

# 11

---

## *Surplus Production Models*

---

---

### 11.1 Introduction

#### 11.1.1 Stock Assessment Modelling Options

Surplus production models are the simplest analytical method available that provides for a full fish stock assessment. Described in the 1950s (Schaefer, 1954, 1957), they are relatively simple to apply, partly because they pool the overall effects of recruitment, growth, and mortality (all aspects of production) into a single production function. The stock is considered solely as undifferentiated biomass; that is, age and size structure, along with sexual and other differences, are ignored. The minimum data needed to estimate parameters for such models are time series of an index of relative abundance and of associated catch data. The index of stock abundance is most often catch-per-unit-effort, but could be some fishery-independent abundance index (e.g., from trawl surveys, acoustic surveys), or both could be used.

To conduct a formal stock assessment it is necessary, somehow, to model the dynamic behaviour of the exploited stock. One objective is to describe how the stock has responded to varied fishing pressure. By studying the impacts on a stock of different levels of fishing intensity it is possible to assess its productivity. If statistics are collected, the process of fishing a stock can provide information about how the stock responds to perturbations (the extra mortality, above natural mortality, imposed by fishing). If a reduction in the stock size cannot be detected reliably (i.e., catch rates or survey results are hyperstable relative to stock size), then stock assessment will be difficult, unreliable, or even impossible.

Given the necessary data, stock dynamics may be modelled using relatively simple surplus production models, also known as biomass dynamic models (Hilborn and Walters, 1992). These have already been briefly introduced in Chapter 2 and more in depth in Chapter 8. An alternative to surplus production models might be to use the more complex and data-demanding age-structured models (e.g., cohort analysis, virtual population analysis, or statistical catch-at-age; see Chapter 12; Quinn and Deriso, 1999). A less common alternative, which tends to be used with those species that are

difficult or impossible to age, would be to use a length- or size-based assessment model (see Chapter 13; Sullivan et al., 1990; Sullivan, 1992; Punt et al., 1997b). Because age-based models follow identifiable cohorts, they suffer from fewer problems than length-based models (given good data, which is not necessarily easy to obtain). Fisheries scientists usually try to collect the data required to produce an age-structured model in preference to the simpler data requirements of a surplus production model. However, Ludwig and Walters (1985, 1989) have shown that this is not always the best strategy, as surplus production models may produce answers just as useful and sometimes better for management than those produced by age-structured models, at a fraction of the cost. The usefulness of any model is directly related to how representative the available data are for a fished stock, and whether the index of relative abundance really does provide a clear index of relative stock size. If the index of relative abundance is informative but a set of ageing data are not necessarily representative, then we might expect a surplus production model to be more useful for the provision of management advice.

In a discussion of model selection, Hilborn and Walters (1992) suggest adopting a pragmatic approach. Assuming the data are available, they imply that one should apply both surplus production and age-structured methods, which, because they are fundamentally different, will provide a test of relative performance. They state: "If biomass dynamic methods provide a different answer than age-structured methods, then the scientist should try to understand why they are different and analyze the management implications of the different predictions, rather than concentrating on deciding which method is correct" (Hilborn and Walters, 1992, p. 329). Essentially, this is recommending that one examines whether the different available data streams (e.g., catch rates, age structure, length distributions, etc.) are internally consistent in their implications for the stock dynamics. Surplus production models, therefore, can be useful, and in this chapter we will be examining their use and properties in some detail.

Despite occasional recent use (Saila et al., 1979; Punt, 1994), the use of surplus production models went out of fashion in the 1980s. This was possibly because early on in their development it was necessary to assume the stocks being assessed were in equilibrium, and this often led to over optimistic conclusions that were not supportable in the longer term. Hilborn (1979) analyzed many such situations and demonstrated that the data used were often too homogeneous; they lacked contrast and hence were uninformative about the dynamics of the populations concerned. For the data to lack contrast means that fishing catch and effort information is only available for a limited range of stock abundance levels. However, a lack of contrast in this way can also lead to inconclusive results from age-structured models. There should also be concern that uncertainty (bias or lack of precision) in

the observed abundance estimates exacerbates this problem by reducing the information content of the data used in relation to the actual stock size.

11.1.2 Surplus Production

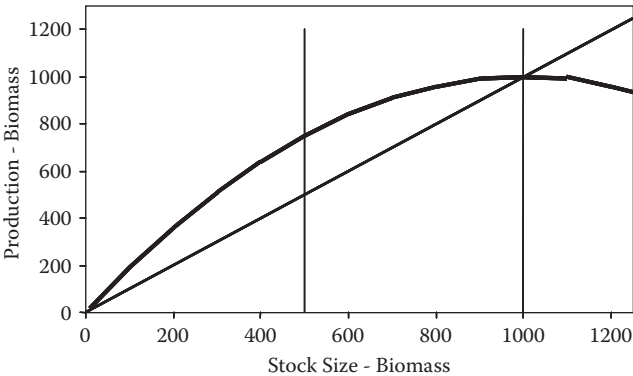
Surplus production, as the name implies, relates to the production from a stock beyond that required to replace losses due to natural mortality. Surplus production, in this case, is the sum of new recruitment and the growth of individuals already in the population minus those dying naturally. Early ideas on surplus production were discussed briefly in Chapter 2, when the logistic population model was introduced. Many of the intuitions regarding surplus production in fisheries derive from the logistic, and we will revisit them here in order that their strengths and weaknesses can be illustrated.

Schaefer (1954) first applied the logistic curve as a description of the production of Pacific halibut and later to eastern Pacific yellowfin tuna (Schaefer, 1957). Until then, no simple method of assessing a fished stock was available. Using the logistic as a foundation, Schaefer (1954, 1957) demonstrated a theoretical link between stock size and expected catch rates. This all related back to the expected level of surplus production produced by particular stock sizes. Thus, given a known stock biomass, the total production could be predicted thus:

$$B_{t+1} = B_t + rB_t \left(1 - \frac{B_t}{K}\right)$$

(11.1)

where  $B_t$  is the stock biomass at time  $t$ ,  $r$  is the population growth rate, and  $K$  is the maximum population size for growth to be positive. Both  $r$  and  $K$  are



**Figure 11.1**  
Production curve for the discrete logistic curve with  $r = 1$  and  $K = 1,000$ . The right-hand vertical line indicates the carrying capacity, while the central line illustrates the point of maximum production at  $K/2$ . The diagonal line is the line of replacement (Example Box 11.1).

© 2011 by Taylor & Francis Group, LLC

Copyright © 2011. CRC Press LLC. All rights reserved.

EXAMPLE BOX 11.1

Different production curves using an asymmetric production equation (Equation 11.2). Name cell B1 as *ri*, cell B2 as *K*, and cell B3 as *p*. Extend the biomass levels in column A in steps of 100 down to 1,300 in row 20. Column B contains the equivalent of Equation 11.2, while column C indicates the production above replacement levels. Plot column B against A, adding a diagonal line as in Figure 11.1. Select D2:E3, copy, and paste special these data as a new series on the same graph to illustrate the stock size at maximum production. Using the solver, maximize C5 by modifying A5. This will automatically reposition the line of maximum production. Modify *r* and *p* to see their relative effects. When *p* = 1, the equation is equivalent to the simpler Schaefer equation. The effect of the asymmetry parameter is to make the effects of density-dependent regulation nonlinear (see Chapter 2).

	A	B	C	D	E
1	<i>ri</i>	1			
2	<i>K</i>	1000		=A5	0
3	<i>p</i>	1		=D2	1200
4					
5	500	=A5+( <i>ri</i> / <i>p</i> )*A5*(1-(A5/ <i>K</i> )^ <i>p</i> )	=B5-A5		
6	Biomass	Production	Surplus		
7	10	=A7+( <i>ri</i> / <i>p</i> )*A7*(1-(A7/ <i>K</i> )^ <i>p</i> )	=B7-A7		
8	100	=A8+( <i>ri</i> / <i>p</i> )*A8*(1-(A8/ <i>K</i> )^ <i>p</i> )	=B8-A8		
9	200	Copy down to row 20	Copy down		
10	300	510	210.0		
11	400	640	240.0		
12	500	750	250.0		

parameters of the logistic equation (see Chapter 2). A property of this equation is that the maximum production occurs at *K*/2 (Figure 11.1, Example Box 11.1).

Irrespective of the stock size, it should be possible to take the excess production, above the equilibrium line of replacement, and leave the stock in the condition it was in before production and harvesting (Figure 11.1). An obvious management strategy deriving from this theory would be to bring the stock to a size that would maximize the surplus production and hence the potential yield. This supports the intuition that it is necessary to fish a stock down in size so that it becomes more productive. There are, of course, many problems with this simplistic view of fisheries production. It assumes that the population is in equilibrium with all of its inputs and outputs (a poor assumption). It also implies that, while one may begin to fish inefficiently, one would not be able to crash a fish stock through overfishing (clearly a very poor assumption).

© 2011 by Taylor & Francis Group, LLC

One problem that was easily solved was the fact that the logistic curve generated a symmetrical production curve, which was felt to be overly constraining. Pella and Tomlinson (1969) solved this by introducing an asymmetry term  $p$ , which modifies the logistic as follows:

$$B_{t+1} = B_t + \frac{r}{p} B_t \left( 1 - \left( \frac{B_t}{K} \right)^p \right) \quad (11.2)$$

Equation 11.2 is from Polacheck et al. (1993) but is still referred to as the Pella–Tomlinson model because of the potential asymmetry in the production curve. This means that with this model, the stock size at which maximum production occurs is not necessarily at  $K/2$  (Example Box 11.1).

## 11.2 Equilibrium Methods

Wherever possible, equilibrium methods should be avoided in fisheries assessments. With surplus production models, if a fish stock is in decline, then equilibrium methods consistently overestimate the sustainable yield. Their use in the past undoubtedly contributed to a number of major fishery collapses (Boerema and Gulland, 1973; Larkin, 1977). While equilibrium methods are no longer recommended, how they were used will be demonstrated so that the literature may be more easily understood.

Equilibrium methods are best described using the Schaefer or Pella–Tomlinson model of stock dynamics; this is just Equation 11.2 minus any catch:

$$B_{t+1} = B_t + \frac{r}{p} B_t \left( 1 - \left( \frac{B_t}{K} \right)^p \right) - C_t \quad (11.3)$$

where  $C_t$  is the total catch in year  $t$ ; if  $p = 1$ , this is equivalent to the Schaefer model (Equation 11.1).

Equilibrium methods rely on the assumption that for each level of fishing effort there is an equilibrium sustainable yield. The stock is assumed to be at some equilibrium level of biomass producing a certain quantity of surplus production. If the fishing regime is changed, the stock is assumed to move immediately to a different stable biomass with its associated surplus production. This is patently wrong, as it ignores the difference in standing crop between the two different biomass levels and the time it takes the system to respond to changed conditions. At heart, the assumption is that the yield taken is always surplus production from a population in equilibrium. From this assumption it is possible to estimate the maximum sustainable yield (MSY) and the associated effort that will give rise to the MSY ( $E_{MSY}$ ) given the

© 2011 by Taylor & Francis Group, LLC

appropriate biomass ( $B_{\text{MSY}}$ ). It is necessary to assume that the rate of change of biomass is zero for all years, i.e.,  $B_{t+1} = B_t = \text{constant}$ , and that the relationship between the index of abundance and stock biomass (Equation 11.4) is exact. One requires a set of data to fit this model, and in the case of fisheries data, this is usually a time series of catch rates,  $I_t$ , from either fishery-independent surveys or commercial catch rates. In fact, because of the assumption of equilibrium, the time series nature of the data is ignored (another of the flaws in the methodology). Real contrast in the data is required (meaning that information on catch rates from widely different effort and stock size levels provide the best information for fitting fisheries models). To connect the population dynamics model (Equation 11.3) to reality, we connect the catch rates to the stock biomass,  $B$ , via  $q$ , the catchability coefficient (= proportion of the total stock taken by one unit of effort). Given that  $C$  is the catch and  $E$  is the associated effort, then

$$I = \frac{C}{E} = qB \quad (11.4)$$

Note the lack of any  $t$  subscripts. This is to emphasize that the time series nature of the data is ignored in this equilibrium method. At equilibrium,  $B_{t+1}$  will equal  $B_t$ , and so the year when each data point was generated becomes irrelevant. We can solve Equation 11.3 for  $C$  after assuming  $B_{t+1} = B_t$ , and after substituting

$$\frac{C}{(qE)} = B \quad (11.5)$$

for  $B$  in Equation 11.3 to give

$$C = \frac{rC}{pqE} \left[ 1 - \left( \frac{C}{qEK} \right)^p \right] \quad (11.6)$$

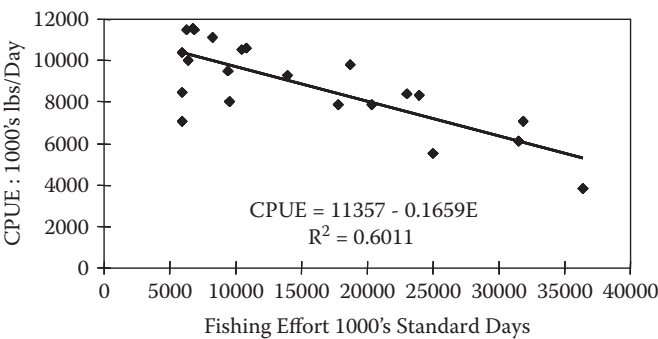
which in turn can be solved for  $C/E$  or  $I$  (see Appendix 11.1), so that

$$I = \frac{C}{E} = \left( (qK)^p - \frac{pq^{p+1}K^pE}{r} \right)^{\frac{1}{p}} \quad (11.7)$$

If we reparameterize the constants, by defining  $(qK)^p$  to be a new parameter  $a$ , and the second term,  $(pq^{p+1}K^p)/r$ , to be a new parameter  $b$ , this would lead to the form (Polacheck et al., 1993)

$$C/E = (a - bE)^{1/p} \quad (11.8)$$

© 2011 by Taylor & Francis Group, LLC



**Figure 11.2**  
Relationship between CPUE and effort for eastern tropical Pacific ocean yellowfin tuna. (Data from Schaefer, 1957; see Table 11.1.) Given  $p = 1$  and the parameters  $a$  (11,357) and  $b$  (0.1659) permits the calculation of an equilibrium  $MY$  and  $E_{MSY}$  (194,366,258 lb and 34,228 class 4 clipper days, respectively).

and therefore, the equilibrium catch is

$$C = E(a - bE)^{1/p} \tag{11.9}$$

The sum of squares estimates of the parameters  $a$ ,  $b$ , and  $p$  can then be obtained by minimizing the quantity

$$\sum (I - \hat{I})^2 \tag{11.10}$$

where  $(C/E) = I$  is the observed catch rate from which we subtract the predicted or expected catch-per-unit-effort from the model (denoted by the  $\hat{\phantom{x}}$  symbol). If we assume  $p = 1$ , the model simplifies to the original Schaefer stock production model. Estimates for the two parameters  $a$  and  $b$  can then be obtained using standard linear regression techniques (Figure 11.2, Example Box 11.2).

