

# THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY  
OF NORTH CAROLINA



DATA TRANSFORMATIONS IN REGRESSION ANALYSIS WITH APPLICATIONS TO STOCK -  
RECRUITMENT RELATIONSHIPS

David Ruppert

and

Raymond J. Carroll

Department of Statistics  
University of North Carolina  
Chapel Hill, NC 27514/USA

Mimeo Series #1578

May 1985

DEPARTMENT OF STATISTICS  
Chapel Hill, North Carolina

DATA TRANSFORMATIONS IN REGRESSION ANALYSIS WITH APPLICATIONS TO STOCK - RECRUITMENT  
RELATIONSHIPS

David Ruppert

and

Raymond J. Carroll

Department of Statistics  
University of North Carolina  
Chapel Hill, NC 27514/USA

Abstract

We propose a methodology for fitting theoretical models to data. The dependent variable (or response) and the model are transformed in the same way. Two types of transformations, power transformation and weighting, are used together to remove skewness and to induce constant variance. Our method is applied to the stock-recruitment data of four fish stocks. Also discussed are estimates of the conditional mean and the conditional quantiles of the original response.

1. Introduction

In regression analysis one seeks to establish a relationship between a response  $y$  and independent variables  $\underline{x} = (x_1, \dots, x_k)'$ . Often the physical or biological system generating the data suggest that in the absence of random error  $y = f(\underline{x}, \underline{\theta})$  where  $f$  is a known function and  $\underline{\theta}$  is an unknown parameter. In reality random variability, modeling errors, and measurement errors will cause  $y$  to deviate from  $f(\underline{x}, \underline{\theta})$ .

Usually  $\underline{\theta}$  is estimated by the (possibly nonlinear) least-squares estimator  $\hat{\underline{\theta}}$  which minimizes

$$\sum_{t=1}^N (y_t - f(\underline{x}_t, \hat{\underline{\theta}}))^2$$

where  $y_t$  and  $\underline{x}_t = (x_{1t}, \dots, x_{kt})'$  are the  $t$ -th observations of the response and the independent variables,  $t=1, \dots, N$ . The method of least squares is not uniquely determined in the following sense. If  $h(y)$  is a monotonic function then  $y = f(\underline{x}, \theta)$  implies that  $h(y) = h(f(\underline{x}, \theta))$ . In the absence of random error, there is an infinity of possible models, one for each  $h$ . Taking  $h(y)$  to be the new dependent variable and  $h(f(\underline{x}, \theta))$  to be the new regression model, the least-squares estimate  $\hat{\theta}(h)$ , depending on  $h$ , minimizes

$$\sum_{t=1}^N \{h(y_t) - h(f(\underline{x}_t, \hat{\theta}))\}^2 .$$

This paper addresses the question "how should  $h$  be chosen?" In the past  $h$  was often chosen so that the model was linear in the parameters, but with the wide availability of nonlinear least squares software linearization is no longer necessary. Instead  $h$  should be selected so that  $\hat{\theta}(h)$  has high statistical efficiency in some sense, say low asymptotic mean square error. For nonlinear models finite-sample mean square errors are seldom known and large-sample, or asymptotic, mean square errors must be used.

Least-squares estimation is asymptotically efficient when the errors are additive, normally distributed, and homoscedastic, that is, if

$$y_t = f(\underline{x}_t, \theta) + \epsilon_t , \quad (1)$$

where  $\epsilon_1, \dots, \epsilon_N$  are independent  $N(0, \sigma^2)$  random variables for some  $\sigma^2 > 0$ . If (1) fails to hold for the original response and model, it may hold after these have been transformed. For example, if the errors are multiplicative and lognormal, then

$$\log(y_t) = \log(f(\underline{x}_t, \theta)) + \epsilon_t$$

so that  $h(y) = \log(y)$  is the appropriate transformation.

In general, it is impossible to know a priori how the random errors affect  $y$ . It is, however, often reasonable to assume as an approximation that for an unknown  $h$

$$h(y_t) = h(f(\underline{x}_t, \theta)) + \epsilon_t \text{ and } \epsilon_t \sim N(0, \sigma^2),$$

and then to estimate  $h$  from the data.

Carroll and Ruppert (1984) introduced a methodology where  $h$  is assumed to belong to a parametric class such as the class of power functions. Specifically,  $h(y) = h(y, \lambda)$  where  $h(y, \lambda)$  is a known parametric family and  $\lambda$  is an unknown parameter. Then  $\theta$ ,  $\sigma$ , and  $\lambda$  are estimated simultaneously, for example by maximum likelihood.

Box and Cox (1964) used a modified power transformation family

$$\begin{aligned} y^{(\lambda)} &= (y^\lambda - 1)/\lambda & \lambda \neq 0 \\ &= \log(y) & \lambda = 0, \end{aligned}$$

which includes the log transformation in a natural, i.e., continuous, fashion. It should be mentioned that Box and Cox were concerned with a quite different transformation methodology. They transform only the response, not the model; see Carroll and Ruppert (1984).

Another transformation family which, as we will see, is useful in stock-recruitment analysis is division by  $(u_t)^\alpha$  where  $u_t$  is a known constant and  $\alpha$  is an unknown parameter. Since  $u_t$  is a constant which does not depend on  $y_t$ , a transformation of this kind has no effect on skewness but it induces homoscedasticity if  $(u_t)^\alpha$  is the standard deviation of  $y_t$ . The constant  $u_t$  can be one of the independent variables, a variable not depending on  $x_t$  or  $y_t$ , or a function of such variables. Division by  $(u_t)^\alpha$  will be called a "power weighting transformation". Note that  $u_t^\alpha$  denotes ordinary, not Box-Cox, power transformation of  $u_t$  so that  $u_t^\alpha = 1$  when  $\alpha = 0$ .

This paper is restricted to a model combining a Box-Cox power transformation with a power weighting transformation:

$$y_t^{(\lambda)} / u_t^\alpha = f^{(\lambda)}(\underline{x}_t, \underline{\theta}) / u_t^\alpha + e_t. \quad (2)$$

However, the methodology that is developed here can be applied to other transformation families.

