PN-ABH-807
71312

WORKING PAPER SERIES

FISHERIES STOCK ASSESSMENT

TITLE XII
Collaborative Research Support Program

FSA/CRSP
Management Office
1109 Symons Hall
College Park, MD 20742

In cooperation with the United States Agency for International Development (Grant No. DAN-4146-G-SS-5071-00) the Fisheries Stock Assessment CRSP involves the following participating institutions:

The University of Maryland—Center for Environmental and Estuarine Studies
The University of Rhode Island
The University of Washington
The University of Costa Rica
The University of the Philippines—Marine Science Institute (Diliman)
—College of Fisheries (Visayas)

In collaboration with The University of Delaware; The University of Maryland—College Park, College of Business and Management; The University of Miami; and The International Center for Aquatic Resources Management (ICLARM).

-1

PN-ABH-807

Working Paper Series

Working Paper No. 1
"An Empirical Approach to Multispecies
Stock Assessment"
by

Saul B. Saila, Chief Scientist
Karim Erzini, Research Assistant
Graduate School of Oceanography
University of Rhode Island

Fisheries Stock Assessment
Title XII
Collaborative Research Support Program

The Fisheries Stock Assessment CRSP (sponsored in part by USAID Grant No. DAN-4146-G-SS-5071-00) is intended to support collaborative research between U.S. and developing countries' universities and institutions on fisheries stock assessment and management strategies.

This Working Paper has been produced by the University of Rhode Island research component in collaboration with the University of the Philippines and in association with the International Center for Living Aquatic Resources Management. Additional copies are also available from:

International Center for Marine Resource Development
The University of Rhode Island
126 Woodward Hall
Kingston, Rhode Island 02881

Abstract.-A method for the analysis of observed changes in catch rates by similar species or groups from southeast Asia and the northwest Atlantic region is developed and described. The aggregated time series of catches per standard tow from research vessel surveys were examined as Markov chain models after transition probability matrices were derived from them using quadratic programming. The matrices were projected to provide some insight about stability and changes in the multispecies assemblages. Results from these analyses generally support the inferences previously made from the data and provide some additional information regarding stability.

NOTE

This manuscript has been submitted to the "Transactions of the American Fisheries Society" for publication. It represents the initial phase of work on an empirical multispecies model. This work is being continued and further testing of the model with Thai and Indonesian trawl data is in progress.

We believe that this empirical approach to multispecies stock assessment represents a viable technique for the preliminary analysis of research vessel survey data which has been carried out over time.

Comments and suggestions for further work are encouraged.

Species composition of fish assemblages tends to vary in both space and time. Examples of dramatic temporal changes include transitions in relative species abundance in newly impounded lakes and reservoirs, fluctuations in relative abundance in farm-pond populations of mixed species, and changes in large marine fish communities, as exemplified by demersal fish on Georges Bank in the Northwest Atlantic and tropical demersal fish in the Gulf of Thailand. The rate and nature of change are functions of external and internal factors which influence different species populations in different ways.

The objectives of the study were to describe briefly and illustrate with applications an empirical model for a multispecies fishery. More specifically, we reviewed some of the applications that have been made of multispecies fishery models, including Markov models. Then we provided a more detailed explanation and application of a Markov model of fish community dynamics. This model was based on aggregated time series of catch-per-unit-effort (CPUE) data from surveys by research vessels. The model was then applied to multispecies fisheries, with the goal of obtaining information on their short-term dynamics by projection of the multispecies transition probability matrix.

Background and Method.-Ours is but one of several studies which have attempted to address multispecies fishery problems. Mercer (1979) edited a review of multispecies approaches to management advice, and Saila and Roedel (1980) also included some of the early background for tropical multispecies management. The volume edited by Pauly and Murphy (1982) represents a good summary of the recent status of tropical multispecies models. The proceedings of the workshop at the Bedford Institute of Oceanography, edited by Mahon (1985) reviews approaches to

including fisheries interactions in management advice, and includes a paper on virtual population analysis of multispecies assemblages, sometimes called legion analysis. Several important contibutions to multispecies systems have been described recently in the report of the Dahlem workshop on Exploitation of Marine Communities (May 1984). Pope (1979) has contributed significant information to tropical multispecies models as well as to extensions of age-structured assessment models.

With respect to Markov process applications in fisheries, perhaps the earliest was by Riffenberg (1969) for a multispecies fishery. Thereafter, Saila (1982) provided some background for Markov model use for fisheries applications, Getz and Swartzman (1981) utilized a Markov submodel for recruitment, and Rothschild and Mullen (1985) considered the stock-recruitment relationship in the context of a first-order Markov process.

In order to make effective use of a Markov process model for a multispecies application, time-ordered observations that reflect the movements or changes in species composition must generally be available. That is, we should know in each time period the number of species moving from one classified outcome state to another. A tally matrix of the number of transitions by species (or species groups) could then be made and used to generate a transition probability matrix (TPM).

Unfortunately, what is available from fishery research surveys is aggregated data. That is, we are given the proportions (weights or numbers) by species or species groups in each sample taken by a standardized piece of fishing gear fished for similar periods of time. It is fortunate, however, that there are ways (albeit somewhat tedious) by which a TPM can be estimated from aggregated data. Methods include

introducing prior information into a linear model as described by Cooke (1982) for plant succession, and techniques developed by Lee et al. (1970) for economic time series analysis, which include quadratic programming and restricted least squares. The latter approach has been utilized for our application, but because it has not been described previously for aggregated fisheries data, it will be examined briefly.

Our primary problem involves the estimation of transition probabilities of a Markov process from sets of observations on one-dimensional frequency distributions. This can be illustrated with some fishery data utilized in this paper. Consider, as an example, the average annual catch composition of major species or species groups caught by trawl surveys in the Gulf of Thailand as shown in Table 1.