Referring to Equation 11.8 makes it clear why plotting CPUE against effort was such a popular analytical tool in the past (Figure 11.2). When  $p = 1$ ,  $a = qK$ , and  $b = (q^2K)/r$ , differentiating Equation 11.9 with respect to  $E$  gives

$$E_{MSY} = \frac{a}{-2b} = \frac{r}{2q}, \text{ if } p < 1 \quad E_{MSY} = \frac{a}{b(p+1)} = \frac{r}{q(p+1)} \tag{11.11}$$

Substituting Equation 11.11 into Equation 11.6 gives

$$MSY = \frac{(a/2)^2}{b} = \frac{rK}{4} \tag{11.12}$$

and if  $p < 1$ , then  $a = (qK)^p$  and  $b = (pq^{p+1}K^p)/r$ , so that

EXAMPLE BOX 11.2

An equilibrium surplus production model fitted to peruvian anchovy data (after Pitcher and Hart, 1982). The equations used in B4 and B5 are Equations 11.11 and 11.13. Select B1:C1, type =linest(D7:D15,C7:C15, true, true), then press <Ctrl><Shift><Enter> to enter the array function for linear regression (see Excel help). The gradient is in B1 and the intercept in C1. Column F is Equation 11.9, but for the Schaefer model (i.e.,  $p = 1$ ). Plot the regression data (columns D against C) and add a linear trend line, showing the equation (cf. Figure 11.2). In a separate graph, plot column F against E as a line. Add columns C vs. B as a scatter of points to mimic the type of graph shown in Figure 11.3 (select and copy them, select the graph, paste special as a new series, edit the series so that column C is on the x axis and column B on the y axis). Replace the data in columns A to C with the eastern Pacific yellowfin tuna data from Schaefer (1957), found in Table 11.1. Alter the Linest function in B1:C1, and the effort values in column E appropriately. Do the calculated values for  $a$ ,  $b$ ,  $MSY$ , and  $E_{MSY}$  match those given in Figure 11.2?

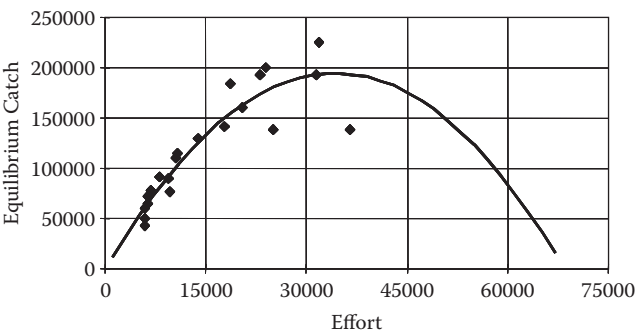
	A	B	C	D	E	F
1	Linest	-0.01409	0.782203			
2	a	=C1				
3	b	=B1				
4	MSY	=((B2/2)^2)/-B3				
5	Emsy	=B2/(-2*B3)				
6	Year	Catch	Effort	CPUE	E	Equil C
7	60	6.00	10.1	0.594	2	=(B\$2+B\$3*E7)*E7
8	61	8.40	12.0	0.700	6	=(B\$2+B\$3*E8)*E8
9	62	10.10	21.0	0.481	10	6.413
10	63	10.60	24.0	0.442	14	8.188
11	64	11.00	28.0	0.393	18	9.513
12	65	9.00	25.0	0.360	22	10.387
13	66	9.20	22.0	0.418	26	10.809
14	67	10.75	21.5	0.500	30	10.781
15	70	12.70	30.0	0.423	34	10.301

$$MSY = \frac{p}{b} \left( \frac{a}{p+1} \right)^{\frac{p+1}{p}} = \frac{rK}{(p+1)^{\frac{(p+1)}{p}}}$$

(11.13)

The best way to visualize these analyses is to plot the expected equilibrium catches and the observed catches (as in Figure 11.3) against the imposed effort. The outcome of this equilibrium analysis can appear dangerously





**Figure 11.3**  
Equilibrium analysis of eastern tropical Pacific ocean yellowfin tuna. (Data from Schaefer, 1957.) The expected equilibrium catches for particular effort levels (thousands of class 4 clipper days) are represented by the curve's solid line. The observed catch levels (thousands of pounds) are the scattered points. The maximum sustainable yield (MSY) is the top of the curve, and the optimum effort ( $E_{MSY}$ ) is directly below the MSY peak of equilibrium catch ( $C = (a - bE)E$ ; see Example Box 11.2).

**Table 11.1**

Total Eastern Pacific Catch of Yellowfin Tuna, Catch per Standard Day Fishing, and Calculated Relative Fishing Effort

Year	Catch	Effort	CPUE	Year	Catch	Effort	CPUE
1934	60,913	5,879	10,361	1945	89,194	9,377	9,512
1935	72,294	6,295	11,484	1946	129,701	13,958	9,292
1936	78,353	6,771	11,572	1947	160,134	20,381	7,857
1937	91,522	8,233	11,116	1948	200,340	23,984	8,353
1938	78,288	6,830	11,462	1949	192,458	23,013	8,363
1939	110,417	10,488	10,528	1950	224,810	31,856	7,057
1940	114,590	10,801	10,609	1951	183,685	18,726	9,809
1941	76,841	9,584	80,18	1952	192,234	31,529	6,097
1942	41,965	5,961	7,040	1953	138,918	36,423	3,814
1943	50,058	5,930	8,441	1954	138,623	24,995	5,546
1944	64,094	6,397	10,019	1955	140,581	17,806	7,895

Source: Data from Schaefer, M. B., *Bulletin, Inter-American Tropical Tuna Commission*, 2, 247–85, 1957.

Note: Catch is in thousands of pounds, effort is in standardized class 4 clipper days, and catch rates are in pounds per class 4 clipper day. These data are used in Example Boxes 11.2 and 11.3 and are illustrated in Figures 11.3 and 11.5.

convincing, but this is to be resisted, as it usually overestimates the safe catch levels (Example Box 11.2). Such analyses can be conducted with remarkable speed, but these should be considered examples where the approximations and assumptions used (especially the assumption of equilibrium) mean the analyses cannot be used validly.

© 2011 by Taylor & Francis Group, LLC

Surplus production models no longer need the assumption of equilibrium to enable them to be fitted to fisheries data. The nonequilibrium approach to fitting the models means they are better able to represent the dynamics of fished populations. Nevertheless, the inherent simplicity of surplus production models means there are limits to how far their development may be taken. We will investigate some recent developments after a discussion of the different methods that can be used to fit nonequilibrium surplus production models.

## 11.3 Surplus Production Models

### 11.3.1 Russell's Formulation

Surplus production models relate directly to Russell's (1931) verbal formulation of stock dynamics and, in difference equation or discrete form, have the general structure

$$B_{t+1} = B_t + f(B_t) - C_t \quad (11.14)$$

with

$$\hat{I}_t = \frac{C_t}{E_t} = qB_t \quad (11.15)$$

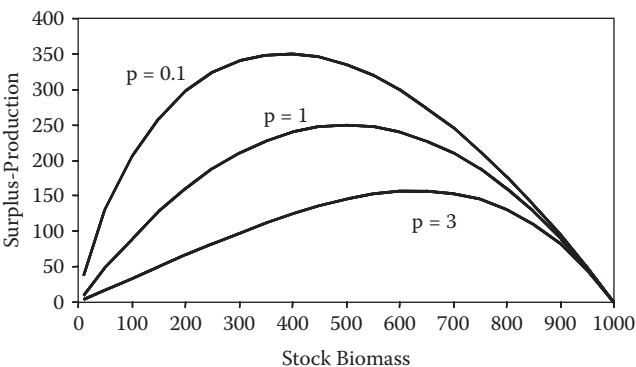
where  $B_{t+1}$  is the exploitable biomass at the end of year  $t$  or the beginning of year  $t + 1$ ,  $B_t$  is the exploitable biomass at the start of year  $t$ ,  $f(B_t)$  is the production of biomass, as a function of biomass at the start of year  $t$ ,  $C_t$  is the biomass caught during year  $t$ ,  $I_t$  is the an index of relative abundance for year  $t$ , and  $q$  is the catchability coefficient. The  $\hat{\phantom{x}}$  symbol above a parameter denotes a value estimated from the model. The function describing the production of biomass in any year can take many forms; here are three that all derive from a form of the logistic equation. This is especially clear in the Schaefer model:

$$f(B_t) = rB_t \left( 1 - \frac{B_t}{K} \right) \quad \text{classic Schaefer (1954) form} \quad (11.16)$$

$$f(B_t) = \ln(K)rB_t \left[ 1 - \left( \frac{\ln(B_t)}{\ln(K)} \right) \right] \quad \text{modified Fox (1970) form} \quad (11.17)$$

$$f(B_t) = \frac{r}{p} B_t \left[ 1 - \left( \frac{B_t}{K} \right)^p \right] \quad \text{modified Pella and Tomlinson (1969)} \quad (11.18)$$

© 2011 by Taylor & Francis Group, LLC



**Figure 11.4**  
Influence of the parameter  $p$  (with values 0.1, 1, and 3) on the discrete Pella–Tomlinson version of the biomass dynamic model. When  $p = 1$ , the equation is equivalent to the Schaefer model, and thus has a symmetrical production curve around the midpoint in biomass. Values of  $p < 1$  skew the curve to the left and values  $> 1$  skew it to the right.

where  $r$  is a growth rate parameter (derived from the intrinsic rate of natural increase), and  $K$  is the virgin biomass ( $B_0$ ) or the average biomass level prior to exploitation (derived from the idea of carrying capacity). If  $p = 1$ , then Equation 11.18 is equivalent to Equation 11.16. As  $p$  tends to zero, Equation 11.18 becomes equal to Equation 11.17. The usual manner of writing the Fox model is without the  $\text{Ln}(K)$  at the front of the production equation (Equation 11.17). However, without its inclusion, while the two equations take on an equivalent form as  $p$  tends to zero, they do not produce exactly the same value for production. The particular form of Equation 11.18 was given by Polacheck et al. (1993), and alternative forms for the Pella–Tomlinson model exist (the most common alternative using an asymmetry parameter  $m - 1$  instead of the  $p$  parameter). The key property of the Pella–Tomlinson model is its ability to express an asymmetric production curve with potentially many shapes.

The linear density-dependent effects built into the logistic Schaefer model become nonlinear with the addition of the  $p$  parameter, leading to the asymmetry of the production curve (Figure 11.4). Note that by modifying  $p$  the absolute level of production alters (Figure 11.4, Example Box 11.1), which implies that adding the asymmetry term has altered the interpretation that can be placed upon the other parameters. For example, as  $p$  decreases below a value of 1, and all other parameters do not change, the absolute production increases. This increase could be offset during the model fitting process by a decrease in the value of  $r$ . The point is that it is not possible to directly compare the parameter values of the three different models (although if the  $\text{Ln}(K)$  term is included in the Fox model, then Equations 11.17 and 11.18 are directly comparable if  $p$  is very close to zero).

© 2011 by Taylor & Francis Group, LLC

Equation 11.15 captures how the expectations from the model are compared against reality. The equation constitutes a very strong assumption, which is that catch rates are linearly related to stock biomass. The catchability coefficient,  $q$ , is often referred to as a nuisance parameter whose job is simply one of scaling the modelled stock biomass to match the trends in catch rates. Assessment becomes very difficult without some time series of relative abundance estimates. Surplus production methods tend to be used when large amounts of good quality data are not available. The assumption that catch rates relate to the stock biomass must be considered carefully when such analyses are conducted.

### 11.3.2 Alternative Fitting Methodology

In the development of fisheries methods a number of contrasting algorithms or strategies have been used for fitting stock production models to observed data on catches and catch rates. These methods differ with regard to whether or not they assume the population to be in equilibrium with the exploitation rate. They also differ with regard to where the residual errors are attributed between the model and the data.

The earliest method used two estimated equilibrium management targets, the MSY and  $E_{MSY}$ . For this method to be valid, the fishery was assumed to be in equilibrium. It was recognized, however, that fish stocks are rarely in equilibrium. An early solution to this problem was to use a weighted average of a number of years' fishing effort for each year instead of just observed effort for that year. Unfortunately, if this is done, the interpretation of the relationship between catch and effort becomes problematical. While effort averaging certainly improved the analyses in a limited way, it is an *ad hoc* solution that can be thought of as spreading an invalid assumption of equilibrium across a number of years. Here, it will not be pursued further.

Process error estimators were then developed, which assumed all observations, such as catch rates, were made without error (Equation 11.15 was exact or without error), and that all error was in the equation describing changes in population size (Equation 11.14 was imprecise). Thus, in any one year, to obtain a close match between the predictions of the model and the data, residual errors would need to be added to some or all of the parameters (Polacheck et al., 1993).

Alternatively, observation error estimators have been developed that assume all residual errors are in the catch rate or biomass observations (Equation 11.15), and that the equation describing the time series of biomass values (Equation 11.14) is deterministic and without error. In this chapter we will focus our efforts on this method of fitting surplus production models.

More recently, attempts have been made to create estimators that use both forms of error, most notably using a technique borrowed from control engineering termed the Kalman filter (Meinhold and Singpurwalla, 1983). However, currently, there are no generally available methods that can make

estimates where both types of error are being modelled without constraints (Quinn and Deriso, 1999).

