ICES J. mar. Sci., 49: 379–387. 1992

Combining information from commercial catches and research surveys to estimate recruitment: a comparison of methods

A. A. Rosenberg, G. P. Kirkwood, R. M. Cook, and R. A. Myers

Rosenberg, A. A., Kirkwood, G. P., Cook, R. M., and Myers, R. A. Combining information from commercial catches and research surveys to estimate recruitment: a comparison of methods. – ICES J. mar. Sci., 49: 379–387.

Three basic methods for estimating year-class strength given several research surveys or commercial catch indices of recruitment are described. Two are regression methods – calibration regression and predictive regression. The third method is factor analysis, in which the covariance between the indices is modelled as a function of the relationship to the underlying true, but unobservable, recruitment. All three of the methods estimate recruitment as an inverse variance weighted average of the estimates from each of the index series.

In many cases, the commercial catch information will yield an estimate of recruitment for years 1 . . . T through some procedure such as virtual population analysis (VPA). This gives an estimate of absolute abundance (AA), while research surveys can give estimates of relative abundance (RA) or absolute abundance of recruits. The regression methods make specific assumptions concerning the relative magnitude of the errors in the estimates from different sources, e.g. that (AA) is most precise (calibration) or of no greater precision than the other indices (prediction). The difficulty arises in choosing between these methods with real data where the relative sizes of the errors in the available estimates of year-class strength are unknown.

Simulation tests were constructed to test the three basic methods as well as a calibration regression where the resulting weighted average includes a term for "shrinking" the estimate towards the mean of the AA series. The tests indicate that factor analysis and calibration with shrinkage perform best overall. Calibration can be quite sensitive to missing data, however, and may break down if the most recent year's recruitment is far from the mean of the AA series. Under these conditions, factor analysis performs better in simulation trials.

Key words: year-class strength, commercial catches, research surveys, simulation tests.

Received 18 January 1990; accepted 29 May 1992.

A. A. Rosenberg, and G. P. Kirkwood: Renewable Resources Assessment Group, Imperial College, 8 Prince's Gardens, London SW7 1NA, England. A. A. Rosenberg present address: National Marine Fisheries Service, Woods Hole, Massachusetts 02543, USA. R. M. Cook: DAFS Marine Laboratory, P.O. Box 101, Victoria Road, Aberdeen AB9 8DB, Scotland. R. A. Myers: Science Branch, Department of Fisheries and Oceans, P.O. Box 5667, St Johns, Newfoundland, Canada A1C 5X1.

Introduction

A vital element of many fish stock assessments is the estimation of year-class abundance. These estimates are used to produce catch forecasts, set quotas, and monitor stock status. A primary source of data for absolute abundance (AA) estimates of year-class strength is the commercial catch, often analysed using virtual population (VPA) or cohort analysis (Gulland, 1965; Pope, 1972). For many North Sea fish stocks additional information is available from annual research surveys carried out by national and international groups which produce indices of relative abundance (RA). VPA estimates of absolute abundance are usually taken to be of high precision compared to survey indices of relative abundance. In practice, however, the relative magnitude of the error terms is unknown. The assumption of higher precision of absolute abundance estimates from VPA or other procedures based on commercial catch data is not assured, particularly in the most recent years due to the convergent properties of the procedure (Pope, 1972; Anon., 1991). In addition, estimates of year-class strength from commercial catch data may only be available up to year T, while research survey estimates often are available for year T+1. Given the key role of year-class strength estimates in management advice, the properties of estimation methods which use all

1054–3139/92/040379+09 $08.00/0

available observations on both AA and RA, and which are robust to data of varying precision and can produce estimates in the most recent year, need to be explored.

The statistical methodology presented in this paper pools the information from AA and RA estimates to calculate the best estimate of year-class strength in the current year. We present three general approaches to the problem: calibration regression, predictive regression, and factor analysis, and present both simulation results and an example from the North Sea haddock fishery to demonstrate the advantages of the various techniques. We conclude that factor analysis provides the most robust estimates of year-class size, but calibration regression with a term to shrink the estimates toward the mean performs nearly as well and is computationally simpler.

The ICES Stock Assessment Methods Working Group (Anon., 1987) has considered the methods discussed here, along with some other related techniques. They compared methods using a real data set. Currently, ICES Working Groups use a program for predicative and calibration regressions written by Dr J. Shepherd of MAFF Fisheries Laboratory, Lowestoft, UK. Other simulation tests and descriptions of methodology for combining information from several sources to estimate recruitment are given in Rosenberg and Kirkwood (1987) and Cook (1987).