The fish community is considered as a collection of species or species groups (units), each of which is changing independently according to a finite Markov process. By this is meant a stochastic process whereby the outcome at any particular time $t(t=1,2,3,\ldots T)$ depends only on the outcome of a previous time $(t-1)$, with this dependency the same at all stages in the time sequence. The proportion of fish in species (state) i at time t is $s_i(t)$. The state of the community as a whole is described by a vector of proportions $(s_1(t), s_2(t), \ldots s_r(t))$, assuming r states, which are the major species or species groups in our case. As shown in Table 1, the vector of species or species groups is known at a number of equal time intervals (eight years in this illustration) of which seven were utilized for analysis. However, there is no information on the sequence of states or relative proportions through which the individual

catches-per-unit-effort pass. In this situation we have only aggregated data regarding the process.

Lee et al. (1970) described various estimators of transition probabilities from aggregated data. These include unrestricted and restricted least squares, generalized least squares, minimum chi-square, maximum likelihood, Bayesian and absolute deviations estimators. A restricted estimator refers to limitions $0 \leq p_{ij} \leq 1$ on elements of the TPM. An explanation which is included in the Appendix provides a summary which attempts to describe the statistical model used and the development of some of the estimators. The interested reader is referred to this and Lee et al. (1970) for further details.

Once the transition probability matrix has been derived for a given time series, it can be analyzed as a Markov chain model by methods of matrix algebra. The elements of the derived TPM are interpreted as probabilities that a given species or species group from the total assemblage will be represented by another of its own kind or by some other after a unit time interval. The transition probability matrix is a positive, square, stochastic matrix. It is stochastic because all elements sum to unity by rows. Matrix theory as illustrated by Cox and Miller (1965) has shown that the greatest characteristic root (eigenvalue) of such a matrix has value unity. It is further known that a characteristic root is associated with a characteristic (stationary) vector. The characteristic vector associated with the greatest characteristic root provides a maximum likelihood estimate of the asymptotic (equilibrium) probability distribution of the process. Another way to derive the equilibrium probability distribution vector is to multiply the vector by the transition probability matrix a large

number of times. Prior to convergence to the stable equilibrium probability distribution vector, the dynamics of the system are determined by the eigenvalue spectrum. Transient dynamics can be examined by considering the subdominant eigenvalues. Caswell (1986) has demonstrated that the convergence of a projection matrix to a stable equilibrium is more rapid, the larger the dominant eigenvalue is in relation to the other eigenvalues. He has defined a damping ratio:

$$d = \lambda_1/\lambda_2$$

when λ_1 is the largest eigenvalue and λ_2 is the second largest (in magnitude) eigenvalue.

The equilibrium vector effectively indicates the percent of time that the process is in each state. It is directly related to the mean recurrence time, which measures the length of time taken for the process to return to its present state, once it has left it. The mean recurrence time to state j is $1/s_j$, where s_j is the jth entry from the equilibrium vector.

The mean first passage time f_k is a function whose value is the number of steps required before entering s_k for the first time after moving from its present state. This value is always finite for a process such as described herein. The matrix M of first passage times can be calculated as follows:

$$M = (I - Z + E\,A_{dg})D$$

where the matrix E has all entries one, D is a diagonal matrix with the jth entry $1/s_j$ (as shown for the mean recurrence time). Z is the fundamental matrix: $Z = (I - P + A)^{-1}$, where P is the transition probability matrix and A is the limiting matrix. See Kemeny and Snell (1966) for details regarding mean first passage and recurrence times.

Our interpretation of the properties of the TPM involves projecting successive changes in the multispecies assemblage. By projection we mean determining what would happen to the assemblage if the present conditions were to be maintained over a protracted period. That is, we are trying to establish some relationship between present conditions and the species or groups experiencing these conditions. It is recognized that there are limitations to this method, especially with the data at hand. Two of the major limitations are that this assemblage must be classified into a relatively small number of groups (species) when the time series is fairly short, and it is assumed that the transition probabilities do not change with time.

The equilibrium vector indicates the average proportion of time spent in each state in the long term. The concept of equilibrium is considered to be statistical in nature for the assemblage and dynamic for individual species or species groups. In our case this vector does not imply, for example, that there are no changes in CPUE values among species. On the contrary, the stochastic concept of equilibrium explicitly requires movement into and out of each state. However, on average forces acting to increase CPUE for a given species are exactly counter-balanced by forces tending to decrease it. High values of specific elements of the equilibrium vector indicate relatively high proportions of time spent in that state.

<u>Results and Discussion</u>.-The properties of the transition probability matrix described above are interpreted in a fishery management context in this study. The analysis of the aggregated time series data to derive transition probability matrices and the results of related

statistics have been obtained for four distinct data sets. These include the following:

(1) Gulf of Thailand annual demersal trawl survey data from 1966-1972;

(2) Gulf of Maine annual groundfish trawl survey data from 1963-1973;

(3) Northwest Atlantic annual groundfish trawl survey data from 1967-1974;

(4) Northwest Atlantic annual groundfish trawl survey data from 1975-1982.

All of the above consist of research vessel cruise data in which the catch-per-unit-effort by species or species groups is expressed in kilograms per standard tow. The selection of data sets was based on availability and a desire to include a comparative study of results between tropical and boreal fish assemblages. In all of the above data sets an additional year of data was excluded from analysis to permit comparison of the projected values with observational data.

The Gulf of Thailand is regarded as an example of a tropical multispecies fishery that is very highly exploited and for which an excellent survey record exists. The Northwest Altantic stratified mean-catch-per-tow data systematically gathered by the U.S. National Marine Fisheries Service represents some of the best available multispecies fishery data in north temperate regions. Generally, the Gulf of Maine area was regarded as a less highly exploited region than Georges Bank and other parts of the Northwest Atlantic during the 1960s and early 1970s. Thus, another broad within-regions comparison seemed possible. The Northwest Atlantic data set was divided into two time

series--one prior to the enactment of the Fishery Conservation and Management Act when exploitation was predictably high and another for a similar period after the Act was invoked. This was done to provide a further comparison of results which might reflect some changes in the fishery.