According to model (2),  $y_t^{(\lambda)}$  is symmetrically distributed about  $f^{(\lambda)}(\underline{x}_t, \underline{\theta})$ . Therefore,  $f^{(\lambda)}(\underline{x}_t, \underline{\theta})$  gives both the conditional (given  $\underline{x}_t$ ) mean and median of  $y_t^{(\lambda)}$ . This implies further that the untransformed model  $f(\underline{x}_t, \underline{\theta})$  gives the conditional median of the untransformed response  $y_t$ . However, the conditional mean of  $y_t$  is not  $f(\underline{x}_t, \underline{\theta})$  except if  $\lambda = 1$ . The conditional mean is discussed below. We see then that  $f(\underline{x}_t, \underline{\theta})$  has two interpretations; it gives the value of  $y_t$  if there is no error, and otherwise it gives the conditional median of  $y_t$ .

In our examples,  $\underline{x}_t = x_{1t}$  is the size of a spawning stock  $S_t$  and the response  $y_t$  is the total return  $R_t$ . Commonly used model functions  $f(\underline{x}, \underline{\theta})$  include the Ricker (1954) model

$$R_t = S_t \exp(\theta_1 + \theta_2 S_t), \quad (3)$$

and the Beverton-Holt (1957) model

$$R_t = 1/(\theta_1 + \theta_2/S_t). \quad (4)$$

We will let  $u_t = S_t$ . Then (3) can be transformed to

$$R_t/S_t = \exp(\theta_1 + \theta_2 S_t) \quad (5)$$

by letting  $\lambda=1$  and  $\alpha=1$ . By using  $\lambda=0$  and  $\alpha=0$  on (5), one obtains

$$\log(R_t/S_t) = \theta_1 + \theta_2 S_t. \quad (6)$$

Equation (6) is often favored since it is linear in the parameters.

Equation (4) can be linearized to

$$1/R_t = \theta_1 + \theta_2 (1/S_t) \quad (7)$$

or

$$S_t/R_t = \theta_2 + \theta_1 S_t \quad (8)$$

corresponding to  $\lambda=1$  and  $\alpha=0$  and  $\lambda=1$  and  $\alpha=-1$ , respectively.

By estimating  $\lambda$  and  $\alpha$  we can tell when these linearizing transformations are appropriate for a given set of data, and we can find more suitable transformations when the linearization is not appropriate.

Although we advocate transforming the response to achieve an error structure where least-squares is efficient, often one must estimate characteristics (such as the conditional mean) of the original response. In section 5 estimates of the conditional mean and the conditional quantiles of  $y_t$  given  $x_t$  are presented.

## 2. Power transformations, skewness, and heteroscedasticity

In this section, we discuss how power transformations affect, and in particular remove, skewness and heteroscedasticity. Suppose that the random variable  $y_t$  has mean  $m_t$  and variance  $\sigma_t^2 = g(m_t)$ ; the same function  $g$  applies for each  $t$ . If  $y_t$  is transformed to  $h(y_t)$ , then by a Taylor approximation as in Bartlett (1947), the variance of  $h(y_t)$  is

$$\begin{aligned} E\{h(y_t) - E h(y_t)\}^2 &\approx \\ E\{h(y_t) - h(m_t)\}^2 & \\ \approx (\dot{h}(m_t))^2 E\{y_t - m_t\}^2 & \\ = (\dot{h}(m_t))^2 g(m_t) & \end{aligned}$$

where  $\dot{h}(y) = d/dt h(y)$ . The variance of  $h(y_t)$  is approximately constant if  $\dot{h}(y)$  is proportional to  $g^{-1/2}(y)$ .

In many cases  $g$  is a power function. For example,  $g(m) = m$  for Poisson-distributed data and  $g(m)$  is proportion to  $m^2$  if the coefficient of variation (CV) is constant. Also,  $g(m) = m^2$  for exponentially distributed data. When  $g(m) \propto m^{2(1-\lambda)}$  then  $\dot{h}(m) \propto g^{-1/2}(m)$  if  $h(y) = y^{(\lambda)}$ . In the case  $\lambda = 0$  (constant CV), the log transformation stabilizes the variance, and for Poisson data,  $\lambda = 1/2$ , the square-root transformation is indicated.

The effect of  $h$  on skewness is also revealed by  $\dot{h}$ . Suppose that  $y$  is positively skewed. The extended right tail in the distribution of  $y$  is reduced through transformation to  $h(y)$  if large values of  $y$  are "compressed together" more

than small values; more precisely, if for all  $\Delta > 0$ ,  $|h(y_1 + \Delta) - h(y_1)| < |h(y_2 + \Delta) - h(y_2)|$  whenever  $y_1 > y_2$ . Such compression occurs if  $h(y)$  is a decreasing function, in which case  $h$  is called concave.

When  $h(y)$  is a Box-Cox power transformation then  $h(y) = (d/dy) \lambda^{-1} (y^\lambda - 1)$  when  $\lambda \neq 0$  or  $h(y) = (d/dy) \log y$  when  $\lambda = 0$ . In either case,  $h(y) = y^{\lambda-1}$ , and consequently  $h$  is concave if  $\lambda \leq 1$ . Box-Cox transformations reduce right-skewness when  $\lambda < 1$  and the amount of reduction increases as  $\lambda$  decreases.

Left skewness is reduced by transformations that are convex, that is, which have an increasing derivative. If  $\lambda \geq 1$ , then  $y^{(\lambda)}$  is convex.

### 3. Simultaneous power transformation and weighting by maximum likelihood.

Although power transformations can reduce both skewness and heteroscedasticity, the value of  $\lambda$  inducing normality, or at least a reasonably symmetric distribution, need not be the same  $\lambda$  which transforms to constant variance. A more flexible approach to modeling combines a Box-Cox power transformation to normality with a power weighting transformation to homoscedasticity as in equation (2). In this section we discuss the estimation of  $\lambda$ ,  $\alpha$ ,  $\theta$ , and  $\sigma$  by maximum likelihood and the construction of a confidence region for  $\lambda$  and  $\alpha$  by likelihood-ratio testing.