Process errors assume that the observations used to fit the model are made without error; thus, all error occurs in the predicted change in population size (i.e., observations are made without error, but the model does not exactly mimic reality). Observation error is the inverse of process error, i.e.,  $B_{t+1} = \hat{B}_{t+1}$ , with no error term (once again, the  $\hat{\phantom{x}}$ , or caret symbol, denotes the expected or estimated parameter value), so that observations are made with error but the model exactly describes the population dynamics. Observation errors imply

$$C_t/E_t = qB_t e^{\varepsilon_t}$$

where  $C_t$  is catch and  $E_t$  is the fishing effort in year  $t$ . The term  $e^{\varepsilon}$  represents log-normal residual errors. Alternatively, in this model process errors would imply

$$r = \hat{r} + \varepsilon_r \quad \text{or} \quad K = \hat{K} + \varepsilon_K$$

However, the form of the residual errors here could be other than normally distributed.

The approach of observation error estimation is the method now most commonly recommended because simulations have demonstrated that it can more closely reflect the circumstances underlying the observations (Hilborn and Walters, 1992; Polacheck et al., 1993; Punt, 1990, 1995). Generally, if both types of error are present (seems generally likely), then so far it has been found to be more efficient to assume only the presence of observation error (Ludwig and Walters, 1989). There has been some work attempting to model some restricted types of fisheries data using both kinds of residual error via the Kalman filter (Sullivan, 1992; Reed and Simons, 1996). But this remains generally intractable. One way in which both types of error could be modelled would arise if there were an estimate of the ratio of the respective variances of the two processes (Ludwig et al., 1988). In practice, it is extremely difficult to work with both forms of error in an estimation model.

## 11.4 Observation Error Estimates

### 11.4.1 Outline of Method

If we assume that the equation describing the stock dynamics is deterministic (i.e., zero process residual error), then all residual errors are assumed to occur in the relationship between stock biomass and the index of relative abundance. The stock biomass time series can therefore be estimated

© 2011 by Taylor & Francis Group, LLC

by projecting forward the initial biomass,  $B_0$ , at the start of the catch series, using the selected biomass dynamic model and the historic annual catches (Example Box 11.3, Equation 11.19):

$$B_{t+1} = B_t + \frac{r}{p} B_t \left( 1 - \left( \frac{B_t}{K} \right)^p \right) - C_t \quad (11.19)$$

Using Equation 11.19, and given the observed  $C_t$ s, the parameters  $r$ ,  $K$ , and  $p$ , and an initial starting biomass  $B_0$ , a series of expected  $B_t$ s, can be produced. In the following example, the asymmetry parameter  $p$  is set to 1, so the model simplifies to the Schaefer model. The discrete version of the surplus production model is used to produce the predicted series of  $B_t$ s, and these, given a  $q$  value, are used to produce a predicted series of CPUE values ( $C/E = qB$ ; see Equation 11.20) that can be compared with those observed, using either maximum likelihood or least squares.

In summary, one makes predictions about the deterministic trajectory of the system from a hypothesized set of parameters and initial starting conditions, and then compares the observed values with the predictions (Figure 11.5). Connecting the deterministic series of biomass levels to observed catch rates is implemented using

$$\hat{I}_t = \frac{\hat{C}_t}{E_t} = qB_t e^\varepsilon \quad \text{or} \quad \hat{I}_t = q \frac{(B_{t+1} + B_t)}{2} e^\varepsilon \quad (11.20)$$

where the  $e^\varepsilon$  indicates that the residual errors are assumed to be lognormally distributed (a standard assumption with catch rate data). Taking the average of two biomass levels relates to using the average biomass at the start and end of year  $t$  so that the catches relate to the biomass more realistically. In effect, the expected catch rates are related to midyear biomass. In the process of fitting the observed data to the model, it will be necessary to log-transform both the observed and the predicted catch rates to normalize the residual errors.

#### 11.4.2 in Theory and Practice

Assuming that the error in Equation 11.20 is multiplicative and lognormal with a constant variance (i.e.,  $I_t = qB_t e^\varepsilon$ , where  $\varepsilon = N(0; \sigma^2)$ ), the estimates of the model parameters ( $B_0$ ,  $r$ ,  $q$ , and  $K$ ) are obtained by maximizing the appropriate likelihood function:

$$L(\text{data} | B_0, r, K, q) = \prod_t \frac{1}{I_t \sqrt{2\pi\hat{\sigma}}} e^{-\frac{(\ln I_t - \ln \hat{I}_t)^2}{2\hat{\sigma}^2}} \quad (11.21)$$

© 2011 by Taylor & Francis Group, LLC

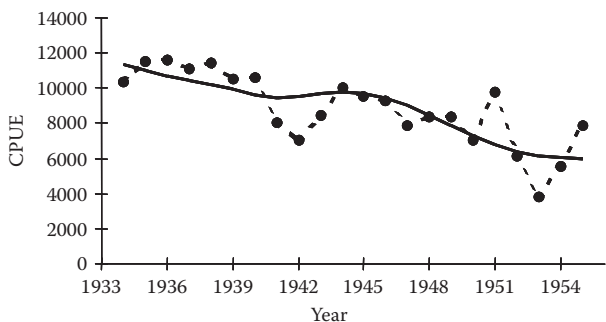
EXAMPLE BOX 11.3

A nonequilibrium Schaefer surplus production model fitted to Schaefer's (1957) original eastern Pacific yellowfin tuna data (Table 11.1). Copy the data into columns A to C, down to row 31. In D11 put Equation 11.19 as  $=\max(D10+\$B\$1*D10*(1-(D10/\$B\$2))-B11,100)$ . The max function ensures that the stock biomass cannot go extinct when using the solver. In E10 put  $=((D10+D11)/2)*\$B\$4$ , the midyear biomass in a given year multiplied by the estimate of  $q$  (in E31 put  $=D31*\$B\$4$ ). Put  $=(C10-E10)^2$  into F10 to obtain the squared normal residual errors. In G10 put  $=\ln(C10/D10)$ , to generate the contributions to the  $q$  estimate. In H10 put  $=(\ln(C10)-\ln(E10))^2$  to generate the residuals for lognormal random errors. Select D11:H11 and copy down to row 31. Complete the worksheet by putting  $=\exp(\text{average}(\$G\$10:\$G\$31))$  into B4 to calculate the closed form of  $q$  (see Equation 11.23). Count the number of observations by putting  $=\text{count}(C10:C31)$  into B6, and finally, to estimate the standard deviation of the residuals, put  $=\text{sqrt}(\text{sum}(H10:H31)/B6)$  into B7 (see Equation 11.22). Plot columns C and E against column A (use different colours). The parameters shown are close to the initial optimum. Use the solver to maximize the log-likelihood in E7 by changing B1:B3. Compare the results obtained by minimizing the sum of squared residuals in E8. They should be the same. How different are the answers when you minimize E6, which contains normal random residual errors instead of lognormal? How stable is the answer? Start the model from different starting points (e.g.,  $r_i$ ,  $K_i$ ,  $B_0 = 0.05, 1,500,000, 1,500,000$ ; or  $0.5, 3,500,000, 1,000,000$ ). In the solver options, turn on automatic scaling and increase the required precision and decrease the convergence limits. Try removing the max function from column D. Try estimating  $q$  directly (just vary the values of B1:B4 when using the solver). Put the  $p$  parameter into B5. Alter column D to match Equation 11.19 and E3 and E4 to match Equations 11.11 and 11.12. How does permitting an asymmetric production curve affect the results? Are the results biologically sensible?

	A	B	C	D	E	F	G	H
1	<b>ri</b>	0.17830		<b>Bcurr</b>	=D31			
2	<b>Ki</b>	2525547		<b>Bcurr/K</b>	=E1/B2			
3	<b>B0</b>	2748326		<b>MSY</b>	=B1*B2/4			
4	<b>qi</b>	0.00436		<b>Emsy</b>	=B1/(2*B4)			
5	<b>p</b>	1		<b>FFinal</b>	=B31/E31			
6	<b>n</b>	22		<b>SSQ_I</b>	=sum(F10:F31)			
7	<b>Sigma</b>	0.1726		<b>LogLik</b>	=-(B6/2)*(Ln(2*pi))+2*Ln(B7)+1)			

continued

EXAMPLE BOX 11.3 (continued)								
	A	B	C	D	E	F	G	H
8				SSQ_Ln(I)	=sum(H11:H32)			
9	Year	Catch	O(I)	ExpectB	E(I)	SSQ_I	Expt_q	Ln(I-I)
10	1934	60913	10361	=\$B\$3	11763.4	1966371.8	-5.542	0.0161
11	1935	72294	11484	2549745	11330.2	23763.4	-5.403	0.0002
12	1936	78353	11572	2467036	10943.8	394501.1	-5.362	0.0031
13	1937	91522	11116	2385705	10585.9	281500.8	-5.369	0.0024



**Figure 11.5**  
Observed index of relative abundance (dotted line) vs. year, with best fit initial predicted index of relative abundance superimposed (smooth line) for the eastern tropical Pacific ocean yellowfin tuna. (Data from Schaefer, 1957; Table 11.1). The model fitted is the Schaefer surplus production model (Equation 11.19 with  $p = 1$ ). Because the index of relative abundance (CPUE) is only  $qB_t$ , where  $q$  is a constant, the thick line is also a representation of the time series of predicted stock biomass levels (Example Box 11.3).

where  $L(\text{data}|B_0,r,K,q)$  is the likelihood of the data given the parameters, the product is over all years ( $t$ ) for which CPUE data are available, and (Neter et al., 1996, p. 34)

$$\hat{\sigma}^2 = \sum_t \frac{(\text{Ln } I_t - \text{Ln } \hat{I}_t)^2}{n} \tag{11.22}$$

and  $n$  is the number of observations (maximum likelihood estimate of the variance uses  $n$  rather than  $n - 1$ ). An estimate of  $q$ , which maximizes Equation 11.21, is given by the geometric average of the time series of individual  $q$  estimates (see Appendix 11.2 for the derivation); thus,

$$\hat{q} = e^{\frac{1}{n} \sum \text{Ln} \left( \frac{I_t}{\hat{B}_t} \right)} \tag{11.23}$$



Alternatively, one could estimate the  $q$  value directly using the solver, but the value thus determined should differ from that derived from Equation 11.23 by only a small fraction. Such closed-form estimates, as in Equation 11.23, are valuable because, on average, the model tends to be quicker to fit and more robust when it has fewer directly estimated parameters.

With a model as indicated in Equations 11.19, 11.20, and 11.23 (with  $p = 1$ ), the actual fitting process would be as follows: given a time series of catches ( $C_t$ ), and guesses at  $r$ ,  $K$ , and  $B_0$ , the model produces a series of expected biomass values,  $\hat{B}_t$ . Then, given a catchability coefficient  $q$  (see Equation 11.23), the  $\hat{B}_t$  are used to produce a series of expected catch rates  $\hat{C}/E = \hat{I} = q\hat{B}_t$  (Equation 11.20), and these are compared with the observed catch rates (Equation 11.21; in fact, Equation 11.25).

There is more than one method available for determining  $B_0$ . It can be set equal to  $K$  (i.e.,  $B_0/K = 1$ ), or it can be estimated by  $\hat{B}_0 = \hat{I}_0/q$  (forces the first point of the fitted curve to coincide with the first point of abundance index, one degree of freedom is lost), or  $B_0$  can be estimated directly as a separate parameter. These alternatives can give very different results, and which method is used might depend upon whether fishing had occurred before records of catch were available. Punt (1990) found, using simulations, that even with situations where  $B_0/K$  was substantially different from unity, estimation performance was better when  $B_0/K$  was set at unity than when attempts were made to estimate  $B_0$  separately. However, there are situations, especially in fisheries for shorter-lived, more recruitment-driven species, where a direct estimation of  $B_0$  tends to be more efficient (Haddon, 1998).

In a similar example to the one illustrated here, Hilborn and Walters (1992) used a least squares criterion of fit using normal random residual errors between the observed CPUE,  $I_t$ , and the expected CPUE,  $\hat{I}_t$ .

$$\sum \left( \frac{C_t}{E_t} - \frac{\hat{C}_t}{E_t} \right)^2 = \sum (I_t - \hat{I}_t)^2 \quad (11.24)$$

An obvious alternative is to use a log-likelihood approach (Polacheck et al., 1993). In addition, instead of using normal random errors (implied in Equation 11.24), one generally uses lognormal residuals because we are dealing with catch rate data. Equation 11.21 can be converted to a log-likelihood and greatly simplified (see Appendix 11.3 and Example Box 11.3) so that

$$LL = -\frac{n}{2} \left( \ln(2\pi) + 2\ln(\hat{\sigma}) + 1 \right) \quad (11.25)$$

where  $LL$  refers to log-likelihood,  $n$  is the number of observed catch rates, and  $\sigma$  is the square root of Equation 11.22.

© 2011 by Taylor & Francis Group, LLC

### 11.4.3 Model Outputs

A general objective of fisheries modelling is to generate outputs in terms useful to fisheries management. There are many outputs possible from most fishery models, but focus tends to be placed on those that can act as fishery performance measures or that inform about selected limit thresholds. The two classical performance measures that derive from surplus production modelling are the maximum sustainable yield (MSY) and the effort,  $E_{MSY}$ , that, given  $B_{MSY} = K/2$  (for the logistic), should lead to the MSY (as in Equations 11.11 and 11.13):

$$E_{MSY} = \frac{r}{2q} \quad MSY = \frac{rK}{4} \quad (11.26)$$

In addition, we could consider such outputs as the current estimated biomass, the ratio of the current biomass with  $K$  or  $B_0$ , and possibly include an estimate of fishing mortality rate,  $F$  (see Example Box 11.3). A number of alternative methods exist for calculating various model parameters and outputs.

