Short Course on Uncertainty and Projections

(*AZTI, San Sebastian, January 22-23, 2002*)

Prepared by V. Restrepo and L. Kell

1. OVERVIEW

A stock projection basically requires a **population model**, a **management strategy** to be applied into the future, and a **characterization of uncertainty**.

There are many population models in use in fisheries science: Production models, age-structured sequential population analyses, statistical catch-age models, etc.

Management strategies tend to be either constant-catch ones, or those that attempt achieving a target linked to a "biological reference point" (Fmsy, Fmax, F0.1, Bmsy, etc.).

Uncertainty plays an important role in all aspects of projections. It is present in the form of the population model, because no model can mimic reality 100%. It is also present in our estimate of the current status of the stock due to observation errors, i.e. even if we had a perfect model we could not measure the fisheries and stock variables with perfect accuracy. And uncertainty in the future is especially important: it can refer to random processes that affect the population in the future (e.g., recruitment variability) or to the implementation of future management regimes (e.g., quota over-harvests).

When making projections, it is very important to maintain consistency. For example, one would obviously want to use the same type of population model to estimate current stock status and to make projections. Similarly, it is important to maintain consistency in the characterization of uncertainty between the estimation of current stock status and the projection of future stock status.

2. PRODUCTION MODEL

This section introduces a stock production model that will be used as an example throughout the course. We have chosen this model for ease of use in a spreadsheet.

The basic population model is:

$$(3.1)\quad B_{t+1} = B_t + \frac{r}{p} B_t \left(1 - \left(\frac{B_t}{K} \right)^p \right) - C_t$$

where r is the growth rate, K is the carrying capacity, and p is a shape parameter. When p=1, the model becomes the Schaefer (logistic) type in which MSY is obtained at a biomass equal to K/2.

The equilibrium statistics for the model are computed as:

$$(3.2)\quad E_{msy} = \frac{r}{q(p+1)}$$

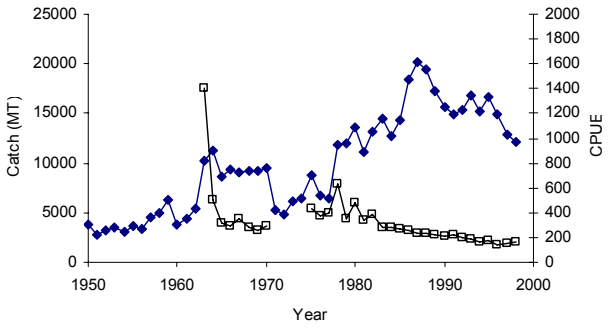
$$(3.3)\quad F_{msy} = qE_{msy}$$

$$(3.4)\quad MSY = \frac{rK}{(p+1)^{(p+1)/p}}$$

$$(3.5)\quad B_{msy} = K(p+1)^{-1/p}$$

An example dataset that will be used is that for North Atlantic swordfish (*Xiphias gladius*).

Year	Catch	CPUE	Year	Catch	CPUE
1950	3746		1975	8839	433.51
1951	2781		1976	6696	372.42
1952	3193		1977	6409	392.43
1953	3503		1978	11835	627.49
1954	3134		1979	11937	355.6
1955	3602		1980	13558	478.38
1956	3358		1981	11180	333.95
1957	4578		1982	13215	385.78
1958	4904		1983	14527	280.03
1959	6232		1984	12791	276.91
1960	3828		1985	14383	266.85
1961	4381		1986	18486	256.75
1962	5342		1987	20236	235.11
1963	10180	1398.86	1988	19513	232.53
1964	11258	501.05	1989	17250	222.13
1965	8652	314.35	1990	15672	209.34
1966	9349	293.67	1991	14937	217.24
1967	9107	351.31	1992	15394	201.22
1968	9172	284.72	1993	16772	183.19
1969	9203	258.54	1994	15235	166.63
1970	9495	296.39	1995	16618	176.71
1971	5266		1996	14921	142.68
1972	4766		1997	12913	147.61
1973	6074		1998	12175	165.24
1974	6362				



3. MODEL FITTING

There are several approaches for fitting models to observed data. All of the approaches aim to minimize discrepancies between the observed data and the data that are predicted conditional on a given model with a given set of parameters.