To find the likelihood, we first note that the density of  $\epsilon_t$  is

$$(2\pi \sigma^2)^{-1/2} \exp(-\epsilon^2 / (2 \sigma^2))$$

and the Jacobian of  $\epsilon_t \rightarrow y_t^{(\lambda)}$  is  $y_t^{\lambda-1} / u_t^\alpha$ . Therefore, the conditional density of  $y_t$  given  $x_t$  and  $u_t$  is

$$f(y_t | x_t, u_t, \theta, \sigma, \alpha, \lambda) = (2\pi \sigma^2)^{-1/2} (y_t^{\lambda-1} / u_t^\alpha) \exp(-[y_t^{(\lambda)} - f^{(\lambda)}(x_t, \theta)]^2 / (2 \sigma^2 u_t^{2\alpha})).$$

If  $x_t$  and  $u_t$  are fixed constants, not depending on  $y_1, \dots, y_{t-1}$ , then the log-likelihood of  $y_1, \dots, y_N$  is

$$\begin{aligned} L(\lambda, \alpha, \theta, \sigma^2) &= -N/2 \log(2\pi \sigma^2) \\ &+ (\lambda-1) \sum_{t=1}^N \log(y_t) - \alpha \sum_{t=1}^N \log(u_t) \\ &- (1/2 \sigma^2) \sum_{t=1}^N (y_t^{(\lambda)} - f^{(\lambda)}(x_t, \theta))^2 / u_t^{2\alpha}. \end{aligned}$$

If  $x_t$  and  $u_t$  depend on previous values of  $y$ , then  $L(\lambda, \alpha, \theta, \sigma^2)$  is still the

conditional log-likelihood of  $y_1, \dots, y_u$  given  $x_1, x_0, x_{-1}, \dots$  and  $u_1, u_0, u_{-1}, \dots$ , and maximization of  $L(\lambda, \alpha, \theta, \sigma^2)$  should still produce good estimates of  $\lambda, \alpha, \theta, \sigma^2$ . The distinction between conditional and unconditional estimates from time series data is discussed in Box and Jenkins (1976, Chapter 7). The dependence of  $x_t$  and  $u_t$  on past  $y$  may, however, produce biases. Walters (1985) found sizeable biases in a Monte Carlo study with small sample sizes,  $N = 10$ . His study assumed that  $\lambda$  was known and equal to 0. His formulas show that the bias should decrease as  $N$  increases.

Following Box and Cox (1964), we maximize  $L(\lambda, \alpha, \theta, \sigma^2)$  in two stages. First, for fixed  $\lambda$  and  $\alpha$  the MLE of  $\theta$  is the nonlinear, weighted least-squares estimator  $\hat{\theta}(\lambda, \alpha)$  which minimizes

$$SS(\lambda, \alpha, \theta) = \sum_{t=1}^N \{y_t^{(\lambda)} - f^{(\lambda)}(x_t, \theta)\}^2 / u_t^{2\alpha}$$

and the MLE of  $\sigma^2$  is  $\sigma^2(\lambda, \alpha) = N^{-1} SS(\lambda, \alpha, \hat{\theta}(\lambda, \alpha))$ . Define

$$L_{\max}(\lambda, \alpha) = L(\lambda, \alpha, \hat{\theta}(\lambda, \alpha), \sigma^2(\lambda, \alpha))$$

to be the log-likelihood maximized with respect to  $\theta$  and  $\sigma$  for fixed  $\lambda$  and  $\alpha$ . An approximate MLE is found by computing  $L_{\max}(\lambda, \alpha)$  on a grid; in section 4 we use  $\alpha = -1(.25)1$  and  $\lambda = -1(.25)1$ . If the exact MLE is needed then  $L_{\max}(\lambda, \alpha)$  can be maximized by a numerical optimization technique using the approximate MLE as an initial value. However, an exact MLE is probably unnecessary for most applications.

The function  $L_{\max}(\lambda, \alpha)$  can be used to test hypotheses and to construct confidence regions for  $(\lambda, \alpha)$ . For a given null value  $(\lambda_0, \alpha_0)$ , one can test  $H_0: (\lambda, \alpha) = (\lambda_0, \alpha_0)$  against  $H_1: (\lambda, \alpha) \neq (\lambda_0, \alpha_0)$  by the likelihood ratio test. The log-likelihood ratio is

$$LR(\lambda_0, \alpha_0) = L_{\max}(\hat{\lambda}, \hat{\alpha}) - L_{\max}(\lambda_0, \alpha_0).$$

Here  $(\hat{\lambda}, \hat{\alpha})$  is the MLE or approximate MLE.  $H_0$  is rejected at level  $\epsilon$  if  $2 \ln LR(\lambda_0, \alpha_0)$  exceeds  $\chi^2(1, 1-\epsilon)$ , the  $100(1-\epsilon)$  percentile of the chi-square distribution with one degree of freedom (Rao, 1973, section 6e). This test can be applied to each value  $(\lambda_0, \alpha_0)$  on the grid, and a  $100(1-\epsilon)$  percent confidence region for  $(\lambda, \alpha)$  consists of all null values which are not rejected.

#### 4. Examples

##### 4.1 Population A

These data and the population B data discussed later were obtained through Professor Carl Walters (pers. comm.). Permission to identify the stocks has been refused by the original source. There are twenty-eight years of data. The

variables,  $R_t$  = total return and  $S_t$  = spawner escapement, are plotted in figure 1.