The statistical model

In this section we present the basic statistical model for combining sets of indices of year-class strength from several sources, e.g. VPA estimates, survey results, etc. Suppose for T years there are K series of indices of year-class strength:

$$X_{kt};\ k = 1,2,\ldots,K;\ t = 1,2,\ldots T.$$

The data may be incomplete such that all series are not observed in all years. We discuss this problem in relation to the specific estimation procedures described below. However, the development of the model here is valid even for series of different lengths or incomplete series.

The first of the series, X_{1t}, we take to be an index of absolute abundance such as would be obtained from VPA or another procedure using commercial catch data. If X_{1t} is an unbiased index of the true underlying year-class strength, then:

$$X_{1t} = P_t + e_{1t};\ t = 1,2,\ldots,T,$$

where P_t is the true year-class size at time t and the e_{1t} are independent errors with mean zero and constant variance, v_1^2. The $\{e_{1t}\}$ and other error terms in this paper are assumed to be normally distributed. In practice this means transformation of the indices (e.g. working with the log or square root of the series) may be necessary.

The remaining $K-1$ series represent relative abundance indices from the various research surveys. We assume that they are linearly related to abundance:

$$X_{kt} = a_k + b_kP_t + e_{kt};\ t = 1,2,\ldots,T;\ k = 2,3,\ldots,K,$$

where a_k and b_k are constants and the $\{e_{kt}\}$ are independent normally distributed errors with mean zero and constant variance v_k^2. This assumption of independence of the error terms implies that the survey measurement errors are unrelated. This greatly simplifies the analysis, although more complicated models can be accommodated (see, for example, Myers and Cadigan, 1991). In practice, many research surveys use similar gear and sampling protocol, such that the assumptions of independent error terms and constant variance may be untenable. These assumptions need careful inspection as part of the analysis of residuals from the fitted models.

In the current year $(T+1)$ we assume that we also have relative abundance indices:

$$X_{k,T+1} = a_k + b_kP_{T+1} + e_{k,T+1};\ k=2,3,\ldots,K,$$

but no estimate of absolute abundance from VPA or other procedures. Thus, given data $\{X_{1,t},\ t=1,2,\ldots T\}$ and $\{X_{k,t},\ t=1,2,\ldots T+1;\ k=2,3,\ldots,K\}$, we seek to develop estimators of the unknown $\{P_t\}$, especially P_{T+1}.

Estimation

It would appear at first glance from the above equations that it should be possible to derive estimates of all the unknown parameters: the mean (a_k), the slopes (b_k), the measurement error variances (v_k^2), and the true year-class strength (P_t). Unfortunately, this is not possible because these models are in the form of the classical functional regression problem with unknown variances. This problem has been well studied and is known to have no maximum likelihood solution unless the variances (or at least the ratios between them) are known (Kendall and Stuart, 1979; Bard, 1974; Anderson, 1984), which is unlikely to be the case in practice.

We describe here three ways of estimating year-class strength from a set of indices. The first two are classical regression approaches and one of these (calibration regression), modified to include a term for shrinkage toward the mean, is currently used in many ICES Working Group assessments. In applying regression methods P_t is not estimated directly. Rather, the AA series $\{X_{1t}\}$ is related directly to the survey series and this relationship is used to predict $X_{1,T+1}$ from the $\{X_{k,T+1},\ k=2\ldots K\}$ observations.

The third method described here uses the statistical methodology of factor analysis (Lawley and Maxwell, 1971), which retains the true abundances $\{P_t\}$ as unobserved random variables and estimates the year-class strengths by making assumptions regarding the distribution of the P_t.

In the next three subsections these statistical procedures are described. This is followed by simulation results comparing the methods. Finally, a real data example is presented along with comments on the application of this methodology.

Calibration regression estimates

In many assessment analyses, virtual population analysis, or some related procedure, is considered, implicitly, to estimate year-class strengths retrospectively with relatively high precision. Of course, the estimates of AA from VPA are less precise in the most recent year (Pope, 1972; Anon., 1991) and can usually only be made up until year T, while here we focus on the year-class strength in year T+1. However, if AA is well estimated such that so it can be assumed that $v_1^2=0$ and then estimating $X_{1,T+1}$ is equivalent to estimating P_{T+1}. This is a reasonable approach if the measurement error variance of AA is very much smaller than the error variance of the other indices of year-class strength. In this situation, the problem is then to obtain an estimate to fill in year T+1 for the AA series, X_{1T+1}, and a calibration regression approach is appropriate.

