Northwest Atlantic Fisheries Organization

Serial No. N2422

NAFO SCR Doc. 94/51

SCIENTIFIC COUNCIL MEETING - JUNE 1994

The Statistical Analysis of Catch-at-Age Data With Correlated Errors

Ransom A. Myers and Noel G. Cadigan
Northwest Atlantic Fisheries Centre, Science Branch
P.O. Box 5667, St. John's, NF A1C 5X1, Canada

Abstract

We extend the statistical model used to estimate abundance from commercial catch-at-age data for many of the major commercial fish species in the world. The model we consider combines commercial catch-at-age data and independent survey estimates of fish abundance; we extend the model to allow correlated errors among ages within a year for the independent survey estimates of fish abundance. We also formulate a method for modeling the fishing mortality on the oldest ages of the fish caught. Estimates are obtained using maximum likelihood. We conclude that the level of correlation among ages is sufficiently large to produce large biases in the standard methods. The statistical model that includes correlated errors greatly reduces bias and increases efficiency if the correlation in the estimation error is large.

Introduction

The management of the most important marine fisheries of the world rely on catch-at-age models to make estimates of population abundance using commercial catch-at-age data and independent estimates of abundance. However, there are severe problems with the models presently in use because they can be very biased (ICES 1991). Abundance is usually overestimated and the fishing mortality underestimated in the most recent years. This can have disastrous consequences for a fishery because it can lead to overexploitation and depletion of the population before assessment biologists are aware of the problem. The purpose of this paper is to investigate some of the statistical and population dynamics modelling assumptions used in the analysis of catch-at-age data.

We examine two aspects of the formulation of statistical catch-at-age models. First, we examine the model error structure. In most models the errors in the estimates of abundance at age are assumed to be independent. However, Myers and Cadigan (1993) demonstrated that this assumption is not valid for many research surveys of fish abundance. They found from analyses of multiple surveys of the same population that there are generally positive correlations in the errors among ages for a given year. For example, if the number of three year olds were overestimated in 1990, the number of four year olds were usually overestimated as well. Myers and Cadigan (submitted) introduced a model that explicitly considers such an error structure. However, their analysis mostly focused on testing for changes in natural mortality over time in northern cod. The method is extended here to several stocks and its utility is assessed using simulations.

The second problem we address is how to formulation the catch at age model. Normally an assumption about the fishing mortality at the oldest age is made that yields an expression for the numbers at that age. We consider a general formulation because in the past several *ad hoc* methods have been used and they may yield very different estimates of abundance. For example, two alternative formulations of the fishing mortality at the oldest age for cod on the southern Grand Banks of Newfoundland produced different estimates of biomass by a factor of two (NAFO 1994).

Our approach for both of these problems is to construct statistical models in a hierarchical fashion. That is, our models allow simple and more complex models to be tested using standard likelihood ratio methods. We apply our models to four data sets, and use data based simulations to test our methods and to determine the conditions under which the correlated error model can be expected to improve estimates of abundance

Methods

Commercial catch-at-age data is analyzed jointly with research vessel estimated numbers of age using statistical virtual population analysis (VPA) models (Hilborn and Walters 1992). In these models, population numbers at age are estimated in the last year for which survey data are available, and then the age composition for all years is constructed. Although a variety of these models have been proposed (Gavaris 1988; Hilborn and Walters 1992; Mohn and Cook 1993), they share many common assumptions. We will base our approach on the the methods used in the current assessments in eastern Canada, the eastern USA, and by ICES.

Population model

Some notation is first developed. $N_{a,y}$ is the number of age a fish in the population at the beginning of year y. $C_{a,y}$ is the number of age a fish in year y that are caught by the fishery. Lowercase letters are used to denote log transformations, e.g. $n_{a,y} = \log(N_{a,y})$. The number of ages and years modeled are A and Y, respectively.

The model for numbers at age is

$$N_{a,y} = C_{a,y}e^{m/2} + N_{a+1,y+1}e^{m}, \qquad (1)$$

where m is the natural mortality rate which is usually assumed to be constant for all ages and years (Pope 1972). The catches are assumed to occur halfway through the year in (1).

Let $F_{a,y}$ denote the fishing mortality rate and $Z_{a,y} = m + F_{a,y}$ denote the total mortality. $F_{a,y}$ is defined as:

$$F_{a,y} = \log(\frac{N_{a,y}}{N_{a+1,y+1}}) - m.$$

Numbers at age A for all other years than Y are computed using $F_{A,y}$ and the solution for $N_{A+1,y+1}$ in (1):

$$\begin{aligned} N_{A,y} &= \frac{C_{A,y}\exp(M_{A,y}/2)}{1-\exp(-F_{A,y})}, \\ F_{A,y} &= \alpha_y \sum_{a=1}^{A-1} w_a F_{a,y}. \end{aligned} \qquad (2)$$

where α_y is a parameter to be estimated and w_a is a predetermined weight constrained such that $\sum_{a=1}^{A-1} w_a = 1$. The weights, w_a, are chosen over the ages in which the fishing mortality is well estimated. We investigate placing reasonable constraints on the α_y's.

The above formulation allows the population to be reconstructed, i.e. estimates made of $N_{a,y}$, if the numbers at age at the last year and the α_y's are estimated. We will call the numbers at age in the last year "survivors" and they will be denoted by $S_a \equiv N_{a,Y}$ and $s_a \equiv n_{a,Y}$. Given estimates of the S_a's, the numbers at age for all the cohorts represented in the last year can be reconstructed using Eq. (1). Given the α_y's, all other numbers at age can be constructed using Eq. (2) and (1). Thus, the goal of the procedure is to estimate the survivors, S_a's, and the α_y's.

The key assumptions in the population dynamics model are as follows.

1. The population dynamics is described by Eq. 1.

2. Catch-at-age is known without error. There is unknown misreporting and aging errors in the catches that could lead to considerable error in their reporting. We do not address this problem here.

3. Fishing mortality on the oldest age is a weighted average of the fishing mortality at younger ages multiplied by α_y.

4. Natural mortality, m, is known and does not change with time. For the cod stocks consider, we will follow the usual assumption that $m = 0.2$ for all years and ages.

Statistical models

Let $R_{a,y}$ be a random index of abundance from a research survey and let $r_{a,y} = \log(R_{a,y})$. We will not distinguish between a random variable and its observation by using the usual notation (normally upper and lower case is used for this purpose but the convention is in use here for logarithms); however, the distinction will be made when necessary.

Independent errors

The usual statistical model used in the analysis of commercial catch-at-age data assumes that the deviations in the log survey estimates from the population model are distributed as uncorrelated, normal random deviates with constant variance, i.e.

$$\begin{aligned} r_{a,y} &= q_a + n_{a,y} - tZ_{a,y} + \epsilon_{a,y}, \\ \epsilon_{a,y} &\overset{\text{ind}}{\sim} \mathrm{N}(0, \sigma^2), \end{aligned} \tag{3}$$

where q_a is the log catchability of the research surveys at age a and $\overset{\text{ind}}{\sim}$ denotes independently distributed over years and ages. The parameter t is the proportion of the year that has been completed when the survey takes place; for example, if the surveys take place in November, we would set $t = \frac{11}{12}$. Eq. (3) applies to some subset of the ages and years the population model covers.