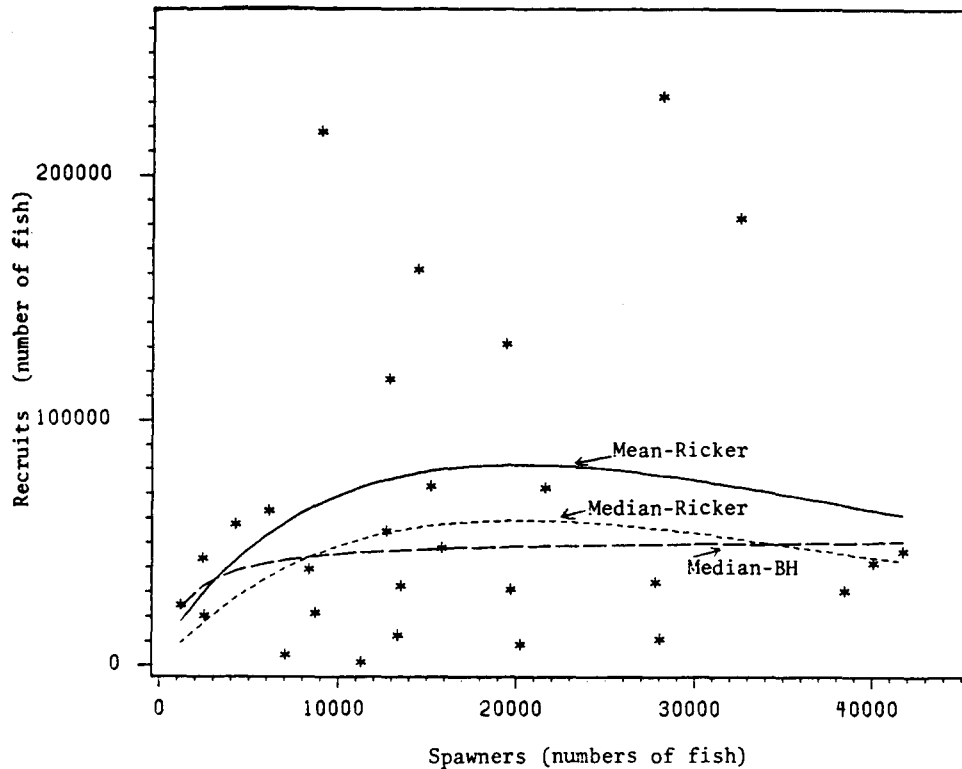


Fig. 1: Population A. Actual recruitment from 1940 to 1967, estimated median recruitment based on the Ricker and Beverton-Holt (BH) models, and estimated mean recruitment based on the Ricker model.

We first fit the (transformed) Ricker model:

$$R_t^{(\lambda)} / S_t^\alpha = (S_t \exp(\theta_1 + \theta_2 S_t))^{(\lambda)} / S_t^\alpha + \epsilon_t . \quad (9)$$

The approximate MLE (on the grid  $\lambda = -1(.25)1$  and  $\alpha = -1(.25)1$ ) is  $(\hat{\lambda}, \hat{\alpha}) = (.25, 0)$  and the maximized log-likelihood is  $-338.3106$ . With  $(\lambda, \alpha)$  equal to the MLE, model (9) becomes

$$\sqrt[4]{R_t} = \sqrt[4]{S_t} \exp((\theta_1 + \theta_2 S_t)/4) + \epsilon_t .$$

Confidence regions for  $(\lambda, \alpha)$  are given in Table 1. The 95% univariate confidence regions for  $\lambda$  and  $\alpha$  are  $\{0, .25, .5\}$  and  $\{-.25, 0, .25, .5, .75\}$ , respectively.



Lambda	Alpha				
	-.25	0	.25	.5	.75
0	*	*	*		
	+	+	+	+	
.25	*	**	*	*	
	+	+	++	+	+
.5		*	*	*	*
		+	+	+	+

\* = 95% confidence region assuming the Ricker model.

\*\* = MLE assuming the Ricker model.

+ = 95% confidence region assuming the Beverton-Holt model.

++ = MLE assuming the Beverton-Holt model.

Table 1: Population A. Confidence regions and maximum likelihood estimates of lambda and alpha.

Next the transformed Beverton-Holt model

$$R_t^{(\lambda)}/S_t^\alpha = \{1/(\theta_1 + \theta_2/S_t)\}^{(\lambda)}/S_t + \epsilon_t$$

was fit. The MLE is  $(\hat{\lambda}_1, \hat{\alpha}) = (.25, .25)$  and the maximum log-likelihood is -337.0919. The difference between the maximum log-likelihood for the Beverton-Holt and Ricker models is only 1.2187 which indicates that both models fit almost as well, though by this criterion the Beverton-Holt model does provide a slightly better fit.

The estimated median return, calculated as described in section 5, is plotted in figure 1 for both the Ricker and Beverton-Holt models, as is the "smearing estimate (section 5) of the mean assuming the Ricker model.

A researcher with only linear regression software might be tempted to linearize both the Ricker and Beverton-Holt models and then to compare them on their linearizing scales. When the linearized Ricker model

$$\log(R_t/S_t) = \theta_1 + \theta_2 S_t$$

is fit  $R^2 = 0.207$  and the F-value for testing overall significance of the model is 6.79 ( $p = 0.015$ ). If the linearized Beverton-Holt model

$$1/R_t = \theta_1 + \theta_2/S_t$$

is fit then  $R^2 = 0.000549$ ,  $F = 0.01$ ,  $p = .9058$ , and if the alternative linearized model

$$S_t/R_t = \theta_2 + \theta_1 S_t$$

is fit then  $R^2 = 0.110$ ,  $F = 0.29$ , and  $p = .5950$ . A researcher comparing these models only on their linearizing scales might easily be tempted into concluding that the Ricker model provides a far better fit. On the contrary, we believe that the Beverton-Holt model is very slightly better fitting, but the log transformation is vastly superior to the inverse transformation.

It is interesting to see how well the MLE transformation achieves both homoscedasticity and near normality. In table 2 the skewness and kurtosis are given for the three transformed Ricker models:

$$(I) \quad R/S = \exp(\theta_1 + \theta_2 S) \quad (\lambda=1, \alpha=1)$$

$$(II) \quad \log(R) = \log(S) + \theta_1 + \theta_2 S \quad (\lambda=0, \alpha=0)$$

$$(III) \quad \sqrt[4]{R} = \sqrt[4]{S} \exp\{(\theta_1 + \theta_2 S)/4\} \quad (\lambda=.25, \alpha=0),$$

that is, no power transformation, log transformation and the MLE, respectively.

Lambda	Alpha	Skewness	Kurtosis	Spearman Correlation
1	1	.99	2.69	.39
0	0	-1.04	1.22	.12
.25	0	-.27	-.23	.14

Table 2: Population A. Skewness and kurtosis of residuals and Spearman rank correlation between the absolute residuals and the predicted values. The Ricker model was used.

For model (III) both skewness and kurtosis are closest to 0, their theoretical value under normality. Also in table 2 are the Spearman rank correlations between the absolute residuals and the fitted values. Since the fitted values are an increasing function of S, the rank correlation is unchanged if the fitted values are replaced by S. Clearly,  $\lambda=0$  and  $\lambda=.25$  both transform to near homoscedasticity where there is little or no relationship between the mean and variance of R. However, the MLE  $\lambda=.25$  is preferable to  $\lambda=0$  since  $\lambda=0$  "overtransforms" to negative skewness.

Normality and homoscedasticity of the residuals can be checked graphically by a normal plot of the residuals and a plot of the residuals against the fitted values. The graphical analysis of residuals should be done routinely for transformation models as for ordinary regression models. Draper and Smith (1981) provide an excellent account of graphical residual analysis. We plotted the residuals from the MLE and found no evidence of non-normality or heteroscedasticity. A normal

probability plot of the residuals is roughly linear, though the very slight negative skewness (table 2) is evident. For these data, a plot of the residuals versus the predicted values is not easy to interpret; the Ricker and Beverton-Holt curves are nearly constant over the observed range of S so the predicted values are quite similar except for those corresponding to the smallest values of S.

