Working document No. 32

Time series models in fish stock analysis

by

Guðmundur Guðmundsson

The time series methodology employed in present approach to fish stock assessment is described in the book by Harvey (1989). The adaptation to fish stock assessment and some applications were presented by Gudmundsson (1994 and 1999). ICES (1993 and 1995) have some comparisons of estimation by this method with ADAPT and XSA.

Model

The analysis is based on observations of catch-at-age. Each of them is connected with other quantities by a measurement equation as follows:

$$C_{at} = \frac{F_{at}}{Z_{at}}(1-e^{-Z_{at}})N_{at} + \varepsilon_{at}. \qquad (1)$$

The measurement equations can be extended by auxiliary information like catch per unit effort, CPUE,

$$U_{at} = \varphi_a \psi_t N_{at}^{Q(a)} e^{-\tau Z_{at}} + \varepsilon_{u,at}. \qquad (2)$$

C_{at} = observed number of fish of age a caught in year t,
U_{at} = observed CPUE in a catchable age from a survey, carried out at time τ in the year, $(0 \leq \tau \leq 1)$,
F_{at} = fishing mortality rate,
Z_{at} = total mortality rate, $Z_{at} = F_{at} + M_{at}$,
M_{at} = natural mortality rate,
N_{at} = number of fish of age a in the stock at the beginning of year t,
$\phi_a \psi_t$ = catchability (separable into age- and year effects),
$Q(a)$ = exponent describing stock-dependent catchability,
ε_{at} = measurement errors, serially uncorrelated, $N(\mathbf{0};\sigma^2\mathbf{\Omega_t})$, $\mathbf{\Omega_t}$ predetermined,
$\varepsilon_{u,at}$ = measurement errors, serially uncorrelated, $N(\mathbf{0};\sigma_U^2\mathbf{\Omega_{Ut}})$, $\mathbf{\Omega_{ut}}$ predetermined.

The rate of mortality of the stock is described by the equation

$$N_{at} = N_{a-1,t-1}e^{-Z_{a-1,t-1}}. \qquad (3)$$

This equation relates the stock at each age with last year's values except the youngest age, here 4 years, where we use the model

$$N_{4t} = N_0 + \theta_0 r_t + \delta_{0t},. \quad (4)$$

N_0 is a constant value, r_t a recruitment index, θ_0 is a parameter and the residuals are modelled as uncorrelated $N(0;\sigma_o^2)$. When no observed recruitment index is available the simplest initial estimate of N_{4t} after the first year is N_0 for each cohort. This produces bias in the estimated fishing mortality rates if there is trend in the actual recruitment. We have therefore included an option where r_t is –1 for the first three values and +1 for the last three and these values are connected by a linear trend.

The fishing mortality rates are modelled as

$$lnF_{at} = \sum_{j=1}^{4} \zeta_{jt}\Gamma_{ja} + \delta_{5at} + \delta_{6t}. \quad (5)$$

The functions Γ_{ja} are predetermined as follows: $\Gamma_{1a} = 1$ for all ages and Γ_{2a} is 1 for the youngest age and zero for other ages. The functions Γ_{3a} and Γ_{4a} are polynomials of 2nd and 3rd degree in the age with zero derivatives at $a=a_m$ which is predetermined and do not change after this age.

The series ζ_{jt} are modelled as random walk:

$$\zeta_{jt} = \zeta_{j,t-1} + \delta_{jt}. \qquad (t > 1)$$

The residuals δ_{5at} are defined as serially uncorrelated $N(\mathbf{0};\sigma_5^2\,\mathbf{\Omega_5})$ with $\mathbf{\Omega_5}$ predetermined. Other residuals are defined in the usual way with variances σ_j^2. The modelling of persistent changes in selectivity entailed in the present formulation is less flexible than the models in the papers from 1994 and 1999 but more sensitive to shifts between younger and older fish.

Transitory variations in ln F at each age and year are represented by δ_{5at} and joint transitory variations by δ_{6t}.

Permanent variations in lnF are produced by the random walk residuals. If the variances of all random walk residuals except δ_{1t} are zero the fishing mortality rates follow the separable model apart from the transitory variations δ_{5at}.

The selectivity of the CPUE is estimated as a linear combination of the same kind of functions of age as used in equation (5), i.e.

$$\varphi_a = \sum_{j=1}^{4} \theta_j\Gamma_{ja}.$$

The selectivity is assumed to be constant, but joint variations in catchability are modelled as a sum of random walk, linear trend and transitory variations:

$$\psi_t = \eta_{t-1} + \delta_{7t}$$

$$\eta_t = \eta_{t-1} + \lambda + \delta_{8t}.$$

Transitory variations in catchability, common to all ages are represented by δ_{7t}. Surveys are designed to avoid permanent variations in catchability although persistent changes in natural conditions could upset this. In the analysis of survey data it is therefore normal to test the hypothesis that λ and σ_8 are zero and only include other values of these parameters if they are highly significant. But there is no reason to assume the absence of permanent variations in CPUE from commercial fishing fleets.

