suggested. In the example of silky shark, both CPUE year trends obtained from these two methods are similar.

KEY WORDS: **CPUE standardization, delta method, delta-type two-step model, generalized linear model, interval estimate, normal approximation, Taylor expansion, zero-catch.**

INTRODUCTION

Catch per unit effort (CPUE) standardization is currently an important and essential concept for fish stock analysis because the year trends of standardized CPUE enable us to not only grasp the relative abundance but also to use the estimates for stock assessment models as a tuning index.[1] As a statistical method, analysis of covariance (ANCOVA) called the CPUE–log-normal model, where the logCPUE model (natural logarithm of CPUE is set to the response variable) with normal error structure (normal distribution is assumed as the observation error) is applied, has been usually used for CPUE analysis.[2] However, this traditional CPUE–log-normal model cannot be applied to zero-catch data because the natural logarithm of zero-catch data is minus infinity. In such a case, the following three methods have often been applied: (i) *ad hoc* method to add a small constant value to all CPUE values in the above CPUE–log-normal model; (ii) generalized linear model (GLM) such as catch model with a Poisson or negative binomial error structure;[3] and (iii) so-called delta-type two-step model, where the ratio of zero-catch is estimated using logistic or probit regression in the first step and the typical model such as CPUE–log-normal or catch–negative-binomial is applied to the CPUE without zero data (i.e. CPUE of positive catch) in the second step.[4]

The delta-type two-step model generally has good performance and is easy to handle by common statistical software such as SAS (SAS Institute, Cary, NC, USA) or R (The R Project for Statistical Computing). However, it is difficult to estimate the confidence intervals of the year trend of standardized CPUE. In this paper, we suggest a procedure for estimating the approximate confidence intervals of the CPUE year trend obtained from this delta-type two-step model using a combination of the delta-method[5] and normal approximation based on the outputs of typical statistical packages, and also describe a case example of this simple method to real fishery data.

MATERIALS AND METHODS

Procedure and concept for estimating confidence interval of CPUE year trend

In this delta-type two-step model for CPUE standardization, the first step for estimating the zero-catch rate is formulated as:

*Corresponding author: Tel: 81-54-336-6000. Fax: 81-54-335-9642. Email: hshono@affrc.go.jp

Received 6 June 2007. Accepted 15 January 2008.

doi:10.1111/j.1444-2906.2008.01581.x

$$\mathrm{g}(p) = z = \text{Intercept} + \text{Year} + \text{Area} + \text{Season} + \ldots + \text{Interactions}, \mathrm{E}[X] = p, X \sim \text{Binomial}(\theta)$$

where E, expectation; g, link function (g^{-1}, link inverse function) (Table 1); *p*, predicted mean (i.e. zero-catch rate); *z*, linear predictor (composite function); Year, effect of year; Area, effect of area; Season, seasonal effect (e.g. month, quarter); Interactions, two–way interaction expressed by the product of main effects (e.g. Year × Area); and

$$X = \begin{cases} 1 & (\text{if Catch} > 0) \\ 0 & (\text{Otherwise}) \end{cases}$$

Common forms of the link function (g) and link inverse function (g^{-1}) including the logistic model are expressed in Table 1.

The second step for computing the CPUE without zero-catch data (i.e. CPUE in the positive catch) is shown as:

$$\mathrm{h}(U) = u = \text{Intercept} + \text{Year} + \text{Area} + \text{Season} + \ldots + \text{Interactions} + \text{error}, \text{error} \sim \mathrm{N}(0, \sigma^2)$$

where h, scaled function in the second step (h^{-1}, scaled inverse function), e.g. h(*U*) = log(*U*); *U*, CPUE of the non-zero part; *u*, linear predictor (scaled CPUE); and N(0,σ^2), normal distribution with mean 0 and variance σ^2.