#### 4.2 Skeena River sockeye salmon

For this stock, effective escapement (S) and total return (R) are given for years 1940 to 1967 in Ricker and Smith (1975). These data are plotted in figure 2 along with the estimated medians (section 5) for the Ricker and Beverton-Holt models. Compared with the Population A stock, the Skeena River data show less positive skewness but more heteroscedasticity. Figure 2 shows several particularly low values of R associated with high values of S but only one particularly high value. This may indicate overcompensation, or it may simply be due to the high variability in R when S is large.

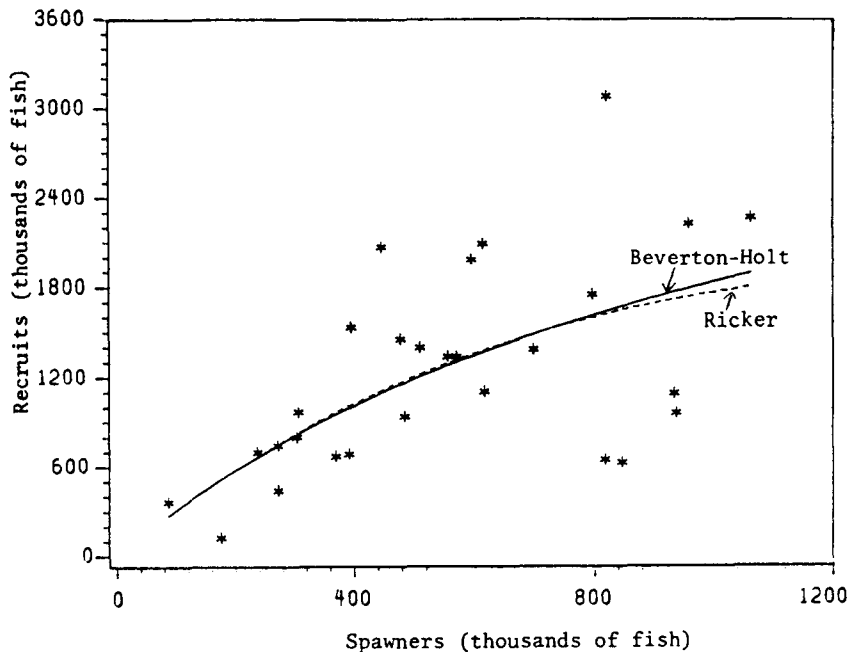


Fig. 2: Skeena River sockeye. Actual recruitment from 1940 to 1967 and estimated median recruitment based on Beverton-Holt and Ricker models.

The 95% confidence regions and the MLE of  $(\lambda, \alpha)$  are the same for the Ricker and Beverton-Holt models and they are given in table 3. The maximum log-likelihoods for the Ricker and for the Beverton-Holt models differ by only 0.164, so there is little to suggest one model over the other. In particular, the data provide no strong evidence of overcompensation.

Lambda	Alpha					
	-.25	0	.25	.5	.75	1
0	*	*				
.25	*	*	*	*		
.5		*	*	*	*	
.75		*	*	**	*	*
1			*	*	*	*

\* = 95% confidence region.

\*\* = MLE

Table 3: Skeena River sockeye salmon stock. 95% confidence region and maximum likelihood estimate assuming either the Ricker or Beverton-Holt model.

Compared with the Population A data, the MLE here results in less transformation ( $\hat{\lambda} = .75$  instead of  $\hat{\lambda} = .25$  as before) but more weighting ( $\hat{\alpha} = .5$  instead of  $\hat{\alpha} = 0$  or  $.25$ ). This is consistent with the observation that the untransformed Skeena data exhibit little skewness but considerable heteroscedasticity.

#### 4.3 Pacific Cod, Necate Strait (Walters, et al., 1982)

The twenty-one observations, 1959 to 1979, on this stock are plotted in figure 3 along with predicted medians for two models.

Recruitment is virtually independent of spawning stock except that the two unusually large recruitments occur when S is rather small. For this reason the Beverton-Holt model fits poorly. For this stock we will consider the Ricker model and the power model

$$R_t = \theta_1 S_t^{\theta_2}$$

With Box-Cox and weighting transformations the power model becomes

$$R_t^{(\lambda)} / S_t^{\alpha} = (\theta_1 S_t^{\theta_2})^{(\lambda)} / S_t^{\alpha} + \epsilon_t$$

Confidence regions for  $(\lambda, \alpha)$  are given in table 4. The maximum log-likelihood is -34.9690 and -37.2541 for the power and Ricker models, respectively, and the large difference (2.2851) indicates lack-of-fit for the Ricker model.

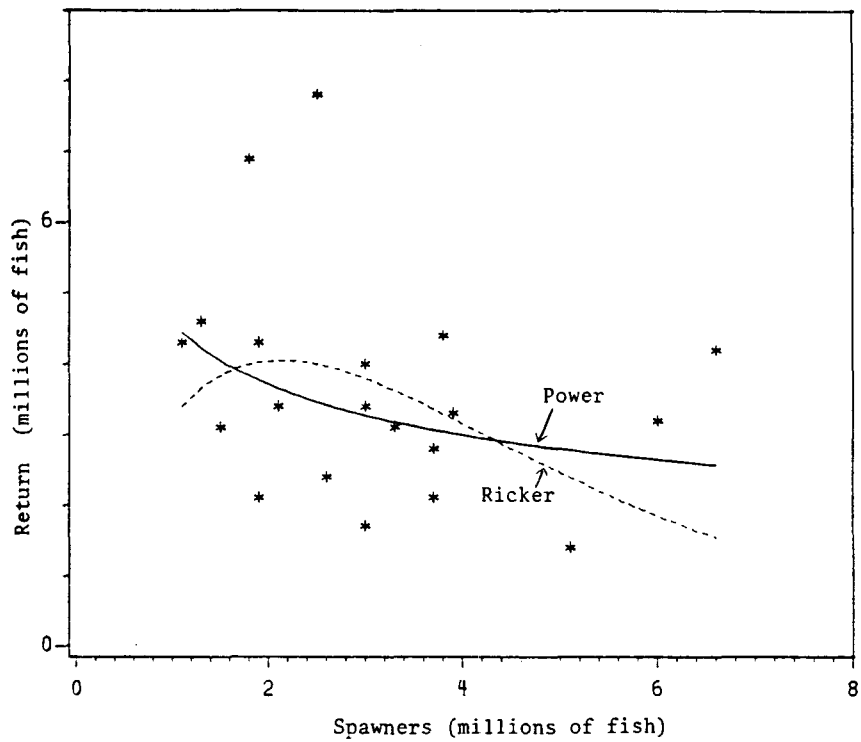


Fig. 3: Necate Strait Pacific Cod. Actual recruitment from 1959 to 1979 and estimated median recruitment based on the Power and Ricker models.

