USING BOX-JENKINS MODELS TO FORECAST FISHERY DYNAMICS: IDENTIFICATION, ESTIMATION, AND CHECKING

ROY MENDELSSOHN[1]

ABSTRACT

Box-Jenkins models are suggested as appropriate models for forecasting fishery dynamics. Unlike standard production models, these models are empirical, dynamic, stochastic models. Box-Jenkins models are not biased when estimating relationships between catch and effort, as are standard production models. The use of these techniques is illustrated on catch and effort data for the skipjack tuna fleet in Hawaii. An actual 12-month forecast is shown to give a reasonable fit to the observed data. Most of the discrepancies are explained by changes in the behavior of the fishermen (i.e., economic factors), rather than by lack of knowledge of the behavior of fish.

Accurate forecasting models would be useful in fishery management because extended jurisdiction and international agreements require preseasonal predictions of the actual catch of a fleet. In addition, improved forecasts of fish availability can lead to improved planning by fishermen or by processing firms. Forecasting techniques have expanded greatly in the last years, but few have been adapted to research in fisheries management. Instead, techniques designed to establish the equilibrium health of the stocks are also being used to attempt dynamic forecasting.

At present, two least squares procedures are being used to estimate the general production model, the search procedure of Pella and Tomlinson (1969) and the weighted least squares of Fox (1970, 1971, 1975). The Fox procedure fits catch per unit effort against a function of lagged effort. Several authors (Chayes 1949; Eberhardt 1970; Atchley et al. 1976) have demonstrated that scaling the dependent variable (i.e., catch) by the independent variables (i.e., effort) biases the fit by introducing artificial correlation into the data. Johnston (1972) showed that ordinary least squares gave biased estimates and an inflated F-statistic when used with variables lagged on themselves. Neither the Fox nor the Pella-Tomlinson procedure accounts for the effect of autocorrelated errors in the estimation procedure which Granger and Newbold (1977) and Newbold and Davies (1978) have demonstrated bias both estimation and tests of fit. An examination of the residuals in Fox (1971, figure 3B) clearly shows them to be autocorrelated. Residuals from many spawner-recruit curves display similar behavior.

In this paper, the use of Box-Jenkins models for modeling and forecasting fisheries dynamics is explored. Box-Jenkins and other related forecasting techniques are specifically designed for estimating and testing models in the presence of autocorrelated errors. The fitted models are stochastic rather than deterministic, thus reflecting the variability found in most fisheries. The models are constructed empirically, and are best suited for forecasting. The models tell us little about the long-term health of the stocks, so that a judicious use of production, yield per recruit, and accurate forecasting models is required to give the best overall picture of the fishery.

My preference for Box-Jenkins models over other forecasting methods now available is due to the good documentation (see for example Anderson 1975; Box and Jenkins 1976; Granger and Newbold 1977) and computer accessibility. The results presented here were obtained using a package originally developed by David Pack at Ohio State University and now available through Automatic Forecasting Systems.[2]

The three-step process of model identification, model estimation, and model diagnostic checking is illustrated by developing a model that makes monthly forecasts of skipjack tuna, *Katsuwonus pelamis*, catches in Hawaii. Experience with the model suggests that for a 12-mo forecast of catch,

[1]Southwest Fisheries Center Honolulu Laboratory, National Marine Fisheries Service, NOAA, Honolulu, HI 96812.

[2]Reference to trade names does not imply endorsement by the National Marine Fisheries Service, NOAA.

Manuscript accepted May 1980.
FISHERY BULLETIN: VOL. 78, NO. 4, 1981.

during peak months the forecast is within 15% of the observed catch (and is usually within 8-10% of the observed catch), most turning points in the catch trend are predicted, and the important feature of a low, flat catch during the summer months or high, peaked catches are accurately predicted. Moreover, the reasons for forecasts with large errors appear to be related more to fishermen's decisions in face of weather and economic factors, than to mispredicting the availability of the fish.

THE DATA AND UNDERLYING MODEL

The data to be analyzed are landings of skipjack tuna by approximately 12 boats from Oahu during 1964 through 1978. The raw data consist of the daily landings (each boat rarely stayed out more than a day or two), broken down by boat, and by four skipjack tuna size classes: large, medium, small, and extra small. For purposes of analysis, the data were aggregated into monthly totals, with the total number of fishing trips used as the measure of fishing effort. For monthly catch and effort during 1964-78 see Figures 1 and 2.

There are several causes for the observed seasonal variability. First, the tuna are only available in large numbers seasonally. Second, price considerations, particularly around Christmas and New Year when there is large demand, tend to spur fishing even when availability is low. Third, with only 12 boats fishing, if 1 or 2 boats are not able to fish for a few weeks, the catch will drop sharply. Finally, environmental factors, particularly weather (such as bad seas) will affect the landings since the boats are unable to fish.