The instantaneous fishing mortality rate can be estimated in two ways. The first is as a conversion of the annual exploitation rate (catch/biomass) to an instantaneous fishing mortality rate:

$$F_t = -Ln\left(1 - \frac{C_t}{(B_t + B_{t+1})/2}\right) \quad (11.27)$$

where  $F_t$  is the instantaneous fishing mortality rate in year  $t$ ,  $C_t$  is the catch in year  $t$ , and  $(B_t + B_{t+1})/2$  is the midyear biomass for year  $t$  (as in Equation 11.20 and Example Box 11.3). Alternatively, we could use the standard catch equation so that instantaneous fishing mortality relates to expected effort and the catchability coefficient,  $q$ :

$$F_t = qE_t = q \frac{C_t}{C_t/E_t} = q \frac{C_t}{\hat{I}_t} \quad (11.28)$$

But in this case the  $F_t$  is an annual harvest rate rather than an instantaneous rate (this relates to when the standard catch equation was a differential equation rather than a difference equation). Given Equation 11.28, and that  $E_{MSY} = r/2q$  (from Equation 11.26), we can see that the harvest rate at MSY,  $F_{MSY}$ , would be

$$F_{MSY} = qE_{MSY} = q \frac{r}{2q} = \frac{r}{2} \quad (11.29)$$

Prager (1994) described many extensions to standard surplus production models, and one of these was to point out that  $F_{0.1}$  (see Chapter 2) is

© 2011 by Taylor & Francis Group, LLC

approximately 90% of  $F_{\text{MSY}}$ . Thus, it would be simple to include both of these potential management targets in the outputs from the model.

The real-world interpretation of management targets is not always straightforward. An equilibrium is now assumed to be unlikely in a fished population, so the interpretation of MSY is more like an average, long-term expected potential yield if the stock is fished optimally. The  $E_{\text{MSY}}$  is only the effort that should give rise to the MSY if the stock biomass is at  $B_{\text{MSY}}$ , the biomass needed to generate the maximum surplus production. Clearly, a fishery could be managed by limiting effort to  $E_{\text{MSY}}$ , but if the stock biomass is depleted, then the average long-term yield will not result. In fact, the  $E_{\text{MSY}}$  effort level may be too high to permit stock rebuilding.

Few of these potential management outputs are of value without some idea of the uncertainty around their values. It would also be very useful to be able to project the models into the future to provide a risk assessment of alternative management strategies.

## 11.5 Beyond Simple Models

### 11.5.1 introduction

Prager (1994) described and reviewed a range of extensions to simple surplus production models. He gave detailed descriptions of some of the fundamental equations relating to the population dynamics and to some of the management targets ( $F_{0.1}$  has already been mentioned). He also discussed the handling of multiple data series and missing data, along with suggestions for dealing with changing catchability through time. He briefly described how to estimate the uncertainty around parameters using bootstrap procedures, and an algorithm, that was an extension of the bootstrap, for using surplus production models in projections. Projections are necessary for conducting risk assessments of different proposed management options.

Prager's (1994) paper is recommended reading, but we will still consider some of these subjects here so as to give more detail and make further extensions. We will focus on the more general surplus production model suggested first by Pella–Tomlinson but developed by Polacheck et al. (1993) (Equation 11.19 and following equations, but with  $p$  constrained to  $=0.00000001$  instead of  $=1$ ; Sections 11.3.1 and 11.4.1).

When  $p = 1$  the Pella–Tomlinson model is algebraically identical to the Schaefer model. This fact, and its symmetrical production curve, has even led to suggestions that the simpler Schaefer model no longer be used (Maunder, 2003). However, as Prager (2003) argues well, it is often difficult to generate an acceptable estimate of the asymmetry parameter,  $p$ , and when data are limited or of marginal quality, the well-known and simpler model has

© 2011 by Taylor & Francis Group, LLC

advantages (the Schaefer model is simpler because it implies linear density dependence). One of Prager's (2003, p. 154) conclusions was: "For any stock analyzed with a production model, it will be useful to conduct sensitivity analyses of management recommendations to assumptions on model shape." This is eminently sensible advice and reflects the fact that any model is likely to have assumptions that may not be met by a natural population. Irrespective of how well a model fits to available data, it remains only an abstraction of reality and it is best to be suspicious of its behaviour. Such suspicions may be alleviated through well-chosen sensitivity tests.

### 11.5.2 Changes in Catchability

One major assumption in the use of surplus production models is that the relationship between catch rates and stock biomass is constant ( $C/E = qB$ ). This relationship implies that the catchability coefficient,  $q$ , remains constant through time. In fact, because fishers tend to be good at what they do, there tend to be continual improvements to fishing gear and fishing practices such that the effectiveness of each unit of effort increases through time. This effort creep is often considered in terms of changes in fishing power brought about, for example, by introducing new gear such as radar, coloured echo sounders, and Global Positioning System (GPS) receivers and plotters (Brown et al., 1995). By using general linear models to compare the catch rates of vessels that had adopted GPS and related plotters in different years, Robins et al. (1998) found that vessels in the Australian Northern Prawn Fishery obtained a 4% advantage with the introduction of GPS, and this figure grew to 7% if a plotter was also installed. Over the subsequent two years there were further improvements of between 2 and 3% per year (i.e., learning was a factor). Overall, once the complete fleet had adopted the technology (a matter of three to four years), the increase in fishing power accorded to this alteration alone was 12%. Multiplying the units of effort by 1.12 is a possibility, but such an approach would make the units of effort confusing. For example, if effort were in hours fished, then it would become necessary to refer to effort as hours standardized relative to some reference year (100 hours in 1998 might be 112 hours in 1994). Instead, perhaps the simplest interpretation to place on increases in fishing power is to consider them as changes to the catchability coefficient. In numerical terms, because  $C = qEB$ , it does not matter whether the  $E$  or the  $q$  changes.

Clearly, the assumption that  $q$  is a constant is rather an oversimplification. Prager (1994) pointed out that if it were suspected that the catchability coefficient had changed rather suddenly, then the nonequilibrium model could be applied as if there were two time series of catches and catch rates. The same parameters ( $r$ ,  $K$ ,  $B_0$ , and perhaps  $p$ ) would apply to each time series and would be fitted together. However, there would need to be as many  $q$  parameters as there were separate time series, and these would need to be fitted separately. Alternatively, two or more sets of closed-form calculations could be produced, but if the number of observations in each time series becomes very

© 2011 by Taylor & Francis Group, LLC

low, then the closed-form calculations may become suboptimal and direct estimation might be more robust. Each suspected major change in catchability would entail the addition of a further parameter. Naturally, as the number of parameters increases, one would expect the quality of model fit to improve. Prager (1994) suggests using an  $F$  ratio test to compare the simple models with the more complex. This would be equivalent to using a likelihood ratio test.

Prager (1994) also considers a linear increase in catchability through time. This would be equivalent to a constant absolute improvement each year:

$$q_t = q_0 + q_{add} \times t \quad (11.30)$$

where the  $t$  subscript denotes the particular year,  $0$  to  $n - 1$ ,  $q_0$  is the catchability in the first year, and  $q_{add}$  is the constant increase added to the catchability each year. Prager (1994) suggested this could be parameterized by estimating the first and last year's  $q$  and interpolating for the intervening years. Perhaps this would be most easily implemented by using Equation 11.30, directly estimating the  $q_0$  for the first year and then the  $q_{add}$  that provides the best fit. Alternatively, a closed-form estimate of  $q_0$  and  $q_{add}$  can be generated by implementing the appropriate regression analysis (see Appendix 11.2 for the derivation).

In some fisheries it has been suggested that there is a constant proportional increase in catchability each year. For example, in the Australian northern tiger prawn fishery the annual proportional increase in the effectiveness of effort previously accepted by managers and industry (for purposes of discussing effort reduction targets) was 5% per annum (Pownall, 1994). Thus, instead of Equation 11.30, we would need

$$q_t = q_0 \times q_{inc}^t \quad (11.31)$$

where  $q_t$  is the catchability in year  $t$  and  $q_0$  is the catchability in the first year. In year 0, the  $q_{inc}$  would be raised to the power zero and hence equal 1. For a 5% per annum increase,  $q_{inc}$  would = 1.05. As with the additive form of catchability increase, closed-form estimates of  $q_0$  and  $q_{inc}$  can be obtained if we log-transform Equation 11.31 to give it the form of a linear regression (see Appendix 11.2 for the derivation):

$$\text{Ln}(q_t) = \text{Ln}(q_0) + t \times \text{Ln}(q_{inc}) \quad (11.32)$$

### 11.5.3 The Limits of Production Modelling

We will consider the Australian northern tiger prawn fishery and illustrate some of the variations possible when implementing nonequilibrium surplus production models. This fishery extends across the top of Australia from the Gulf of Carpentaria to the west of Joseph Bonaparte Gulf (Pownall, 1994). The fishery has been operating for over thirty-eight years with significant tiger prawn catches since 1970 (Table 11.2, Figure 11.6). Management is via

© 2011 by Taylor & Francis Group, LLC

TABLE 11.2

Published Catch, in Tonnes, and Effort, in Fishing Days, for the Northern Australian Tiger Prawn Fishery

Year	Catch	Effort	Year	Catch	Effort	Year	Catch	Effort
1970	1,138	5,818	1980	5,124	30,594	1990	3,550	25,525
1971	1,183	6,057	1981	5,559	31,895	1991	3,987	20,744
1972	1,380	7,380	1982	4,891	32,956	1992	3,084	21,789
1973	1,672	7,362	1983	5,751	34,551	1993	2,515	16,019
1974	666	3,439	1984	4,525	32,447	1994	3,162	18,592
1975	973	6,010	1985	3,592	26,516	1995	4,125	16,834
1976	1,118	6,660	1986	2,682	26,669	1996	2,311	16,635
1977	2,900	11,673	1987	3,617	22,478	1997	2,694	15,385
1978	3,599	18,749	1988	3,458	26,264	1998	3,250	18,003
1979	4,218	17,791	1989	3,173	27,036			

Source: From Pownall, 1994; AFMA, 1999.

Note: Data are for both the brown (*Penaeus esculentus*) and grooved tiger prawns (*P. semisulcatus*) combined. Catch rates for each year can be determined by dividing the catch by the effort for each year.

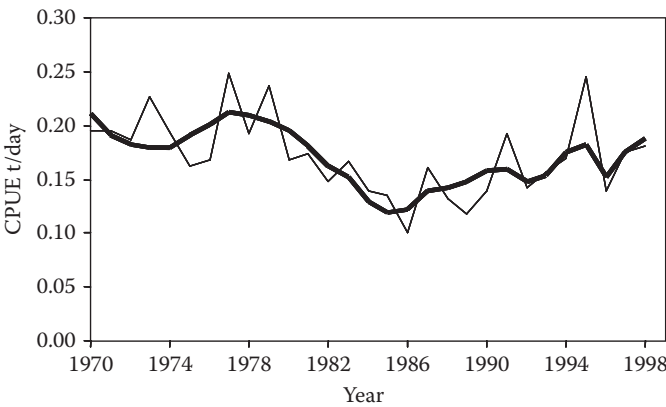


Figure 11.6

Observed catch-per-unit-effort data from the northern Australian tiger prawn fishery (see Table 11.2) as the fine irregular line, with an optimal model fit as the thicker smoother line (Example Box 11.4). The optimal model was constrained to be equivalent to the modified Fox model ( $p = 0.0000001$ ; Equation 11.17; Fox, 1975).

input controls (being a mixture of limited entry, closed seasons, closed areas, and gear controls), and stock assessment uses a relatively complex model (Wang and Die, 1996). Nevertheless, because effort creep has been identified as a major issue in this fishery, it provides an opportunity to apply the techniques described earlier involving the estimation of effort creep levels.

© 2011 by Taylor & Francis Group, LLC

## EXAMPLE BOX 11.4

Nonequilibrium surplus production model of the northern Australian tiger prawn fishery (Table 11.2, Figure 11.6). The row numbers are omitted to save space, but they start in row 1. The manipulations in C1:C4 assist the solver by keeping the values to vary similar in value; alternatively, logs could be used. Name C1 as *re*, C2 as *K*, and C4 as *p*. Select B5:C5 and put `=linest(F9:F37,I9:I37,true,false)` using <Ctrl><Shift><Enter> to enter the array function (cf. Equation 11.32). The closed-form estimates of *q* and *qinc* are in B6:C6. Column I contains numbers 0 to 28 representing *t*-1 in the closed-form calculations. In B7 put `=(F6/2)*(Ln(2*pi())+2*Ln(F7)+1)`, which is Equation 11.25. Enter the data into columns A to C from Table 11.2 (CE\_Obs is Catch/Effort). D9 is B0 (=C3). D10 is `=max(D9+D9*(re/p)*(1-(D9/K)^p)-B9,100)`, which is Equation 11.33 with the max function preventing the possibility of negative biomass (which would halt the solver). G9 is `=(Ln(C9)-Ln(E9))^2`, which represents the lognormal squared residuals. Select E9:G9 and copy down to row 10. Select D10:H10.

Plot columns C and E against A to mimic Figure 11.6. Optimize the model fit by maximizing B7 by varying B1:B3, leaving the *p* value as it is. Save the parameter and model output values somewhere else on the sheet. Now solve by varying B1:B4 to see the impact on the *p* value when it is free to vary. Are the values for *qinc*, *p*, and B0/*K* reasonable? Would the shape of the production curve make biological sense (cf. Figure 11.1)? Alter the structure of the sheet to estimate a constant *q* (put `=exp(average(F9:F37))` into C6, and copy H9 down to row 37). Re-solve, for either B1:B3 or B1:B4. How big an impact would this have on the model outputs? Because the multiplicative effects of *qinc* lead to an exponential impact on catchability, the results are very sensitive to the model of *q* used. Try implementing the additive model of *q* and *qadd*. When solving this model, it may be necessary to search carefully for the optimal solution. Try solving for individual parameters and moving toward the correct solution if you lose the optimum. Try varying the parameters to see how sensitive the solution is to the starting position. Can you find any false minima? If you do, does altering the options in the solver alter the solution? Is adding a constraint that *p* be greater than zero a good idea? Try the different possibilities listed for generating B0 (independent fit, *K*, *=CE/q*). How much do they affect the results? Plot the predicted biomass history for the stock. Are things looking good for the northern Australian tiger prawn fishery according to this model? Compare the solution obtained for the maximum likelihood with that obtained from the summed squared residuals. Implement normally distributed residuals (put `=(C9-E9)^2` into G9 and copy down), and solve by minimizing these. How big a difference does the residual error structure make?