The methodology was programmed for a microcomputer in BASIC in order to maximize its utility and portability in the developing world. This created some limitations in the size of data sets and increases in computer time involved. For example, initially we were limited to a maximum of six species and 18 years due to the size of the sparse matrix to be inverted. Other limits to the methodology are that the time series must be taken at approximately equal time intervals, and clearly the time series cannot be smaller than the number of species or species groups included. In practice it was found that three to six species or species groups could be handled reasonably from a computational and interpretational point of view. Therefore, only the most significant species or species groups (in terms of yield) were selected for analysis. No particular attention was given at this time to determining prior ecological or technological interactions between or among the selected species.

The raw data used for analysis, the results of the fit of the model to the data and step-ahead predictions, the transition probability matrix, the eigenvalues of the transition probability matrix, the steady state probability vector, and a matrix of expected first passage and recurrence times are illustrated for each of the four data sets.

Gulf of Thailand.-Table 1 represents the portion of the Gulf of Thailand data which was analyzed in this report. Actual weights caught

per standard tow can easily be derived by multiplying the total weight caught per tow (element in last column) by the respective row elements which are expressed in decimal fractions. Note that the fitting procedure does not include the first row of observational data but does include a one-step ahead prediction by species or species groups. The chi-square statistic for testing the fit of the observed to the calculated data matrix was 9.175, which was tested with (n-2) degrees of freedom. This was not signifiant (P = 0.18, d.f = 5), and the element by element comparison is displayed in Table 1. Some new data were provided for 1973 (D. Pauly, personal communication) to permit a comparison of the predicted values with data not included in the input matrix. These are shown in the last row of the table, and the model projection corresponds fairly well with the data.

The derived transition probability matrix (Table 2) shows that the last two elements on the main diagonal have relatively high probability values, suggesting that squid and other species (in this case) tend to remain in the somewhat similar proportions from one interval to the next. There also seems to be a relatively high probability for Leiognathidae to be replaced by other species as indicated by the value of 0.75 for element 3 in column 5. There are no imaginary parts to the eigenvalues (Table 2). However, one of the real roots is negative suggesting some oscillation in the movement to a final steady state. Of most importance is the large size of the second largest eigenvalue (0.9365), indicating an extremely slow movement toward the steady state probability vector. This suggests a high degree of instability in this system. The elements of the vector indicating the average proportion of time spent in each state are high for squid and other species (Table 2),

suggesting movement toward these states under the exploitation regime operating at the time the data set was obtained. For expected first passage and recurrence times, the main diagonal elements (the recurrence times) are lowest for squid and for other species as shown by elements 4,4 and 5,5 (Table 2). In general, the transition times to column 5 (other species) seem to be lower than for the specific species or groups indicated in the other four columns.

Gulf of Maine.-For the Gulf of Maine area, the species selected for analysis were those for which there was no significant directed fishery at the time the data were collected (Table 3). These included goosefish, *Lophius americanus*, cusk, *Brosme brosme*, and white hake, *Urophycis tenuis*. A chi-square value of 6.07 ($P = 0.30$, d.f. = 5) indicates an acceptable fit to the data. Note that there are also predicted values for 1974 as well as the actual observations which were not included in developing the model. This permits another empirical comparison of the model projections for one year ahead with the actual values obtained at that time. In general, the predicted values were excellent for goosefish, very poor for cusk, and reasonable for white hake and the "other fish" category.

The derived transition probability matrix shown in Table 4 indicated a strong main diagonal element for white hake, suggesting that the probability of relatively similar proportions of white hake was high in successive steps. There also seemed to be a high probability of cusk being replaced by goosefish as indicated by the element 2,1 of the matrix. All of the eigenvalues of the transition probability matrix were real (Table 4). However, two of them were negative suggesting some oscillation in the relative abundance of species. The moderate size for

the second largest eigenvalue (0.5706) indicated that these oscillations were moderately damped, and that the steady state probability vector was relatively rapidly achieved. The largest element in the steady state probability vector was 0.692 for the white hake, suggesting that the CPUE for this species was considerably higher than for other species. This inference is similar to that made directly from the transition probability matrix. The expected recurrence time shown as element 3,3 (Table 4) was shortest for the white hake. Furthermore, the expected first passage times to white hake from the other species were lower than for the other species. This suggests that the white hake quickly returns to its present state after leaving it, and that it also quickly moves from its present state to another state.

Northwest Atlantic Groundfish 1967-1974.-An assumption implied in the use of the Markov model is that the transition probability matrices are the result of a process that is stationary in time or space, i.e., the transition probabilites are constant through time (in our case). For a long-time series of data, it is desirable to test to see if transition probabilities calculated from sub-intervals of the time series are equal to the lumped transition probability matrix obtained by estimation over the whole interval (null hypothesis). The alternative hypothesis is that the Markov process is non-stationary, that is: $P_{ij}(t) = P_{ij}$. A decision was made to divide the entire data set into equal periods assumed to reflect conditions prior to and after the imposition of the Fishery Conservation and Management Act, but with adequate time series in each period. A test of the null hypothesis was made using the data from 1967-74 and 1975-82 against the entire set consisting of 1967-82. The test statistic was developed by Anderson and Goodman (1957), and the

test statistic is distributred as chi-square with $(T-1)[m(m-1)]$ degrees of freedom, where m = numbers of states, and T = number of subintervals. The null hypothesis was rejected ($P < .01$), and the subsequent analyses were carried out on the two separate time series of data.

The period 1967-1975 represents a period of very high exploitation. We used data on five major species of groundfish: Atlantic cod *Gadus morhua*, haddock, *Melanogrammus aeglefinus*, redfish, *Sebastes marinus*, pollock, *Pollachius virens*, and silver hake, *Merluccius bilinearis* (Table 5). Again, the fit of the model to the data was very good, with a chi-square value of 4.59 ($P > 0.5$, d.f. = 7). These data permitted another opportunity to test the predictive power of the model in a step-ahead forecast mode because data were available for 1975; they were not included in model development. The forecast for these species (Table 5, last line) was not as good as that obtained for the Gulf of Maine area. All species showed some deviations from observed values but the forecasts for Atlantic cod and pollock were particularly bad in this case.