3.1 Maximum Likelihood

The Maximum Likelihood (ML) method seeks to find the parameter values that maximize the likelihood of having observed the data at hand (conditional on the model form).

The ML method requires a definition of how residual errors are distributed around the model's expectations. That is, it requires a probability density function (pdf) for the observations.

The pdf for the Normal Distribution is typically assumed in models with a simple error structure:

(3.1) $$pdf = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(X-\mu)^2}{2\sigma^2}\right)}.$$

Here, μ is the mean and σ is the standard deviation; these are the 2 parameters that define the pdf.

For a given set of observations, Likelihood is defined as the product of the individual pdfs for the observations, given a set of parameter values:

(3.2) $L(\theta) = \prod_{i=1}^{n} pdf(X_i|\theta)$. In the normal distribution example,

(3.3) $$L(\theta) = \prod_{i=1}^{n} pdf(X_i|\mu,\sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(X_i-\mu)^2}{2\sigma^2}\right)}.$$

To obtain ML estimates of μ and σ for a given set of observations, one would look for the values of these parameters that result in a maximum value of $L(\theta)$. Equivalently, one could look for the maximum value of the natural logarithm of $L(\theta)$, which is easier to compute and is less affected by machine precision (due to multiplying probability values that are often much smaller than 1.0). The Log-Likelihood is:

(3.4) $$LL(\theta)=\sum_{i=1}^{n}\ln[pdf]=\sum_{i=1}^{n}\ln\left[\frac{1}{\sigma\sqrt{2\pi}}e^{\left(\frac{-(X_i-\mu)^2}{2\sigma^2}\right)}\right]=n\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)+\frac{1}{2\sigma^2}\sum_{i=1}^{n}-(X_i-\mu)^2 .$$

Finding a maximum for $LL(\theta)$ on the above example is simple. In model applications, one won't be estimating the mean μ and standard deviation σ of a distribution, but rather a set of parameters that give model predictions, $\hat{X}_i$, which can be compared against the observations, X_i.

For example, consider a linear regression of the form $Y_i=a+bX_i$. The log-likelihood to be maximized would be

(3.5) $$LL(\theta)=n\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)+\frac{1}{2\sigma^2}\sum_{i=1}^{n}-(Y_i-(a-bX_i))^2 .$$

Simplifying,

(3.6) $$LL(\theta)=-\frac{n}{2}[\ln(2\pi)+2\ln(\sigma)+1],$$

where (3.6.1) $$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-(a+bX_i))^2.$$

Another type of error structure assumed often is the Lognormal Distribution. In this case, the pdf is

(3.7) $$pdf=\frac{1}{x\ \sigma\sqrt{2\pi}}e^{\left(\frac{-(\ln(x)-\ln(\mu))^2}{2\sigma^2}\right)}$$

the log-likelihood in a linear regression example will be similar to the one above:

(3.8) $$LL(\theta)=-\frac{n}{2}[\ln(2\pi)+2\ln(\sigma)+1]+\sum_{i=1}^{n}\frac{1}{X_i},$$

with (3.8.1) $$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(\ln(Y_i)-\ln(a+bX_i))^2.$$

Fitting the Production Model

The production model shown earlier can be fitted to data using the ML approach. One way of doing this is to assume that

(1) The catches are known exactly, and
(2) Errors in CPUE (denoted by U) are log-normally distributed.

The link between data and model is accomplished by setting the predicted CPUE as a function of, e.g., mid-year biomass:

$$\hat{U}_t=q\overline{B}_t e^{\varepsilon}=q\frac{(B_{t+1}+B_t)}{2}e^{\varepsilon},$$

where q is the catchability coefficient and ε is normally-distributed with mean 0 and variance σ^2. Then, the log-likelihood to be maximized by estimating the set of parameters (B_1, K, r, p, σ^2 and q) is

(3.9) $LL(B_1, K, r, p, q, \sigma^2) = -\frac{n}{2}\left[\ln(2\pi) + 2\ln(\sigma) + 1\right] + \sum_{i=1}^{n} \frac{1}{U_i}$,

with (3.9.1) $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\ln(U_i) - \ln(\hat{U}_i)\right)^2$.