*continued*



EXAMPLE BOX 11.4 (continued)								
A	B	C	D	E	F	G	H	I
r	32.965	=B1/100		B98	=D37			
K	27.3301	=B2*1000		B98/K	=F1/C2			
B0	42.1005	=B3*1000		MSY	=(C1*C2)/((C4+1)^((C4+1)/C4))			
p	1	=B4/1E9		B0/K	=C3/B2			
q	0.0763	-12.2195	Ln(q0)	Ssq	=sum(G9:G37)			
qinc	=exp(B5)	=exp(C5)	q0	n	=count(G9:G37)			
LL	17.2676			Sigma	=sqrt(average(G9:G37))			
Year	Catch	CE_Obs	PredB	Pred_I	Pred_q	SSQ	q	Yr
1970	1138	0.1956	=C3	=D9*H9	=Ln(C9/D9)	0.0072	=\$C\$6	0
1971	1183	0.1953	35774	0.1905	-12.118	0.0006	=H9*\$B\$6	1
1972	1380	0.1870	31593	0.1816	-12.037	0.0009	0.0000058	2
1973	1672	0.2271	28840	0.1789	-11.752	0.0569	0.0000062	3
1974	666	0.1937	26767	0.1792	-11.837	0.0060	0.0000067	4
1975	973	0.1619	26377	0.1906	-12.001	0.0267	0.0000072	5

The dynamics of the model are described by the same nonequilibrium model as used with the Schaefer data (Equation 11.19), along with the relationship between catch rates and biomass illustrated by Equation 11.20. The estimation of  $q$  will involve either the constant catchability (Equation 11.23) or the additive or the multiplicative incremental increases in  $q$ , as in Equations 11.31 and 11.32. The precise relationship between catch rates and stock biomass is

$$\hat{I}_t = \frac{\hat{C}_t}{E_t} = qB_t e^{\epsilon}$$

(11.33)

In this case the biomass is not averaged across two years, as the prawns are almost annual in their life cycle, with very few animals surviving from one year to the next.

When fitting such an array of options it is obviously best to start simple and progress to the more complex.

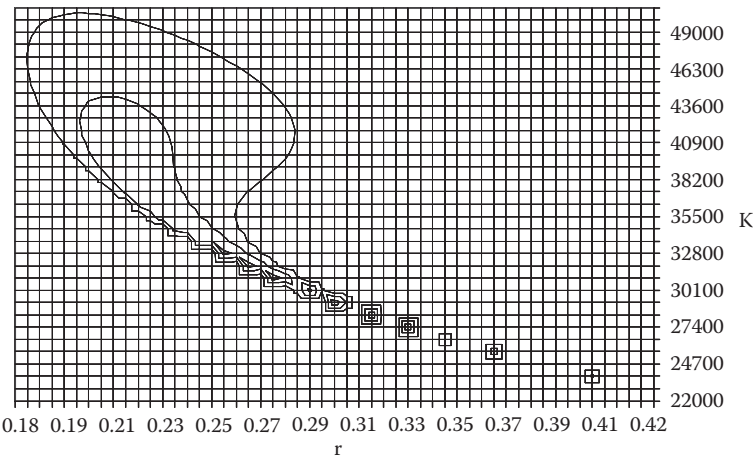
## 11.6 Uncertainty of Parameter Estimates

### 11.6.1 Likelihood Profiles

Polacheck et al. (1993) used the log-likelihood criterion, even though it provides the same estimates as the least squares estimates (as long as the

© 2011 by Taylor & Francis Group, LLC





**Figure 11.7**  
Approximate 95% confidence intervals from the two-dimensional likelihood profile, for the  $r$  and  $K$  parameters in the nonequilibrium surplus production model for the northern Australian tiger prawn fishery (as in Example Box 11.4). The optimum solution was at  $r = 0.32965$  and  $K = 27,330$ . The target log-likelihood was 14.273, which is the outer curve (Example Box 11.5). The gaps in the tail are unreal, and the resolution of the graph is insufficient to show the detail of connected contours.

$I$  values are log-transformed first so as to keep the same error structures). They did this because they also suggested using Venzon and Moolgavkar's (1988) approximate likelihood profile method to produce confidence intervals around the parameter estimates. The methodology behind this was discussed in Chapter 3 on parameter estimation (Section 3.4.14) and in Chapter 8 (Section 8.4). For single parameters the results are essentially the same as standardizing the log-likelihoods so they all add to 1, and then finding the confidence intervals by using the parameter limits that contain 95 or 99% (or whatever confidence interval chosen) of the likelihood curve. Polacheck et al. (1993) found that the likelihood profiles obtained when using observation error estimators were much smaller than those deriving from process error estimators.

A maximum of two parameters can be visualized at one time. To determine the likelihood profile confidence intervals, as shown in Chapters 3 and 8, involves subtracting  $5.9915/2 (= \chi^2/2, \text{ for } 2 \text{ degrees of freedom})$  from the maximum likelihood and searching for the parameter combinations that generate this likelihood (Figure 11.7, Example Box 11.5).

11.6.2 Bootstrap Confidence intervals and estimates of Bias

There are disadvantages to using likelihood profiles. These include the complexity of implementing the method when there are many parameters

EXAMPLE BOX 11.5

The generation of confidence intervals using likelihood profiles. First construct the table of likelihoods from the array of values for the two parameters of interest, in this case,  $r$  and  $K$ . The sixth row relates to the  $K$  value, and column U7 downwards, to the  $r$  parameter. Copy W6 across to column BB, then copy U8 down to row 59. Cell U6 points to the likelihood value in B7. Select U6:BB59 (or however big you have made the table), then choose the Data/Table menu item. In the dialog box the row input cell is \$B\$2 and the column input cell is \$B\$1. This should complete the Excel table. To plot this as a contour surface plot it is best to convert the borders to text so that the axes are automatically labeled. In V66 put =text(V6\*100,"#0") and copy across to column BB. In U67 put =text(U7/100,"#0.000") and copy down to row 119 to match the table above. In V67 put =V7 and copy down, and then copy the column across to regenerate the table. Select the whole table, including the borders (U66:BB119), and generate the contour surface plot by selecting the appropriate menu item (Insert/Chart/Surface). The target log-likelihood for the approximate 95% confidence intervals for two parameters would be the optimum  $\pm 5.99/2$  ( $=B7-\text{chiinv}(0.05,2)$ ), which in this case = 14.273. Double-click on the legend and alter the scale to a minimum of 13.273 and a major unit (tick value) of 1.0, which will lead to a graph similar to Figure 11.7. Drag the graph up near to U1:U4 and experiment with the table ranges and the scale shown. Be wary of selecting a major unit that is too small, else the poor hardworking hardware can reach its limits and you will need to crash out of the program. Do *not* run the solver if a table is active unless you have a great deal of time to spare (deleting the core of the table—V7:BB59—will fix that problem).

	T	U	V	W	X
1	K	22			
2		0.9			
3	r	17.5			
4		0.48			
5					
6		=B7	=U1	=V6+\$U\$2	26.2
7		=U3	=table(B2,B1)	=table(B2,B1)	=table(B2,B1)
8		=U7+\$U\$4	=table(B2,B1)	=table(B2,B1)	=table(B2,B1)

involved (parameter correlation usually increases the width of confidence intervals), and likelihood profiles give no indication of any bias in the parameter estimates and bias can be an important aspect of uncertainty. Fortunately, alternative methods exist for fitting confidence intervals around parameter estimates. A common approach is to use a bootstrap strategy as introduced in Chapter 6. This approach resamples the residuals from the optimum fit to generate new bootstrap samples of the observed time series. The model is fitted to many replicate bootstrap samples and the outputs stored so that percentile confidence intervals can be determined, as is usual with bootstrap methods. The confidence intervals generated can be asymmetric and synthesize the effects of all the parameters varying at once. A further advantage of the bootstrap strategy is that one can estimate whether or not the parameter estimates are biased.

Surplus production models are fitted to time series of relative abundance indices (CPUE). To obtain confidence intervals and bias estimates using bootstrap procedures, it is important that the time series nature of the data not be disrupted. So that the time series nature of the CPUE data is retained, one should not bootstrap the raw data but instead bootstrap the residuals between the observed and expected values, i.e., randomly sample from the original best fit residuals, with replacement, to generate a new vector of bootstrapped residuals. This vector of bootstrapped residuals is combined with the optimum vector of expected CPUE data to obtain each new bootstrap sample of CPUE data (Prager, 1994; Haddon, 1998).

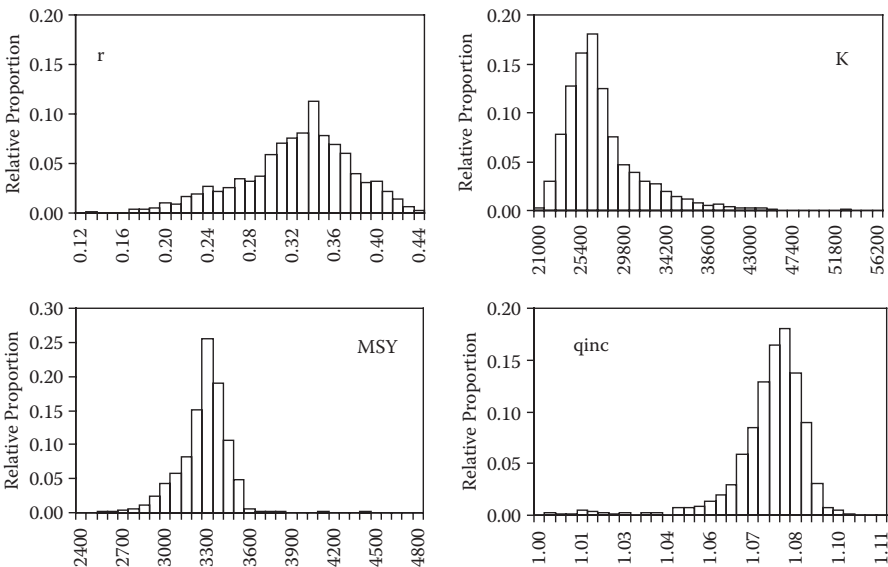
If we had been using normal, additive errors, then the residuals would simply be (observed CPUE – expected CPUE), and after bootstrapping, these would each have been added to the sequence of original expected CPUE value (Equation 11.34):

$$I_t^* = \hat{I}_t + (I_t - \hat{I}_t)^* \equiv \hat{I}_t + \varepsilon^* \quad (11.34)$$

where  $I_t^*$  is the bootstrapped CPUE value that equals the expected CPUE value,  $\hat{I}_t$ , plus a bootstrapped normal residual  $(I_t - \hat{I}_t)^*$ , which is equivalent to combining each expected catch/effort value with a bootstrapped error or residual term. In this way a new time series of bootstrapped “observed” CPUE data is generated to which the model may be refitted and the parameters may be reestimated. However, with lognormal, multiplicative errors (see Equation 11.33), we must use the ratio of the CPUE values (observed/expected) to calculate the residuals that are to be bootstrap sampled. To obtain the bootstrapped CPUE values the residuals are multiplied with their respective original expected CPUE values (Equation 11.35):

$$I_t^* = \hat{I}_t \times \left( \frac{I_t}{\hat{I}_t} \right)^* \quad (11.35)$$

© 2011 by Taylor & Francis Group, LLC



**Figure 11.8** Bootstrap distributions for four of the parameters and model outputs from Example Box 11.6, for tiger prawns until 1998. By comparing the panels for the  $r$  and  $K$  parameters and Table 11.3, the first-order bias correction can be seen to shift the confidence intervals in the direction of the skew of the distribution of bootstrap values (Example Box 11.7). The lower two panels represent values of interest to management. The MSY can be interpreted as the long-term average yield expected from the stock when it is at its optimum size. The value of  $qinc$  suggests the level of effort creep, and this indicates the urgency of any measures to limit effort in the fishery.

Confidence intervals can be estimated by generating thousands of bootstrap samples, refitting the model, and generating an equal number of parameter estimates (these could include outcomes such as MSY). The central 95% of these (the 0.025 and 0.975 percentile values of the sorted estimates) would represent the bootstrap percentile confidence intervals (Figure 11.8, Example Box 11.6).

If we wished to take into account any bias in the parameter estimates, we would do best to calculate bias-corrected percentile confidence intervals (Efron and Tibshirani, 1993). Percentile confidence intervals are determined by using the 25th and 975th ordinal values (out of one thousand replicates for 95% intervals). Bias correction leads one to use different percentile values, depending on whether the parameter estimates are positively or negatively biased. The procedure begins by determining what proportion ( $LT$ ) of the bootstrap replicates are less than the original optimal fit estimate of the parameter or output of interest, and this value is transformed via the inverse cumulative standard normal distribution ( $\Phi^{-1}$ ):

$$z = \Phi^{-1}(LT) \tag{11.36}$$

**EXAMPLE BOX 11.6**

Bootstrap implementation for the surplus production model from Example Box 11.4. The procedure replaces the observed catch rates with a bootstrap sample; it is therefore prudent to store a copy of the original catch rate data into column J. Select E9:E37, copy as values into K9 (Paste Special/Values). Given the optimum fit put  $=C9/E9$  into L9 and copy down L37. These are the residuals around the catch rates calculated as (observed/expected). An easy error is to forget to convert these to values. Select L9:L37, copy, and paste as values onto themselves so they are ready for bootstrap sampling. The resampling is conducted in column M using the offset function. The  $\text{trunc}(\text{rand}()*29)+1$  will provide a random integer between 1 and 29. Copy M9 down to row 37. The bootstrap sample is generated in column N by multiplying the selected residual with the expected catch rate in that year. To conduct the bootstrap select N9:N37 and copy as values into C9 and re-solve for an optimum to provide a single set of bootstrap estimates, which will need to be stored as values somewhere else on the worksheet. The cells in column P have been arranged to make this copying and storage a simpler process. Of course, to do the bootstrapping in a sensible way, one needs to write a macro to do the necessary copying, solving, and storing of results. Cells L1:L3 and M1:M3 all relate to a suitable macro to do the bootstrapping. Before constructing the macro, carry out the bootstrap a few times manually. Under the solver options be sure to provide generous time and iteration limits, and refine the precision and convergence criteria. For this problem the other options that seem to provide stable answers are Estimate: Tangent, Derivatives: Forward, and Search: Conjugate, but do try the alternatives. Do you think it is a good idea to always start the search from the original optimum? To provide a convenient source of the optimum value is why they were copied into L1:L3 before bootstrapping started. Do you ever have to run the solver twice to find a stable optimum? It would be a good idea to run it twice for each bootstrap sample, giving the solution a slight shift before the second solve (the reason for M1:M3). Add a few of the bootstrap samples to the plot of the observed and expected catch rates (retaining the original observed values). How closely do they compare? Create a macro to do the bootstrapping. Start with only a few bootstraps to see how long it takes and whether there are problems. Then set it going for one thousand bootstraps. See the following macro for details that cannot be recorded.