The main diagonal elements of the transition probability matrix (Table 6) showed no dominant values. However, there seemed to be some indication of a transition from pollock and Atlantic cod to redfish as shown by the relative sizes of elements 1,3 and 2,3 of the matrix. The eigenvalues (Table 6) were all real, but there were some negative values suggesting oscillation. The second eigenvalue (-0.5669) was moderate in value indicating a moderate rate of convergence of the system to a stable vector. There were no strong elements in the vector of steady state probabilities, suggesting that the probabilities of observing any of the various states were reasonably similar. The expected recurrence

times (main diagonal) showed the smallest value for redfish. Similarly, all the values in column three (redfish) were lower than for the other species, suggesting some transition to this species, under the assumption that the observed conditions would persist in time.

Northwest Atlantic Groundfish Data, 1975-1982.-As indicated previously this data set was partitioned from the entire available data in order to meet the Markov process assumption of stationarity and also to met minimum sample requirements. The period 1975-1982 included the period when the Fishery Conservation and Management Act was invoked, and might, therefore, reflect some possible recovery from the highly exploited phase prior to this. Table 7 illustrates the data used for model construction and the fit of the model to the observed data. The chi-square value of 7.38 (P = .20, d.f. = 7) indicates a good fit of the model to the data. These data also permitted a test of the one step-ahead predictive ability of the model (Table 7, last line). There was some real disparity in the predicted proportions for redfish, silver hake, and pollock, suggesting some transient condition in the fishery or highly selective fishing.

The transition probability matrix (Table 8) indicated reasonably high values for the main diagonal elements 1,1 and 2,2 suggesting that Atlantic cod and haddock were more stable than the other species. There also seemed to be some reasonable probability for transition from redfish to pollock as shown by the size of element 3,5. The eigenvalues of the transition probability matrix had imaginary parts, and the modulus of the complex roots was determined to be 0.3774, suggesting moderate oscillations but with more rapid damping than that found for the earlier data set. The steady state probability vector had no

extremely dominant elements. However, the data for Atlantic cod and haddock seemed to be more stable than the other species in this case. The expected first passage times and recurrence probabilities indicated that pollock had the largest recurrence time. Transitions from other species to pollock indicated in the last column of the table were also the largest.

Management Implications.- Although this work was designed primarily to demonstrate the methodology for deriving transition probability matrices from aggregated time series of multispecies fishery data and various analyses of the transition probabilty matrix, some tentative management-related conclusions are indicated from the specific data which have been analyzed.

The overall results of this analysis of the Thai trawl data support prior assessments of the fishery by others. That is, changes which have been documented (Pauly 1979, 1984) and supported by us include rapid declines of large fishes such as rays, decreases in the family Leiognathidae, and increases in squid and other species. A unique feature of our analysis involves projection of the TPM to a steady state under the observed conditions. This projection suggests ultimate dominance by squids, and other less-valued species if the present fishing strategy were continued. A highly unstable condition of the fishery is also indicated suggesting that a very long period of time would be required to reach the ultimate steady state probability vector.

The Gulf of Maine data for non-commercial groundfish (goosefish, cusk, white hake and other) was expected to be the most stable of the three northwest Atlantic data sets used herein, because there was no

directed commercial fishery for these species at the time. This was supported by the results of some of our analyses. The step-ahead forecasts for this non-directed fishery were the most accurate and precise of the three. The rate at which the system appeared to approach a steady state was moderate with relatively little oscillation. For the northwest Atlantic region, the combined data for selected groundfish species, which was divided into two periods of 1966-1974 and 1975-1982, provided less precise predictions then the Gulf of Maine data. Because the 1966-1974 period was one of very high exploitation and the 1975-1982 period included a new fishery conservation regime, it might be expected that the fishery would show less stability during the earlier period than was demonstrated from the analyses we performed. However, a more rapid convergence to a steady state was indicated during the 1975-1982 period than in the earlier period.

In summary, we believe that the approach described herein for the analysis of mulitispecies fishery data has demonstrated a potential for usefulness. The examples provided are but a small subset of the possible applications of this new methodology to multispecies fisheries.

References

Anderson, T. W., and L. A. Goodman. 1957. Statistical inference about Markov chains. Annals of Mathematical Statistics, 28:89-110.

Clark, S. H., and B. E. Brown. 1977. Changes in biomass of finfishes and squids from the Gulf of Maine to Cape Hatteras, 1963-74, as determined from research vessel survey data. Fishery Bulletin, 75:1-8.

Cooke, D. 1982. A Markov chain model of plant succession. Pages 231-247 in R. W. Hiorns and D. Cooke, editors. The mathematical theory of the dynamics of biological populations II. Academic Press, London.

Caswell, H. 1986. Life cycle models of plants. Lectures on mathematics in the life sciences, 18:171-233.

Cox, D. R., and H. D. Miller. 1965. The theory of stochastic processes. Metheun, London.

Getz, W. M., and G. L. Swartzman. 1981. A probability transition matrix model for yield estimation in fisheries with high variable recruitment. Canadian Journal of Fisheries and Aquatic Sciences. 38:847-855.

Kemeny, T. G., and J. L. Snell. 1966. Finite Markov chains. Springer-Verlag, Berlin.

Lee, T. C., G. G. Judge, and A. Zelner. 1970. Estimating the parameters of the Markov probability model from aggregate time series data. North Holland Publishing Company, Amsterdam.

Mahon, R., editor. 1985. Towards the inclusion of fishery interventions in management advice. Canadian Technical Report on Fisheries and Aquatic Sciences, Number 1347.

May, R. M., editor. 1984. Exploitation of marine communities. Springer-Verlag, Berlin.

Mercer, M. C., editor. 1979. Multispecies approaches to fisheries Management advice. Canadian Special Publication on Fisheries and Aquatic Science, Number 59.

Pauly, D. 1979. Theory and management of tropical multispecies stocks: a review, with emphasis on the southeast Asian dermersal fisheries. ICLARM Studies and Reviews, 1. International Center for Living Aquatic Resources Management. Manila, Philippines.