Given the assumption that the AA estimate is of higher precision, for each survey series we then consider a model of the form:

$$X_{kt} = a_k + b_k X_{1t} + e_{kt};\ t = 1,2,\ldots T,$$

where the $\{e_{kt}\}$ are independent, normally distributed random variables with mean zero and constant variance Z_k^2. Having obtained estimates of $\hat{a}_k$, $\hat{b}_k$, and $\hat{Z}^2$ from standard theory (Neter *et al.*, 1983), we then need to estimate $X_{1,T+1}$ using an observation $X_{k,T+1}$. The appropriate estimator is:

$$\hat{X}_{1,T+1} = \frac{(X_{k,T+1} - \hat{a}_k)}{\hat{b}_k}.$$

For later pooling of estimates of $X_{1,T+1}$ the variance of $X_{1,T+1}|X_{k,T+1}$ is needed. This can be obtained from the standard method of Taylor series approximation which gives:

$$\mathrm{Var}(\hat{X}_{1,T+1}|X_{k,T+1}) = \frac{\hat{Z}_k^2}{\hat{b}_k^2}\left(1 + \frac{1}{T} + \left\{\frac{(\hat{X}_{1,T+1} - \bar{X}_1)^2}{\sum_{t=1}^{T} (X_{1t} - \bar{X}_1)^2}\right\}\right).$$

To apply the calibration procedure, regressions are performed for each survey series on the AA series and estimates of $X_{1,T+1}|X_{k,T+1}$ are calculated. The estimates can be pooled by taking a weighted average, with inverse variance weighting, over all series. Writing:

$$\hat{V}_k^2 = \mathrm{Var}(\hat{X}_{1,T+1}|X_{k,T+1}).$$

Then:

$$\hat{X}_{1,T+1} = \frac{\sum_{n=2}^{N} (X_{k,T+1} - \hat{a}_k)/(\hat{b}_k \hat{V}_k^2)}{\sum_{n=2}^{N} 1/\hat{V}_k^2}.$$

Note that, because of the assumption that the measurement errors in the surveys are independent identically distributed random variables, the problem of incomplete series, e.g. missing data, does not affect this procedure but is simply accounted for in the regressions where T (the effective number of data points) may not be the same for each survey.

It has been suggested (Anon., 1987) that, when using this type of calibration procedure, information on recruitment contained in the mean of the VPA of other estimates of absolute abundance should be incorporated in the estimate of year-class strength in year T+1. In other words, that there should be "shrinkage" toward the mean AA estimate of recruitment. One simple form for such a shrinkage term is to include in the weighted sum above a term for the mean of the AA series weighted by its inverse sample variance.

Predictive regression estimates

The critical assumption made for the calibration approach described above is that the measurement error variance of the estimates of absolute abundance, v_k^2 is small relative to the other measurement error variances. In some assessments this assumption is untenable, or at least suspect. If one of the research survey series of estimates of RA, for example, was more precise, then the appropriate statistical procedure would be to use a predictive regression of X_{1t} on each of the X_{kt} to estimate abundance. Even if nothing were known about the relative sizes of the error variances, then some statisticians have argued that if prediction is the aim of the analysis, predictive regression should be used (Krutchkoff, 1967), although this argument has been criticized (Lwin and Maritz, 1982).

The procedure for using predictive regression to estimate year-class strength is to perform a set of simple predictive regressions of X_{1t} on each of the available X_{kt} and then combine the estimates of $X_{1,T+1}|X_{k,T+1}$ with inverse variance weighting as with the calibration approach.

For K series the model is:

$$X_{1t} = a_k + b_k X_{kt} + e_{kt},$$

for $t = 1,2, \ldots, T$, where the e_{kt} are distributed as independent normal random variates with mean 0 and variances Z_k^2. From standard regression analysis (Neter *et al.*, 1983) estimates of $\hat{a}_k$, $\hat{b}_k$ and $\hat{Z}_k^2$ can be obtained, and:

$$\bar{X}_{1,T+1}|X_{n,T+1} = \hat{a}_n + \hat{b}_n X_{n,T+1}$$

The variance of the estimates is given by:

$$\mathrm{Var}(\hat{X}_{1,T+1}|X_{n,T+1}) = Z_n^2\left\{1 + (1/T) + \frac{(X_{n,T+1} - \bar{X}_n)^2}{\sum_{t=1}^{T+1} (X_{nt} - \bar{X}_n)^2}\right\}.$$

The final pooled estimate is given by:

$$\hat{X}_{1,T+1} = \frac{\sum_{n=2}^{N} (\hat{X}_{1,T+1}|X_{n,T+1})/\mathrm{Var}(\hat{X}_{1,T+1}|X_{n,T+1})}{\sum_{n=2}^{N} 1/\mathrm{Var}(\hat{X}_{1,T+1}|X_{n,T+1})}.$$

Factor analysis

In practice, for the majority of fish stocks it may not be possible to determine *a priori* which regression method is most appropriate for estimating year-class strength because the relative magnitude of the measurement error variances of each series is unknown. Choosing between calibration and prediction regression methods is therefore problematic and the decision is usually made in a rather arbitrary manner.