Figure 1 shows the functional relationships in the two processes. In common statistical packages such as SAS or R, it is usually outputted that the point estimate of *z* and *u*, $\hat{z}$ and $\hat{u}$, and the variance, Var($\hat{z}$) and Var($\hat{u}$), obtained from the variance–covariance matrix of estimate values, $\hat{z}$ and $\hat{u}$. Therefore, we suggest an approximate method to estimate the confidence intervals of the estimated CPUE based on the values of *z* and *u* in both steps. The point estimate of standardized CPUE is shown in equation 1 under the assumption that *q* (=1 − *p*, non-zero catch rate, first step) and *U* (CPUE of positive data, second step) are independent because the estimated CPUE is usually defined as *q* multiplied by *U*:

Table 1 Specific form of link and link inverse function for common statistical model in the first step of the delta-type two-step model

Model	Link function g()	Link inverse function g^{-1}()
Logistic	log[p/(1 − p)]	1/[1 + exp(−Z)]
Probit	Φ^{-1}(p)	Φ(Z) = Φ[(x − μ)/σ]
c-log-log	log[−log(1 − p)]	1 − exp[−exp(Z)]

Φ, cumulative distribution function of standard normal distribution; c-log-log, complementary log-log.

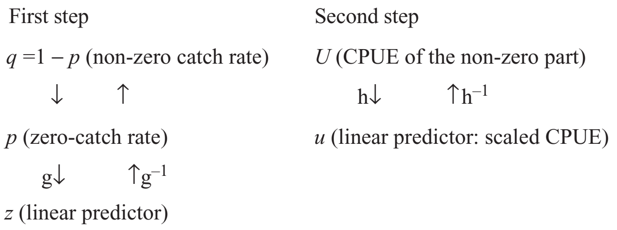
First step
Second step
q =1 − p (non-zero catch rate)
U (CPUE of the non-zero part)
h↓
↑h⁻¹
p (zero-catch rate)
u (linear predictor: scaled CPUE)
g↓
↑g⁻¹
z (linear predictor)

Fig. 1 Relationships between variables and functions in the two steps (g and h show the link and scaled functions, respectively, function *q* = 1 − *p*).

$$\widehat{\mathrm{CPUE}} = \hat{\mathrm{q}}\,\hat{\mathrm{U}} \tag{1}$$

where $\widehat{\mathrm{CPUE}}$, $\hat{\mathrm{q}}$ and $\hat{\mathrm{U}}$, are the point estimates of CPUE, *q* and *U*, respectively.

Estimated CPUE is shown in equation 2 using the functions g and h in Figure 1.

$$\mathrm{CPUE} = (1 - \mathrm{g}^{-1}(z))\mathrm{h}^{-1}(u) \tag{2}$$

where g^{-1} is the inverse function of g and h^{-1} is the inverse function of h.

The variance of CPUE, Var(CPUE), is developed into a Taylor series up to the first order with respect to the point estimates of *z* and *u* shown in equation 3.[5]

$$\mathrm{Var}(\mathrm{CPUE}) = \left[\mathrm{h}^{-1}(\hat{\mathrm{u}})\left(\frac{\partial}{\partial z}\{1 - \mathrm{g}^{-1}(z)\}\Big|_{z=\hat{z}}\right)\right]^2 \mathrm{Var}(z) + \left[\{1 - \mathrm{g}^{-1}(\hat{z})\}\left(\frac{\partial}{\partial u}\mathrm{h}(u)\Big|_{u=\hat{\mathrm{u}}}\right)\right]^2 \mathrm{Var}(u) \tag{3}$$

Thus, we can derive equation 4 from equation 3 and the variance of the point estimate of CPUE, $\widehat{\mathrm{CPUE}}$, is shown as:

$$\mathrm{Var}\left(\widehat{\mathrm{CPUE}}\right) = \left[\mathrm{h}^{-1}(\hat{\mathrm{u}})\left(\frac{\partial}{\partial z}\{1 - \mathrm{g}^{-1}(z)\}\Big|_{z=\hat{\mathrm{z}}}\right)\sigma(\hat{\mathrm{z}})\right]^2 + \left[\{1 - \mathrm{g}^{-1}(\hat{\mathrm{z}})\}\left(\frac{\partial}{\partial u}\mathrm{h}(u)\Big|_{u=\hat{\mathrm{u}}}\right)\sigma(\hat{\mathrm{u}})\right]^2 \tag{4}$$

where σ(*z*) and σ(*u*) are the standard error of *z* and *u*.

We can estimate the (100–α)% confidence intervals of the point estimate of the estimated CPUE based on normal approximation as follows:

$$\left[\widehat{\mathrm{CPUE}} - \mathrm{Z}\left(\frac{\alpha}{2}\right)\sigma\left(\widehat{\mathrm{CPUE}}\right), \widehat{\mathrm{CPUE}} + \mathrm{Z}\left(\frac{\alpha}{2}\right)\sigma\left(\widehat{\mathrm{CPUE}}\right)\right] \tag{5}$$

where $Z\left(\frac{\alpha}{2}\right)$ is the two-sided α-percentile of the standard normal distribution and $\sigma(\widehat{\mathrm{CPUE}})$ is the standard deviation of the point estimate of the CPUE, $\widehat{\mathrm{CPUE}}$.

