

7 ASSESSING RELIABILITY

7.1 SENSITIVITY ANALYSIS AND THE EFFECT OF HIGHLY INFLUENTIAL OBSERVATIONS

The estimates of different parameters are not equally sensitive to all observations. A 10% change in an observation does not affect all elements of the estimated stock status equally. This is explored in sensitivity analysis. Sensitivity analysis in fish stock assessment is usually used in two contexts:

- Exploring how a change in an observation, e.g. a survey result, affects the estimate of the elements in the stock status.
- Exploring how a change in a parameter, e.g. the natural mortality will affect both the estimates and the projection of the stock.

The sensitivity of a function is defined as $\frac{\partial \log y(\beta)}{\partial \log \beta}$. This is the expression for the relative change of the function $y(\beta)$ from a relative change in the independent variable β that follows from the identity $\frac{\partial \log y(\beta)}{\partial \log \beta} = \frac{\partial y(\beta)/y(\beta)}{\partial \beta / \beta}$.

The sensitivity can be calculated numerically by introducing small changes in the observations and re-estimating the parameters. The sensitivity of a stock projection is estimated following the same procedure: a small change in parameters or variables is introduced and the changes in an output quantity (e.g. the projected catch) is calculated and expressed relative to the introduced change. In order not to lose the overview of the sensitivity structure, some strategy for the observation change should be worked out before embarking on such studies. The above definition of sensitivity is then used to present the results.

It is of particular importance that observations that strongly influence the estimated parameters or the projections are identified. It follows in some cases directly from inspection of the data model, which are the most influential observations. For example, the estimate of the strength of the most recent year class will most strongly depend on the most recent survey result.

The influence of observations is assessed by alternately excluding each observation and repeating the estimation each time. When it is an abundance index that is under suspicion, removal is simple. Just ignore the contribution to the sum-of-squares from that index or from that year, and then revised estimates can be compared to those obtained with the full data set. If it is catch data that are suspected to be highly influential (e.g. for a particular year), the procedure will be to treat this as a year for which catch data were not available using the methods discussed in Section 7.4.

Robust regression tries to minimise these effects using lower weighting of observations with the largest residuals. This includes trimmed means regression,

where a proportion of the largest residuals are excluded from the estimation (see Section 6.5).

7.2 THE ANALYTICAL APPROACH TO VARIANCE-COVARIANCE ESTIMATION

An analytical estimate of the variance-covariance matrix can be obtained using calculus. The result can be obtained because at the least-squares minimum, the differential of the model with respect to each parameter is zero, and because overall variance can be estimated directly from the sum of squares calculated during the least-squares procedure. The analytical approach is only valid if errors are normally distributed. If it is suspected that they are not normally distributed then a more robust scheme is recommended.

From least-squares theory, the covariance matrix of the parameters can be defined as:

$$V(\beta) = 0.5 \frac{L(\beta)}{n-p} \{H\}^{-1} \quad (83)$$

where β is the vector of best-fit parameters, $\{H\}^{-1}$ is the inverse Hessian matrix (Equation 74) and $L(\beta)$ is the sum of squares, so $L(\beta)/(n-p)$ is the unbiased estimate of the error variance. As in Equation 74, the Hessian can either be approximated numerically or by partial differential of the model with respect to each parameter, so:

$$V(\beta_i, \beta_j) = \frac{L(\bar{\beta})}{n-p} \left\{ \sum_{obs} \left(\frac{\partial Model(\beta)}{\partial \beta_i} \frac{\partial Model(\beta)}{\partial \beta_j} \right) \right\}_{i,j}^{-1} \quad (84)$$

While these estimates of the parameter standard error can be used to estimate confidence intervals, the procedure only works when errors are normally distributed, otherwise they are only generally indicative of the variation and correlation in parameters. For confidence regions, more robust techniques such as bootstrapping are much preferred.

7.3 ESTIMATING CONFIDENCE INTERVALS USING BOOTSTRAPS

Bootstrapping techniques are also known as “resampling” techniques or “computer intensive” methods (Efron and Tibshirani 1993, Manly 1996). Manly (1996) distinguished between these concepts and defines re-sampling as sampling without replacements, bootstrapping as sampling with replacement and Monte Carlo methods as the generation of random number from a specified theoretical distribution. We restrict our discussion to the sampling with replacement technique, but also note that in many applications other techniques may be extremely useful.