An alternative procedure presented here—factor analysis—circumvents this problem by making additional assumptions about the underlying abundance. For the factor analysis procedure the true year-class strengths, P_t, are assumed to be independent identically distributed normal random variables with mean m and variance ψ^2. In statistical terminology we treat the problem as one of structural as opposed to functional regression (see Kendall and Stuart, 1979, or Anderson, 1984, for details).

The model equations for factor analysis are:

$$X_{1t} = P_t + e_{1t};\ t = 1,2, \ldots T$$

$$X_{kt} = a_k + b_k P_t + e_{kt};\ t = 1,2, \ldots T;\ k = 2,3 \ldots K.$$

In practice, the a_k can be adequately estimated by the sample means and it is easier to work with the deviations from the means writing:

$$Y_{kt} = X_{kt} - \bar{X}_k = b_k P_t + e_{kt},$$

where $\bar{X}_k$ is the sample mean of series k.

As with the other methods the first series X_{1t} is taken as an index of absolute abundance such as VPA and its mean is retained to scale the P_ts. However, v_1^2 is not assumed *a priori* to be smaller than the other measurement error variances. The unknown parameters to be estimated are $b_2, \ldots, b_k$ and $v_1^2, v_2^2 \ldots v_K^2$ and ψ^2.

Estimation

The model equations express linear dependencies of each of the series on the underlying true year-class strength so it follows that the covariances between the observed survey data series result from this underlying factor. Factor analysis fits the expected covariances between the observed series $(X_{1t}, Y_{2t} \ldots Y_{k,t})$ based on the model parameters to the sample covariances calculated from the data. In matrix notation with $B = (1, b_2 \ldots b_K)'$ and $V^2 = (v_1^2, v_2^2 \ldots v_K^2)$, the expected covariance matrix between the series of year-class strength estimates (the Y_{kt}) R is:

$$R = B\psi^2 B' + V^2.$$

If we write the sample covariance matrix of the observed series as S with element s_{ij} for the covariance of series i with series j; $i = 1 \ldots N$; $j = 1 \ldots N$, then the log likelihood function is given by:

$$L = -0.5T\left\{\ln|R| - 0.5T \sum_{i=1}^{N}\sum_{j=1}^{N} (s_{ij} r_{ij}) - \ln|S| - N\right\}.$$

While it is possible to obtain maximum likelihood estimates for this factor analysis model (see Lawley and Maxwell, 1971; Bollen, 1989), the resulting estimator can be sensitive to departures from normality and numerical singularities in the R matrix such as those arising from small sample sizes. In the problem considered here, we usually have very small sample sizes available to estimate the sample covariances, and since the series are highly correlated the problems of numerical singularity can be severe. Because of this it is often easier to work with a simple least squares criterion for estimating the model parameters. This requires minimizing:

$$5\mathrm{tr}(S - R)^2,$$

where tr refers to the trace of the matrix, i.e. the sum of the diagonal elements.

The least squares approach is a simple objective function which does not depend upon a distributional assumption and does not require the inversion of the R matrix. While we forego the statistical properties of maximum likelihood, these are primarily large sample properties which are less relevant for the problem considered here. A direct search algorithm for function minimization can be used to obtain parameter estimates. It is usually necessary to constrain the minimization such that the variance parameters are greater than zero. An additional advantage, discussed below, of the simple least squares criterion is that it can be easily modified to accommodate incomplete data.

Once estimates of B, V^2, and ψ^2 are available the P_t can be estimated from:

$$P_t = \frac{x_{1t}/v_1^2 + \sum_{k=2}^{K} b_k Y_k/v_k^2}{1/v_1^2 + \sum_{k=2}^{K} b_k^2/v_k^2}.$$

This is simply an inverse variance weighted sum as used in the other methods. To estimate P_{T+1} where $X_{1,T+1}$ is missing we use:

$$P_{T+1} = \frac{\sum_{k=2}^{K} b_k Y_k/v_k^2}{\sum_{k=2}^{K} b_k^2/v_k^2}.$$

Example with three series

The factor analysis method is not usable for less than three observed series. This is because there is insufficient information available to estimate all the parameters. To see this and illustrate the method consider the case where there are three observed series, i.e. AA and two RA series of estimates of year-class strength in each of a series of years. The model equations are:

$$X_{1t} = P_t + e_{1t},$$

$$Y_{2t} = b_2 P_t + e_{2t},$$

$$Y_{3t} = b_3 P_t + e_{3t}.$$

Now, the covariance matrix of these three measures of year-class strength in terms of the model parameters (i.e. R as a function of the slopes, b_k, and the variances of the e_{kt}, v_k^2) is obtained by taking cross products. The lower triangular covariance matrix is then:

$$\begin{matrix} \psi^2 + v_1^2, & & \\ b_2\psi^2 & b_2^2\psi^2 + v_2^2, & \\ b_3\psi^2 & b_2 b_3\psi^2 & b_3^2\psi^2 + v_3^2. \end{matrix}$$