Lambda	Alpha								
	-1	-0.75	-0.5	-0.25	0	.25	.5	.75	1
-1						*	*	*	*
-0.75					*	*	*	*	*
-0.5				*	*	*	*	*	*
-0.25			*	*	*	*	*	*	*
0			*	*	*	*	**	*	*
.25			*	*	**	**	*	*	*
.5		*	*	*	*	*	*	*	
.75		*	*	*	*	*	*		
1			*	*	*	*			

\* = 95% Confidence region.

\*\* = MLE or within .1 of maximizing the log-likelihood.

Table 4. Hecate Strait Pacific cod stock. 95% confidence region and maximum likelihood estimate assuming the power model.

We can test the Ricker model as follows. The general model

$$R_t = S_t^{\theta_3} \exp(\theta_1 + \theta_2 S_t)$$

includes the Ricker model ( $\theta_3 = 1$ ) and the power model ( $\theta_2 = 0$ ) as special cases. One tests the Ricker model by testing  $H_0: \theta_3 = 1$  against  $H_1: \theta_3 \neq 1$ . The maximum log-likelihood is -34.9022 and the log-likelihood maximized subject to the constraint  $\theta_3 = 1$  is -37.2541 (see above). Twice the difference is 4.704 which exceeds  $X^2(1, .95) = 3.84$ . The Ricker model is reject at level .05. By the same reasoning the power model is accepted since  $34.9690 - 34.9022 = .0668$  is very small.

The 95% confidence region for  $(\lambda, \alpha)$  assuming the power model is quite large because (a) there are only 21 data points and, more importantly, (b) the variability in both spawning stock and return is small compared to the previously examined stocks. The confidence region for the Ricker model is even larger, but this is to be expected considering the lack of fit.

For the power model, the MLE is  $(\hat{\lambda}, \hat{\alpha}) = (0, .5)$ , but the log-likelihood at  $\lambda = .25$  and  $\alpha = .25$  or  $.5$  is within 0.1 of the maximum.

#### 4.4 Population B (from Professor Carl Walter, pers. comm.)

For this stock,  $R_t$  and  $S_t$  vary over a much wider range than for the three preceding stocks, and it seems preferable to use the return to spawner ratio,  $R_t/S_t$ , as the response. A plot of  $\log(R_t/S_t)$  against  $S_t$  is rather linear, but most of the data are bunched together in the range  $0 \leq S_t \leq 300,000$  while the remainder are scattered over  $300,000 \leq S_t \leq 3,300,000$ , so in figure 4  $\log(R/S)$  is plotted against  $\log(S)$ .

The only model studied here is the transformed Ricker model

$$(R_t/S_t)^{(\lambda)} / S_t^\alpha = \{\exp(\theta_1 + \theta_2 S_t)\}^{(\lambda)} / S_t^\alpha + e_t.$$

Also the grid  $\alpha = -1(1/4)1$  is replaced by  $\alpha = -1/4(1/16)1/4$  because values of  $\alpha$  far from 0 fit poorly and lead to convergence problems when  $\hat{\theta}_1$  and  $\hat{\theta}_2$  are computed. The 95% confidence region for  $(\lambda, \alpha)$  and the MLE are found in table 5. Clearly, the confidence region is quite small. Because  $R_t$  and  $S_t$  vary over extremely wide ranges,  $\lambda$  and  $\alpha$  are well determined by the data.

#### 5. Estimating the Conditional Distribution of y

We have seen how to utilize the model

$$y_i^{(\lambda)} / u_i^\alpha = f^{(\lambda)}(x_i, \theta) / u_i^\alpha + e_i$$

to efficiently estimate  $\theta$ . The model expresses a transformed response as a function

of  $\theta$ ,  $\underline{x}_i$ , and the normally (or approximately normally) distributed  $\epsilon_i$ . Typically interest centers on the untransformed response  $y_t$ . In this section it is shown how to estimate the conditional (given  $\underline{x}_i$ ) mean of  $y_i$  as well as conditional quantiles such as the median.