Bootstrapping is based on the idea that the observations are a random sample from the population under investigation and any random sample from the observations

also forms a random sample of the population. A simple example of the procedure is illustrated by a linear regression between two covariates. In this simple case, the observations are the paired control and response variables. The procedure is implemented as:

- (i) List the observations and number them 1,2, ... n .
- (ii) Using a random number generator (or table) select a sub-sample with k elements using sampling with replacement. In many applications, the size of the sub-sample k is chosen to be equal to the total number of observations in the original sample n .
- (iii) Calculate the statistics under investigation, in this case slope and intercept of the regression, and store the results.
- (iv) Repeat the calculations from point (ii) above 100 to 1000 times, depending on the specific problem and the accuracy needed.
- (v) When enough estimates have been obtained, appropriate statistics can be calculated, typically the mean, variance and 10%, 50% and 90% percentiles. These statistics are indicative of the properties of the estimated parameters.

This procedure works fine and is very robust to model structural errors, unless the observations are dependent. In fisheries models, the sequential dependence between observations forms part of the population dynamics, so this procedure often cannot be used. The alternative approach is to use the fitted model and residuals. Each observation is made up of the model estimate and the residual error. If each of the model estimates are combined with residuals drawn randomly with replacement, a new simulated data set is created. The procedure now becomes:

- (i) Fit the model and obtain expected values for observations, and calculate the residuals.
- (ii) Using a random number generator (or table) add to each expected value a new residual drawn at random with replacement to create a new simulated observation.
- (iii) Fit the model to the new simulated data.
- (iv) Calculate the statistics under investigation (in the case of fisheries, fishing mortalities, selectivity etc.) and store the results.
- (v) Repeat the calculations from point (ii) 100 to 1000 times, depending on the specific problem and the accuracy needed.
- (vi) When enough estimates have been obtained, appropriate statistics can be calculated, typically the mean, variance and 10%, 50% and 90% percentiles.

It is critical before applying bootstrap procedures to have a good understanding of the error structure. Bootstrapping aims to simulate these errors, so the procedure should follow the correct error model as far as possible. For example, if we have two CPUE time series, we may not want to assume that they have the same error

distribution. In this case, the residuals from which each new series is generated are kept separate. The case is a little more complicated if the variance is thought to change during the series. Although it may still be possible to draw random residuals, they will have to be re-scaled according to the estimated variance of each observation.

The real advantage of the bootstrap scheme comes when we need to estimate the parameter variation associated with a complex estimation procedure. It is not unusual to apply sequential analyses for fisheries. For example, we might firstly estimate a growth rate parameter from size frequency data before using that same data to slice catches into cohorts and apply VPA. We can apply the bootstrap procedure in this case by simulating the length frequency sample and catches, and then use these data to estimate the growth rate, cohorts and so on right through the procedure as though on real data. As long as the whole procedure is repeated enough times and the assumptions underlying the errors and models are correct, the estimates of uncertainty should also be correct.

In some cases, the error model is too complex to use real residuals, or perhaps more often the sample size is too small. In these cases, we may use the data to estimate the parameters of an assumed error distribution rather than use observed residuals. Errors are then drawn from this parametric distribution, rather than the set of residuals. This obviously requires an additional assumption on the exact nature of the error distribution, but is preferable to using only a small sample of residuals.

7.4 ESTIMATION WHEN DATA ARE MISSING

It is not always the case that time series data are complete. The estimation of stock status is still possible, but the lack of data means the variance of the estimates increases. There is no problem if there is a year without an abundance index, it is simply left out of the sum-of-squares. It is a little more difficult if catch data are missing. Most VPA methods require that a cohort can be fully traced without gaps in the time series. The virtual population is actually the sum of all catches of a given cohort and this represents a lower bound on the original recruitment.

Where observations are not independent, so they cannot just be excluded (i.e. mainly data used in the population model), the general approach is to use the expected value in place of an observation. The iterative procedure of replacing missing values with their expected values from the model and then refitting the model to generate new expected values until convergence, is known as the EM algorithm. The algorithm is robust, but can be very slow.