Equating this expected covariance matrix in terms of the model parameters to the observed sample covariance matrix and solving we obtain:

$$\psi^2 = S_{13}S_{12}/S_{23},$$

$$b_2 = S_{12}/\psi^2,$$

$$b_3 = S_{13}/\psi^2,$$

$$v_1^2 = S_{11} - \psi^2,$$

$$v_2^2 = S_{22} - b_2^2\psi^2,$$

$$v_3^2 = S_{33} - b_3^2\psi^2.$$

For three observed series then there are six equations to solve for six unknown parameters. If only two observed series were available there would be only three equations for four unknown parameters. If there are more than three series then there is no longer a simple analytical solution for the estimators and a numerical solution must be sought as described above. However, when more than three series are available, additional parameters, such as covariances between the measurement errors (off diagonal elements in the V^2 matrix), can potentially be estimated.

Given the parameter estimates in the three variable cases it is a simple matter to compute the inverse variance weighted estimates of the year-class strength at any time, t.

One further point on this methodology needs to be clarified. While the simple regression methods described above could be used straightforwardly when some of the data points were missing, factor analysis requires some modification. Missing data essentially correspond to using year-class strength index series that are of different lengths. This means that the elements of the sample covariance matrix S will not each be estimated from the same number of data points, and the subsequent least squares estimator which compares S to R needs to be weighted accordingly. The appropriate weighting comes from the estimate of the variance of the sample covariances (Kendall and Stuart, 1979) such that we minimize:

$$0.5 \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(s_{ij} - r_{ij})^2,$$

where the s_{ij} and r_{ij} are the elements of the respective matrices and the w_{ij} are given by:

$$\frac{\eta_{ij}^2}{\eta_{ij} - 1} \sim \eta_{ij} + 1,$$

with η_{ij} equal to the number of times both series i and j are measured.

Simulation tests

To compare the accuracy and precision of the three methods described above, we have performed tests on simulated data which consider the effects of:

1. Different patterns of measurement error variance.
2. Missing values in some of the series.
3. Inclusion of one series of very poor data in the overall analysis.

The simulated data series of recruitment over 10 years was generated as observations from a normal random variable with a mean of 100 and a coefficient of variation of 0.45. Indices of recruitment were then constructed by adding measurement error to this underlying true recruitment series. The measurement error for each index was a normal random variate with mean zero and variance v_k^2 for $k = 1 \ldots K$.

As a test of the ability of a method to estimate year-class strength, the 11th year's true recruitment was generated either around the mean recruitment or around a mean recruitment which was one or two standard deviations above mean recruitment of the first 10 years. Indices for series 2 . . . K were generated accordingly. It is important to examine the ability of the methods to estimate points far from the mean because detecting changes in incoming recruitment can be critical for management.

Sets of 1000 runs were made for each test and the bias and mean squared error (MSE) of the estimated recruitment compared to the true simulated recruitment was computed.

We chose three combinations of measurement error variances. These correspond to the assumptions made for the various methods. When all the measurement variances are equal there should be little difference between the methods. When the VPA series has lower measurement error than the other two indices, calibration regression should perform best. When VPA has higher error variance than the other series, prediction regression should perform well. These three types of error structures have been used below (Figs 1,2).

Results

The results for all of the simulation tests are summarized in Figures 1 and 2. In these figures, the case numbers down the left hand side refer to the various tests: 11th year's recruitment is distributed: (1) around the mean; (2) one standard deviation above the mean; (3) two standard deviations above the mean; (4) one standard deviation above the mean, but with missing data in the surveys; (5) one standard deviation above the mean, but with an additional low precision survey index. Across the top is the measurement error pattern as described above. Figure 1 shows the bias (as a measure of accuracy) for each method under each scenario. A striking feature is that calibration regression without a term for shrinkage towards the mean is usually positively biased. This method has low accuracy when the VPA is not the most precise measure of abundance, i.e. the estimates are not robust to violation of the underlying assumptions of the method. Calibration regression with no shrinkage towards the mean does give the most accurate results when the error structure conforms to the method (VPA most precise) and there are no missing data. The other methods all tend toward the mean and so underestimate recruitment when a large year-class occurs.

Factor analysis is more accurate than the other methods in most of the tests, particularly when data are missing. However, overall the differences between factor analysis and calibration regression with shrinkage toward the mean are small.

Figure 2 gives the root mean square error of the predictions of recruitment (as a measure of precision) for each of the scenarios simulated. Calibration regression without a term for shrinkage toward the mean had low precision when the assumptions for that method were violated. The other three methods had similar precision across all the scenarios. Factor analysis appears to be more robust to missing data than predictive regression or calibration with shrinkage. This was the only apparent difference between factor analysis and calibration with a shrinkage term. These two methods performed slightly better than predictive regression overall.