Notes:

- Maximizing $LL(\theta)$ is the same as minimizing $-LL(\theta)$.
- For the purpose of finding a maximum (or minimum), constants such as 2π can be ignored.
- ML assuming normally-distributed errors with constant variance is equivalent to Least-Squares.

3.2 Search Methods

(((Do examples fitting PM in excel using Solver and Simulated Annealing)))

4. ESTIMATION OF UNCERTAINTY

4.1 Asymptotic statistics

Given a set of assumptions used in fitting a model (e.g., the distribution of the data), theoretical approaches have been developed to evaluate the goodness of fit of a model. A variance is associated with each estimator and a covariance is associated with each pair of estimators; the variance is a measure of the precision or repeatability of the estimator and is a function of the data and other unknown parameters. In a likelihood function, the variance is derived as the negative inverse of the second partial derivative of the log-likelihood function, evaluated at the ML estimate:

$$\text{var}(\theta) = \left[-\frac{\partial^2 LL(\theta)}{\partial \theta^2} \right]^{-1}_{\theta=\hat{\theta}}$$

For more than one parameter, one can compute the variance-covariance matrix for all parameters. This matrix is the inverse of the negative of the matrix of second-order partial derivatives of the log-likelihood function (i.e., the inverse of the "Information Matrix"). Similar to the above, the (ij)th element of the matrix would be

$$\left[-\frac{\partial^2 LL(\theta)}{(\partial\theta_i)(\partial\theta_i)} \right]^{-1}_{\theta_i=\hat{\theta}_i;\theta_j=\hat{\theta}_j}$$

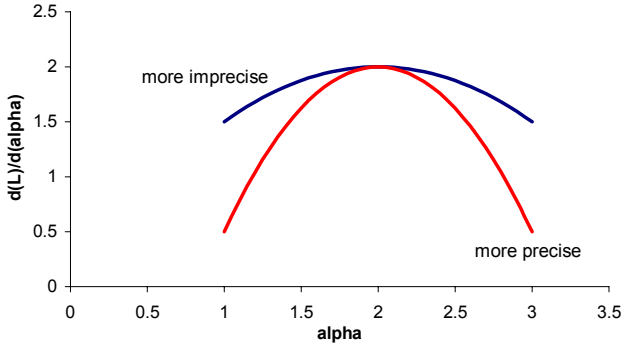
For example, in the ML fit of the linear regression model above, assuming normally-distributed errors, it can be shown that there is an explicit solution for the variances and covariance of *a* and *b*:

$$VAR(b) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

$$VAR(a) = \frac{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + n\bar{x}^2\right)\sigma^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}, \text{ and}$$

$$COV(a,b) = -\frac{\bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sigma^2 .$$

However, analytical solutions such as these are possible mostly for linear models and for relatively-simple nonlinear models that make assumptions about normality. In more complex models, theory allows for a numerical evaluation of these variances and covariances. For example, numerical methods can be used to take the partial derivative of the function being minimized/maximized with respect to the parameters at the solution. The greater the partial derivatives, the larger the variance:



A useful method for use in looking at confidence limits with Maximum Likelihood is the through a deterministic approximation called "Likelihood Profiling". The method relies on the ML theory that the ratio of likelihoods for models with different parameter values follows the χ^2 distribution. Let

$LL(\theta max)$ = The log-likelihood obtained by estimating all parameters,
$LL(\theta^*)$ = The log-likelihood obtained by estimating a subset {*} of p parameters. Then,

$$2\left[LL(\theta^{\max}) - LL(\theta^*)\right] \leq \chi^2_{p,1-\alpha}$$

is used to calculate (1-α)% confidence regions for the parameter set {*}. For example, if one wanted a 95% confidence interval for one parameter, $\chi^2_{1,1-\alpha} = 3.84$, so one would search for the values of the parameter θ_i that result in a log-likelihood:

$$LL(\theta_i) = LL(\theta^{\max}) - \frac{3.84}{2}.$$

Similarly, confidence regions can be constructed for more than 1 parameter.

(((Example)))

4.2 Bootstrapping

Bootstrapping is a useful approach for estimating uncertainty in parameter estimates. The method consists of re-sampling observations, with replacement, and repeating the estimation procedure many times. The final distribution of the estimates is taken to be a picture of the uncertainty in the parameter estimates.