*continued*

EXAMPLE BOX 11.6 (continued)

	K	L	M	N	O	P
1	r	32.969	=B1+B1/100		r	=C1
2	K	27.328	=B2+B2/100		K	=C2
3	B0Est	42.106	=B3+B3/100		B0Est	=C3
4					p	=C4
5					q0	=C6
6					qinc	=B6
7					Bcurr	=F1
8	Exply	Resid	Resample Residuals	Bootstrap	Bcurr/K	=F2
9	0.2117	0.9242	=offset(\$L\$8,trunc(rand()*29)+1,0)	=K9*M9	MSY	=F3
10	0.1897	1.0295	=offset(\$L\$8,trunc(rand()*29)+1,0)	=K10*M10	B0/K	=F4
11	0.1812	1.0320	0.83075	0.209113	LogLike	=B7

Note the solver is used twice and that the predicted biomass time series and the bootstrap sample are all stored along with the results. The times will only be indicative but will determine the approximate time taken for one thousand replicates. You may wish to place labels across the columns starting at S13 with r. All bootstrap analyses should include the original optimum solution (put it in S14:AC14). The bootstrap estimates may be tabulated and percentile intervals calculated (Example Box 11.7, Figure 11.8).

```
Option Explicit
Sub Do_Bootstrap()
Dim i As Integer
Application.ScreenUpdating = False
For i = 1 To 10 ' Set the number of bootstraps
Range("N9:N37").Select
Selection.Copy
Range("C9").Select
Selection.PasteSpecial Paste:=xlValues
Range("L1:L3").Select ' Paste original optimum
solution
Selection.Copy
Range("B1").Select
Selection.PasteSpecial Paste:=xlValues
SolverOk SetCell:="$B$7", MaxMinVal:=1, ValueOf:="0",
ByChange:="$B$1:$B$3"
SolverSolve (True)
```

continued

**EXAMPLE BOX 11.6 (continued)**

```

Range("M1:M3").Select      \ optimum + 1%
Selection.Copy
Range("B1").Select
Selection.PasteSpecial Paste:=xlValues
SolverOk SetCell:="$B$7", MaxMinVal:=1, ValueOf:="0",
    ByChange:="$B$1:$B$3"
SolverSolve (True)
Range("P1:P11").Select    \ Save the answers
Selection.Copy
ActiveCell.Offset(13 + i, 3).Range("A1").Select
Selection.PasteSpecial Paste:=xlValues, Transpose:=True
Range("D9:D37").Select    \ Save the biomass values
Selection.Copy
ActiveCell.Offset(5 + i, 26).Range("A1").Select
Selection.PasteSpecial Paste:=xlValues, Transpose:=True
Range("C9:C37").Select    \ Save the bootstrap data
Selection.Copy
ActiveCell.Offset(5 + i, 56).Range("A1").Select
Selection.PasteSpecial Paste:=xlValues, Transpose:=True
Application.CutCopyMode = False
Next i
Application.ScreenUpdating = True
End Sub

```

A convenient way to do this would be to use the  $z = \text{NORMINV}(LT, 0, 1)$  in Excel, where the mean of 0 and the standard deviation of 1 imply the standard normal distribution. This  $z$  value is used in the cumulative standard normal distribution ( $\Phi$ ) to calculate the appropriate percentile to use instead of the standard 25th and 975th:

$$\begin{aligned}
 P_{\text{lower}} &= \Phi(2z - 1.96) \\
 P_{\text{upper}} &= \Phi(2z + 1.96)
 \end{aligned}
 \tag{11.37}$$

where  $\Phi$  is the cumulative standard normal distribution function, conveniently provided by  $\text{NORMDIST}(P_{\text{index}}, 0, 1, 0.05)$ , where the 0 and 1 define the normal distribution to be standard, and the 0.05 and 1.96 reflect the normal values required for 95% intervals. With first-order bias-corrected confidence intervals, if  $LT$  were 0.5, then  $z$  would be zero and we would, of course, obtain the 25th and 975th percentiles (Haddon, 1998). However, if, for example,  $LT$  were 0.459, then  $z$  would be  $-0.10295$ , which would lead us to

TABLE 11.3

Comparison of Different Approximate Confidence Intervals for the Two Parameters  $r$  and  $K$ , from the Nonequilibrium Surplus Production Model for the Northern Australian Tiger Prawn Fishery (Example Boxes 11.4 to 11.7)

	$r$			$K$		
Interval type	Lower95	Average	Upper95	Lower95	Average	Upper95
Likelihood profile	0.1825	0.32965	0.4050	23,800	27,330	50,400
Bootstrap percentile	0.2303	0.32965	0.4194	23,475	27,330	38,820
BC <sub>1</sub> percentile	0.1954	0.32965	0.4098	24,095	27,330	44,905

*Note:* The first-order bias-corrected bootstrap percentile intervals are closer to the likelihood profile intervals than the straight percentile intervals. These intervals are only approximate and are likely to be underestimates; it is therefore tempting to adopt the widest as providing the best estimates, or having the greatest chance of covering the true value.

use the 16th and 960th percentiles (note they are no longer necessarily symmetrical around the median).

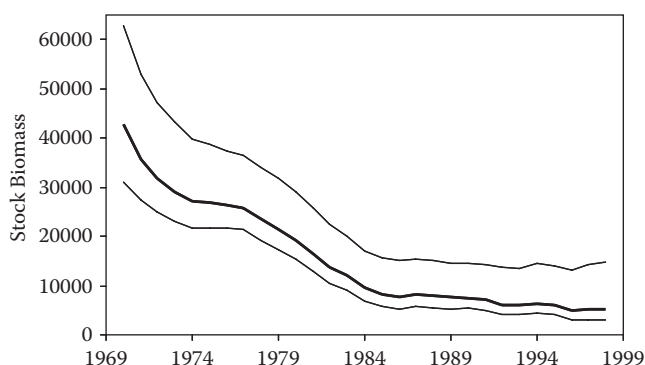
The percentile confidence intervals around the parameter estimates for the northern Australian tiger prawn fishery tend to be narrower than the likelihood profile intervals (Table 11.3). However, all of these confidence intervals are only approximate and only capture the variability inherent in the data, ignoring other sources of variability. These would include the simplicity of the model failing to capture the full dynamics of the population (for example, there are two species of tiger prawns that are lumped in the catch and catch rate information) and the short time series of fisheries data not capturing the full range of environmental variation possible.

The confidence intervals produced by both likelihood profile and bootstrap methods will often be asymmetric and will vary greatly between parameters (Figure 11.8). The predicted biomass estimates can be treated in exactly the same way as the parameter estimates, so that the estimated history of the stock biomass can be illustrated (Figure 11.9). The confidence intervals are relatively wide around  $B_0$ , which would be typical of the uncertainty surrounding this parameter (Figure 11.9). In this fishery, the situation is even more complicated in that the fishery was only developing over the years 1970 to 1975, so it is difficult to be certain as to how meaningful the early catch rates are in terms of stock biomass. Note the asymmetry of the confidence intervals. The lower bound is much closer to the average than the upper. This makes sense, as there must be a certain minimum present to sustain the history of the catches that have been taken from the fishery. In this case, bias correction makes only a slight difference (from Example Box 11.7, try plotting the two data series for comparison). Clearly, the model outputs are consistent with the stock being in

© 2011 by Taylor & Francis Group, LLC

Copyright © 2011. CRC Press LLC. All rights reserved.





**Figure 11.9**

The predicted time series of stock biomass for the northern Australian tiger prawn fishery (Example Boxes 11.6 and 11.7). Most of the severe decline is driven by the continually increasing catchability coefficient. If  $q$  is increasing and catch rates stay stable, then this actually implies that the stock biomass must be declining. This is one reason why unstandardized catch rates are dangerous as a basis for managing a fish stock.

a relatively depressed state. Over the last ten to twelve years the stock has been in a relatively low level, below the size at which we might expect it to be maximally productive.

None of the parameter estimates and model outputs in the tiger prawn fishery are especially certain. It would almost always be a good strategy to obtain further information from the fishery rather than attempting to estimate peripheral parameters within the model. Systematic changes in catchability can have a huge effect, and using such a model form should not be adopted routinely. In the case of the tiger prawns, if changes to catchability could be determined empirically by determining the timing of novel gear changes and introductions, and their relative effect, this would have the potential for providing greater certainty. Of greatest value would be to obtain a direct and relatively precise estimate of abundance in more than one year. This could be used to anchor the model to reality along with the catch rates. More than one data series can be included in the likelihood equation. Each series might best be weighted in accord with its relative precision (standard error or coefficient of variation for each estimate, or some other estimate of its relative precision).

## 11.7 Risk Assessment Projections

### 11.7.1 introduction

Invariably, there will be many sources of error and uncertainty that are not accounted for in the model. Determining the uncertainty in an analysis only

© 2011 by Taylor & Francis Group, LLC

EXAMPLE BOX 11.7

Calculation of the bootstrap percentile confidence intervals and the first-order bias-corrected percentile intervals from the results of the bootstrapping from Example Box 11.6. The macro shown in Example Box 11.6 will deposit the bootstrap estimates in cells S14:CI1013 (assuming one thousand replicates). If more than one thousand are calculated, then the ranges in the various functions listed here will need to be extended. This assumes that the optimum answers were all selected and pasted into row 14 (hence S4 and T4). S1:S3 are just the standard bootstrap percentile intervals and the bootstrap estimate of the parameter. Generally the median in S5 (could be =median(S14:S1014)) will be closer to the optimum model fit when there is any bias. S6 is the count of bootstrap estimates that were less than the optimum value in S4, which must be converted to text for the countif function to work. The many decimal places are to obtain robust answers across the range of values experienced by the different parameters. S7 is Equation 11.36 and S8:S9 is Equation 11.37, which are used in the percentile estimates of S10:S11. S12 is an estimate of the bias shown in the bootstrap estimates. Select S1:S12 and copy across to column BF. The formatting (number of decimal places) will need adjustment in each column. Store these equations in a separate worksheet and delete the originals if you wish to do more bootstrapping (otherwise they slow the calculations a great deal). Your row 15 will differ from that shown here because the bootstrap replicates will differ. Construct a bootstrap for the eastern Pacific yellowfin tuna (Example Box 11.3). How certain is the nonequilibrium estimate of the average long-term yield, MSY? The bias correction is converting the confidence intervals from being centered on the average to being centered more on the median of the distribution.

	R	S	T	U	~	X
1	U95%	=percentile(S14:S1014,0.975)		64250.4		1.09289
2	Average	=average(S14:S1014)		43169.6		1.07631
3	L95%	=percentile(S14:S1014,0.025)		31376.5		1.05388
4	Optimum	=S14	=T14	42105.6		1.07931
5	Median	=percentile(S14:S1014,0.5)		42512.1		1.07902
6	LT_mean	=countif(S14:S1014,"<" & text(S4,"#0.0000000"))				359
7	Z	=norminv(S6/1000,0,1)		-0.60828		-0.91727
8	Pupper	=normdist(2*S7+1.96,0,1,0.05)		0.77139		0.54992
9	Plower	=normdist(2*S7-1.96,0,1,0.05)		0.00075		0.00007
10	U95%	=percentile(S14:S1014,S8)		48094.6		1.07992

continued

EXAMPLE BOX 11.7 (continued)						
	R	S	T	U	~	X
11	L95%	=percentile(S14:S1014,S9)		22299.0		0.99272
12	%Bias	=100*(S2-S4)/S4				
13		r	K	B0Est		Qinc
14		0.32969	27327.694	42105.588		1.079306
15		0.210056	40102.191	45006.252		1.039695

tells us that we need to be careful when attempting to interpret the model outcomes; it cannot inform resource managers about the risk level associated with a particular management option. To answer such questions a risk assessment is required.

Risk assessment implies projecting the population dynamics model into the future under the constraint of different management options (for example, a particular catch or effort regime, or different open and closed areas, etc.). Given the selected catch or effort, we need to be able to model the projected recruitment levels in a stochastic manner, with the variability of that recruitment reflecting the stock dynamics observed in the available time series of data. The projected recruitments would be offset against the catches, and the trajectory of the stock biomass through time could thus be generated. The problem, when using surplus production models, is to generate these stochastic recruitments.

11.7.2 Bootstrap Projections

Prager (1994) suggested that because surplus production models imply a recruitment function, they could be used to make projections based upon hypothetical catch or effort allocations. In the standard operation of surplus production models the stock biomass is projected forward under the constraint of the time series of catches and catch rates (e.g., Equation 11.19). To do this for a risk assessment would be simply to extend this stock biomass projection beyond the years for which data are available. The projected catches or efforts (which, given the catchability and stock biomass, would imply catches) would be dictated by proposed management options. The stock biomass projections are deterministic, so a mechanism for introducing the required stochastic element is still required. Prager’s (1994) suggestion for varying the population projections was to conduct a bootstrap analysis and project each bootstrap forward to obtain a risk assessment at the same time as a determination of the level of uncertainty in the analysis. This mechanism uses the variation inherent in each bootstrap sample to represent the variation likely to occur in the stock dynamics of the species

© 2011 by Taylor & Francis Group, LLC

Copyright © 2011. CRC Press LLC. All rights reserved.

concerned. The variation is represented by the residuals between the observed and predicted catch rates, which are assumed to relate back to stock biomass via the catchability coefficient ( $C/E = qB$ ).