Application to data for North Sea haddock

Simulation is the only way of investigating the effect of different error structures on the performance of the methods since, in the real world, the magnitude and distribution of the measurement errors associated with the indices will not be known. Nevertheless, it is instructive to apply the methods to a set of real data in order to gauge the disparity of performance. All four methods of estimation were applied to the data for North Sea haddock (Anon., 1988), which includes two research survey indices of recruitment and VPA estimates for the period 1970–1984, to estimate the size of the 1982–1984 year-classes at age 1.

Following the convention of the ICES North Sea Roundfish Working Group which regularly assesses this stock, data have been log transformed for the analysis.

The results (Table 1) mirror the pattern of the simulation results. That is, predictive regression, calibration with a shrinkage term toward the mean, and factor analysis give similar results. Calibration without the shrinkage term appears to be more variable, giving lower estimates of small year-classes and higher estimates of the large 1983 year-class. The calibration regression with a shrinkage term toward the mean stabilizes these estimates and gives results close to the VPA estimates of these year-classes made several years later with the benefit of hindsight.

Factor analysis gives similar estimates to predictive regression. In the simulation studies this was often indicative of the underlying error structure. Factor analysis tends to perform like a predictive regression when the precision of the VPA is lower than the other indices and it performs like a calibration regression when the precision of the VPA is higher than the other indices. For this example, there is some indication that the predictive regression model is the most appropriate.