Most types of bootstraps are said to be conditional. That is, the re-sampling is based on the results of the original model fit and proceeds by re-sampling from the model residuals. An un-conditional bootstrap would be based on the original observations instead, without reference to the original model fit.

Another way to define bootstraps refers to whether or not parametric assumptions are made during re-sampling. A conditional non-parametric bootstrap uses the distribution of residuals from the original fit as the

unique set from which bootstrap samples are created. The new bootstrap observations, Y^*, are computed as $Y_i^* = \hat{Y}_i + R^*$, where R^* is one of the n available residuals $(Y_j - \hat{Y}_j)$ chosen at random.

A conditional parametric bootstrap is similar, except that the R^* are taken from a theoretical distribution that is defined by the assumptions made in obtaining the basic model fit. For instance, if the model assumes lognormal errors, then the residuals will be generated randomly from a lognormal distribution with parameters $LN(\ln(\hat{Y}_j), \sigma)$.

(((Example)))

4.3 Sensitivity to Assumptions and Constraints

In fitting models to data and in making projections, it is important to examine the sensitivity of the results to different assumptions. There are generally two ways of doing this, the most common one being by simply changing models. For example, in the production model,

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

one may want to assume either a "Schaefer" (logistic) model by fixing p=1 or a "Fox" model by fixing $p \rightarrow 0$ (similarly, one can fix B_1=K, etc.).

The approach given above is appealing in that it makes the assumptions very clear. However, it may be criticized in cases where the data would be sufficiently informative so as to allow for estimation of the parameter in question. A useful alternative lies somewhere between fixing parameters and estimating them freely, and it consists of adding a "penalty" to the log-likelihood when a parameter estimate wanders too far away from what the analyst thinks it should be. In Bayesian terms, this is adding a "prior" for a parameter. For example, in constraining the estimate of p so that it is close to $\bar{p}$, one could add a term as follows to $LL(\theta)$:

Penalty = $-\frac{1}{\sigma_p^2}(p - \bar{p})^2$.

How strong is the penalty depends upon the value of the variance term: A smaller σ_p^2, indicates that one has a stronger belief that one knows that $p = \bar{p}$; a larger σ_p^2 induces a weaker penalty if the estimate of p wanders away from $\bar{p}$.

((((Likelihood profile examples using production model in Fishlab.))))

5. PROJECTIONS

5.1 Theory

Most projections are intended to examine questions such as "what if the catch is XX" (sometimes called "output controls" or catch controls) or "what if the effort is YY" (sometimes called "input controls" or effort controls). In general, however, management based on effort controls tends to be difficult because catchability can change rapidly. Thus, limiting the number of vessel-days or some other nominal measure of fishing effort will not necessarily limit effective fishing effort.

At the very least, managers will want to know what happens if the current catch is kept constant. But in the last decade, there has been an emphasis on examining catch or effort controls that can result in a desired management objective. Such objectives are often so-called biological reference points, or management targets (or limits) that have some biological significance. For example, if the stated target in a country's fisheries

legislation is Bmsy, managers may be interested in knowing what series of catches could take the stock from Bnow to Bmsy in *y* years.

In the above example, the longer *y* is, the more options that could be examined. For instance, managers could be presented with some options that cut catches once and keep catches constant for the entire time, and other options that change catches gradually, ending up in a much lower value by year *y*. One difficulty that scientists almost always face in this regards, is that managers are not typically explicit about what options they want to examine.

There is another important consideration to keep in mind in terms of the length of *y*: The longer it is, the more dependent the results will be upon things that have not happened yet, and the less dependent on our estimate of current stock status. Consider a stock of a species that has *k* year-classes: a projection *k*+1 years into the future will be based almost 100% on events that have not happened yet. This is a source of frustration for many scientists and managers, but it should not be. When making long-term projections, the question being examined should almost always be viewed as a comparison between different alternatives, and not as a true prediction of the future state of the system. Short-term projections (in the order of *k*/10) are another thing: if scientists have a good assessment, the status of the stock one year or two into the future may be easier to predict.