Next, we describe the procedure using the normal approximation of the natural logarithm of CPUE because the observation error of CPUE seems not to be normally distributed (i.e. constant) but log-normally distributed in many cases. The natural logarithm of the estimated CPUE and its variance are shown in equations 6 and 7.[5]

$$\log(\mathrm{CPUE}) = \log[1 - g^{-1}(z)] + \log[h^{-1}(u)] \qquad (6)$$

$$\mathrm{Var}(\log(\mathrm{CPUE})) = \left[\frac{-1}{1-g^{-1}(z)}\left(\frac{\partial}{\partial z}g^{-1}(z)|_{z=\hat{z}}\right)\right]^2 \mathrm{Var}(z) + \left[\frac{1}{h^{-1}(z)}\left(\frac{\partial}{\partial u}h^{-1}(u)|_{u=\hat{u}}\right)\right]^2 \mathrm{Var}(u) \qquad (7)$$

The $(100 - \alpha)$% confidence intervals of point estimates of the natural logarithm of CPUE, $\log\left(\widehat{\mathrm{CPUE}}\right)$, is shown in equation 8 based on the normal approximation:

$$\left[\log\left(\widehat{\mathrm{CPUE}}\right) - Z\left(\frac{\alpha}{2}\right)\sigma\left(\log\left(\widehat{\mathrm{CPUE}}\right)\right),\ \log\left(\widehat{\mathrm{CPUE}}\right) + Z\left(\frac{\alpha}{2}\right)\sigma\left(\log\left(\widehat{\mathrm{CPUE}}\right)\right)\right] \qquad (8)$$

Thus, we can estimate the $(100 - \alpha)$% confidence interval of estimated CPUE as the following equation 9.

$$\left[\frac{\widehat{\mathrm{CPUE}}}{\exp\left\{Z\left(\frac{\alpha}{2}\right)\sigma\left(\log\left(\widehat{\mathrm{CPUE}}\right)\right)\right\}},\ \widehat{\mathrm{CPUE}}\exp\left\{Z\left(\frac{\alpha}{2}\right)\sigma\left(\log\left(\widehat{\mathrm{CPUE}}\right)\right)\right\}\right] \qquad (9)$$

Data analysis

We calculated the standardized CPUE of catch and effort data (in which the rate of zero-catch is more than 80%) for silky shark *Carcharhinus falciformis* caught by Japanese training vessels in the North Pacific Ocean using the delta-type two-step method as a case example. We applied the logistic regression model (i.e. link function g is the logit) in the first step and a CPUE–log-normal model (i.e. scaled function h is the natural logarithm; the link function is defined as the identity mapping and response variable, which is normally distributed, and is set to the natural logarithm of CPUE) in the second step. The following explanatory variables were chosen based on the results of model selection by the Bayesian information criterion (BIC)[6] because the selection performance of consistent information criterion such as the BIC in large samples is better than that of the Akaike information criterion (AIC) for the variable selection of the ANCOVA model corresponding to the CPUE standardization in many cases.[7] In the example, we computed through the GLM and GENMOD procedures of the SAS/STAT package v9.1 (SAS Institute).

$$g(p) = \log(p/(1-p)) = z = \mathrm{Intercept} + \mathrm{Year}(i) + \mathrm{Area}(j) + \mathrm{Season}(k) + a \times \mathrm{HPB} + b \times [\mathrm{Area}(j) * \mathrm{HPB}] \qquad (10)$$

$$h(U) = \log(U) = u = \mathrm{Intercept} + \mathrm{Year}(i) + \mathrm{Area}(j) + \mathrm{Season}(k) + a \times \mathrm{HPB} + b \times [\mathrm{Area}(j) * \mathrm{HPB}] + \mathrm{error}, \mathrm{error} \sim N(0, \sigma^2) \qquad (11)$$

where Year, effect of year (1992–2003); Area, effect of area (1 = 0–20 N, 2 = 20–30 N, 3 = 30–40 N, 4 = 40–50 N); Season, effect of quarter (1 = Jan–Mar, 2 = Apr–Jun, 3 = Jul–Sep, 4 = Oct–Dec); HPB, effect of gear (i.e. hooks per basket); *a*, *b*, regression coefficients estimated; Area(*j*) × HPB, two–way interaction of effect between area and hooks per basket; *U* (CPUE of the non-zero data in the second step), catch in number per 1000 hooks; $E[X] = p$, $X \sim \mathrm{Binomial}(\theta)$; and $X = \begin{cases} 1 & (\text{if Catch} > 0) \\ 0 & (\text{Otherwise}) \end{cases}$.