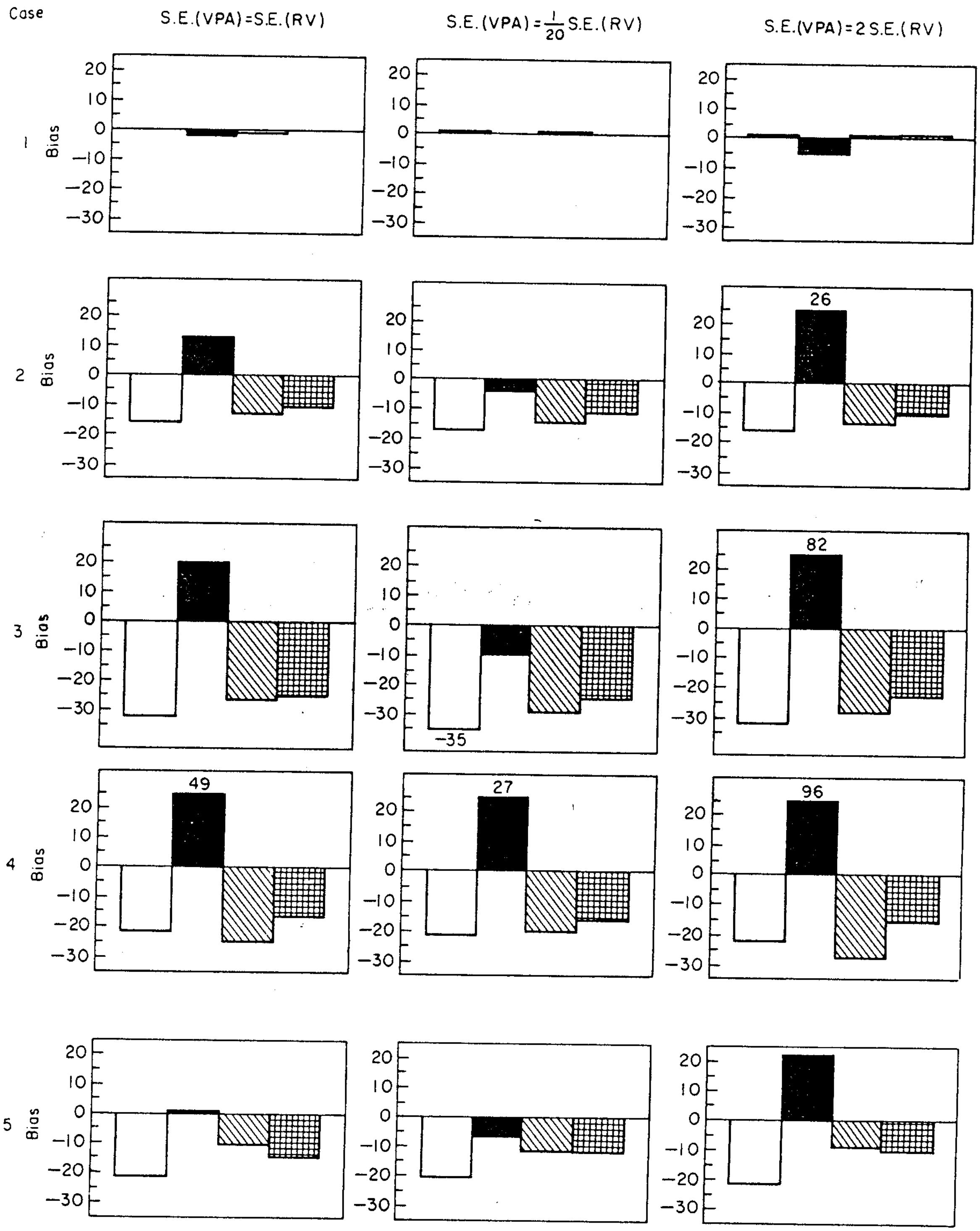

Figure 1. Simulation test results. Bias in estimates of year-class strength made from AA and RA series for five cases under three different measurement error structures. The numbers down the left hand side refer to the case numbers as described in the text. The measurement error structures are indicated across the top of the figure where the standard error of the measurement errors for the AA series is given relative to that for the RA series. In case 5, the additional survey series has a measurement error standard error three times the other surveys. The four methods tested are simple predictive regression (open bars), simple calibration regression (solid bars), calibration regression with a term for shrinkage toward the mean (diagonal lines), and factor analysis (crosshatched). If the bias is greater than the indicated range, the value is indicated above or below the bar. Each value is obtained over 1000 simulation runs.

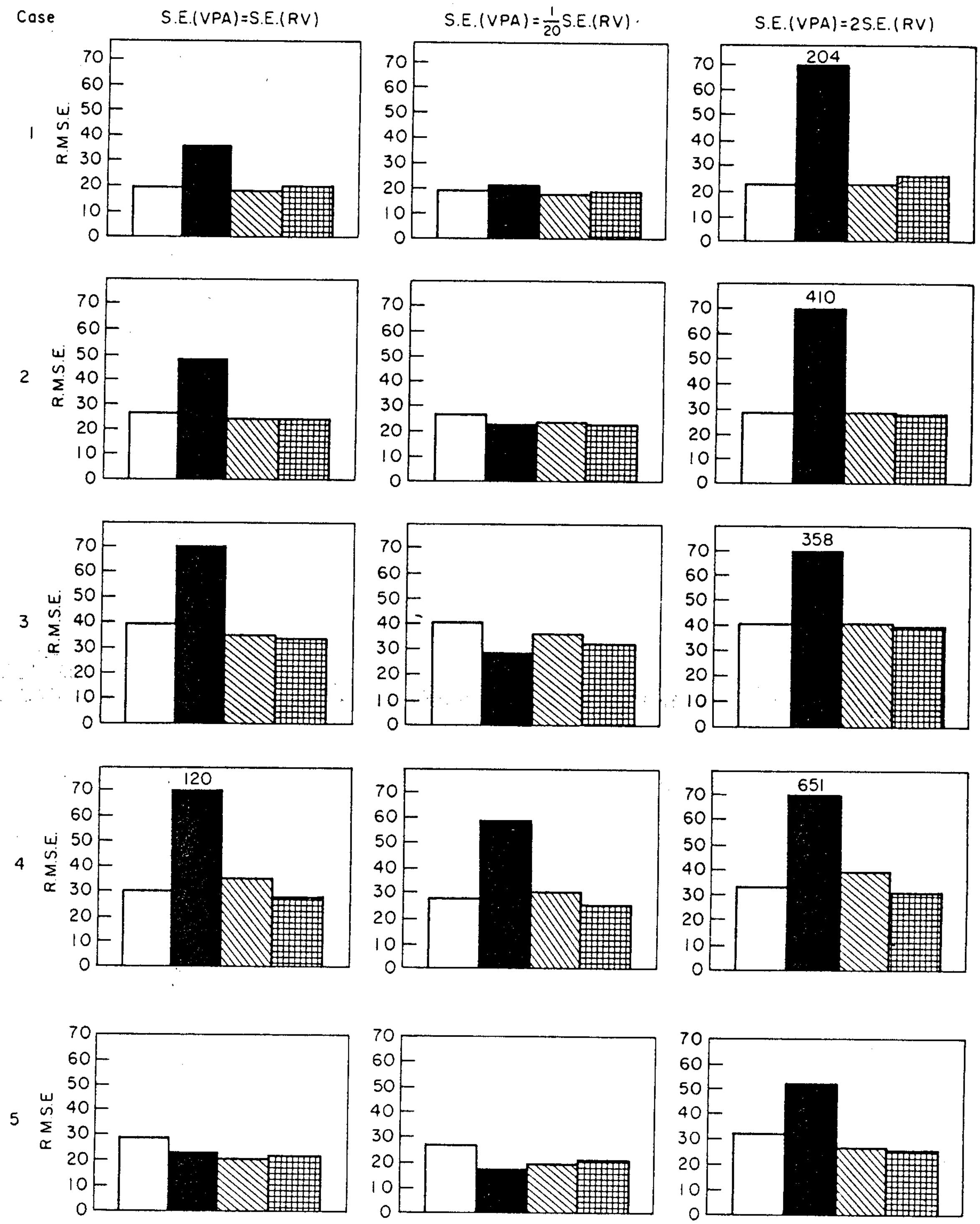

Figure 2. Simulation test results. Root mean square error in estimates of year-class strength made from AA and RA series of estimates. Cases and legend is as for Figure 1.