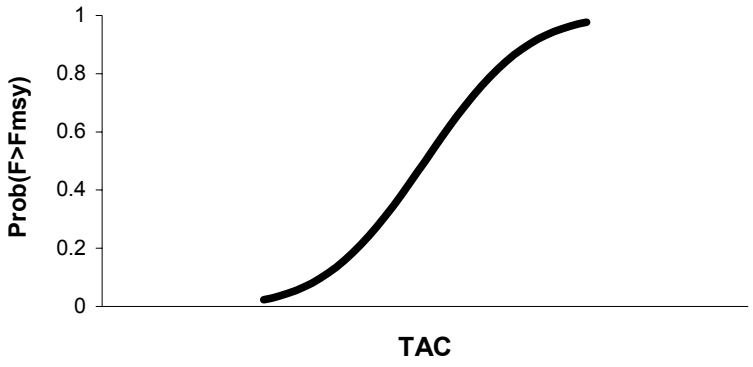
In practice, making projections is relatively simple. What is needed is a starting point, a population model, and the options to be examined. Including uncertainty in the projections (in both the starting point and the future states of the system) provides for one more layer of complexity.

5.2 Outputs

Presenting useful and understandable projection information to the fisheries community is just as important as the manner of making the projections. In relatively simple cases, it is useful to show trajectories of biomass or fishing mortality, or other relevant variables. But when the projections are stochastic, it may be more informative to summarize the information of a set of projections into a table that states the probability of some event taking place, given some management action.

For example, consider the question: what is the probability that F exceeds Fmsy if the TAC is set to 25,000 t during the next 5 years? One way of doing this is to carry out the stochastic projections setting C_t=25,000 (e.g., with equation 3.1) for 5 years, and then computing the proportion of stochastic runs where F>Fmsy.

Managers may also be interested in a similar percentage, but with TAC={5000, 10000, 15000, 20000, 30000, 35000....}. And then perhaps even TAC={22332, 28739, ...}. In such cases, the analyst may save some time and carry out the projection at conditioning on F=Fmsy and then summarize the results in terms of the final cumulative distribution of the resulting catches (a "probability profile"):



(the above practice will result in the same distribution as would be obtained if the analyst carried out stochastic projections in a continuum of TAC selections).

5.3 Examples

Examples based on production models using bootstrapping and Monte Carlo simulations based on Hessian.

==========

EXAMPLES

1) ML Fit of production model with (a) Solver [Victor] and (b) Simulated Annealing [Laurie]
2) Var-Covar matrix for ML fit with Hessian [Laurie]. Likelihood profiles for B_{1999}, r and K.
3) Bootstrapping example for prod model [Victor]. Distribution of B_{1999}, r and K.
4) Sensitivity of estimates to assumptions using Likelihood Profiles [Laurie]:
   How B_{1999} /B_{msy} varies with (estimate 2, 3 or 4 parameters; add constraint for r).
5) 20-year projections at F_{msy} and at Y_{1998}:
   MC projections from Hessian [Laurie]
   MC projections from Bootstrap [Victor]

APPENDIX - Other Useful Information

A1. Maximum Likelihood Point Estimation

The ML approach can be used to obtain analytical solutions in many cases. Consider a likelihood involving only one parameter, α. If $LL(\alpha)$ is unimodal, its maximum occurs where the slope (derivative) of LL with respect to α is equal to zero:

$$\frac{\partial LL(\alpha)}{\partial \alpha} = 0$$

For example, equation (3.5) has 3 parameters (a, b and σ^2):

$$LL(\theta) = n \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + \frac{1}{2\sigma^2}\sum_{i=1}^{n} -\left(Y_i - (a - bX_i)\right)^2$$

One can estimate the value of σ^2 algebraically by deriving the equation with respect to σ^2, setting the result equal to zero and solving for σ^2:

$$\frac{\partial LL(\theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}\left(Y_i - (a - bX_i)\right)^2}{2\sigma^4} = 0$$

$$\frac{n}{2\sigma^2} = \frac{\sum_{i=1}^{n}\left(Y_i - (a - bX_i)\right)^2}{2\sigma^4}$$, which results in the solution given in equation (3.6.1):

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - (a + bX_i)\right)^2.$$

In fact, in this simple linear regression example it is also possible to obtain analytical estimates of a and by through

$$\frac{\partial LL(\theta)}{\partial a} = 0 \text{ and } \frac{\partial LL(\theta)}{\partial b} = 0.$$

This is known as the "normal equations" in regression texts.