The research surveys index of abundance is related to the estimated numbers at age by the catchability of the research gear. Fish at different ages are not caught with the same efficiency in the research surveys; thus the log catchability coefficients (q_a's) are estimated for each age separately.

The parameters we estimate are the A survivors in year Y, the A survey catchability coefficients (by age), and the α_y's. Note that sometimes the α_y's are constrained to be equal to one parameter that is estimated, α, or constrained to be equal to one. Note that $n_{a,y}$ is a nonlinear function of the survivors and α_y's. The variance of the $\epsilon_{a,y}$ (σ^2) is also estimated. Let $\mathbf{r}$ be the random $A \times Y$ matrix of log survey numbers,

$$\mathbf{r} = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdot & \cdot & \cdot & r_{1,Y} \\ r_{2,1} & r_{2,2} & \cdot & \cdot & \cdot & r_{2,Y} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{A,1} & r_{A,2} & \cdot & \cdot & \cdot & r_{A,Y} \end{bmatrix} = [\mathbf{r}_{,1}, \ldots, \mathbf{r}_{,Y}],$$

and let $\underline{\mathbf{r}}$ be the $AY \times 1$ vector of log survey numbers formed by stacking the columns of $\mathbf{r}$:

$$\underline{\mathbf{r}}_{AY\times 1} = \begin{bmatrix} \mathbf{r}_{,1} \\ \mathbf{r}_{,2} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{r}_{,Y} \end{bmatrix}.$$

The variance-covariance matrix is $\boldsymbol{\Sigma} = \text{cov}(\underline{\mathbf{r}}) = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is a $AY \times AY$ identity matrix.

Correlated errors

We consider an alternative statistical model which allows for correlated errors among ages in each year. We add a random effect (ξ) for years:

$$r_{a,y} = q_a + n_{a,y} - tZ_{a,y} + \epsilon_{a,y} + \xi_y, \quad (4)$$
$$\epsilon_{a,y} \overset{\text{ind}}{\sim} \mathrm{N}(0, \sigma^2), \quad (5)$$
$$\xi_y \overset{\text{ind}}{\sim} \mathrm{N}(0, \phi).$$

This is a mixed effects model in which all errors are equally correlated within a year but are independent between years. The correlation is $\rho = \phi/(\sigma^2 + \phi)$. The variance-covariance matrix is $\boldsymbol{\Sigma} = \text{cov}(\underline{\mathbf{r}}) = \sigma^2\mathbf{I} + \phi(\mathbf{I}_\mathrm{Y} \otimes \mathbf{U}_\mathrm{A})$, where $\otimes$ denotes the Kronecker product (defined in the **Appendix 1**), $\mathbf{I}_\mathrm{Y}$ is a $Y \times Y$ identity matrix and $\mathbf{U}_\mathrm{A}$ is an $A \times A$ matrix of 1's.

Linear models with random components, e.g. mixed models, are a well developed branch of statistics (Searle *et al.* 1992). Our models are applications of recent research with non-linear models.

Maximum likelihood estimates

Maximum likelihood methods are used to estimate the model parameters under the two alternative error structures described above. The loglikelihood (l) is given by

$$l = k - \left[\log(|\,\boldsymbol{\Sigma}\,|) + (\underline{\mathbf{r}} - \mathrm{E}(\underline{\mathbf{r}}))'\boldsymbol{\Sigma}^{-1}(\underline{\mathbf{r}} - \mathrm{E}(\underline{\mathbf{r}}))\right]/2,$$

where k is a constant. The covariance matrix is combined with a weighting matrix to accomodate missing surveys and to extend the population beyond the oldest age that survey data is used for estimation. The covariance matrix actually modelled is:

$$\mathbf{W}^{-1}\boldsymbol{\Sigma}\mathbf{W}^{-1}$$

where $\mathbf{W}$ is a diagonal matrix of full rank. In this analysis the elements in $\mathbf{W}$ are either 1's or ∞'s; the ∞ elements correspond to missing surveys and gives these surveys zero weight in parameter estimation. Inferences are based on the marginal likelihoods of surveys with nonzero weights.

Maximum likelihood (or equivalently generalized least-squares) parameter estimates are obtained using the iterative algorithms in Gumpertz and Pantula (1992). The derivatives of (3) with respect to the survivors, the q_a's, and the α_y's required in this algorithm are developed in **Appendix 2**. Gumpertz and Pantula (1992) present a model with the same error structure as assumed here and call it a "onefold nested error structure".

It is useful to compare the two error assumptions on the estimation of the survivors, the q_a's, and the α_y's. The difference can be understood by fixing ϕ and σ^2 and examining the resulting estimating functions. Assuming errors are independent leads to a search for parameter estimates that minimize the total error sum of squares

$$\sum_y \sum_a e_{a,y}^2/\sigma^2, \quad (6)$$

where $e_{a,y} = r_{a,y} - \hat{r}_{a,y}$. Assuming errors are correlated within years leads to minimizing

$$\sum_y \sum_a e_{a,y}^2/\sigma^2 - \sum_y \frac{(\sum_a e_{a,y})^2}{(A + \sigma^2/\phi)\sigma^2}. \quad (7)$$

This equation is derived using the inverse formula in Gumpertz and Pantula (1992): $\boldsymbol{\Sigma}^{-1} = \mathbf{I} \otimes \mathbf{U}(\lambda_2 - \lambda_1)/A\lambda_1\lambda_2 + \mathbf{I}/\lambda_2$ where $\lambda_1 = A\phi + \sigma^2$ and $\lambda_2 = \sigma^2$. If $\phi = 0$ then (6) and (7) are the same. The total sums of squares in (7) is adjusted by removing a component due to the year effects.

The likelihood ratio test is used to test statistical hypotheses. To test that a subset of p parameters is equal to some specified value, models with and without the parameter constraints are fit yielding reduced and full log-likelihoods. Two times the difference between the reduced and full log-likelihoods is asymptotically a chi square random variable with p degrees of freedom (Cox and Hinkley 1974). For example, to test that $\phi = 0$ one computes twice the difference between the log-likelihoods obtained with independent and correlated errors and compares this value with the critical value for a $\chi^2_{(1-\alpha),1}$ where α is the level of the test.

The above formulation allows alternative hypotheses about the fishing mortalities at the oldest ages ($F_{A,y}$'s) to be formally tested. α_y may be estimated separately for each year which is equivalent to estimating a separate $F_{A,y}$ for each year. These estimates can then be examined for trends with time, etc. If there are no trends, then a common parameter $\alpha_y = \alpha, y = 1, ..., Y$, for all y can be estimated. Similarly, α_y can be constrained to be constant for one time period and then estimated separately for another time period.

