

Bias and Bias Correction in Fish Recruitment Prediction

D. G. CHEN*

International Pacific Halibut Commission, Seattle, Washington 98195, USA

Abstract.—A lognormal error is usually assumed in the traditional stock–recruitment (SR) analysis. In this paper, I demonstrate that a positive bias results when an SR model with lognormal error is used for fish recruitment prediction. Not only is this bias exponentially dependent on the structure of the model’s residual variance and the historical SR data, it is also dependent on the specific value of spawner biomass used. I then derive a bias correction that is asymptotically unbiased with a finite-sample bias that is practically zero. Data from two Pacific salmon populations, southeast Alaska pink salmon *Oncorhynchus gorbuscha* and Chilko Lake sockeye salmon *O. nerka*, are used to demonstrate this approach. The results show that the relative bias is about 10% for a spawning biomass within the historical range and that the bias is substantial outside that data range. However, with the proposed bias correction, the bias is negligible both within (<0.1%) and outside of (<0.3%) the data range.

Stock–recruitment (SR) analysis usually begins with the assumption of a functional relationship, denoted by $F(\cdot)$, between spawners and recruits, namely,

$$R_t = F(S_t, \Theta), \quad (1)$$

where S_t is the spawning biomass at year t , $t = 1, \dots, n$, R_t is the recruitment biomass corresponding to S_t , and Θ is a vector of parameters associated with this relationship and usually related to the fishery management policy (Hilborn and Walters 1992; Quinn and Deriso 1999).

For a statistical analysis of the SR relationship, a lognormal error is traditionally included into the deterministic model based on the arguments of Hennemuth et al. (1980), Peterman (1981), and Hilborn (1985). Hennemuth et al. (1980) showed empirically that the variability in stock and recruitment is often lognormally distributed. Peterman (1981) reviewed the theory behind the traditional assumption of lognormal errors and empirically demonstrated the validity of the theory for some salmon stocks. Third, the mathematical and statistical properties of the lognormal distribution are convenient. Lastly, the traditional regression methods used to estimate the SR param-

eters implicitly assume a lognormal distribution (Hilborn 1985). With the inclusion of the lognormal error, the stochastic version of model in equation (1) is

$$R_t = F(S_t, \Theta)e^{\varepsilon_t}, \quad (2)$$

where the ε_t are assumed to be independently and normally distributed with mean 0 and variance σ^2 .

The most commonly used SR models are the Beverton–Holt (Beverton and Holt 1957) and Ricker models (Ricker 1975), which may be represented as follows:

$$R_t = \frac{\alpha S_t}{1 + \beta S_t} e^{\varepsilon_t} \quad (\text{Beverton-Holt model}), \quad (3)$$

$$R_t = S_t e^{\alpha - \beta S_t + \varepsilon_t} \quad (\text{Ricker model}), \quad (4)$$

where α and β are parameters for measures of productivity and capacity, respectively. To unify these two models, Deriso (1980) introduced an additional shape parameter γ and proposed the following SR model:

$$R_t = \alpha S_t (1 - \gamma \beta S_t)^{1/\gamma} e^{\varepsilon_t} \quad (\text{Deriso model}). \quad (5)$$

Therefore, the Beverton–Holt and Ricker models are special cases of the Deriso model with $\gamma = -1$ and $\gamma \rightarrow 0$, respectively. Further extensions of these models have been comprehensively discussed in Hilborn and Walters (1992).

From the regression approach it has long been known that at any stock size arithmetic average recruitment is larger than modal recruitment (Ricker 1975; Hilborn 1985). A bias can be introduced into the parameter estimates that subsequently affects the calculation of fishery management parameters such as the maximum sustainable yield (MSY) spawner and the MSY exploitation rate. However, this bias is dependent only on the estimated model’s residual variance and is easily corrected. Hilborn (1985) proposed a bias correction for the estimated parameters of the Ricker model. Hilborn and Walters (1992:270–272) summarized the bias correction for the SR parameters and the associated management parameters for both the Ricker and Beverton–Holt models.