Pauly, D., and G. I. Murphy, editors. 1982. Theory and management of tropical fisheries. Proceedings of the ICLARM/CSIRO workshop on the theory and management of tropical multispecies stocks, 16-21 January, 1981. Cronulla, Australia.

Pauly, D. 1984. Fish population dynamics in tropical waters: a manual for use with programmable calculators. ICLARM Studies and Reviews, 8. International Center for Living Aquatic Resources Management, Manila, Philippines.

Pope, J. 1979. Stock assessment in multispecies fisheries with special reference to the Gulf of Thailand. SCS/DEY/29/19. South China Sea Fisheries Development and Coordinating Programs, Manila, Philippines.

Riffenburgh, R. H. 1969. A stochastic model of interpopulation dynamics in marine ecology. Journal of the Fisheries Research Board of Canada, 26:2843-2880.

Rothschild, B. J., and A. J. Mullen. 1985. The information content of stock-and-recruitment data and its non-parametric classification. Journal du Conseil 42:116-124.

Saila, S. B. 1982. Markov models in fish community studies - some basic concepts and suggested application. Pages 202-209 in G. R. Huntsman, W. K. Nicholson, and W. W. Fox, Jr., editors. National Oceanic and Atmospheric Administration Technical Memorandum NMFS-SEFC-82, Washington D. C.

Saila, S. B., and P. M. Roedel, editors. 1980. Stock assessment for tropical small-scale fisheries. International Center for Marine Research Development, The University of Rhode Island, Kingston.

Appendix.-Estimation of Transition Probabilities from Aggregated Data

A basic linear statistical model for estimating transition prbabilities assumes that sample observations are generated by the stochastic relationship;

$$y_j(t) = \sum_i y_i(t-1)\, p_{ij} + u_j(t), \qquad (1)$$

where $y_j(t)$ and $y_i(t-1)$ are observed proportions, p_{ij} are transition probabilities, and u_j are random disturbances.

In matrix notation this set of equations, of which Equation (1) is a part, may be written as:

$$y = Xp + u, \qquad (2)$$

with

$$E(u) = 0 \qquad (3)$$

and

$$E(uu') = \Sigma \qquad (4)$$

This multivariable linear statistical model consists of r Markov chain relations of the form of Equation (1). Σ is a covariance matrix of the random disturbances and r is the number of states in the Markov chain. X is an $(rT \times r^2)$ matrix with r blocks of identical X_j, $j = 1, 2, \ldots, r$ on the diagonal of the matrix, and T is the number of time periods.

An example given by Lee et al. (1970) for illustration provides the following proportions:

0.50	0.50
0.75	0.25
0.88	0.12
0.94	0.06
0.97	0.03

In the above, r = 2 and T = 5, and the X matrix of dimensions ($rT \times r^2$), which is of dimension 10 x 4 in this case, is developed as follows:

$$X = \begin{matrix} 0.50 & 0.50 & 0.00 & 0.00 \\ 0.75 & 0.25 & 0.00 & 0.00 \\ 0.88 & 0.12 & 0.00 & 0.00 \\ 0.94 & 0.06 & 0.00 & 0.00 \\ 0.97 & 0.03 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.50 & 0.50 \\ 0.00 & 0.00 & 0.75 & 0.25 \\ 0.00 & 0.00 & 0.88 & 0.12 \\ 0.00 & 0.00 & 0.94 & 0.06 \\ 0.00 & 0.00 & 0.97 & 0.03 \end{matrix}$$

The column vector u has elements u_j, $j = 1,2...,r$. In this example the dimensions are (4 x 1):

$$u = \begin{matrix} u_1 \\ u_2 \\ u_1 \\ u_2 \end{matrix}$$

In a similar manner p is a column vector of dimension (4 x 1) with subvectors p_j, $j = 1,2...,r$:

$$p = \begin{matrix} p_1 \\ p_2 \\ p_1 \\ p_2 \end{matrix}$$

From the basic statistical model described by Equations (1) and (2), a number of unrestricted and restricted estimators of transition probabilities are developed. The interested reader is referred to the flow chart (Figure 1) and to Table 9 for summaries and to Lee et al. (1970) for details.

In general, it has been found that the restricted estimators are better than the unrestricted. Of the non-Bayesian estimators, the GLS and the minimum chi-square (MCS) estimators seem to perform the best in the applications made to date. Estimates with smallest variance were judged to be the best.

Table 1. Average annual catch composition (kilograms per hour of trawling) in the Gulf of Thailand 1966-1972 (modified from Ritragsa 1976). The observed values are decimal fractions of the total weight of all species per tow for each year. O = observed values, X = expected values from model calculations. The 1973 data are from D. Pauly (personal communication).

Year	Species or Species Groups[a]															
	A			B			C			D			E			
	Catch	O	X	Catch	O	X	Catch	O	X	Catch	O	X	Catch	O	X	Sum
1966	9.6	0.07	---	15.3	0.12	---	20.0	0.16	---	8.6	0.06	---	77.8	0.59	---	131.3
1967	4.8	0.04	0.04	11.8	0.12	0.10	10.9	0.09	0.10	9.1	0.08	0.10	78.5	0.69	0.67	115.1
1968	2.2	0.02	0.03	7.5	0.03	0.09	14.4	0.15	0.10	10.6	0.11	0.11	71.5	0.69	0.68	106.0
1969	3.0	0.03	0.03	7.4	0.07	0.08	10.6	0.10	0.09	11.6	0.12	0.12	70.1	0.68	0.68	102.7
1970	2.9	0.03	0.03	8.6	0.09	0.08	10.2	0.10	0.09	8.6	0.09	0.03	67.2	0.67	0.68	97.5
1971	1.4	0.02	0.02	7.3	0.11	0.09	3.0	0.04	0.09	11.0	0.17	0.12	63.6	0.68	0.68	66.3
1972	1.2	0.02	0.02	4.7	0.08	0.08	4.9	0.08	0.09	14.2	0.22	0.20	38.1	0.61	0.61	63.1
1973	1.5	0.03	0.02	3.2	0.06	0.07	6.0	0.12	0.08	9.9	0.19	0.24	31.2	0.60	0.58	51.8

[a]A = rays, B = *Nemipterus* spp., C = Leiognathidae, D = *Loligo* spp., E = all others.