Results

We consider four cod populations in detail. We use the latest data in all cases except for Labrador cod. For this population we terminated the analysis in 1992 because the fishery was closed in that year. The catch data in 1993 is very poorly estimated because it was largely recreational, bycatch, and from-non Canadian boats beyond the 200 mile limit (Bishop et al. 1994). The methods are applied to data sets consisting of catch-at-age and survey estimates-at-age of relative abundance for fish populations in the Northwest Atlantic (Table 1). All surveys used in the analysis are from directed research trawl surveys. In all cases we use the longest survey time series.

Fishing Mortality at the Oldest Age

We first consider different assumptions about the fishing mortality at the oldest age, $F_{A,y}$. For each population, we estimate 6 models that represent the range of assumptions commonly used in practise. We assume that errors are correlated; results presented later suggest that there is little loss of statistical power by assuming random year effects exists when they don't. If q_A is estimated then it is necessary to constrain it to equal q_{A-1} (or some other reasonable assumption) because there is confounding between α and q_A. This constraint generally does not produce significant changes in model fits. The results (Table 2) demonstrate that the assumption $F_{A,y} = F_{A-1,y}$ is generally the most parsimonious. In no case was there evidence that the fishing mortality at age A was significantly different from that at age $A-1$.

Note that for Southern Grand Banks cod the loglikelihood of two constrained models is greater than for models without the constraints. This appears to be related to the resolution of the maximization algorithm (Fisher's scoring method) used; that is, the unconstrained solution is not converging at the exact maximum. This is only a problem when the the parameter estimate is very close to the constrained value and the loglikelihoods are very similar.

We also estimated the ratio of fishing mortality in age A to $A-1$ for each year separately. In doing so, if the maximum age is less than A it is necessary to constrain $\alpha_1, ..., \alpha_d$ ($d = A$-maximum survey age) because there is no survey data for these cohorts. All models with year specific α_y's produced significant increases in loglikelihoods compared to models with constant α's. The estimated α_y's are presented in Fig. 1 with 95% Bonferroni confidence intervals (the level of the test is divided by the number of estimated α_y's). It was necessary to constrain some α_y's because estimates otherwise diverged; for Labrador cod $\alpha_{1979} = \alpha_{1980}$, for Southern Grand Banks cod $\alpha_{1977} = \alpha_{1978} = \alpha_{1979}$, and for Eastern Scotian Shelf cod $\alpha_{1978} = \alpha_{1990} = \alpha_{1992}$ = average of the other α_y's.

The significant increases in loglikelihoods are caused by only a few α_y's for each stock. No patterns are evident in the estimated α_y's indicating that the only reasonable simplification of these models is to estimate a common α. We recommend doing this for several reasons: First, the errors in the catches at ages A and $A-1$ are likely to be relatively large and some of the significant α_y's may in fact not be significant if the variability in the catch data could be included; the estimates of survivors for younger ages did not change very much whether year specific or a constant α is estimated; estimating year specific α's is numerically difficult.

Correlated Errors

We next consider estimates for the model under the assumption $F_{A,y} = F_{A-1,y}$ for the models with independent and correlated estimation errors (Table 3).

In all cases there was a very large increase in the loglikelihood with the correlated error model, i.e. in all cases ϕ was significantly different from zero. Our analysis demonstrates most of the variability in the research surveys for Labrador cod is due to correlated errors, i.e. ϕ is much larger than σ^2 (Table 3). The error components are approximately equal for Southern Grand Banks cod and St. Pierre Bank cod. For Eastern Scotian shelf cod ϕ is less than σ^2 (Table 3).

There is a large difference in the estimated survivors for the different error assumptions. The correlated error model produces lower estimates of survivors in all cases except for 3+ survivors in Eastern Scotian Shelf cod. We will use simulations to help to determine the importance of these differences.

The reason for the increase in the loglikelihood is clear from an examination of the residuals in the models which assume independent errors (Fig. 2a and 2b). The model errors in Eq. (4), i.e. the correlated errors model, should be standardized to compare with those in the independent errors model. This is done by multiplying the vector of differences between observed and predicted survey numbers by the square-root inverse of the estimated covariance matrix (Jones 1993). Including correlated errors greatly reduces the year effects apparent in the model errors. The year effects are much greater for Labrador, St. Pierre Bank, and Southern Grand Banks than for Eastern Scotian shelf cod.

A note of caution is in order. The residuals from the independent errors fit and the unstandardized residuals from the correlated errors fit show a deterministic pattern since 1987 for Labrador and Southern Grand Banks cod and throughout all years for Eastern Scotian shelf cod. This suggests model misspecification and, potentially, that some of the variation due to year effects may be of a deterministic nature rather than a random nature. That is, there are trends in the residuals. If the causes for the apparent deterministic year effects were known then they could be used to improve the analysis.

Simulation Results

We use a simulation study to determine the conditions under wich the correlated error model improves abundance estimates. Our simulation study will use data from four populations, and will remain as close as possilbe to the data.

We estimated the bias and efficiency of estimation with correlated and independent errors. We generated pseudo-research survey estimates, the assumed population dynamics model using the parameter estimates (Ratkowsky 1983, p. 23). Two sets are generated: pseudo estimates under the assumption of independence errors and under the assumption that the estimates are correlated. For each of the two case, 1000 realizations of the pseudo-estimates are obtained for each year and age. We then estimated the model parameters for each realization under the assumption of independent and correlated errors. That is, we obtained 4000 estimates of abundance, from which we calculated estimates of the bias and variance for the estimated numbers at age.

We examine four cod populations in detail because they represent low (Eastern Scotian Shelf cod), medium (St. Pierre Bank and Souther Grand Banks cod), and high (Labrador cod) levels of correlated estimation errors among ages. The variance parameters used to generate the pseudo-research survey estimates are those given in Table 3. The only difference between the two simulated data sets in each iteration is the distributional assumptions used to generate the simulated survey estimates.

We estimate the bias components and efficiency of independent and correlated mles when the true distribution of research surveys are either independent or correlated. We measure the bias component and standard deviation in parameter estimates using the simulationed parameter estimates. We normalize by dividing the bias and standard deviation calculated in the 1000 realizations by the true population numbers used in generating simulation data. We summarize the results for the recruits, at age 3, and the total estimated numbers in each year (Fig. 3a and 3b).

Correlated errors mles have lower variances and are less biased than the independent errors mles if the errors are significantly correlated for both recruits (Fig. 3a) and 3+

numbers (Fig. 3b). The reduction in bias in the final year for the 3+ numbers is important for Labrador cod, from 20% to 9%, and Southern Grand Banks cod, from 50% to 30%. Similarly, there is a large reduction in the normalized standard deviation for these two populations in the last year. This reduction is particularly important for Labrador cod, where the correlated errors mles have reduced the standard deviation of the estimates to almost one third of that from the independent error mles.

If the correlation in the estimation errors are small (Eastern Scotian Shelf cod), or if they are assumed to be independent in the simulations, then both the correlated and the independent error mles perform similarly (Fig. 3a and 3b). This result shows that if the correlated error mles is used on a population with independent errors there is no extra bias or loss in efficiency.