In this paper, I show that there is an additional bias: a positive prediction bias that occurs when

* E-mail: din@iphc.washington.edu

Received March 13, 2003; accepted September 15, 2003

the fitted SR model is used for fish recruitment prediction for specific observations. The structure of this prediction bias is more complicated than that of the bias of the estimated parameters. Not only is the prediction bias exponentially dependent on the estimated model's residual variance, it is also dependent on the historical SR data and the value of spawner biomass to be predicted. I then derive a bias correction and demonstrate analytically that this bias correction is asymptotically unbiased with a finite sample bias that is practically zero. For simplicity, the Ricker model (equation 4) is used in this paper; extension to other SR models can be easily done.

Prediction Bias and Its Correction

The Ricker model can be linearized as

$$\log_e(R_t/S_t) = \alpha - \beta S_t + \varepsilon_t \tag{6}$$

or in matrix form as

$$\mathbf{Y} = \mathbf{X}\Psi + \varepsilon, \tag{7}$$

where

$$\mathbf{X} = \begin{pmatrix} 1 & -S_1 \\ \vdots & \vdots \\ 1 & -S_n \end{pmatrix},$$

an $n \times 2$ matrix with all the elements in the first column being equal to 1 and those in the second column representing the negative spawner time series from $t = 1$ to $t = n$; Ψ = the transpose of (α, β) ; $\mathbf{Y} = (Y_1, \dots, Y_n)' = [\log_e(R_1/S_1), \dots, \log_e(R_n/S_n)]'$, which is an n -dimensional vector for the \log_e recruitment per spawner; and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$, which is an n -dimensional vector for the errors distributed as $N(0, \sigma^2\mathbf{I})$. In this situation, the error variance (σ^2) may include both process and measurement errors if there is no prior information with which to separate them.

The parameter vector Ψ and its variance can be estimated by means of multiple linear regression (Graybill 1976: Chapter 6):

$$\hat{\Psi} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \tag{8}$$

$$\hat{\sigma}^2 = \frac{\mathbf{Y}'\mathbf{Y} - \hat{\Psi}'\mathbf{X}'\mathbf{Y}}{n - 2}. \tag{9}$$

Since ε is normally distributed, the theory of regression (Graybill 1976; Draper and Smith 1981) implies that $\hat{\Psi}$ and $\hat{\sigma}^2$ are unbiased, sufficient, and complete statistics for the SR parameters Ψ and σ^2 , $\hat{\Psi}$ being distributed as $N(\Psi, \sigma^2[\mathbf{X}'\mathbf{X}]^{-1})$ and

$(n - 2)\hat{\sigma}^2/\sigma^2$ being distributed as a chi-square with $n - 2$ degrees of freedom.

Therefore, for any specific spawning biomass S_0 , the predicted \log_e stock production index $\hat{y}_0 = \hat{\alpha} - \hat{\beta}S_0 = \mathbf{X}_0\hat{\Psi}$ is normally distributed, with $\mathbf{X}_0 = (1, -S_0)$. The traditional prediction of the recruitment corresponding to the spawning biomass S_0 (\hat{R}_0) is obtained with the antilog transformation $\hat{R}_0 = S_0e^{\hat{y}_0}$. The bias of this prediction can easily be seen from the fact (Aitchison and Brown 1957) that if a random variable w is normally distributed with mean μ and variance σ^2 , then e^w is lognormally distributed with mean $e^{\mu + \sigma^2/2}$ and variance $e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$. Therefore, with simple algebra, it can be shown that

$$\begin{aligned} E(\hat{R}_0) &= S_0 \cdot E[\exp(\hat{y}_0)] \\ &= S_0 \cdot \exp[E(\hat{y}_0) + V(\hat{y}_0)/2] \\ &= S_0 \cdot \exp[\mathbf{X}_0\Psi + \sigma^2\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_0/2] \\ &= R_0 \cdot \exp[\sigma^2\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_0/2] \\ &= R_0 \cdot \exp\left[\frac{\sigma^2 \sum_{t=1}^n (S_t - S_0)^2}{2n \sum_{t=1}^n (S_t - \bar{S})^2}\right], \end{aligned} \tag{10}$$

where \bar{S} is the mean of all S_t .