Year, Area, Season are assumed as categorical variables and HPB as a continuous variable.

RESULTS

In this case example for silky shark in the North Pacific Ocean caught by Japanese training vessels, the year trends *q* (non-zero catch rate) in the first step and *U* (CPUE of the non-zero-part) in the second step, $\hat{q}$ and $\hat{U}$, are shown in Table 2. These values are derived as the least-square means for type III sum of squares of the year effects. On the basis of these estimated values, we can obtain the point estimates of yearly CPUE using equation 1.

Table 2 Estimated non-zero catch rate (*q*) and CPUE in the second step (*U*) for the example of silky shark

Year	*q*	*U*
1992	0.067	1.427
1993	0.088	1.499
1994	0.057	1.544
1995	0.120	1.881
1996	0.059	1.287
1997	0.087	1.541
1998	0.076	1.511
1999	0.084	1.353
2000	0.067	1.327
2001	0.066	1.295
2002	0.033	1.171
2003	0.049	1.255

Table 3 Estimates of linear predictor in first step (*z*), second step (*u*) and these standard error *σ*(*z*), *σ*(*u*) for the example of silky shark, obtained as outputs from common statistical packages such as SAS or R

Year	*z*	*u*	*σ*(z)	*σ*(u)
1992	2.634	0.355	0.169	0.022
1993	2.334	0.405	0.159	0.023
1994	2.809	0.434	0.164	0.025
1995	1.996	0.632	0.160	0.018
1996	2.768	0.252	0.167	0.018
1997	2.356	0.433	0.166	0.027
1998	2.503	0.413	0.170	0.039
1999	2.385	0.302	0.166	0.020
2000	2.635	0.283	0.169	0.019
2001	2.654	0.258	0.173	0.020
2002	3.377	0.158	0.184	0.020
2003	2.971	0.228	0.176	0.022

Table 3 shows the estimated values of linear predictor *z* and *u*, *ẑ* and *û*, which correspond to the least-squared mean of year effects in the first and second steps and these dispersions (i.e. standard errors), σ(ẑ) and σ(û). We obtain the variance of yearly CPUE (i.e. estimated CPUE year trend) based on equation 4 (Table 3).

Point estimates and variances of the yearly CPUE and the 95% confidence intervals are shown in Table 4 based on the values of Tables 2 and 3 obtained from equations 1–5. The link and scaled inverse functions are defined as $g^{-1}(z) = 1/[1 + \exp(-z)]$ and $h^{-1}(u) = \exp(u)$. We can obtain the variances of the estimated CPUE trend by the derivative of functions g^{-1} and h^{-1} as equation 12.

$$\mathrm{Var}\left(\widehat{\mathrm{CPUE}}\right)=\left[e^{\hat{u}}\left\{-\frac{e^{-\hat{z}}}{\left(1+e^{-\hat{z}}\right)^{2}}\right\}\sigma(\hat{z})\right]^{2}+\left[\frac{e^{-\hat{z}}}{1+e^{-\hat{z}}}e^{\hat{u}}\sigma(\hat{u})\right]^{2} \quad (12)$$

Thus, we can obtain the 95% confidence intervals of the CPUE year trend (Table 4) in which $Z\left(\frac{\alpha}{2}\right)=Z(0.025)$ is set to 1.96. Figure 2 shows the 95% confidence intervals.