Discussion

The assumption that research survey estimates of abundance at age are independent has a large effect on the estimation of population abundance in catch-at-age models if the assumption is violated. We have introduced a model in which the estimation errors are assumed to be correlated. We used a data based simulation study to conclude that the correlated error maximum likelihood model has

- greater accuracy, lower bias and smaller variance, in estimating numbers-at-age if the the true estimation errors are correlated, and
- there is no loss in accuracy if the ture estimation errors are independent.

It is crucial that an appropriate error structure be used in the statistical analysis of catch at age data. The assumption that the estimation errors in the research surveys are independent led to unsupportable inference of an increase in natural mortality during the winter of 1991 of Labrador cod(Myers and Cadigan submitted).

The source of this correlation in the estimation errors is unknown. Environmental conditions that would effect catchability, differences in how the crew handles the fishing gear, or sampling variability are obvious possible causes. The trends in residuals are a cause for concern; they indicate that there is potential misspecification in the model for the expectation of research surveys and that biases other than statistical biases may exist.

It is useful to consider why including correlated errors improves the model performance in estimating abundance. If errors are correlated within a year, then we can use this information when estimating the abundance of the youngest ages in the final year in the analysis. For the oldest ages in the final year, there will have been typically 5 or 6 surveys of these cohorts. Thus, if the survey estimates of these cohorts in the final year is lower than expected, then the model will use this information to lower the estimates of abundance for the younger ages. That is, by incorporating the correlated structure of the errors, improved estimates of abundance are possible because cohorts are observed over several years.

Our simulation results clearly demonstrate that the correlated error model reduces the variability of the estimated survivors and their bias. The tendency of VPA models to overestimate abundance in the last years in which the survey was carried out is well known but not well understood (ICES 1991).

Considerable statistical bias still exists even with the improved modelling of research survey errors and it would be desirable to reduce this bias. Smith and Gavaris (1993) present two bootstrap bias estimation procedures (one is identical to that used here) and a parametric procedure based an a Taylor's series expansion of the loglikelihood formulated in Gavaris (1993) for the bias of some derived parameters in a VPA. However, the authors do not test whether the bias estimation techniques can be used to provide bias reduced estimators and at what cost in terms of increases in standard errors. Also, the procedure Smith and Gavaris (1993) suggest (Taylor's series) has not been developed for mixed effects nonlinear regression models. For these reasons we have not implemented the techniques but future research in this direction may lead to improved estimation procedures.

Acknowledgments

We thank Geoff Evans, Bob Mohn, and Don Stansbury for comments on early versions of the manuscript. This work was supported by the Northern Cod Science Program.

REFERENCES

Bishop, C.A., Murphy, E.F., Davis, M.B., Baird, J.W., and G.A. Rose. 1993. An assessment of the cod stock in NAFO Divisions 2J+3KL. NAFO SRCR Doc. 93/86, Ser. No. N2271: 51 p.

Bishop, C. A., Baird, J. W., and Murphy, E. F. 1991. "An assessment of the cod stock in NAFO Div. 3Ps," Canadian Atlantic Fisheries Scientific Advisory Committee Res. Doc. 91/36.

Cox, D. R., and D. V. Hinkley. 1974. Theoretical Statistics. Chapman and Hall, London, J.K. 210 p.

Deriso, R. B., T. J. Quinn II, and P. R. Neal. 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42: 815-824.

Fournier, D. and C. Archibald. 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39: 1195-1207.

Gavaris, S. 1993. Analytical estimates of reliability for the projected yield from commercial fisheries. p. 185-191. *In* S. J. Smith, J. J. Hunt, and D. Rivard [ed.] Risk evaluation and biological reference points for fisheries management. Can. Spec. Publ. Fish. Aquat. Sci. 120.

Gumpertz, M. L., and S. G. Pantula. 1992. Nonlinear regression with variance components. J. Amer. Stat. Assoc. 87: 201-209.

Hilborn, R., and Walters, C. J. (1992), *Quantitative fisheries stock assessment: choice, dynamics and uncertainty,* New York: Chapman and Hall.

ICES (1991), "Report of the working group on methods of fish stock assessments," International Council for the Exploration of the Sea C.M. 1991/Assess:25.

Jones, R. H. 1993. Longitudinal data with serial correlation: a state-space approach. London: Chapman and Hall.

Lapointe, M. F., Peterman, R. M., and MacCall, A. D. (1989). "Trends in fishing mortality rate along with errors in natural mortality rate can cause spurious time trends in fish stock abundances estimated by virtual population analysis (VPA)," *Canadian Journal of Fisheries and Aquatic Sciences,* 46, 2129-2139.

Megrey, B. A. (1989), "Review and comparison of age-structured stock assessment models from theoretical and applied points of view," *American Fisheries Society Symposium,* 6, 8-48.

Mohn, R. K., and Cook, R. (1993), "Introduction to sequential population analysis," Northwest Atlantic Fisheries Organization Sci. Counc. Stud. No. 17.

Mohn, R. K., and MacEachern, W. J. (1992), "Assessment of Eastern Scotian Shelf cod in 1991," Canadian Atlantic Fisheries Scientific Advisory Committee Res. Doc. 92/54.

Myers, R. A., and Cadigan, N. G. (1993), "Density-dependent juvenile mortality in marine demersal fish," *Canadian Journal of Fisheries and Aquatic Sciences,* 50: 1576-1590.

NAFO 1992 Report of the Special Meeting of Scientific Council, June 1992. NAFO SCS. Doc. 92/20 Ser NO N2112.

NAFO 1994 Assessment of cod in NAFO Div. 3N

Pope, J. G. 1972. An investigation of the accuracy of virtual population analysis. ICNAF Res. Bull. 9, 65-74.

Ratkowsky, D. A. 1983. Nonlinear regression modeling. Marcel Dekker, Inc., New York. 276 p.

Rivard, D. (1989), "Overview of the systematic, structural, and sampling errors in cohort analysis," *American Fisheries Society Symposium,* 6, 49-65.

Smith, S. J., and S. Gavaris. 1993. Evaluating the accuracy of projected catch estimates from sequential population analysis and trawl survey abundance estimates. p. 163-172. *In* S. J. Smith, J. J. Hunt, and D. Rivard [ed.] Risk evaluation and biological reference points for fisheries management. Can. Spec. Publ. Fish. Aquat. Sci. 120.

Searle, S. R., G. Casella, and C. E. McCulloch. 1992. Variance components. John Wiley and Sons, New York, NY. 501 p.

Appendix 1

If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $r \times s$ matrix then the Kronecker product (direct product or tensor product) of $\mathbf{A}$ and $\mathbf{B}$ is

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,n}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \cdots & a_{2,n}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{m,1}\mathbf{B} & a_{m,2}\mathbf{B} & \cdots & a_{m,n}\mathbf{B} \end{bmatrix}.$$

$\mathbf{A} \otimes \mathbf{B}$ is a matrix of order $(mr \times ns)$.