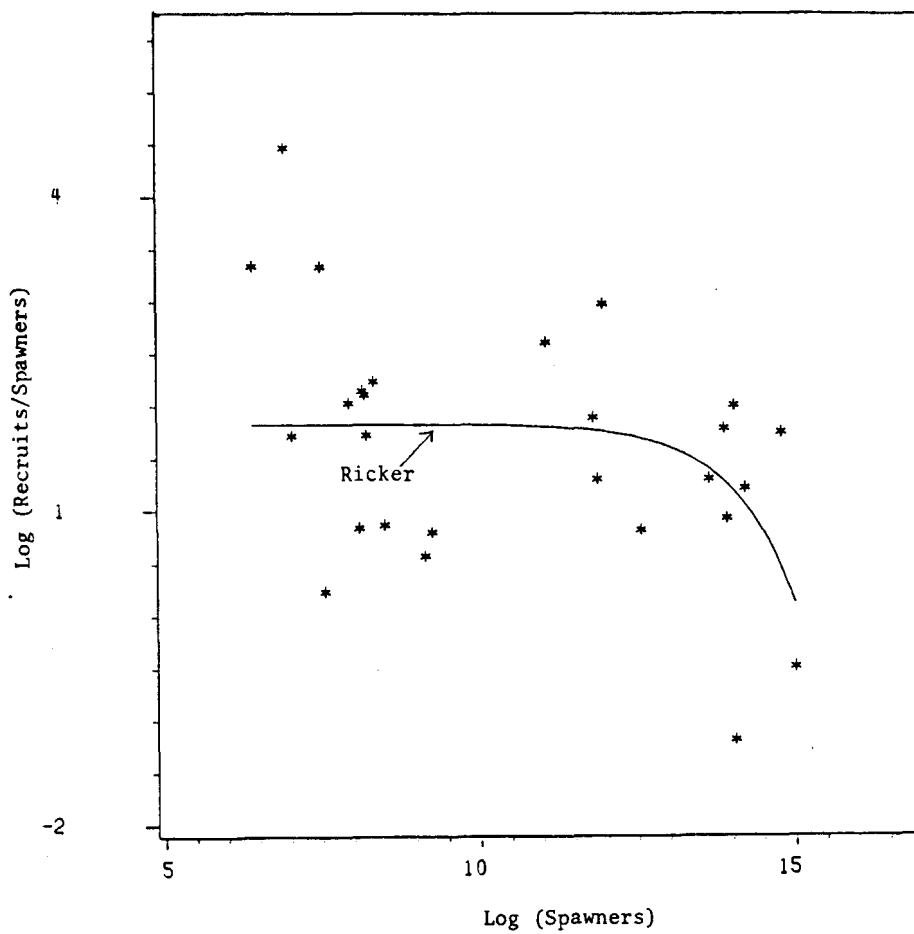


Fig. 4: Population B. Estimated median production ratio based on the Ricker model. Recruits and spawners are in numbers of fish. The production ratio and spawners are expressed in logarithms.

Lambda	Alpha								
	-4/16	-3/16	-2/16	-1/16	0	1/16	2/16	3/16	4/16
-0.25	*	*							
0			*	**	*	*			
.25						*	*	*	*

\* = in 95% confidence regions

\*\* = MLE

Table 5. Population B. 95% confidence region assuming the Ricker model  $R/S = \exp(\theta_1 + \theta_2 S)$ .

Now  $\epsilon_i$  cannot be exactly normally distributed in most cases, for example when  $y_i$  and  $f(x, \theta)$  are non-negative and  $\lambda \neq 0$ . It is better to suppose that  $\epsilon_i$  is nearly normal but with a bounded range. Fortunately, for present purposes we need not make any assumption about the  $\{\epsilon_i\}$  except that they are independent and identically distributed.

If  $y^{(\lambda)} = x$ , then the inverse relationship is  $y = (1 + \lambda x)^{1/\lambda}$  if  $\lambda \neq 0$  and  $y = \exp(x)$  if  $\lambda = 0$ . In what follows we assume that  $\lambda \neq 0$ . If  $\lambda = 0$ , then one simply replaces  $(1 + \lambda x)^{1/\lambda}$  by  $\exp(x)$ . Using (2) we can express the original response  $y_i$  as a function of  $\underline{x}_i$ ,  $u_i$ ,  $\epsilon_i$  and the parameters  $\lambda, \alpha, \theta$ :

$$y_i = \{1 + \lambda f^{(\lambda)}(\underline{x}_i, \underline{\theta}) + \lambda u_i^\alpha \epsilon_i\}^{1/\lambda}.$$

With this representation we can study the untransformed  $y_i$ . Let  $F$  be the distribution function of  $\epsilon_1, \dots, \epsilon_N$ . The conditional mean of  $y_i$  given  $\underline{x}_i$  is

$$E(y_i | \underline{x}_i) = \int \{1 + \lambda f^{(\lambda)}(\underline{x}_i, \underline{\theta}) + \lambda u_i^\alpha \epsilon\}^{1/\lambda} dF(\epsilon). \quad (10)$$

Let  $q_p$  be the  $p$ th quantile of  $F$ , i.e.,  $F(q_p) = p$ . Then the conditional  $p$ th quantile of  $y_i$  given  $\underline{x}_i$  is

$$q_p(y_i | \underline{x}_i) = \{1 + \lambda f^{(\lambda)}(\underline{x}_i, \underline{\theta}) + \lambda u_i^\alpha q_p\}^{1/\lambda}$$

Duan (1983) has proposed the "smearing estimate" of  $E(y_i | \underline{x}_i)$ :



$$\hat{E}(y_i | x_i) = N^{-1} \sum_{t=1}^N \{1 + \hat{\lambda} f^{(\hat{\lambda})}(x_i, \hat{\theta}) + \hat{\lambda} u_i^{\hat{\alpha}} \hat{\epsilon}_t\}^{1/\hat{\lambda}} \quad (11)$$

-where  $\epsilon_t$  is the  $t$ -th residual,

$$\hat{\epsilon}_t = (y_t^{(\hat{\lambda})} - f^{(\hat{\lambda})}(x_t, \hat{\theta})) / u_t^{\hat{\alpha}}.$$

The relationship between (10) and (11) is clear; all parameters are replaced by their estimates and averaging with respect to the theoretical distribution of  $\epsilon_i$  is replaced by averaging over the sample  $\hat{\epsilon}_1, \dots, \hat{\epsilon}_N$ . Even if  $F$  were known, (9) could only be evaluated by numerical integration, but (10) does not require integration. This is a distinct advantage of the smearing estimate.

Let  $\hat{q}_p$  be the  $p$ -th sample quantile of  $\{\hat{\epsilon}_i\}$ . Then we estimate  $q_p(y_i | x_i)$  by

$$\hat{q}_p(y_i | x_i) = \{1 + \hat{\lambda} f^{(\hat{\lambda})}(x_i, \hat{\theta}) + \hat{\lambda} u_i^{\hat{\alpha}} \hat{q}_p\}^{1/\hat{\lambda}}.$$

In the case of the median ( $p = .5$ ), the residuals should have a median close to 0 and we can replace  $\hat{q}_{.5}$  by 0. The estimated conditional median is then

$$\begin{aligned} \hat{m}(y_i, x_i) &= \hat{q}_{.5}(y_i | x_i) \\ &= \{1 + \hat{\lambda} f^{(\hat{\lambda})}(x_i, \hat{\theta})\}^{1/\hat{\lambda}} = f(x_i, \hat{\theta}). \end{aligned}$$

As mentioned before, in figure 1  $\hat{E}(y | x_i)$  and  $\hat{m}(y | x)$  are plotted for Population A assuming the Ricker model, and  $\hat{m}(y | x)$  is plotted for the Beverton-Holt model as well. The mean return is always considerably larger than the median return. This reflects the considerable positive skewness seen in the actual recruitments and evident in the MLE of  $\lambda$ ,  $\hat{\lambda} = .25$ .

Because recruitment is so highly variable any realistic management model will be stochastic, and when a stochastic model is constructed it is vital that the entire conditional distribution of  $R_t$  given  $S_t$  be estimated. This can be done by estimating conditional quantiles as above. Ruppert, Reish, Deriso, and Carroll (1985) use a closely related method for estimating conditional quantiles when constructing a stochastic model of the Atlantic menhaden population. Transformation of the menhaden stock-recruitment data is discussed further in Carroll and Ruppert (1984).