### 11.7.3 Projections with Set Catches

Many fisheries are now managed through output controls in the form of a total allowable catch (TAC). In such fisheries, a vital management control is to set the TAC at a level consistent with stock sustainability and, often, with optimizing production. Investigation of the implications of setting different catch levels is relatively simple with the surplus production models described in this chapter. If the stock dynamics are assumed to be described by the deterministic equation

$$B_{t+1} = B_t + \frac{r}{p} B_t \left( 1 - \left( \frac{B_t}{K} \right)^p \right) - C_t \quad (11.38)$$

or any production function from which a catch is subtracted, then forward projection only requires those catches to be defined and the projections can be implemented (Example Box 11.8).

Exactly what characteristic of the population to consider in the projections can vary, depending on circumstances and what would be most informative. For example, a common performance indicator would be to determine whether the predicted stock biomass in any given projection year is greater than a selected reference year. If many replicate projections are generated, then in any year the proportion in which the stock biomass is greater than that in the reference year can represent the probability that the modelled stock will have increased in size in that year (cf. Figures 11.10 and 11.11). Alternatively, if there is a risk of stock collapse, then the number of replicates under a given set of management constraints that led to collapse (defined as some low biomass level) can also be collated and graphed.

### 11.7.4 Projections with Set effort

The northern Australian tiger prawn is an input-controlled fishery, so the management controls used to constrain the projections will involve considering the impact of different effort levels. There is also the problem of effort creep to attend to in the risk assessment. In theory, it should be possible to constrain the annual proportional increase in the effectiveness of effort. The presently accepted level is 5% per annum, but it is recognized that this level cannot continue, so alternative, lower levels will also need consideration. The best strategy is to conduct a grid analysis, running the projections for each selected effort level and all the levels of the annual increment in catchability ( $qinc$ ) that are to be considered (Figures 11.10 and 11.11, Example Box 11.8).

© 2011 by Taylor & Francis Group, LLC

EXAMPLE BOX 11.8

Additions to the bootstrapping procedure, implemented in Example Boxes 11.6 and 11.7, to permit risk assessment projections based upon setting the future catch or effort levels. Save the workbook from Example Box 11.6 under a new name and make the following changes. The worksheet below reflects the original data and optimal fit, before any bootstrapping has occurred. In this fishery we do not assume the catchability will stay constant, but will need to trial different catch or effort levels against different levels of *qinc* (hence H1:H2). To use allocated catch levels put  $=\$H\$1$  into B38 and copy down to row 44. To use allocated effort levels put  $=D38*H38*\$H\$1$  into B38 and copy down to row 44 (this is just  $C = qEB$ ; Equation 11.39). Select D37:E37 and copy down to row 44. Copy H37 down to row 44. These simple extensions will generate the population projection for a single instance of the combination of catch or effort level and *qinc*. To generate replicates we need to modify the macro controlling the bootstrap to save only the data we are interested in during the projections (see below). The bootstrap percentile limits can be used to describe the projections as with the usual analysis, but the projections permit many other questions to be answered. Beside the columns of projected stock biomass, we can add columns determining whether, for example, the biomass in the given year is greater than in some reference year. Thus, if only years 1985–2005 are stored according to the following macro, then in AN14 we could put  $=if(AG14>\$AF14,1,0)$ , where AG14 is the predicted biomass in 1999 and AF14 is that in 1998, the selected reference year. If AN14 is copied across to column AT and down to however many replicates were run, we can sum the columns relating to each of the projected years and ask about the probability of the stock being larger than the reference year under the catch or effort and *qinc* regime adopted (Figures 11.10 and 11.11). Any reference year can be chosen. In this way, risk-averse management strategies can be developed.

	A	B	C	D	E	F	G	H
1						Future catch or effort		14500
2						Managed qinc		1.02
~								
8	Year	Catch	Obs_CE	Biomass	Pred_CE	Pred_q	Ln(I-I)	q
~								
36	1997	2694	0.1686	4431.5	0.1749	-10.139	0.0000	0.00002167
37	1998	3250	0.2142	4395.4	0.1872	-10.100	0.0013	0.00002287

continued

EXAMPLE BOX 11.8 (continued)

	A	B	C	D	E	F	G	H
38	1999	=D38*H38*\$H\$1		3793.4	0.1648		=H37*\$H\$2	
39	2000	=D39*H39*\$H\$1		3873.3	0.1716		=H38*\$H\$2	
40	2001	=D40*H40*\$H\$1		3879.6	0.1754		=H39*\$H\$2	
41	2002	2563		3833.9	0.1768			0.00004610
42	2003	2559		3753.4	0.1765			0.00004703
43	2004	2539		3650.8	0.1751			0.00004797
44	2005	2507		3534.5	0.1729			0.00004893

```
Option Explicit
Sub Projection()
Dim i As Integer
Calculate
Application.ScreenUpdating = False
For i = 1 to 100
    Range("N9:N37").Select
    Selection.Copy
    Range("C9").Select
    Selection.PasteSpecial Paste:=xlValues
    Range("L1:L3").Select
    Selection.Copy
    Range("B1").Select
    Selection.PasteSpecial Paste:=xlValues
    SolverOk SetCell:="$B$7", MaxMinVal:=1, ValueOf:="0",
        ByChange:="$B$1:$B$3"
    SolverSolve (True)
    Range("M1:M3").Select
    Selection.Copy
    Range("B1").Select
    Selection.PasteSpecial Paste:=xlValues
    SolverOk SetCell:="$B$7", MaxMinVal:=1, ValueOf:="0",
        ByChange:="$B$1:$B$3"
    SolverSolve (True)
    Range("D24:D44").Select
    Selection.Copy
    ActiveCell.Offset(-10 + i, 15).Range("A1").Select
    Selection.PasteSpecial Paste:=xlValues, Transpose:=True
    Application.CutCopyMode = False
```

continued

EXAMPLE BOX 11.8 (continued)

```
If (i Mod 10) = 0 Then      ' Keep a check on progress.
  Range("L6").Value = i
  Application.ScreenUpdating = True
  Application.ScreenUpdating = False
End If
Next i
Application.ScreenUpdating = True
Beep ' Announce completion
End Sub
```

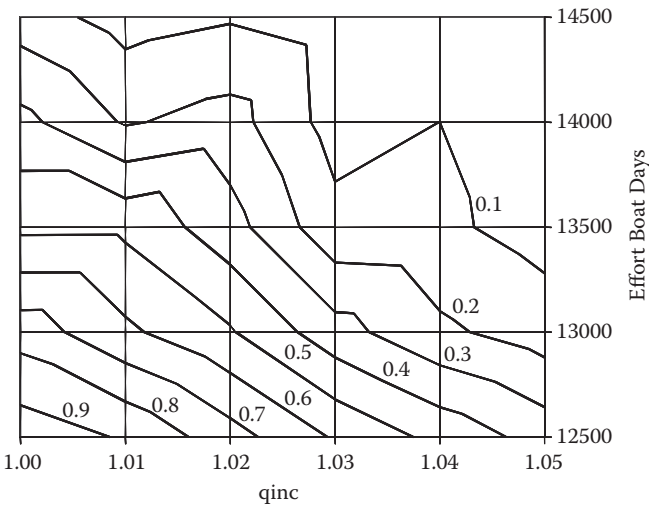


Figure 11.10

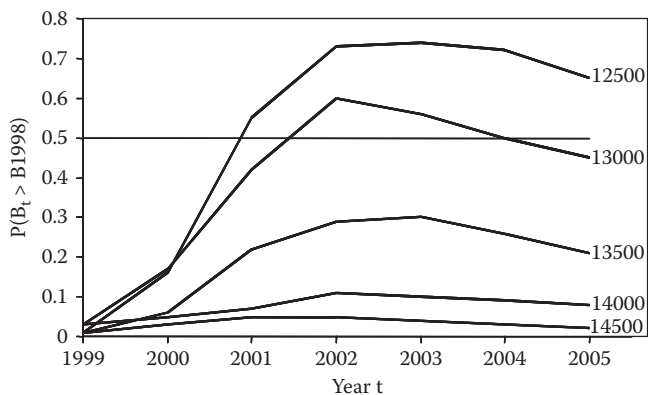
The contours are the probability of stock biomass being greater in the year 2002 than it was in 1998. The contours are not particularly smooth because each intersection is only represented by one hundred bootstrap projections, using a fixed effort strategy in the northern Australian tiger prawn fishery. It is clear that positive stock growth, relative to 1998, will only have a greater than 50% chance of occurring if effort creep (*qinc*) is kept below 3.5% per annum, and effort is less than 13,500 fishing days.

Given a recommended effort level, the model can still be projected into the future by calculating the catch implied by the stock biomass, the catchability in that year, and the effort imposed:

$$C_t = q_t E_t B_t \tag{11.39}$$

Of course, the catches in each projected year are likely to vary from year to year. It is harder to make the stock crash with a constant effort management

© 2011 by Taylor & Francis Group, LLC



**Figure 11.11**  
An alternative view of risk assessment output for the northern Australian tiger prawn fishery. The vertical axis is the probability of the stock biomass in a projected year being greater than the stock biomass in 1998. Each curve relates to a given effort level depicted at the end of each line. The annual increase in catchability was set at 2% ( $qinc = 1.02$ ) for all series. The fine horizontal line is at 50%, indicating the desirability of fixing effort at less than thirteen thousand fishing days to encourage stock rebuilding if  $qinc = 1.02$ .

scenario than with a constant catch level. This is because as stock biomass declines, so does the catch from a certain effort. But a constant catch can be taken until there is no biomass left to take (until an infinite effort is implied).

## 11.8 Practical Considerations

### 11.8.1 introduction

Despite only having minimal data requirements, surplus production models purport to provide an assessment of the state of a given fishery at a particular time. The assumption that the stock and its dynamics can be described purely in terms of its biomass exposes this class of models to some problems peculiar to themselves. Because stock biomass can be either recruitment or standing crop, these possibilities can be confounded in the model outputs.

If the available data exhibit a steady decline in catch rate through time, with no contrast in the ranges of effort imposed over the different stock biomass levels, then the data have little information with respect to the trajectory taken by the stock biomass. The model can have trouble distinguishing between a stock with a high population growth rate and low standing crop and a stock with almost no production but an enormous standing crop from which all the catches have been taken. All the model can do is present outputs that are consistent with the input data, and both of these interpretations

© 2011 by Taylor & Francis Group, LLC



can provide an internally consistent description of events. As long as one is aware of the possibility of the catch rate data having little useful information about the stock biomass, care can be taken in the interpretation of the results from surplus production models.

Equilibrium surplus production models will almost always provide apparently workable management advice, whereas nonequilibrium models can lead to the conclusion that the data provide no information. This latter situation may not appear to be useful, but it is far better to know that the information one has is noninformative than to follow model results blindly.

### 11.8.2 Fitting the Models

Many options for the implementation of surplus production models have been described in this chapter, and it is unlikely that only a single set of options will suit a particular problem. If the results of the modelling are similar irrespective of the options selected, then at least there is a consistent story to tell. If different options lead to different results, then how they differ should inform us about the value of the data and the relative sensitivity to the different parameters. The point is that, wherever possible, multiple options should be used.

In the case of the northern Australian tiger prawns, the constant proportional increase in the catchability coefficient ( $q_{inc}$ ) has an overwhelming impact on the assessment provided by the surplus production model. This parameter has important management implications. If it is assumed equal to 1 (i.e., constant catchability), then the stock does not appear to be as depleted. These simple assessment models can emphasize where research effort should be focused to improve an assessment.

Where there are sufficient years of data available, it is worthwhile to conduct hind-casting trials to compare the assessment results for different series of years. Thus, if data are available from 1970 to 1998, then tabulating any changes derived from comparing years 1970–1995, with 1970–1996, with 1970–1997, and so on, can provide insight into the stability of the model and its outputs.

---

## 11.9 Concluding Remarks

Now that surplus production models have moved away from their equilibrium-based origins they provide a useful tool in the assessment of stocks for which there is only limited information available. Their simplifying assumption implies that any conclusions drawn from their outputs should be treated with caution. Nevertheless, given the constraints of only considering the

© 2011 by Taylor & Francis Group, LLC

total stock biomass, they can provide insights as to the relative performance of the stock through time.

Surplus production models now have surprising flexibility and can be used in risk assessments and to produce management advice that goes well beyond the old traditional performance measure notions of MSY and  $E_{MSY}$ .

Surplus production models have now been developed to a point where even if more information is available and more complex and realistic models can be implemented, it would be sensible to implement a simpler model if only to act as a contrast.

---

### Appendix 11.1: Derivation of Equilibrium-Based Stock Production

The steps between Equation 11.3 and Equation 11.9 are as follows:

$$B^* = B^* + \frac{r}{p} B^* \left( 1 - \left( \frac{B^*}{K} \right)^p \right) - C^* \quad (\text{A11.1})$$

where the \* superscript denotes an equilibrium level. The left-hand  $B^*$  and first  $B^*$  on the right-hand side can be cancelled, and the  $C^*$  can be moved across:

$$C^* = \frac{r}{p} B^* \left( 1 - \left( \frac{B^*}{K} \right)^p \right) \quad (\text{A11.2})$$

Substituting  $B^* = C^*/qE^*$  (Equation 11.5), by assuming equilibrium at all times, we obtain

$$C^* = \frac{rC^*}{pqE^*} \left[ 1 - \left( \frac{C^*}{qE^*K} \right)^p \right] \quad (\text{A11.3})$$

$$pqE^* = \frac{rC^*}{C^*} \left( 1 - \left( \frac{C^*}{qKE^*} \right)^p \right) \quad (\text{A11.4})$$

$$\frac{pqE^*}{r} = 1 - \left( \frac{C^*}{qKE^*} \right)^p \quad (\text{A11.5})$$

$$\left( \frac{C^*}{qKE^*} \right)^p = 1 - \frac{pqE^*}{r} \quad (\text{A11.6})$$

© 2011 by Taylor & Francis Group, LLC

$$\frac{(C^*)^p}{(qK)^p (E^*)^p} = 1 - \frac{pqE^*}{r} \quad (\text{A11.7})$$

$$\frac{(C^*)^p}{(E^*)^p} = (qK)^p - \frac{pq q^p K^p E^*}{r} \quad (\text{A11.8})$$

$$\left( \frac{C^*}{E^*} \right)^p = (qK)^p - \frac{pq^{p+1} K^p E^*}{r} \quad (\text{A11.9})$$

and finally

$$\frac{C^*}{E^*} = \left[ (qK)^p - \frac{pq^{p+1} K^p E^*}{r} \right]^{\frac{1}{p}} \quad (\text{A11.10})$$

If we reparameterize by defining  $(qK)^p$  to be a new parameter  $a$ , and the second term,  $(pq^{p+1}K^p)/r$  to be the new parameter  $b$ , this would lead to the form

$$\frac{C^*}{E^*} = \left( a - bE^* \right)^{\frac{1}{p}} \quad (\text{A11.11})$$

If  $p = 1$ , Equation A11.11 collapses to  $C/E = (a - bE)$ .