Appendix 2

Statistical estimation procedures for both independent and correlated errors require the formulation of a model for the expectation and covariance of research surveys. We will reformulate Eq. 3 and 4 in matrix notation. Let $\mathbf{n}$ be the matrix of log numbers at year and age:

$$\mathbf{n} = \begin{bmatrix} n_{1,1} & n_{1,2} & \cdots & n_{1,Y} \\ n_{2,1} & n_{2,2} & \cdots & n_{2,Y} \\ \vdots & \vdots & & \vdots \\ n_{A,1} & n_{A,2} & \cdots & n_{A,Y} \end{bmatrix} = [\mathbf{n}_{,1}, \ldots, \mathbf{n}_{,Y}],$$

and let

$$\underline{\mathbf{n}}_{AY \times 1} = \begin{bmatrix} \mathbf{n}_{,1} \\ \mathbf{n}_{,2} \\ \vdots \\ \mathbf{n}_{,Y} \end{bmatrix}.$$

Similarly, let $\mathbf{Z}$ be the matrix of total mortalities at age, and let $\underline{\mathbf{Z}}$ is an $AY \times 1$ vector of total mortalities. We previously defined $\underline{\mathbf{r}}$ as the random vector variable for research surveys. Note that $\mathbf{r}$ (the matrix version of $\underline{\mathbf{r}}$) may have fewer than A rows but to simplify notation we assume that $\mathbf{r}$ has A rows. The following results are easily modified if this is not the case. The expectation of $\underline{\mathbf{r}}$ is:

$$E(\underline{\mathbf{r}}) = \mathbf{1} \otimes \mathbf{q} + \underline{\mathbf{n}} - t\underline{\mathbf{Z}},$$

where $\mathbf{1}$ is a $Y \times 1$ vector of 1's and $\mathbf{q}$ is an $A \times 1$ vector of survey catchability coefficients.

The estimation algorithms used require expressions for the derivatives of $E(\underline{\mathbf{r}})$ with respect to the survivors (denoted as $\mathbf{s}$, an $A \times 1$ vector), the α_y's ($\boldsymbol{\alpha}$, a $(Y-1) \times 1$ vector), and the catchabilities ($\mathbf{q}$). For convenience, let $\boldsymbol{\theta}' = [\mathbf{s}', \boldsymbol{\alpha}']$. The derivatives are:

$$\begin{aligned} \frac{\partial E(\underline{\mathbf{r}})}{\partial \mathbf{q}'} &= \mathbf{1} \otimes \mathbf{I}_{A \times A}, \\ \frac{\partial E(\underline{\mathbf{r}})}{\partial \boldsymbol{\theta}'} &= t\frac{\partial \underline{\mathbf{Z}}}{\partial \boldsymbol{\theta}'} - \frac{\partial \underline{\mathbf{N}}}{\partial \boldsymbol{\theta}'} Diag(1/\underline{\mathbf{N}}). \end{aligned}$$

$Diag(1/\underline{\mathbf{N}})$ is a diagonal $AY \times AY$ matrix with diagonal elements equal $1/N_{a,y}$. $\partial E(\underline{\mathbf{r}})/\partial \mathbf{q}'$ is an $AY \times A$ matrix and $\partial E(\underline{\mathbf{r}})/\partial \boldsymbol{\theta}'$ is an $AY \times (A+Y-1)$ matrix. The following results are necessary to compute $\partial E(\underline{\mathbf{r}})/\partial \boldsymbol{\theta}'$:

$$\frac{\partial N_{a,y}}{\partial \theta_i} = \begin{cases} \frac{\partial N_{a+1,y+1}}{\partial \theta_i} \exp(M_{a,y}) & \text{if } a < A, y < Y, \\ \frac{\partial F_{a,y}}{\partial \theta_i} N_{a,y}/(1-\exp(F_{a,y})) & \text{if } \begin{cases} a = A, & y < Y, \quad \theta_i = s_i, \text{ or} \\ a = A, & i < Y, \quad \theta_i = \alpha_i, \end{cases} \\ \frac{F_{a,y}}{\alpha_y} N_{a,y}/(1-\exp(F_{a,y})) & \text{if } a = A, i = y, \theta_i = \alpha_i, \\ 1 & \text{if } a = i, y = Y, \theta_i = s_i, \\ 0 & \text{otherwise.} \end{cases}$$

$$\frac{\partial F_{a,y}}{\partial \theta_i} = \begin{cases} \frac{\partial N_{a,y}}{\partial \theta_i} \frac{1-\exp(F_{a,y})}{N_{a,y}} & \text{if } a < A, \\ \sum_{a=1}^{A-1} w_a \frac{\partial F_{a,y}}{\partial \theta_i} & \text{otherwise.} \end{cases}$$

Table 1. Fish populations analyzed. Ages and years refer to minimum -maximum ages and first-last years used in the analyses. The proportion of the year that has been completed when the survey took place is t.

Population	Source	Survey	t	Ages	Years
Labrador cod	Bishop *et al* (1993)	Autumn	$\frac{11}{12}$	3-12	78-92
Southern Grand Banks cod	Bishop *et al* (1991)	Spring	$\frac{5}{12}$	3-12	78-93
Eastern Scotian Shelf cod	Mohn and MacEachern (1992)	July	$\frac{7}{12}$	3-11	71-93
St. Pierre Banks cod	Bishop *et al* (1991)	Ferburary	$\frac{2}{12}$	3-12	80-94

Table 2. Model estimates under different assumptions about the fishing mortality of the oldest age, $F_{A,y}$. Estimates constrained to be 1 are denoted by "1*". The first line for each population is the assumption traditionally used in assessments. The weights used to compute $F_{A,y}$ are zero for the ages not listed. The survey catchability coefficients for the two oldest ages are constrained to be equal, i.e. $q_{A-1} = q_A$.

Population	Ages	α_y	Log likelihood	P value
Labrador Cod	7,8,9	1*	-24.766	
	7,8,9	0.911	-24.308	0.339
	9,10,11	1*	-24.237	
	9,10,11	0.839	-21.462	0.018
	11	1*	-22.504	
	11	1.096	-21.149	0.100
Southern Grand Banks Cod	7,8,9	1*	-127.604	
	7,8,9	1.045	-127.638	-
	9,10,11	1*	-126.628	
	9,10,11	1.085	-126.641	-
	11	1*	-128.654	
	11	1.229	-128.250	0.369
Eastern Scotian Shelf Cod	6,7,8	1*	-204.982	
	6,7,8	0.422	-199.454	0.001
	8,9,10	1*	-193.691	
	8,9,10	0.594	-187.257	0.000
	10	1*	-185.732	
	10	0.941	-185.227	0.315
St. Pierre Banks Cod	7,8,9	1*	-118.460	
	7,8,9	0.665	-117.638	0.120
	9,10,11	1*	-116.763	
	9,10,11	0.829	-116.332	0.353
	11	1*	-116.312	
	11	0.883	-115.760	0.293

Table 3. Model estimates for under the alternative assumptions for estimation errors. Estimates constrained to be zero are denoted by "-". Numbers are in millions at the beginning of the year. Standard errors are in parentheses. The parameters are: the variance of the random year effects, ϕ, the correlation of the estimation errors among ages in a year, ρ, and the residual variance of the estimation errors, σ^2.