In this case example, the standard error of the natural logarithm of yearly CPUE is expressed in the following equation 13 derived from equation 7 in which the link and scaled functions are defined as $g^{-1}(z) = 1/[1 + \exp(-z)]$ and $h^{-1}(u) = \exp(u)$.

$$\sigma\left[\log\left(\widehat{\mathrm{CPUE}}\right)\right]=\sqrt{\left(\frac{\sigma(\hat{z})}{1+\exp(-\hat{z})}\right)^{2}+(\sigma(\hat{u}))^{2}} \quad (13)$$

Hence, we can estimate the 95% confidence intervals of CPUE year trend (i.e. least-square means of year effect) based on equation 9 using the values of equations 7 and 13 where $Z\left(\frac{\alpha}{2}\right)$ is set to 1.96. Table 5 and Figure 3 show the 95% confidence intervals of the estimated yearly CPUE on the basis of the normal approximation of log(CPUE). The interval estimates do not show left–right symmetry but these values are similar to those in Table 4 and Figure 2. In our case example, these two confidence intervals shown in Figures 2 and 3 (obtained from eqns 5 and 9, respectively) seem to be rather similar.

DISCUSSION

As described, this simple method (i.e. two procedures shown in eqns 5 and 9) has many advantages for interval estimation of the CPUE trend in the delta-type two-step model with a lot of zero-catch data. Because this method is based on the asymptotic normality of CPUE by the central limit theorem, it is expected that we estimate reasonable confidence intervals in many cases especially for large samples. It is also practical for the computation to use the outputs directly (e.g., ẑ, û, and the standard error, σ(ẑ) σ(û)) from common statistical packages such as SAS or R.

We recommend the use of the first procedure based on the normal approximation of CPUE shown in equation 5. However, the assumption of an asymptotic log-normal distribution of observed CPUE seems to be more natural especially in small samples or in the case of fat-tail distributions. The choice of the two methods (shown in eqns 5 and 9)

Table 4 Standardized CPUE and its variance, and 95% confidence intervals obtained from simple method based on the normal approximation of CPUE for the example of silky shark

Year	CPUE	Var(CPUE)	σ(CPUE)	Lower 5%	Upper 5%
1992	0.096	0.0002	0.015	0.066	0.125
1993	0.133	0.0004	0.019	0.094	0.171
1994	0.088	0.0002	0.014	0.061	0.115
1995	0.225	0.0010	0.032	0.162	0.288
1996	0.076	0.0001	0.012	0.052	0.100
1997	0.133	0.0004	0.021	0.093	0.174
1998	0.114	0.0003	0.018	0.078	0.151
1999	0.114	0.0003	0.018	0.080	0.148
2000	0.089	0.0002	0.014	0.061	0.116
2001	0.085	0.0002	0.014	0.058	0.112
2002	0.039	0.0001	0.008	0.024	0.054
2003	0.061	0.0001	0.010	0.041	0.082

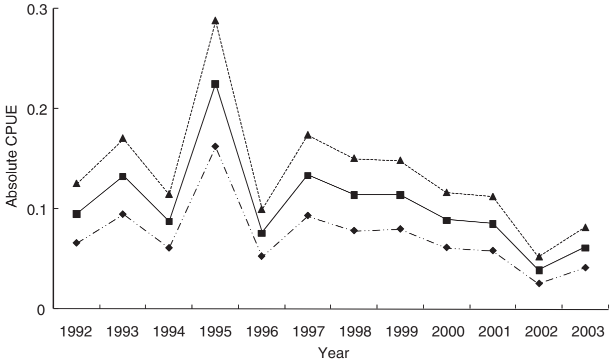

Fig. 2 Point estimate of the year trend of standardized CPUE and its 95% confidence intervals as lower 5% (◆), point estimate (■) and upper 5% (▲), which are obtained from the asymptotic normality of CPUE.

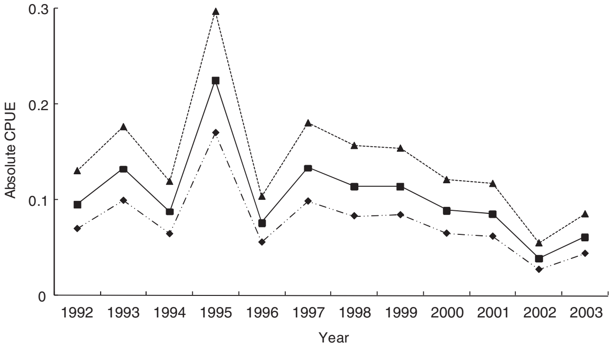

Fig. 3. Point estimate of the year trend of standardized CPUE and its 95 % confidence intervals as lower 5% (◆), point estimate (■) and upper 5% (▲), which are obtained from the asymptotic normality of natural logarithm of CPUE, log(CPUE).

is largely dependent on which assumption, i.e. whether CPUE is normally distributed or log-normally distributed, is more adequate. It is useful to illustrate the histogram of the observed CPUE to decide the model to be used and some information criteria such as the AIC or BIC can be applied for model selection.