(Obviously random walk is not a realistic model of long term variations of logF or CPUE. However, it is a convenient model to estimate highly correlated variations in the short time series which we are commonly working with in this subject).

Estimation

The objective function is the joint likelihood function of the prediction errors of observed series. This differs from methods like CAGEAN and Coleraine where parametric forms of F are fitted by non-linear least squares or corresponding criteria for multinomial distributions. A linear approximation to the Kalman filter is employed to predict, first F and N and subsequently catches and CPUE. The requirement of predictability of F is the key to the method's ability to estimate stocks and fishing mortality rates from catch-at-age data without auxiliary information. Widely different forms of F in the last year(s) can fit the observed catches equally well, but the requirement of predictability is sufficient to choose between them. It is not necessary that the predictions are fairly accurate as long as there is system in the madness (ICES 1995)

When auxiliary information such as an effort series or CPUE from surveys is included, the observed values are used in the same way as the catch-at-age observations, i.e. they are predicted and the likelihood function is calculated from the multivariate distribution of all predicted series.

The estimated parameters include the initial values of ζ_{jt} and the variances of the residuals. The ratio between σ^2 and $\sigma_5{}^2$ is is often badly determined by the data and may be sensitive to moderate departures from normality. We usually fix this ratio so that the variances of $\log C_{at}$ and δ_{5at} are similar in the best observed ages. fish (CAGEAN and Coleraine assume that $\sigma_5 = 0$ and XSA that $\sigma=0$). All other variances are estimated freely which differs from various other statistical methods applied in this subject. The initial specification of $\mathbf{\Omega_t}$ and $\mathbf{\Omega_{Ut}}$ takes into account the magnitude of respective age and cohort. The quantities estimated as parameters in this approach are the following:

σ, σ_u, σ_0, $\sigma_1,\ldots,\sigma_8$.
N_0,
ζ_{j1}, j=1,2,3,4
θ_j, j=0,1,..4

Analysis of residuals is valuable in the initial specification of time series models and as a means of detecting serious misspecifications. Errors in the prediction of catch-at-age and CPUE values constitute the residuals here. Notice that the residuals are not to be regarded as estimates of ε_{at}. They are normalized by dividing by the estimated standard deviation of respective residual. The average sum of squared residuals over each age and year is examined to look for excessive variation but no formal tests are calculated for this. The average third-and fourth moments of the standardized residuals are calculated, standardized to ~N(0;1) variables, called γ_3 and γ_4 and used as tests for normality. Failure on some of these points are commonly produced by outliers. If they are not found to be caused by corrigible errors in the data or migrations a practical way to deal with them in this subject is to increase respective element of $\mathbf{\Omega_t}$ or $\mathbf{\Omega_{Ut}}$ so that single suspectable values do not exert too strong influence on the results.

We calculate the first order correlation of standardized residuals along cohorts, age and time, denoted as r_c, r_a and r_t. Unless all residuals representing joint variations at all ages are zero positive correlation along age is to be expected and is not an indication of misspecification. In our experience correlation within cohorts tends to be positive and could result from various misspecification, e.g. omitted migration and variation in natural mortality. Positive correlation with time is in our experience the most important indicator of misspecifications.

Residual analysis usually indicates bigger variations in the oldes and youngest ages. This is accommodated by defining the matrices $\mathbf{\Omega_t}$ and $\mathbf{\Omega_5}$ so that relative measurement errors are have bigger variance for the oldest fish and transitory variations of lnF for the youngest fish.

The covariance matrix of estimated parameters is obtained from the likelihood function and the covariance matrix of stocks and fishing mortality rates from the Kalman filter. It represents the uncertainty produced by the stochastic elements as represented by the estimated variances σ_i^2 in the model. As they are estimated by maximum likelihood they are biased downwards, but this is negligible. The effect of the uncertainty in other parameters cannot be calculated directly (as in regression).

Some important sources of errors do not appear in these calculations at all. In multivariate (time series) analysis with very few observations many fairly arbitrary specifications have to be imposed and there are limited possibilities of testing them. The uncertainty in the weight at age for the last estimate of the total stock is not negligible but can presumably be quantified. There is considerable uncertainty about how to prepare indices from the survey data.

References:

Gudmundsson, G. (1994). Time series analysis of catch-at-age observations. Applied Statistics *43* (1), 117-126.

Gudmundsson, G. (1999) Joint time series analysis of catch-at-age and CPUE data. *In* T.J.Quinn et al. (Eds). Fishery stock assessment models. Alaska Sea Grant College Program, University of Alaska, Fairbanks, AK-SG-98: 199-217.

Harvey, A.C. (1989). Forecasting structural time series models and the Kalman filter. Cambridge University Press, Cambridge U.K.

ICES (1993). Report of the Working Group on Methods of Fish Stock Assessment. C.M.1993/Assess:12. Copenhagen

ICES (1995). Report of the Working Group on Methods of Fish Stock Assessment. Nr. 199. Copenhagen.