#### Summary and Conclusions

A theoretical model relating a response  $y$  to independent variables  $x$  and parameters  $\theta$  may not be suitable in its original form for least-squares estimation. This is the case if the response exhibits skewness or nonconstant variance.

However, if the response and the model function are transformed in the same way, then the transformed response may be approximately normally distributed with a nearly constant variance and then least-squares estimation will be efficient.

In this paper we propose combining the Box-Cox power transformation with weighting by  $u_t^\alpha$ , where  $u_t$  is a variable not depending on  $y_t$  and  $\alpha$  is a parameter to be estimated. Another possibility, one that is not explored here, is to have  $u_t$  be the predicted value  $f(\underline{x}_t, \hat{\theta})$ . Weighting the untransformed response by  $|f(\underline{x}_t, \hat{\theta})|^\alpha$  is studied in Box and Hill (1974), Pritchard, Downie, and Bacon (1977) and Carroll and Ruppert (1982).

The Box-Cox parameter  $\lambda$  and the power weighting parameter  $\alpha$  can be estimated simultaneously with  $\theta$  and  $\sigma$  by maximum likelihood. A confidence region for  $(\lambda, \alpha)$  can be constructed by likelihood-ratio testing.

Four stock-recruitment data sets were analyzed by our transformation methodology. The original response, return, is in each case skewed or heteroscedastic. Only for the Pacific cod stock is  $\lambda=1$  and  $\alpha=0$ , corresponding to no transformation, in the 95% confidence region. On the other hand  $\lambda=0$  and  $\alpha=0$ , which is the linearizing transformation for the Ricker model, is in the 95% confidence region for all four stocks. Also, the inverse transformation ( $\lambda = -1$ ) which linearizes the Beverton-Holt is not in the 95% region for any of the four stocks. If both the Ricker and Beverton-Holt models are linearized, then the Ricker model will appear better fitting, not because it is necessarily superior but because the log transformation is more suitable than the inverse transformation. There are examples where  $\lambda=-1$  is quite suitable, for example the Atlantic menhaden stock (Carroll and Ruppert, 1984).

After  $\lambda$  and  $\alpha$  have been estimated by maximum likelihood, the fitted model should be checked by residual analysis as described in, for example, Draper and Smith (1981). Maximum likelihood estimation is highly sensitive to outlying observations. Such influential points may be evident from the residuals. Robust estimators for our transformation model is an important area for future research. At present, robust estimation has been studied only for the rather different methodology where only the response, not the model, is transformed; see Carroll and Ruppert (1985), Carroll (1980), and Bickel and Doksum (1981).

The model function gives the conditional median of the response, and the conditional mean and the other conditional quantiles of the response can be easily estimated. By estimating conditional quantiles one in effect builds a model of the skewness and heteroscedasticity in the untransformed response. Such a model is a crucial part of a realistic stochastic management model of the stock.

We have used four fish stocks as examples of our proposed statistical methodology, but our analyses should not be considered definitive. For example, we did not consider the effects on the Skenna River stock of the 1951-2 rock slide, changes in exploitation rates, an artificial spawning channel opened in 1965, or

interactions between year classes; see Ricker and Smith (1975). We believe, however, that power and weighting transformations will be equally as useful for more elaborate models as for the basic ones that we have employed for illustrative purposes.

Acknowledgments: Rod Reish and Rick Deriso introduced us stock-recruitment analysis. Carl Walters supplied the data, and we benefitted from his helpful discussions.

#### REFERENCES

- Bartlett, M.S., 1947. The use of transformations. *Biometrics*, 3, 39-52.
- Beverton, R.J.H., and Holt, S.J., 1957. On the dynamics of export fish populations. Her Majesty's Stationery Office, London, 533 p.
- Bickel, P.J., and Doksum, K.A., 1981. An analysis of transformations revisited. *Journal of the American Statistical Association*, 76 296-311.
- Box, George E.P. and Cox, David R., 1964. An analysis of transformations (with discussion). *Journal of the Royal Statistical Society, Series B*, 26, 211-246.
- Box, George E.P. and Hill, William J., 1974. Correcting inhomogeneity of variance with power transformation weighting. *Technometrics*, 16, 385-389.
- Box, George E.P. and Jenkins, G.M., 1976. *Time Series Analysis: Forecasting and Control, Revised Edition*. Holden-Day: San Francisco.
- Carroll, R.J., 1980. A robust method for testing transformations to achieve approximate normality. *Journal of the Royal Statistical Society, Series B* 42, 71-78.
- Carroll, R.J. and Ruppert, D., 1982. Robust estimation in heteroscedastic linear models. *Annals of Statistics* 10, 429-441.
- Carroll, R.J. and Ruppert, D., 1984. Power transformations when fitting theoretical models to data. *Journal of the American Statistical Association*, 79, 321-328.
- Carroll, R.J. and Ruppert, D., 1985. Transformations in regression: a robust analysis. *Technometrics*, 27, 1-12.
- Draper, Norman and Smith, Harry, 1981. *Applied Regression Analysis, 2nd Edition*. Wiley: New York.
- Duan, Naihua, 1983. Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*, 78, 605-610.
- Pritchard, D.J., Downie, J., and Bacon, D.W., 1977. Further consideration of heteroscedasticity in fitting kinetic models. *Technometrics*, 19, 227-236.
- Ricker, W.E., 1954. Stock and recruitment. *J. of Fish. Res. Board Can.*, 11, 559-623.
- Ricker, W.E. and Smith, H.D., 1975. A revised interpretation of the history of the Speena River sockeye salmon (*Oncorhynchus nerka*). *J. the Fish. Res. Board of Can.*, 32, 1369-1381.

Ruppert, D., Reish, R.L., Deriso, R.B., and Carroll, R.J., 1985. A Stochastic population model for managing the atlantic manhaden fishery and assessing managerial risks. Can. J. Fish. and Aqu. Sc. - to appear.

Walters, C.F., 1985. Bias in the estimation of functional relationships from time series data. Can. J. Fish. and Aqu. Sc., 42, 147-149.

Walters, C.F., Hilborn, R., Staley, M.J., Wong, F., 1982 Report to the Pearse Commission on Pacific Fisheries Policy, Vancouver, B.C.