It can easily be seen from equation (10) that the traditional recruitment prediction \hat{R}_0 is biased; the relative bias, $\exp[\sigma^2/2n \sum_{t=1}^n (S_t - S_0)^2/\sum_{t=1}^n (S_t - \bar{S})^2]$, is always greater than 1 because the term inside the brackets is always positive and its exponentiation is therefore greater than 1. Also, the bias is dependent not only on the model's residual variance σ^2 but also on the observed SR data, as indicated by the term $\sum_{t=1}^n (S_t - S_0)^2/\sum_{t=1}^n (S_t - \bar{S})^2$. In addition, the relative bias is dependent on the specific spawning biomass to be predicted, with no bias at $S_0 = \bar{S}$ since $\sum_{t=1}^n (S_t - S_0)^2/\sum_{t=1}^n (S_t - \bar{S})^2 = \sum_{t=1}^n (S_t - \bar{S})^2/\sum_{t=1}^n (S_t - \bar{S})^2 = 1$. As S_0 deviates from \bar{S} , the term $\sum_{t=1}^n (S_t - S_0)^2/\sum_{t=1}^n (S_t - \bar{S})^2$ becomes larger. At some point, therefore, the relative bias could become highly significant. This means that the prediction bias is not constant over the data range.

Based on equation (10), a bias-corrected value for the predicted recruitment would be

$$\begin{aligned} \tilde{R}_0 &= \hat{R}_0 \cdot \exp[-\hat{\sigma}^2\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_0/2] \\ &= S_0 \cdot \exp\left[\hat{\alpha} - \hat{\beta}S_0 - \frac{\hat{\sigma}^2 \sum_{t=1}^n (S_t - S_0)^2}{2n \sum_{t=1}^n (S_t - \bar{S})^2}\right]. \end{aligned} \tag{11}$$

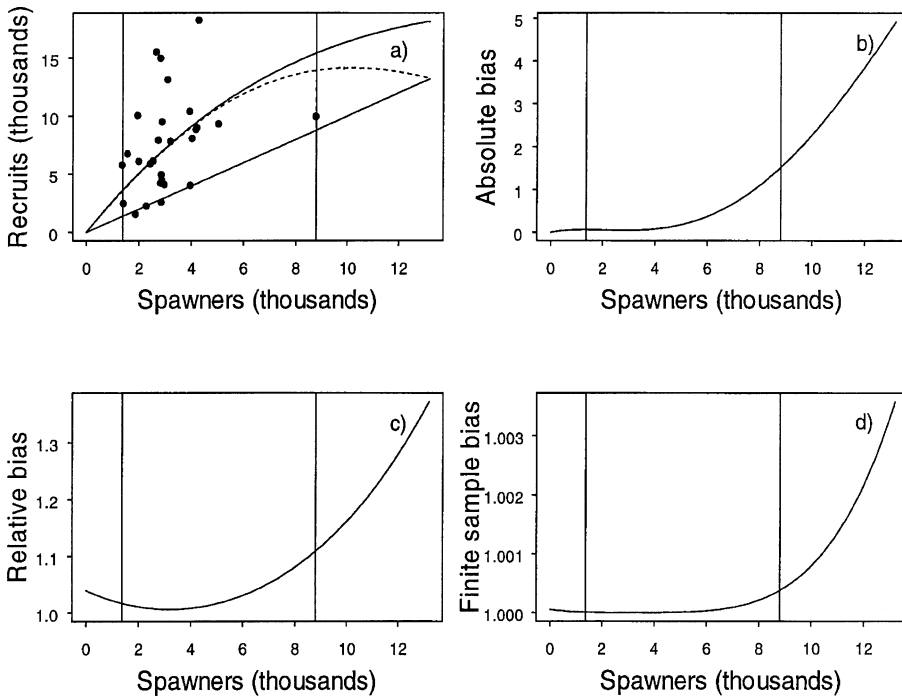


FIGURE 1.—Panel (a) shows the stock–recruitment (SR) data for southeast Alaska pink salmon from Quinn and Deriso (1999). The three curves are the traditional predicted SR relationship (solid curve), the bias-corrected prediction (dashed curve), and the replacement line (straight line). Panel (b) shows the absolute bias ($\hat{R} - \bar{R}$, where \hat{R} is the traditional prediction and \bar{R} is the bias-corrected prediction from equation 11). Panel (c) shows the relative bias (\hat{R}/\bar{R}) and panel (d) the finite sample bias resulting from the finite sample in equation (11). The two vertical lines in the panels indicate the historical data range; values to the right of the second line represent a 50% extrapolation of the maximum observed spawner biomass.