population	ϕ	ρ	σ^2	Log likelihood	Survivors at age 3	Survivors at ages 3+
Labrador cod	-	-	0.311	-114.448	77 (41)	334 (69)
	0.259	0.838	0.050	-18.241	65 (23)	192 (40)
S. Grands Banks cod	-	-	0.631	-160.970	103 (79)	163 (85)
	0.216	0.464	0.250	-127.623	66 (40)	134 (25)
E. Scotian Shelf cod	-	-	0.434	-189.228	39 (25)	110 (32)
	0.056	0.149	0.322	-183.330	36 (22)	112 (34)
St. Pierre Banks cod	-	-	0.422	-148.064	13 (9)	80 (21)
	0.216	0.500	0.215	-115.706	12 (7)	63 (22)

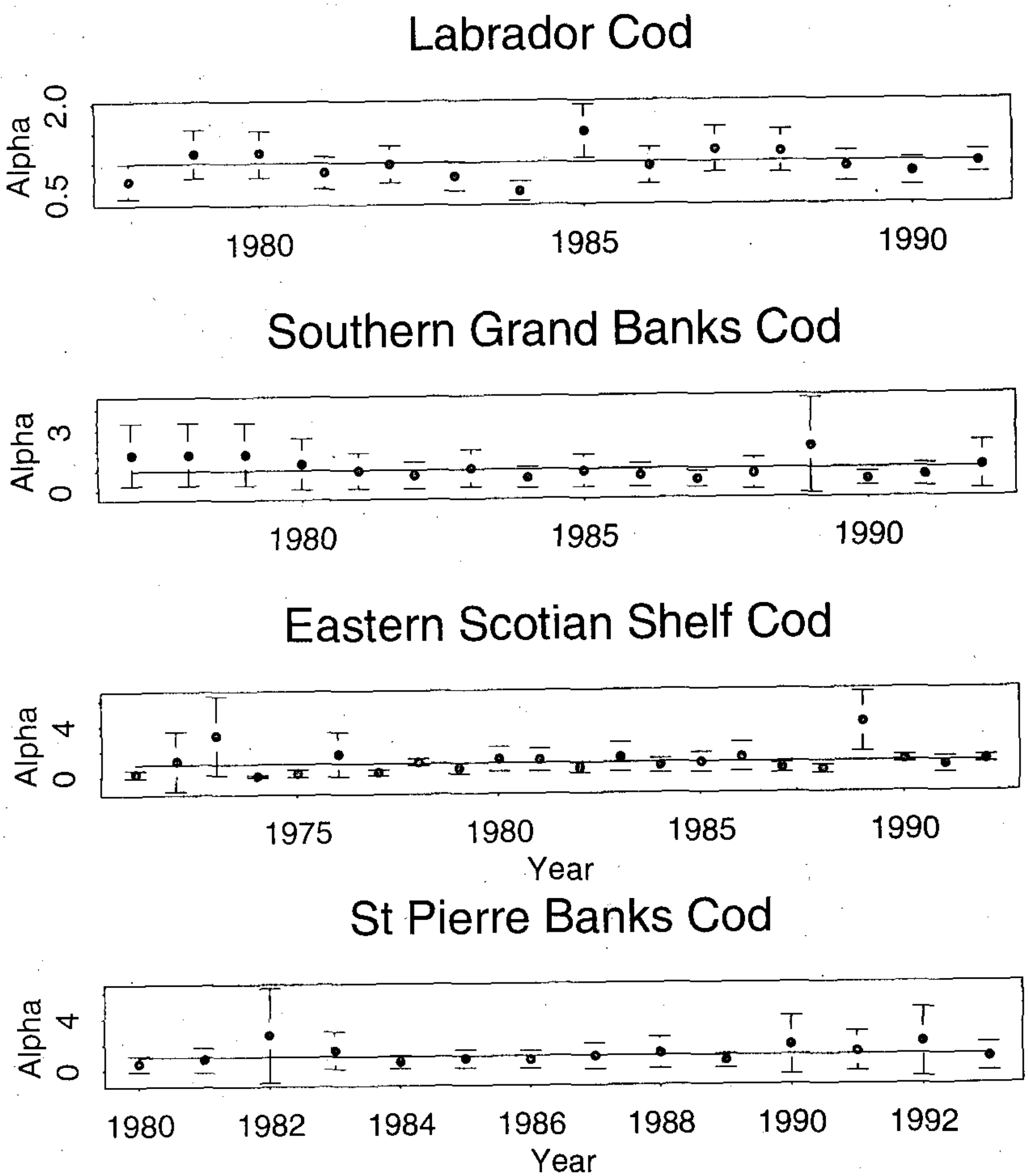

Fig. 1. The estimated ratio of the fishing mortality at the oldest age to the next oldest age, α_y, with 95% Bonferroni confidence intervals (the level of the test is divided by the number of estimated α_y's).