## Appendix 11.2: The Closed Form of the Estimate of the Catchability Coefficient

### Version 1: Constant $q$

Derivation of the statement that we can directly estimate the value of  $q$ , which relates to the maximum likelihood fit of the model, by using the geometric average of the time series of  $q$  estimates  $I_t/\hat{B}_t$ :

$$\hat{q} = e^{\frac{1}{n} \sum \ln \left( \frac{I_t}{\hat{B}_t} \right)} \quad (\text{A11.12})$$

By definition we have

$$\hat{I}_t = \hat{q}_t \hat{B}_t \quad (\text{A11.13})$$

where  $\hat{I}_t$  is the expected CPUE in a given year  $t$ ,  $\hat{q}_t$  is the expected catchability coefficient in year  $t$ , and  $\hat{B}_t$  is the predicted biomass in year  $t$ . However, the assumption is that the catchability coefficient is a constant and each  $\hat{q}_t$  is only an estimate of the overall  $\hat{q}$ . We can either directly estimate this expected

© 2011 by Taylor & Francis Group, LLC

catchability coefficient by using nonlinear estimation, or we can modify Equation A11.13 to include observed data instead of purely expected values. In this way we can generate the expected catchability coefficient; this is known as a closed form of the equation.

Using observation errors that are lognormal, multiplicative, and with a constant variance, we could fit the model using the sum of squared residuals criterion. The model residuals are related to the observed data in the usual way for lognormal errors:

$$I_t = \hat{I}_t e^\varepsilon \quad \text{or} \quad \frac{I_t}{e^\varepsilon} = \hat{I}_t \quad (\text{A11.14})$$

where  $I_t$  is the observed CPUE in a given year  $t$ . In order to obtain the closed form of Equation A11.13, we can substitute the right-hand version of Equation A11.14 into Equation A11.13 to include the observed CPUE values instead of the expected values:

$$\frac{I_t}{e^\varepsilon} = \hat{q}_t \hat{B}_t \quad (\text{A11.15})$$

which is equivalent to

$$I_t = \hat{q}_t \hat{B}_t e^\varepsilon \quad \text{or} \quad \frac{I_t}{\hat{B}_t} = \hat{q}_t e^\varepsilon \quad (\text{A11.16})$$

and log-transforming this gives

$$\text{Ln} \left( \frac{I_t}{\hat{B}_t} \right) = \text{Ln}(\hat{q}_t) + \varepsilon \quad (\text{A11.17})$$

The value of  $\hat{q}$  that minimizes the residuals,  $\varepsilon$ , in Equation A11.17 (remember because of the normalized error term this would be the same as maximizing the likelihood) would be the value that minimized the residuals of Equation A11.17 for all the observed values of catch effort ( $I_t$ ) and biomass  $B_t$ . If there are  $n$  observations, then the best estimate of the log of the constant,  $\hat{q}$ , is simply the mean of the  $t$  estimates from the set of observed catch effort values with associated expected biomass values:

$$\text{Ln}(\hat{q}) = \frac{\sum_{t=1}^n \text{Ln}(\hat{q}_t)}{n} = \frac{\sum \text{Ln} \left( \frac{I_t}{\hat{B}_t} \right)}{n} \quad (\text{A11.18})$$

To obtain the expected value of  $q$  we clearly need to antilog the outcome of Equation A11.18, which is, in fact, the geometric mean of the original estimates of  $q_t$ :

$$\hat{q} = e^{\frac{1}{n} \sum \text{Ln}(\hat{q}_t)} = e^{\frac{1}{n} \sum \text{Ln} \left( \frac{I_t}{\hat{B}_t} \right)} \quad (\text{A11.19})$$

© 2011 by Taylor & Francis Group, LLC

### Version 2: Additive increment to Catchability

In the case where the catchability is assumed to increase by a constant absolute amount each year, the  $q$  value for each year,  $q_t$ , can be determined using a simple linear equation:

$$q_t = q_0 + t \times q_{add} \quad (\text{A11.20})$$

where  $q_t$  is the catchability in year  $t$ ,  $q_0$  is the catchability in the first year for which data are available (time zero), and  $q_{add}$  is the constant absolute amount by which the catchability is incremented each year. Estimation of the two parameters involves finding the gradient,  $q_{add}$ , and intercept,  $q_0$ , of a linear regression between  $q_t$  and time  $t$ , where  $t$  ranges from 0 to  $n - 1$  years (a total of  $n$  years).

In the model, for each year, the implied estimate of  $q_t$  is obtained by dividing each observed catch rate ( $I_t$ ) by the estimated biomass for that year:

$$\hat{q}_t = \frac{I_t}{\hat{B}_t} \quad (\text{A11.21})$$

In the maximum likelihood fit one would have a time series of expected catchability coefficients, which would be described in the model by Equation A11.20 or A11.21. Equation A11.20 has the form of a linear regression, so the equations to find the closed-form parameter estimates are thus

$$q_{add} = \frac{\sum_{t=0}^{n-1} \left( (t - \bar{t}) \left[ \left( \frac{I_t}{B_t} \right) - \left( \sum \left( \frac{I_t}{B_t} \right) / n \right) \right] \right)}{\sum (t - \bar{t})^2} \quad (\text{A11.22})$$

and

$$q_0 = \frac{\sum \left( \frac{I_t}{B_t} \right)}{n} - (q_{add}) \bar{t} \quad (\text{A11.23})$$

where  $n$  is the number of years of data and  $\bar{t}$  is the mean of the  $t$  values representing the 0 to  $n - 1$  years of data (i.e., with none years of data  $\bar{t}$  would equal 4.0, i.e., the mean of 0 ... 8).

By estimating these parameters using the closed form, the number of parameters directly estimated by the fitting procedure is reduced, which simplifies the procedure and speeds the process.

### Version 3: Constant Proportional increase— $q_{inc}$

In the case where the catchability is assumed to increase annually by a fixed proportion, the  $q$  value for each year  $q_t$  is determined as in exponential growth or compound interest:

© 2011 by Taylor & Francis Group, LLC

$$q_t = q_0 \times q_{inc}^t \quad (\text{A11.24})$$

which, when log-transformed, takes the form

$$\text{Ln}(q_t) = \text{Ln}(q_0) + t \times \text{Ln}(q_{inc}) \quad (\text{A11.25})$$

where  $t$  ranges from 0 to  $n - 1$  years. In the final maximum likelihood fit one would have a time series of expected catchability coefficients that would be described in the model by Equation A11.24 or A11.25 (cf. Figure A11.1). Thus, the estimation of the two parameters involves finding the gradient ( $\text{Ln}(q_{inc})$ ) and intercept ( $\text{Ln}(q_0)$ ) of a linear regression between  $\text{Ln}(q_t)$  and time  $t$ , where time  $t$  ranges from 0 to  $n - 1$  years.

The closed-form equations are thus

$$\text{Ln}(q_{inc}) = \frac{\sum_{t=0}^{n-1} \left( (t - \bar{t}) \left[ \text{Ln}\left(\frac{I_t}{B_t}\right) - \left( \sum \text{Ln}\left(\frac{I_t}{B_t}\right) \right) / n \right] \right)}{\sum (t - \bar{t})^2} \quad (\text{A11.26})$$

and

$$\text{Ln}(q_0) = \frac{\sum_{t=0}^{n-1} \text{Ln}\left(\frac{I_t}{B_t}\right)}{n} - \text{Ln}(q_{inc}) \bar{t} \quad (\text{A11.27})$$

where  $n$  is the number of years of data and  $\bar{t}$  is the mean number of years of data. The final parameter estimates are determined by antilogging the values from Equations A11.26 and A11.27.

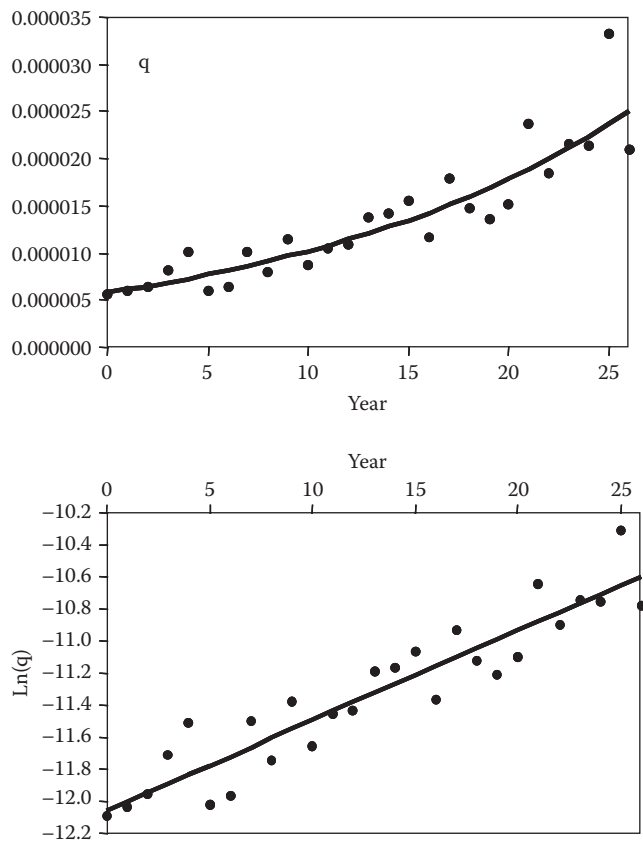
---

### Appendix 11.3: Simplification of the Maximum Likelihood Estimator

Showing the simplification of the maximum likelihood estimator for lognormal random errors. Given Equation 11.21,

$$L(\text{data} | B_0, r, K, q) = \prod_t \frac{1}{I_t \sqrt{2\pi\hat{\sigma}^2}} e^{\frac{-(\text{Ln } I_t - \text{Ln } \hat{I}_t)^2}{2\hat{\sigma}^2}} \quad (\text{A11.28})$$

© 2011 by Taylor & Francis Group, LLC



**Figure A11.1**  
The top panel is the final distribution of expected catchability coefficients with the fitted curve of form Equation A11.24. In order to estimate the two parameters  $q_0$  and  $q_{inc}$ , instead of a simple geometric average of the expected  $q$  values, we have to fit the curve. By log-transforming each  $q_t$  value and plotting these against the number of times the  $q_{inc}$  is applied to the starting value a straight line is obtained, as in the lower panel. The straight line is defined by Equation A11.25, so the two parameters may be determined by antilogging the two parameters from the linear regression. Data are a bootstrap sample from the northern Australian tiger prawn fishery (see Example Box 11.6).

we can convert this to a log-likelihood:

$$LL = \sum L_n \left[ \frac{1}{I_t \sqrt{2\pi\hat{\sigma}}} e^{-\frac{\left[ \ln(I_t) - \ln(\hat{I}_t) \right]^2}{2\hat{\sigma}^2}} \right] \tag{A11.29}$$

Simplifying this by removing constants from the summation and cancelling the  $\ln$  and  $e$ , we obtain

© 2011 by Taylor & Francis Group, LLC

$$LL = nLn\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right) + \sum Ln\left(\frac{1}{I_t}\right) + \dots$$

$$\frac{1}{2\hat{\sigma}^2} \sum \left[ - \left[ \left( Ln(I_t) - Ln(\hat{I}_t) \right)^2 \right] \right] \quad (A11.30)$$

where the maximum likelihood estimator of the standard deviation  $\sigma$  is given by

$$\hat{\sigma} = \sqrt{\frac{\sum \left[ \left( Ln(I_t) - Ln(\hat{I}_t) \right)^2 \right]}{n}} \quad (A11.31)$$

Note the division by  $n$  instead of  $n-1$  to give the maximum likelihood estimate (Neter et al., 1996). Given Equation A11.31, we can simplify Equation A11.30 much further by substituting one into the other:

$$LL = nLn\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right) + \frac{-\sum \left( Ln(I_t) - Ln(\hat{I}_t) \right)^2}{2 \sum \left( Ln(I_t) - Ln(\hat{I}_t) \right)^2} + \sum Ln\left(\frac{1}{I_t}\right) \quad (A11.32)$$

$n$

All terms can be further simplified:

$$LL = nLn\left(\left[\sqrt{2\pi}\hat{\sigma}\right]^{-1}\right) + \left(\frac{-1}{2/n}\right) + \sum Ln(I_t^{-1}) \quad (A11.33)$$

which simplifies again to become

$$LL = -nLn(\sqrt{2\pi}\hat{\sigma}) - \frac{n}{2} - \sum Ln(I_t) \quad (A11.34)$$

A little algebra leads to an alternative final version in A11.38:

$$LL = -n \left[ Ln\left(\left[2\pi\right]^{\frac{1}{2}}\right) + Ln(\hat{\sigma}) \right] - \frac{n}{2} - \sum Ln(I_t) \quad (A11.35)$$

$$LL = -n \left[ \frac{1}{2} Ln(2\pi) + Ln(\hat{\sigma}) \right] - \frac{n}{2} - \sum Ln(I_t) \quad (A11.36)$$

© 2011 by Taylor & Francis Group, LLC



$$LL = -\frac{n}{2} \left( Ln(2\pi) + 2Ln(\hat{\sigma}) \right) - \frac{n}{2} - \sum Ln(I_t) \quad (A11.37)$$

$$LL = -\frac{n}{2} \left( Ln(2\pi) + 2Ln(\hat{\sigma}) + 1 \right) - \sum Ln(I_t) \quad (A11.38)$$

The final summation is a constant and is usually omitted. If the initial negation is omitted, this would provide the negative log-likelihood.