Table 2. The transition probability matrix and some of its properties derived from the data of Table 1 concerning the Gulf of Thailand trawl survey results form 1966-1972.

Derived transition probability matrix

Species[a]	A	B	C	D	E
A	0.13	0.11	0.35	0.00	0.41
B	0.04	0.22	0.00	0.34	0.40
C	0.14	0.11	0.00	0.00	0.75
D	0.03	0.00	0.00	0.97	0.00
E	0.00	0.07	0.12	0.00	0.81

Eigenvalues

Real part	Imaginary part
- 0.2112	0
0.2263	0
1.0000	0
0.9365	0
0.1779	0

Steady State Probabilities

Species	Probability
A	0.03
B	0.05
C	0.05
D	0.51
E	0.36

Expected 1st Passage and Recurrence Times

Species	A	B	C	D	E
A	34.88	11.97	12.84	31.21	4.95
B	36.98	21.28	30.59	19.24	18.18
C	36.89	12.11	18.37	31.35	3.25
D	39.22	45.20	46.07	1.98	38.18
E	41.59	12.62	16.73	31.85	2.74

[a] A = rays, B = Nemipterus spp., C = Leiognathidae, D = Loligo spp, E = all others.

Table 3. Average annual catch composition of selected groundfish from trawl survey data in the Gulf of Maine area 1963-1973 (from Table 5 of Clark and Brown 1977). The observed values are decimal fractions of the total weight of all species per tow for each year. 0 = observed values, X = expected values from model calculations.

Year	Species and Species Groups								Total Weight
	Goosefish		Cusk		White Hake		Others		
	0	X	0	X	0	X	0	X	
1963	0.26	-	0.18	-	0.56	-	0.02	-	14.0
1964	0.19	0.18	0.14	0.15	0.62	0.62	0.05	0.04	8.4
1965	0.16	0.18	0.11	0.13	0.68	0.65	0.05	0.03	11.7
1966	0.20	0.17	0.21	0.12	0.53	0.68	0.06	0.03	17.9
1967	0.24	0.22	0.15	0.14	0.58	0.60	0.03	0.04	7.2
1968	0.20	0.18	0.18	0.14	0.57	0.63	0.05	0.04	10.1
1969	0.19	0.20	0.07	0.14	0.74	0.62	0.04	0.04	24.0
1970	0.14	0.13	0.09	0.11	0.74	0.72	0.03	0.04	22.0
1971	0.19	0.15	0.08	0.10	0.72	0.72	0.01	0.03	21.4
1972	0.67	0.14	0.14	0.12	0.76	0.71	0.04	0.03	22.2
1973	0.17	0.18	0.06	0.07	0750	0.73	0.02	0.02	21.2
1974	0.14	0.13	0.03	0.11	0.82	0.72	0.02	0.03	17.1

Table 4. The transition probability matrix and some of its properties derived from the data of Table 3 concerning the Gulf of Maine trawl survey results for 1963-1973.

Derived transition probability martrix

	Goosefish	Cusk	White hake	Other
Goosefish	0.09	0.48	0.28	0.15
Cusk	0.61	0.00	0.39	0.00
White hake	0.09	0.03	0.87	0.01
Other	0.48	0.51	0.00	0.01

Eigenvalues

Real part	Imaginary part
- .4698	0
- .1348	0
1.0000	0
.5706	0

Steady state probabilities

Species	Probability
Goosefish	0.16
Cusk	0.11
White hake	0.69
Other	0.03

Expected first passage and recurrence times

	Goosefish	Cusk	White hake	Other
Goosefish	6.15	5.34	3.41	29.56
Cusk	4.47	8.78	3.09	32.95
White hake	8.95	11.70	1.44	35.73
Other	3.33	3.56	4.25	31.08

Table 5. Average annual catch composition of selected groundfish from trawl survey data in the northwest Atlantic region 1967-1974 (from Table 8 of Clark and Brown 1977). The observed values are decimal fractions of the total weight of those species for each year. O = observed values, X = expected vaues from model calculatins.

Year	Species										Total
	Atlantic Cod		Haddock		Redfish		Silver hake		Pollock		Weight
	O	X	O	X	O	X	O	X	O	X	
1967	0.18	-	0.33	-	0.34	-	0.10	-	0.05	-	24.3
1968	0.17	0.20	0.20	0.16	0.47	0.45	0.09	0.09	0.07	0.10	28.8
1969	0.20	0.23	0.17	0.16	0.36	0.38	0.08	0.10	0.19	0.13	22.1
1970	0.22	0.22	0.17	0.15	0.47	0.43	0.09	0.10	0.05	0.10	23.4
1971	0.24	0.22	0.12	0.15	0.41	0.38	0.12	0.12	0.11	0.12	19.5
1972	0.28	0.25	0.10	0.14	0.36	0.38	0.16	0.12	0.12	0.12	23.2
1973	0.32	0.27	0.16	0.12	0.28	0.38	0.13	0.13	0.11	0.10	20.1
1974	0.19	0.22	0.08	0.12	0.49	0.44	0.12	0.13	0.12	0.09	15.6
1975	0.16	0.28	0.16	0.15	0.49	0.33	0.15	0.11	0.04	0.14	27.6

Table 6. The transition probability matrix and some of its properties derived from the data of Table 5 concerning the northwest Atlantic trawl survey results from 1967-1974.