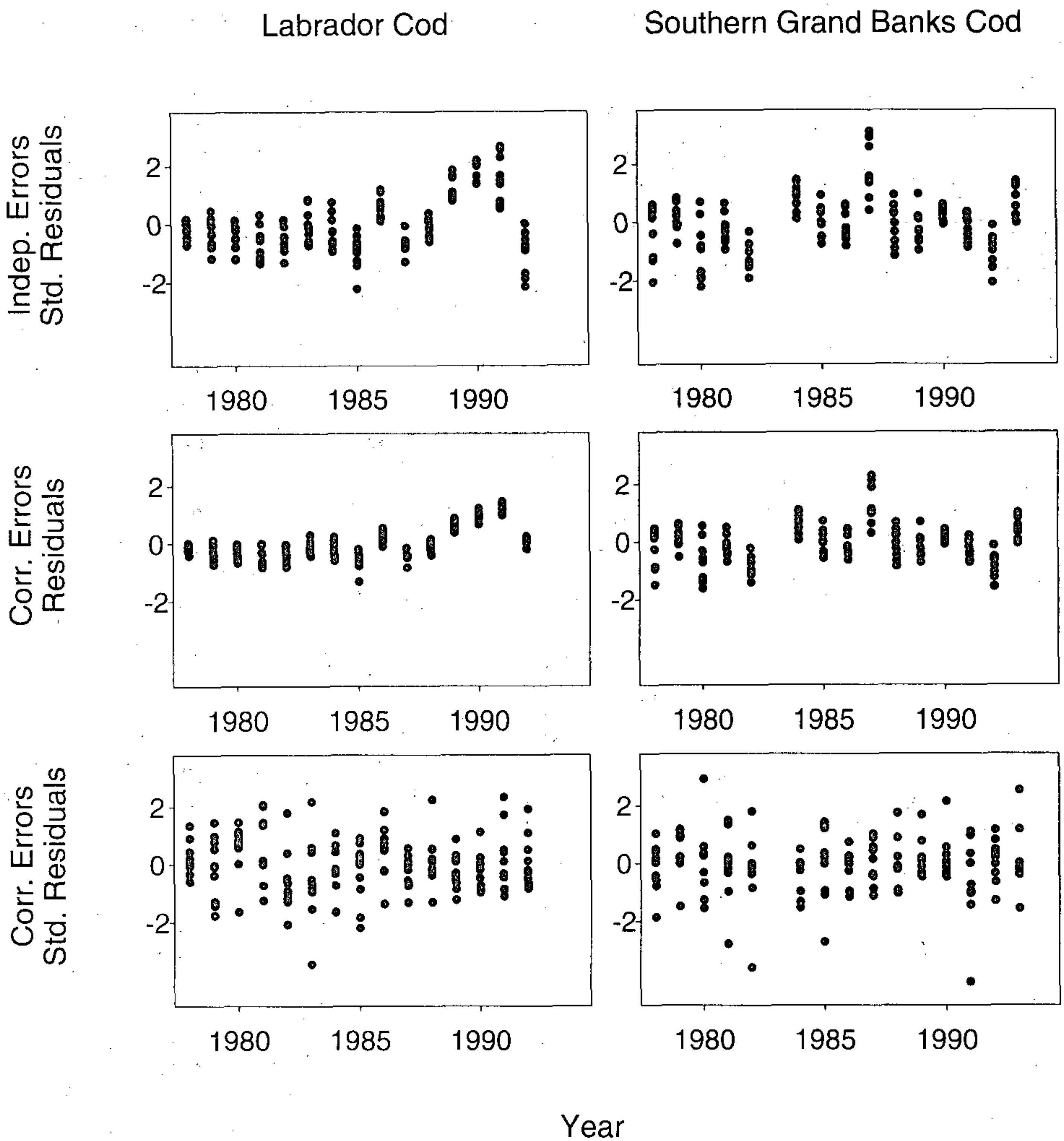

Fig. 2a. Model residuals for Labrador and Southern Grand Banks cod under alternative statistical models: independent and correlated estimation error mles. In all models we assume that $F_{A,y} = F_{A-1,y}$.

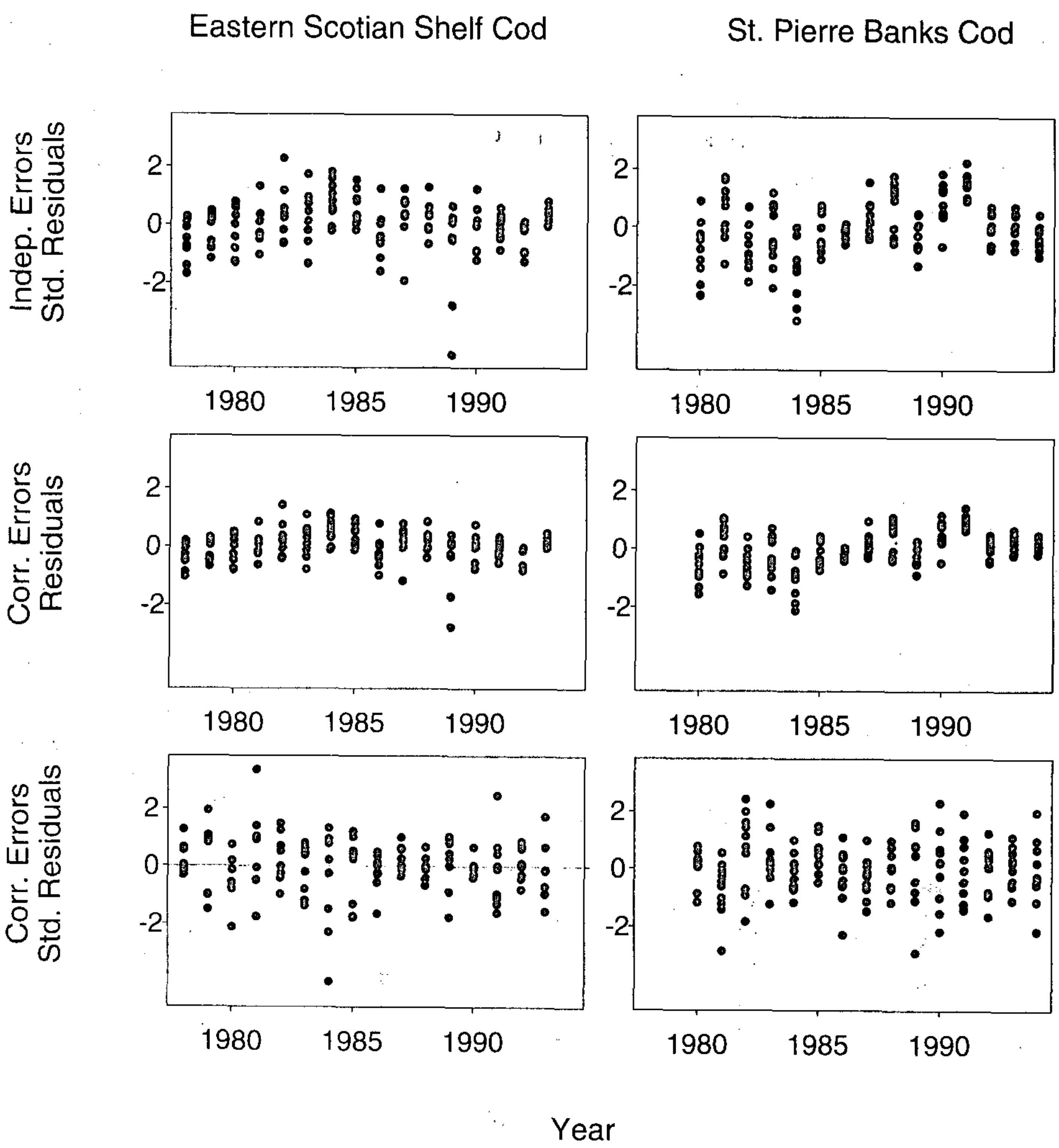

Fig. 2b. Same as Fig. 2a except for Eastern Scotian Shelf and St. Pierre Banks cod.