In the appendix, I show analytically that this value is asymptotically unbiased and that the finite sample bias is negligible. Since equation (11) indicates that the magnitude of the bias is dependent on the historical SR data, I have applied this correction

to two data sets from the literature to illustrate the amount of bias that can typically be expected.

Results

Southeast Alaska Pink Salmon

For the first example, SR data for southeast Alaska pink salmon *Oncorhynchus gorbuscha* from 1960 to 1989 (Figure 1a; Quinn and Deriso 1999: 104–123) are used since these data are readily available and one can easily apply the methods proposed in this paper. Table 1 summarizes the required quantities for the SR analysis.

The linearized Ricker model (equation 6) was fitted to these data (Table 1). The model is not statistically significant ($F = 0.49$), and the R^2 is also very poor (0.017).

An examination of the residual diagnostics for the model indicates that the residuals are normally distributed and the P -value from the Kolmogorov-Smirnov test is greater than 0.5, which confirms the assumption of normality. The autocorrelation

TABLE 1.—Summary of the stock–recruitment data and Ricker model fit. Values for $\mathbf{X}'\mathbf{X}$ and $\mathbf{Y}'\mathbf{Y}$ are given so that readers can repeat this analysis; $\hat{\sigma}^2$ is the estimated residual variance; $\hat{\alpha}$ and $\hat{\beta}$ are the elements of the vector $\hat{\Psi}$ in equation (6), the values in parentheses are standard errors; asterisks indicate statistical significance at the 95% confidence level.

Parameter	Pink salmon	Sockeye salmon
$\mathbf{X}'\mathbf{X}$	$\begin{pmatrix} 30 & -96.764 \\ -96.764 & 371.927 \end{pmatrix}$	$\begin{pmatrix} 48 & -12.724 \\ -12.724 & 5.760 \end{pmatrix}$
$\mathbf{Y}'\mathbf{Y}$	33.288	181.305
$\hat{\sigma}^2$	0.373 (df = 28)	0.461 (df = 46)
$\hat{\alpha}$	1.047 (0.278)*	2.243 (0.152)*
$\hat{\beta}$	0.055 (0.079)	1.728 (0.439)*
R^2	0.017	0.252
F	0.490	<0.001

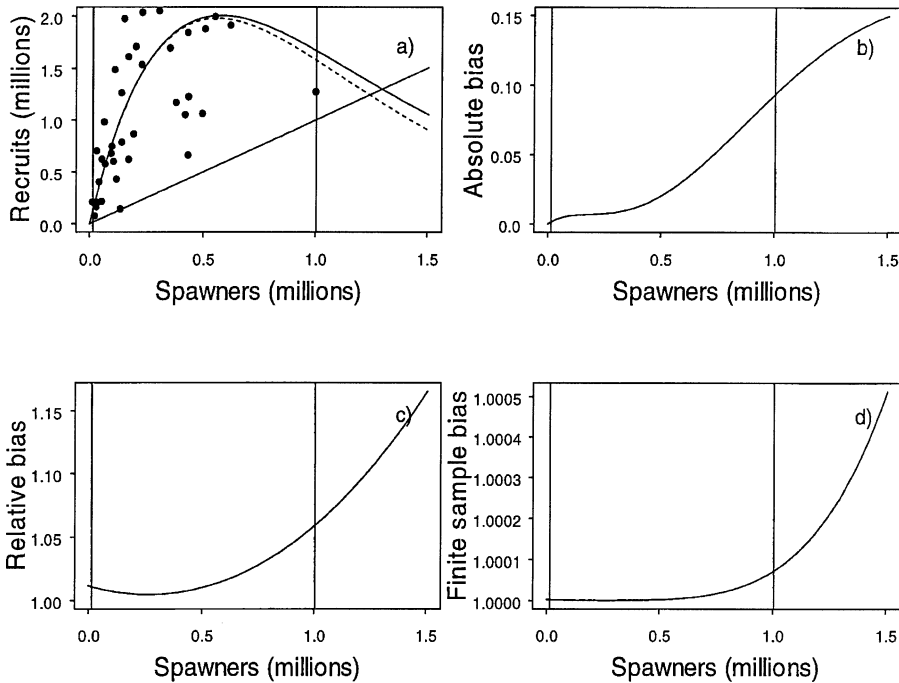


FIGURE 2.—The same information as in Figure 1 only for Chilko Lake sockeye salmon.

function plot for the residuals demonstrates that there is no autocorrelation among the residuals. The residual plot between the residuals and the predicted values shows homogeneity. Therefore, no assumptions are violated when using the Ricker model to fit the SR data.