Derived transition probability matrix

Species[a]	A	B	C	D	E
A	0.00	0.00	0.67	0.31	0.02
B	0.00	0.25	0.75	0.00	0.00
C	0.28	0.21	0.15	0.09	0.27
D	0.95	0.00	0.00	0.05	0.00
E	0.21	0.20	0.59	0.00	0.00

Eigenvalues

Real part	Imaginary part
1.0000	0
-0.5669	0
0.4384	0
0.0029	0
-0.4244	0

Steady state probabilities

Species	Probability
A	0.24
B	0.14
C	0.39
D	0.12
E	0.11

Expected first passage and recurrence times

Species	A	B	C	D	E
A	4.11	10.18	1.67	7.99	9.14
B	5.36	7.20	1.33	11.44	8.76
C	4.02	8.26	1.56	10.11	7.43
D	1.05	11.27	3.02	8.59	10.19
E	4.45	8.01	1.68	10.93	9.05

[a]A = Atlantic cod, B = haddock, C = redfish, D = silver hake, E = pollcck.

Table 7. Average annual catch composition of selected groundfish from trawl survey data in the northwest Atlantic region 1975-1982 (from Table 8 of Clark and Brown 1977 and from S. Clark (personal communication)). The observed values are decimal fractions of the total weight of the species for each year. O = observed values and X = expected values from model calculations.

Year	Species										Total Weight
	Atlantic Cod		Haddock		Redfish		Silver Hake		Pollock		
	O	X	O	X	O	X	O	X	O	X	
1975	0.16	-	0.16	-	0.49	-	0.15	-	0.04	-	24.3
1976	0.12	0.15	0.25	0.27	0.15	0.16	0.17	0.17	0.31	0.75	28.8
1977	0.20	0.21	0.31	0.32	0.20	0.22	0.11	0.12	0.18	0.13	22.1
1978	0.28	0.24	0.28	0.31	0.22	0.19	0.12	0.14	0.10	0.13	21.4
1979	0.24	0.27	0.28	0.29	0.18	0.17	0.11	0.15	0.09	0.13	19.5
1980	0.24	0.27	0.34	0.31	0.19	0.18	0.17	0.14	0.07	0.10	23.2
1981	0.30	0.26	0.28	0.30	0.21	0.19	0.13	0.15	0.08	0.10	20.1
1982	0.34	0.28	0.17	0.29	0.09	0.16	0.19	0.15	0.12	0.12	15.6
1983	0.21	0.27	0.18	0.28	0.09	0.21	0.41	0.18	0.11	0.06	----

Table 8. The transition probability matrix and some of its properties derived from the data of Table 7, concerning the northwest Atlantic trawl survey results from 1975-1982.

Derived transition probability matrix

Species[a]	A	B	C	D	E
A	0.58	0.22	0.00	0.20	0.00
B	0.31	0.39	0.24	0.06	0.00
C	0.00	0.24	0.07	0.19	0.50
D	0.00	0.26	0.48	0.26	0.00
E	0.21	0.36	0.22	0.04	0.17

Eigenvalues

Real Part	Imaginary Part
1.0000	0.0000
-0.2705	0.0000
0.3483	-0.1454
0.3483	0.1454
0.0440	0.0000

Modulus of complex root = 0.3774

Steady state probabilities

Speciess	Probability
A	0.270
B	0.294
C	0.108
D	0.149
E	0.106

Expected first passage and recurrence times

	A	B	C	D	E
A	3.70	4.16	6.70	6.57	12.54
B	5.27	3.40	5.31	7.97	11.14
C	7.13	3.65	5.55	7.65	5.84
D	7.84	3.73	3.21	6.71	9.05
E	5.77	3.41	5.33	8.33	9.40

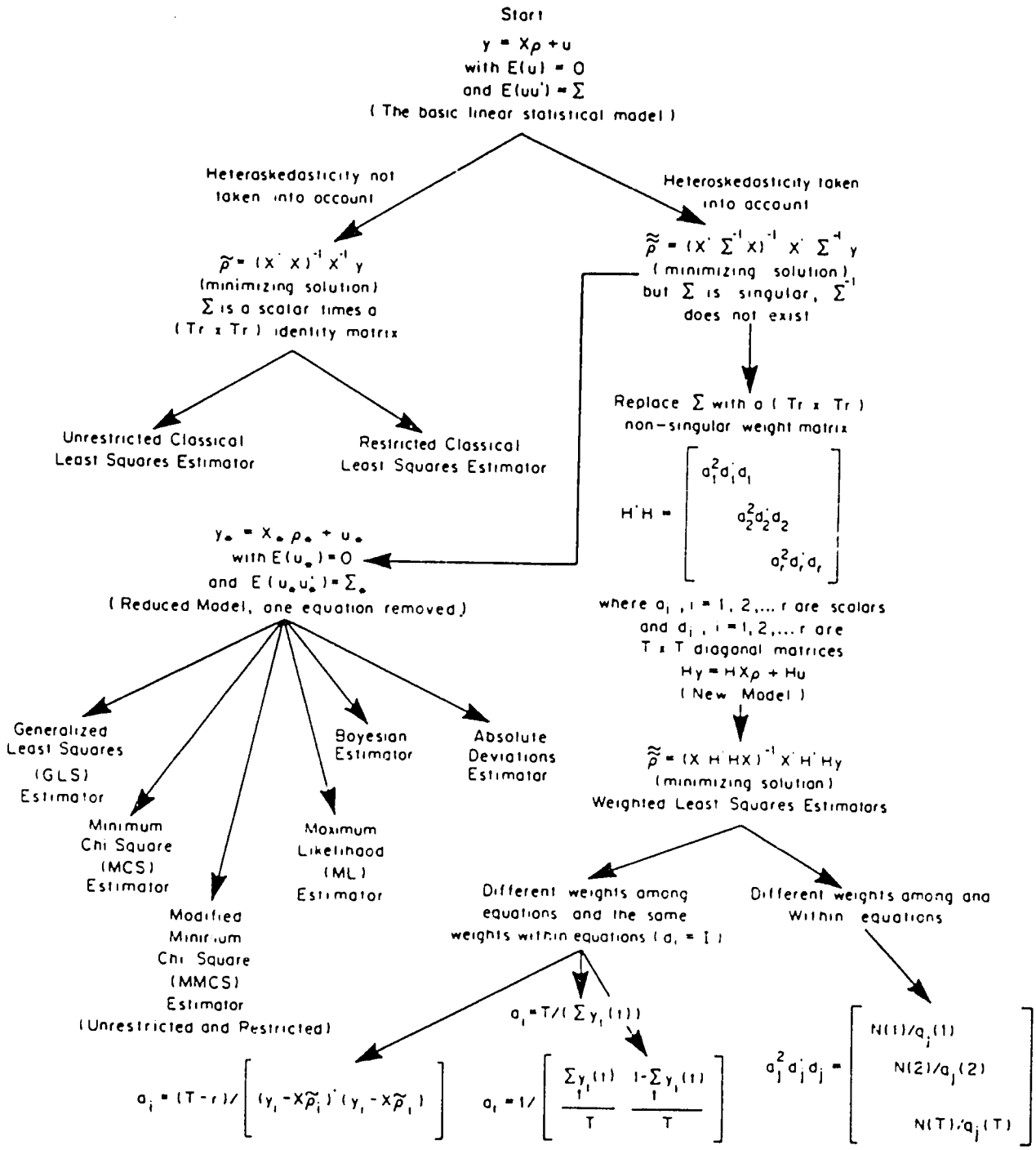
[a]A = Atlantic cod, B = haddock, C = redfish, D = silver hake, E = pollock.