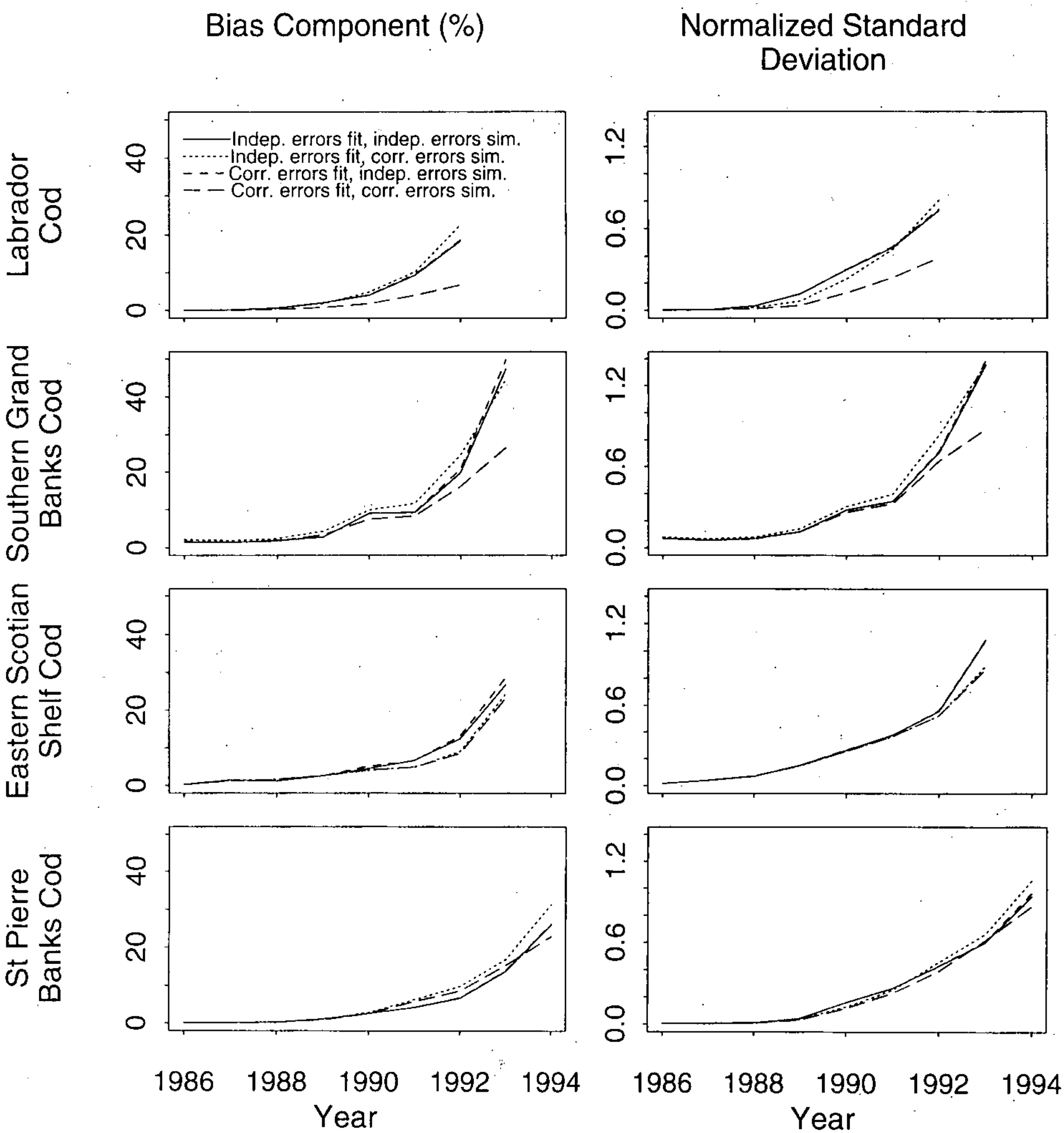

Fig. 3a. Simulated percentage bias and coefficient of variation for independent and correlated error maximum likelihood estimates of recruits at age 3 assuming survey estimates are independent or correlated.

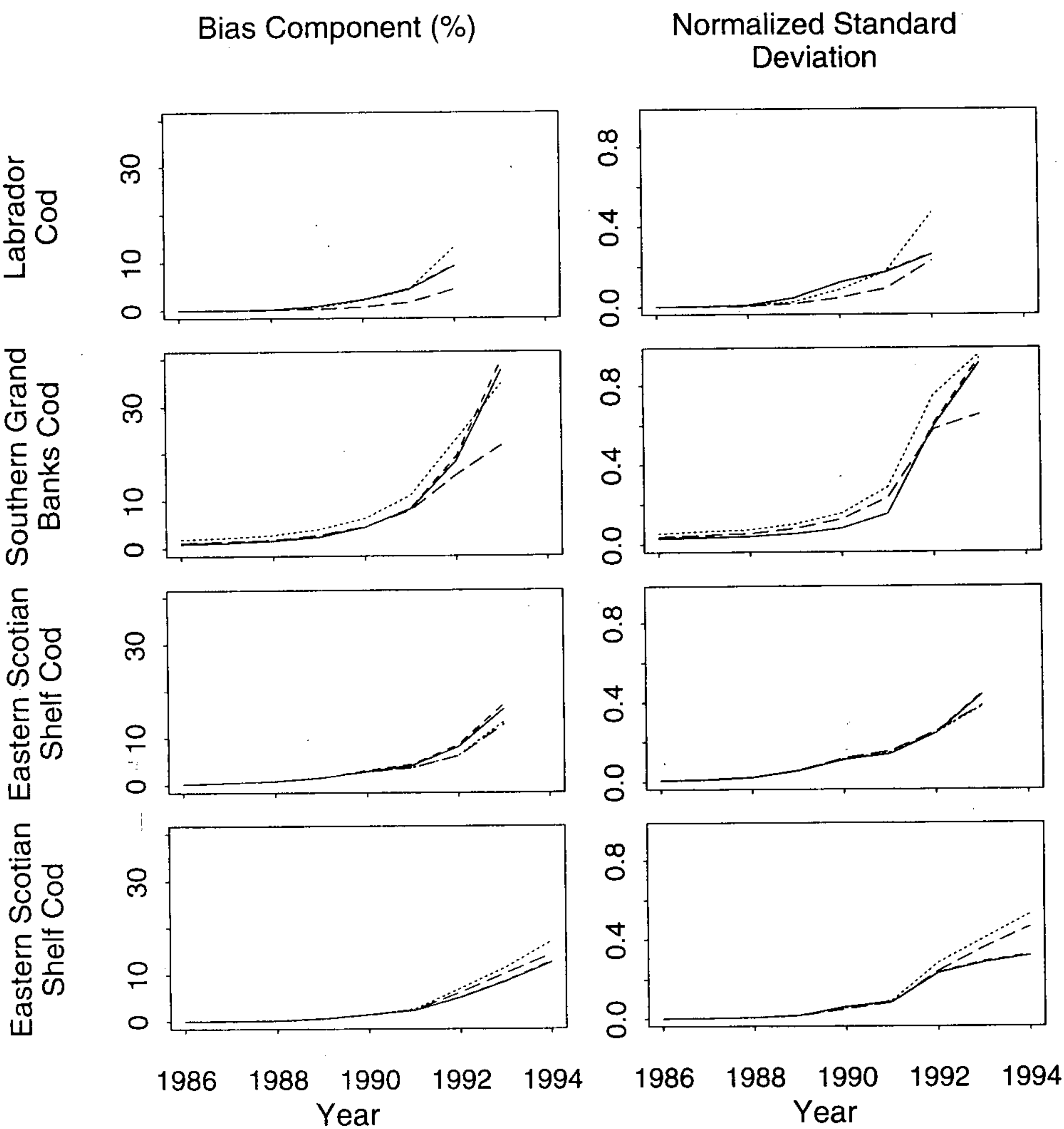

Fig. 3b. Same as Fig. 3a except for total numbers (3+).