Figure 1a shows the traditional prediction \hat{R}_0 in equation (10) and the bias-corrected prediction \tilde{R}_0 defined in equation (11). It can be seen that \hat{R}_0 is always biased upwards. When the prediction is extrapolated beyond the historical data range, the bias worsens very quickly

In Figure 1b, the absolute bias for the prediction is calculated as $\hat{R} - \tilde{R}$. For values in the historical data range, the absolute bias can be as high as 1,500–2,000 fish; the extrapolation could result in a bias as high as 3,000–5,000 fish. The relative prediction bias (\hat{R}/\tilde{R}) is given in Figure 1c; it can be seen to be about 10% for values within the historical data range and 20–30% for values beyond that range. The lowest bias is for values close to the historical mean.

As indicated in the appendix, the bias-corrected value is asymptotically unbiased with a finite sample bias of $[\sigma^4/4n(n-2) \sum_{t=1}^n (S_t - S_0)^2 / \sum_{t=1}^n (S_t - \bar{S})^2]$. To investigate the magnitude of this bias, I plotted the computation in the previous sentence against the values of spawning biomass (Figure

1d). It is seen that the higher-order bias resulting from the finite sample is negligible: less than 0.1% within the historical data range and less than 0.3% for the extrapolation. In this example, therefore, the traditional prediction \hat{R}_0 is biased upwards and the proposed correction \tilde{R}_0 is very satisfactory.

Chilko Lake Sockeye Salmon

Spawner and recruit data for Chilko Lake sockeye salmon *O. nerka* are available for brood years 1948–1995 (Figure 2a). A more detailed description of these data can be found in Chen et al. (2001; the data can also be requested from the author).

When the Ricker model is fitted to these data, the result is statistically significant ($F < 0.001$; Table 1) and the R^2 is relatively good (0.252). An examination of the residual diagnostics indicates that there are no violations of model assumptions. The summary information for these data and the model fit are given in Table 1.

The same bias as in the case of the pink salmon can be seen in Figure 2–d. The traditional prediction \hat{R}_0 is always biased upwards (Figure 2a), especially during extrapolation. The absolute bias for the prediction is as high as 0.1 million fish within the historical data range and increases significantly for the extrapolation (Figure 2b). The relative prediction bias for this stock is about 5–

7% (Figure 2c) within the historical data range, and it increases rapidly for the extrapolation. Figure 2d shows that the finite sample bias from the prediction bias correction equation is much less than 0.001% for the data within the historical data range and less than 0.04% for the extrapolation. This example also demonstrates that the finite sample bias is negligible.

Discussion

The literature on lognormal errors in statistics is substantial. Finney (1951) studied the estimation of the parameters of a lognormal distribution. Bradu and Mundlak (1970) provided a comprehensive discussion of lognormal linear models. The log-linear model has also been discussed in econometrics (Meulenberg 1965; Goldberger 1968). However, these authors used an infinite Taylor series as a bias correction function—a representation that is not obvious and that is difficult for fisheries scientists to apply.

In this paper, I derived a simple correction for the prediction bias that is inherent in traditional SR analysis. Case studies showed that a 5–10% bias can occur. This bias correction was analytically shown to be asymptotically unbiased with a finite sample bias of practically zero.

The estimated variance of the prediction \tilde{R}_0 can be easily obtained mathematically by following the theory of the lognormal distribution. However, it is tedious and therefore I have not reproduced it here. Like the prediction bias, the prediction variance grows quickly for recruitment predictions outside the range of the spawning observations.

The methodology developed in the paper is easy to apply. First, use linear regression to fit the Ricker SR model (equation 6) and obtain estimates for $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}^2$, as in the usual SR analysis. Then, for any specific value of spawner biomass S_0 , an essentially unbiased prediction of the recruitment biomass \tilde{R}_0 can be calculated from equation (11).

The methodology in this paper can be extended to the Beverton–Holt and Deriso SR models with aid from the theory of nonlinear regression. The analysis can also easily be extended to include environmental variables, as in the equation $R_t = S_t \cdot \exp(\alpha - \beta S_t + \tau \mathbf{X}_t + \varepsilon_t)$, where \mathbf{X}_t is a vector of environmental variables and τ is a vector of the associated parameters (Chen and Ware 1999). Furthermore, the bias correction process could be combined with nonparametric (Jacobson and MacCall 1995) and semiparametric (Chen and Irvine 2001) SR models to give more accurate fish recruitment predictions.