Table 9. Objective functions and constraints for the calculations of some transition probabilities estimators by linear and quadratic programming.

Estimator	Objective function and constraints
Unrestricted classical least squares	min: $(y - Xp)'(y-Xp)$ or $p = (X'X)^{-1}Xy$ subject to: $0 \leq p_{ij} \leq 1$ and $\sum_j p_{ij} = 1$ for all i
Restricted classical least squares	max: $(X'y - X'Xp)'p - \lambda_1'\eta_r + \lambda_2'\eta_r = - \lambda_1 a_1 - \lambda_2' a_2$ $- \beta' p \leq 0$ subject to: $Gp = \eta_r$, $G'\lambda_1 - G'\lambda_2 + (X'X)p - \beta = X'y$, and $p, \lambda_1, \lambda_2, a_1, a_2, \beta \geq 0$ where G is an $(r \times r^2)$ matrix with r identity submatrices I_r , η_r is an $(r \times 1)$ column vector with all elements 1, λ_1 and λ_2 are vectors of Lagrangian multipliers and a_1, a_2, and β are vectors of slack variables.
Unrestricted and restricted weighted least squares	Same as classical least squares except for the introduction of a weight matrix $H'H$ (see Figure 1).

Generalized unrestricted least squares	min: $(y_* - X_* p_*)' \Sigma_*^{-1}(y_* - X_* p_*)$ or $p_* = (X_*' \Sigma_*^{-1} X_*)^{-1} X_*' \Sigma_*^{-1} y_*$ subject to: $Rp_* \leq \eta_r$, and $p_* \geq 0$ where R is a submatrix of G $(I_1, I_2, \ldots, I_{r-1})$ with each I a (r x r) identity matrix.
Generalized restricted least squares	max: $(X_* \Sigma_*^{-1} y_* - X_*' \Sigma_*^{-1} X_* p_*)' p_* - \lambda' \eta_r = -\lambda' p_r - \beta' p_* \leq 0$ subject to: $Rp_* + p_r = \eta_r$, $R'\lambda + (X_*' \Sigma_*^{-1} X_*) p_* - \beta = X_*' \Sigma_*^{-1} y_*$, and $p_*, p_r, \lambda, \beta \geq 0$, where p_r and β are vectors of slack variables.
Unrestricted minimum chi-square	Same as unrestricted generalized least squares.
Restricted minimum chi-square	Same as restricted generalized least squares.
Unrestricted modified minimum chi-square	$p = X'DX^{-1}X'Dy$ where D is a (Tr x Tr) diagonal matrix with elements $N(t)/q_j(t)$, $(j = 1, 2, \ldots r;\ t = 1, 2, \ldots T)$.
Restricted modified minimum chi-square	min: $\chi^2 = (y - Xp)'D(y - Xp)$ subject to: $Gp = \eta_r$, and $p \geq 0$ where G is an $(r \times r^2)$ matrix with r identity matrices in a row $(I_1, I_2, \ldots, I_r)$, where each I is an (r x r) identify matrix.

Figure 1. Flow chart of methodology for calculation of transition probability matrices based upon Lee et al. (1970).


Start
y = Xρ + u
with E(u) = 0
and E(uu') = Σ
(The basic linear statistical model)
Heteroskedasticity not taken into account
Heteroskedasticity taken into account
ρ̃ = (X' X)^-1 X^-1 y
(minimizing solution)
Σ is a scalar times a (Tr x Tr) identity matrix
ρ̃̃ = (X' Σ^-1 X)^-1 X' Σ^-1 y
(minimizing solution),
but Σ is singular, Σ^-1 does not exist
Unrestricted Classical Least Squares Estimator
Restricted Classical Least Squares Estimator
Replace Σ with a (Tr x Tr) non-singular weight matrix
H'H =
y* = X* ρ* + u*
with E(u*) = 0
and E(u*u*') = Σ*
(Reduced Model, one equation removed)
where a_i , i = 1, 2,... r are scalars
and d_i , i = 1, 2,... r are T x T diagonal matrices
Hy = HXρ + Hu
(New Model)
Generalized Least Squares (GLS) Estimator
Minimum Chi Square (MCS) Estimator
Modified Minimum Chi Square (MMCS) Estimator (Unrestricted and Restricted)
Maximum Likelihood (ML) Estimator
Bayesian Estimator
Absolute Deviations Estimator
ρ̃̃ = (X' H'HX)^-1 X' H'Hy
(minimizing solution)
Weighted Least Squares Estimators
Different weights among equations and the same weights within equations (a_i = 1)
Different weights among and within equations
a_i = T/(Σ y_i (t))
a_i = (T - r)/[(y_i - Xρ̃_i)' (y_i - Xρ̃_i)]
a_i = 1/[Σ_t y_i(t) / T 1 - Σ_t y_i(t) / T]
a_j^2 d_j' d_j =
N(1)/a_j(1)
N(2)/a_j(2)
N(T)/a_j(T)