Similar procedures have been proposed on the basis that the stock and fishing mortality can be calculated from:

$$\begin{aligned}
 N_{ay} &= Cpue_{ay} / q_a \\
 \text{and} \\
 F_{ay} &= \ln \left(\frac{Cpue_{ay} / q_a}{Cpue_{a+1,y+1} / q_{a+1}} \right) - M_a
 \end{aligned}
 \tag{85}$$

Which give the expected catch based on the Baranov equation (7). Calculating the expected catch in this way can be built into the estimation procedures for those years when the catch data were not available or unreliable. A simple approach is to use the above equations to derive the catches for the year(s) when data were not available, complete the catch-at-age and year matrix with these data and perform the parameter estimation, but exclude these “data” from the model fit (see also Patterson 1998b).

If the separability model applies (Equation 11), the problem reduces to the estimation of the exploitation level E_y for the year for which catch data are missing. In that case the simpler solution is to ignore the contribution to the sum-of-squares from the year for which the catch data are missing and fit the $E(y)$ array as in the standard procedure.

If the total catches are available, but there is no age information for a specific year then the age-length keys must be interpolated between years (Hoenig and Heisey 1987). This is built on estimates of growth and recruitment for the stock in question. For those length groups where the recruiting year-class overlaps significantly with older fish, the ALK becomes very uncertain. Fortunately, such overlaps do not occur frequently. The procedure most often used, simply applying an average ALK or the ALK of the previous or next year, can lead to quite large errors in the estimated mortality. These procedures should be used with caution.

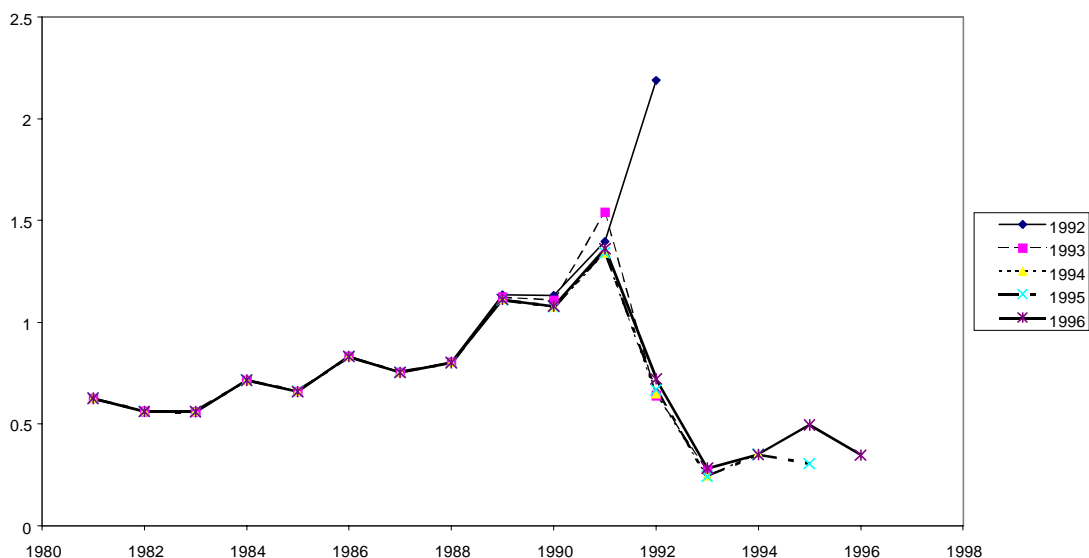


Figure 7.1 Retrospective analysis on F for age group 4 eastern Baltic cod. Each analyses uses one more year's data: 1981-1992, 1981-1993,..., 1981-1996. It is apparent that the estimate for 1992 when that was the final year in the analysis was inaccurate. Otherwise, the estimate of the fishing mortality for the most recent year seems to be stable. Although in 1995, there was some change when another years' data have been added.

7.5 RETROSPECTIVE ANALYSIS

The idea behind retrospective analysis is very simple. Drop one or several of the most recent years' data, repeat the analysis and compare with the analysis including those data (Figure 7.1). The analysis will identify years, which lead to poor projections and test the models general capability of prediction.

An alternative approach to retrospective analysis is to restrict the data to, say, 10 years and estimate the stock status based on such a moving window across the data. The results will not differ significantly particularly as one is most interested in the behaviour of the estimate for the most recent year.

Retrospective analyses have been used extensively to investigate the performance of particular assessments. For example, it is part of the standard ICES assessment procedure.