We conducted a simple simulation to check the performance of equations 5 and 9 as an approximate confidence interval of CPUE. A set of data with 10 000 realizations was generalized from the delta log-normal model, where p (zero-catch rate) is set to 0.25, 0.5 and 0.75 in the first step and U (CPUE of positive data) is the independent and identical log-normally distributed random variables with with a mean 0 and variance 1. We compared the values of maximum log-likelihood (MLL), which is average of 100 replications, between the normal and log-normal distributions (Table 6). Judging from the experiment, the confidence interval shown in equation 5 based on the normal distribution is better than that in equation 9 from the log-normal distribution if p (zero-catch ratio) is high, such as 0.75, and the opposite outcome is obtained if p is low. In our study, MLL is equivalent to information criteria such as the AIC or BIC, since the number of parameters and observations is the same in the two distributions.

In this research, we assumed the independence of the linear predictors z and u, which means they are independent of the first and second steps, for the simplification of calculation. The correlation coefficient of z and u seem very low (not zero) in many cases. If we compute the magnitude (i.e. degree) of the correlation, then we can include the values of covariance between z and u into equations 3 and 7. However, it is generally difficult to integrate the correlation into the model unless we use a different method for simultaneous estimation of the unknown parameters in the first and second steps, such as the zero-inflated model.[8] Therefore, we should be careful about the

Table 5 Standardized log CPUE and its variance, and 95 percent confidence intervals obtained from method based on normal approximation of natural logarithm of CPUE, log CPUE for the example of silky shark

Year	log CPUE	Var(log CPUE)	σ(log CPUE)	Lower 5%	Point estimate	Upper 5%
1992	–2.348	0.025	0.159	0.070	0.096	0.131
1993	–2.021	0.021	0.147	0.099	0.133	0.177
1994	–2.433	0.025	0.157	0.065	0.088	0.119
1995	–1.491	0.020	0.142	0.170	0.225	0.298
1996	–2.576	0.025	0.158	0.056	0.076	0.104
1997	–2.014	0.024	0.154	0.099	0.133	0.181
1998	–2.169	0.026	0.162	0.083	0.114	0.157
1999	–2.171	0.024	0.154	0.084	0.114	0.154
2000	–2.422	0.025	0.159	0.065	0.089	0.121
2001	–2.463	0.026	0.162	0.062	0.085	0.117
2002	–3.253	0.032	0.179	0.027	0.039	0.055
2003	–2.793	0.029	0.169	0.044	0.061	0.085

Table 6 Values of maximum log-likelihood of normal and log-normal distributions obtained from simulation based on delta log-normal model

p	0.25	0.5	0.75
Normal	163 452	211 289	193 720
Log-normal	175 661	213 281	188 056

possibility of underestimating the confidence intervals for CPUE trends obtained from this method.

ACKNOWLEDGMENTS

Thanks to Mr H Matsunaga, National Research Institute of Far Seas Fisheries (NRIFSF), for providing processed catch and effort data. Useful comments of two anonymous referees and Dr M Ichinokawa (NRIFSF) are also appreciated.

REFERENCES

1. Gavaris S. Use of a multiplicative model to estimated catch rate and effort from commercial data. *Can. J. Fish. Aquat. Sci.* 1980; **37**: 2272–2275.
2. Robson DS. Estimation of the relative fishing power of individual ships. *Res. Bull. Int. Comm. North-West Atlantic Fish.* 1966; **3**: 5–14.
3. Reed WJ. Analyzing catch-effort data allowing for randomness in the catching process. *Can. J. Fish. Aquat. Sci.* 1996; **43**: 174–186.
4. Lo NCLD, Jacobson LD, Squire JL. Indices of relative abundance from fish spotter data based on Delta-Lognormal models. *Can. J. Fish. Aquat. Sci.* 1992; **49**: 2515–2526.
5. Seber GAF. *The Estimation of Animal Abundance and Related Parameters*. Oxford University Press, New York, NY. 1982.
6. Schwarz G. Estimating the dimension of a model. *Ann. Stat.* 1978; **6**: 461–464.
7. Shono H. Is model selection using Akaike's information criterion appropriate for catch per unit effort standardization in large samples? *Fish. Sci.* 2005; **71**: 978–986.
8. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 1992; **34**: 1–14.