This methodology can also be extended to SR relationships that have a dependent time series structure. Walters (1985) showed that such a structure “can lead to severely biased parameter estimates.” Obviously, the biased parameter estimates introduce bias into the recruitment prediction. To deal with this situation, ε in equation (7) can be generalized to have the distribution $N(0, \sigma^2 \mathbf{\Omega})$, where $\mathbf{\Omega}$ is the variance–covariance matrix of errors for the time series dependence. Note that $\mathbf{\Omega}$ is the identity matrix if ε is independent. It has been demonstrated (Arnold 1981) that $\mathbf{X}'\mathbf{X}$ in equations (10) and (11) should be replaced by $\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}$. The subsequent development is the same.

This can be also applied to the heterogeneity of variance under log transformation, as discussed in MacCall and Ralston (2002), where $\mathbf{\Omega}$ is the matrix describing the heterogeneity. For example, the variance–covariance matrix $\mathbf{\Omega}$ for l -heterogeneity discussed in equation (2) of MacCall and Ralston (2002) would be a diagonal matrix with the diagonal elements as $\sigma_{0.5} + \delta(S_t - 0.5)$.

Since the prediction bias in equation (10) and the proposed bias correction \tilde{R}_0 in equation (11) depend on the SR data as indicated by the bias $\exp[\sigma^2/2n\sum_{t=1}^n (S_t - S_0)^2/\sum_{t=1}^n (S_t - \bar{S})^2]$, I would recommend that the SR quantities

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & -\sum_{t=1}^n S_t \\ -\sum_{t=1}^n S_t & \sum_{t=1}^n S_t^2 \end{pmatrix} \quad \text{and} \quad \mathbf{Y}'\mathbf{Y} = \sum_{t=1}^n Y_t^2$$

be included in any future publications (e.g., Table 1) so that readers can learn the techniques by repeating the analysis.

In future research, it will be worthwhile to examine the robustness of this bias correction for alternatives to lognormal errors, such as the gamma distribution proposed in the fisheries literature. In addition, further studies using simulation or theoretical development could address several outstanding problems in SR analysis, such as small sample size, limited dynamic range, and errors in variables.

Acknowledgments

I sincerely thank Richard Leickly and Jim Irvine for their constructive suggestions and comments on this paper. I also thank three anonymous reviewers, the editor, and an associate editor for their helpful comments.