A less trivial example is useful in the fitting of the production model. Writing out the log-likelihood in a longer fashion than in equation (3.9):

$$LL(\theta) = \sum_{i=1}^{n} \ln\left(\frac{1}{U_i\sigma\sqrt{2\pi}}\right) + \frac{1}{2\sigma^2}\sum_{i=1}^{n} -\left(U_i - q\overline{B}_i\right)^2$$, one can set $\frac{\partial LL(\theta)}{\partial q} = 0$

in order to solve analytically for q. This exercise results in the equation

$$\hat{q} = e^{\frac{1}{n}\sum(\ln(U) - \ln(\overline{B}))}.$$

Thus, of the set of parameters (B_1, K, r, p, σ^2 and q) it is only necessary to search numerically for B_1, K, r and p because σ^2 and q can be estimated analytically. This provides for faster, more efficient searches, especially when using slow programs like a spreadsheet.

A2. The Delta Method

The Delta Method is a useful approach for finding approximate variances and covariances of parameters that are not directly estimated. For example, in the production model fit, we may obtain estimates of the covariance matrix for (B_1, K, r and p). But what if we want to estimate the variance of B_t in any given year? The delta method can be used for a quick approximation. Suppose that we are interested in the variance of some function of parameters $\boldsymbol{x}$ which we are estimating directly, $g(x)$. According to the delta method:

$$\mathrm{var}(g(x)) = \sum_i \mathrm{var}(x_i)\left(\frac{\partial g}{\partial x_i}\right)^2 + 2\sum_{i<}\sum_j \mathrm{cov}(x_i, x_j)\left(\frac{\partial g}{\partial x_i}\right)\left(\frac{\partial g}{\partial x_j}\right)$$

Continuing with the example, suppose that we want the variance of B in year 2 and that we have estimated (B_1, K, r and p) and their covariance matrix. Then

$$B_2 = g(x) = B_1 + \frac{r}{p}B_1\left(1 - \left(\frac{B_1}{K}\right)^p\right) - C_1 \quad \text{, and}$$

$$\mathrm{var}(B_2) = \mathrm{var}(B_1)\left(\frac{\partial g}{\partial B_1}\right)^2 + \mathrm{var}(K)\left(\frac{\partial g}{\partial K}\right)^2 + \mathrm{var}(r)\left(\frac{\partial g}{\partial r}\right)^2 + \mathrm{var}(p)\left(\frac{\partial g}{\partial p}\right)^2 +$$

$$\mathrm{K}\ 2\,\mathrm{cov}(B_1, K)\left(\frac{\partial g}{\partial B_1}\right)\left(\frac{\partial g}{\partial K}\right) + 2\,\mathrm{cov}(B_1, r)\left(\frac{\partial g}{\partial B_1}\right)\left(\frac{\partial g}{\partial r}\right) + 2\,\mathrm{cov}(B_1, p)\left(\frac{\partial g}{\partial B_1}\right)\left(\frac{\partial g}{\partial p}\right) +$$

$$\mathrm{K}\ 2\,\mathrm{cov}(K, r)\left(\frac{\partial g}{\partial K}\right)\left(\frac{\partial g}{\partial r}\right) + 2\,\mathrm{cov}(K, p)\left(\frac{\partial g}{\partial K}\right)\left(\frac{\partial g}{\partial p}\right) +$$

$$\mathrm{K}\ 2\,\mathrm{cov}(r, p)\left(\frac{\partial g}{\partial r}\right)\left(\frac{\partial g}{\partial p}\right)$$

This procedure can be repeated then for other estimates. For instance, if one then wants to estimate var(B_3), one would express $g(x)=B_3$ as a function of B_2 and the other parameters and follow the same steps. In this fashion, one can propagate the variance of the point estimates into additional estimates.

In the example above, there are analytical solutions to the equations. But, as the calculations propagate into new formulations, it becomes tedious to keep track of all equations. For this reason, numerical approximations are often used in practice.