Table 1. Results for four estimation methods applied to the data for North Sea haddock (Anon., 1988). Estimates for 1982–1984 year-classes at age 1 are given, along with VPA results for these years from the 1988 assessment.

Method	1982 year-class	1983 year-class	1984 year-class
Prediction	2864	8778	2123
Calibration	2566	11 499	1755
Calibration w shrinkage	2807	8955	1978
Factor analysis	3103	8778	2345
Mean VPA (1970–current)	5893	5621	5795
VPA estimate from 1988 assessment	2336	7893	2033

Discussion

Of the four methods tested here, factor analysis and calibration with a term for shrinkage toward the mean seem to be the best overall in terms of both accuracy and precision. For a given structure of measurement error, predictive, or calibration regression may outperform factor analysis. With real data, however, it is not possible to partition the variance of the index series *a priori* into measurement error versus real underlying variability in recruitment. In other words, it is never truly clear whether the structure corresponds to a calibration or a prediction situation. If the wrong choice is made between these two methods the estimates can be very inaccurate and very imprecise.

Factor analysis seems to provide a middle-ground between calibration and prediction. In a calibration situation it is nearly as good as calibration regression and in a prediction situation it is nearly as good as predictive regression. This means that it can be used to indicate the type of measurement error structure in the data. In fact, estimates of the measurement error variance can be obtained directly from factor analysis.

Using shrinkage toward the mean in a calibration regression is a good option in many situations. When the year-class strength is far from the series mean shrinkage can be a poor option for obvious reasons. However, even in this case, the benefits of stabilizing the calibration estimates seem to outweigh the problem of increased bias. Missing data also adversely affects both the calibration procedures. This is worrying since index series of different lengths are probably the rule rather than the exception. Calibration regression with shrinkage does have a large advantage over factor analysis in that the computations are simple.

Acknowledgements

We gratefully acknowledge the support of UK MAFF Chief Scientists Group for this research. Dr John Shepherd commented and advised on the work throughout. G.P.K. participated while an academic visitor at RRAG, Imperial College, and we thank CSIRO for their support. We acknowledge M. Bravington for technical assistance, M. Carlo for providing data, and Al B. Tross for incentive to complete the paper.

References

Anon., 1987. Report of the Working Group on Methods of Fish Stock Assessment. ICES CM 1987/Assess: 24, 106pp.

Anon., 1988. Report of the North Sea Roundfish Working Group. ICES C.M. 1988/Assess: 21.

Anon., 1991. Report of the Working Group on Methods of Fish Stock Assessments. ICES C.M. 1991/Assess: 146pp.

Anderson, T. W. 1984. Estimating linear statistical relationships. Annals of Statistics, 12: 1–45.

Bard, Y. 1974. Nonlinear Parameter Estimation. Academic Press, New York. 314pp.

Bollen, K. A. 1989. Structural Equations with Latent Variables. Wiley & Sons, New York.

Cook, R. M. 1987. A note on the use of research survey indices in VPA-based assessments. ICES CM 1987/G: 5, 5pp.

Gulland, J. A. 1965. The estimation of mortality rates. Annex to Arctic Fish Working Group ICES. ICES CM 1965/G: 3, 10pp.

Kendall, M. G., and Stuart, A. 1979. The Advanced Theory of Statistics, Vol. II, 4th edn. Griffin, London.

Krutchkoff, R. G. 1967. Classical and inverse methods of calibration. Technometrics, 9: 525–539.

Lawley, D. N., and Maxwell, A. R. E. 1971. Factor Analysis as a Statistical Method, 2nd edn. Butterworths, London.

Lwin, T., and Maritz, J. S. 1982. An analysis of the linear calibration controversy from the perspective of compound estimation. Technometrics, 24: 235–242.

Myers, R. A., and Cadigan, N. 1991. The reliability of recruitment time series. ICES Research Doc. D.M. D: 34, 11pp.

Neter, J., Wasserman, J. W., and Kutner, M. H. 1983. Applied Linear Regression Models. R. D. Irwin, Homewood, Illinois.

Pope, J. 1972. An investigation of the accuracy of virtual population analysis using cohort analysis. Tes. Bull. Int. Comm. Northw. Atl. Fish., 9: 65–74.

Rosenberg, A. A., and Kirkwood, G. P. 1987. Integrated analysis of research survey and commercial catch data. Contract Report, MAFF Chief Scientists Group Contract CSA 1135, 36pp.