References

- Aitchison, J., and J. A. C. Brown. 1957. The lognormal distribution. University Press, London.
- Arnold, S. F. 1981. The theory of linear models and multivariate analysis. Wiley, New York.
- Beverton, R. J. H., and S. J. Holt. 1957. On the dynamics of exploited fish populations. Fisheries Investigations, Series II, Marine Fisheries, Great Britain Ministry of Agriculture, Fisheries and Food 19.
- Bradu, D., and Y. Mundlak. 1970. Estimation in lognormal linear models. Journal of the American Statistical Association 65:198–211.
- Chen, D. G., and J. R. Irvine. 2001. A semiparametric model to examine stock–recruitment relationships incorporating environmental data. Canadian Journal of Fisheries and Aquatic Sciences 58:1178–1186.
- Chen, D. G., J. R. Irvine, and A. J. Cass. 2001. Incorporating Allee effects in fish stock–recruitment models and applications for determining reference points. Canadian Journal of Fisheries and Aquatic Sciences 59:242–249.
- Chen, D. G., and D. M. Ware. 1999. A neural network model for forecasting fish stock recruitment. Canadian Journal of Fisheries and Aquatic Sciences 56:2385–2396.
- Deriso, R. B. 1980. Harvesting strategies and parameter estimation for an age-structured model. Canadian Journal of Fisheries and Aquatic Sciences 37:268–282.
- Draper, N. R., and H. Smith. 1981. Applied regression analysis, 2nd edition. Wiley, New York.
- Finney, D. J. 1951. On the distribution of a variate whose logarithm is normally distributed. Journal of the Royal Statistical Society 7(Supplement):155–161.
- Goldberger, A. S. 1968. The interpretation and estimation of Cobb-Douglas functions. Econometrica 35:464–472.
- Graybill, F. A. 1976. Theory and application of the linear model. Wadsworth and Brooks/Cole, Pacific Grove, California.
- Hennemuth, R. C., J. E. Palmer, and B. E. Brown. 1980. A statistical description of recruitment in eighteen selected fish stocks. Journal of Northwest Atlantic Fisheries Science 1:101–111.
- Hilborn, R. 1985. Simplified calculation of optimum spawning stock size from Ricker's stock recruitment curve. Canadian Journal of Fisheries and Aquatic Sciences 42:1833–1834.
- Hilborn, R., and C. J. Walters. 1992. Quantitative fisheries stock assessment and management: choice, dynamics, and uncertainty. Chapman and Hall, New York.
- Jacobson, L. D., and A. D. MacCall. 1995. Stock–recruitment models for Pacific sardine (*Sardinops sagax*). Canadian Journal of Fisheries and Aquatic Sciences 52:566–577.
- MacCall, A. D., and S. Ralston. 2002. Is logarithmic transformation really the best procedure for estimating stock–recruitment relationships? North American Journal of Fisheries Management 22:339–350.
- Meulenbergh, M. T. G. 1965. On the estimation of an exponential function. Econometrica 33:863–868.
- Peterman, R. M. 1981. Form of random variation in salmon smolt-to-adult relations and its influence on production estimates. Canadian Journal of Fisheries and Aquatic Sciences 38:1113–1119.
- Quinn, T. J., and R. B. Deriso. 1999. Quantitative fish dynamics. Oxford University Press, New York.
- Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Fisheries Research Board of Canada Bulletin 191.
- Walters, C. J. 1985. Bias in the estimation of functional relationships from time series data. Canadian Journal of Fisheries and Aquatic Sciences 42:147–149.

Appendix follows

Appendix: Asymptotic Unbiasedness of the Bias Correction

Since $\hat{\Psi}$ and $\hat{\sigma}^2$ are independent, the expectation of equation (11) is

$$E(\tilde{R}_0) = E(\hat{R}_0) \cdot E \left\{ \exp \left[-\frac{\hat{\sigma}^2 \mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0}{2} \right] \right\}.$$

From equation (10),

$$\begin{aligned} E(\tilde{R}_0) &= S_0 \cdot \exp \left[\mathbf{X}_0 \Psi + \frac{\sigma^2 \mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0}{2} \right] \\ &\times \exp \left\{ E \left[-\frac{\hat{\sigma}^2 \mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0}{2} \right] \right. \\ &\quad \left. + \frac{1}{2} V \left[-\frac{\hat{\sigma}^2 \mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0}{2} \right] \right\} \\ &= S_0 \cdot \exp \left[\mathbf{X}_0 \Psi + \frac{\sigma^2 \mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0}{2} \right] \\ &\times \exp \left\{ -\frac{\sigma^2 \mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0}{2} \right. \\ &\quad \left. + \frac{1}{8} V(\hat{\sigma}^2) [\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0]^2 \right\} \end{aligned}$$

$$\begin{aligned} &= S_0 \cdot \exp(\mathbf{X}_0 \Psi) \\ &\quad \times \exp \left\{ \frac{1}{8} V(\hat{\sigma}^2) [\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0]^2 \right\} \\ &= R_0 \cdot \exp \left\{ \frac{1}{8} V(\hat{\sigma}^2) [\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0]^2 \right\}. \end{aligned}$$

Since $(n - 2)\hat{\sigma}^2/\sigma^2$ has a chi-square distribution with $n - 2$ degrees of freedom, based on the properties of that distribution $V[(n - 2)\hat{\sigma}^2/\sigma^2] = 2(n - 2)$ and $V(\hat{\sigma}^2) = 2\sigma^4/(n - 2)$. Substituting the latter into the above equation gives

$$\begin{aligned} E(\tilde{R}_0) &= R_0 \cdot \exp \left\{ \frac{\sigma^4}{4(n - 2)} [\mathbf{X}_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'_0]^2 \right\} \\ &= R_0 \cdot \exp \left[\frac{\sigma^4}{4n(n - 2)} \frac{\sum_{t=1}^n (S_t - S_0)^2}{\sum_{t=1}^n (S_t - \bar{S})^2} \right] \\ &\rightarrow R_0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$