Folklore in Hawaii has it that the catch remains similar each year, no matter how many boats fish. Comitini[3] examined the fishery using dummy variables and ordinary least squares to estimate

[3]Comitini, S. 1977. An economic analysis of the state of the Hawaiian skipjack tuna fishery. Sea Grant Tech. Rep. UNIHI-SEAGRANT-TR-78-01, 46 p.

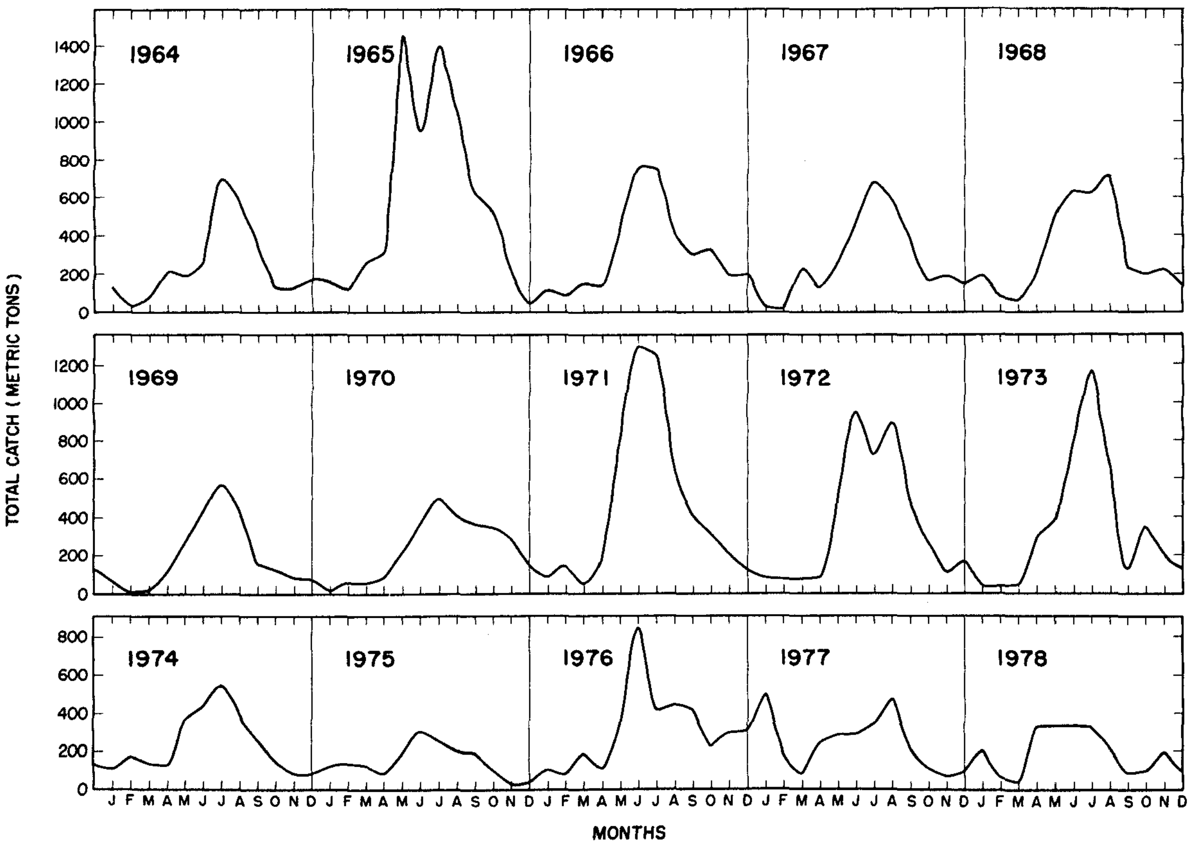

FIGURE 1.—Level of Hawaiian skipjack tuna catch by month, 1964-78.

a Cobbs-Douglas production function. He concluded, among other things, that natural fluctuations in resource availablity are significant, but did not include them in his analysis, nor did he provide a means for forecasting future catch. The National Marine Fisheries Service, using a regression model based on the previous year's catch, water temperature, and salinity at the start of the year, makes yearly predictions that have been mixed in accuracy.

Box-Jenkins models are autoregressive-integrated-moving-average models, or ARIMA models. These are linear, stochastic models that can describe fairly complex behavior, in contrast to Parrish and MacCall (1978) who use highly nonlinear equations to model the fluctuations in fishery data.

The modeling is based on the properties of stationary time series. A time series x_t is stationary if it has a constant mean, and if the covariance between events x_t, x_{t-s} depends only on s and not on t. Many series are stationary after removing a deterministic trend. Others are differenced in order to achieve stationary. Also, transforming the time series, particularly using the Box-Cox family of transformations, often improves the behavior of the time series. The initial step then is to transform and difference the data as necessary to achieve stationary. It is convenient to use the backshift operator B^j, where $B^j x_t = x_{t-j}$, to denote lagged variables. Given the new series $z_t = (1 - B^d)x_t$, a mixture of autoregressive and moving average models are sought. Autoregressive models are models that depend on the past history of the time series:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + a_t$$

in terms of the backshift operator:

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p) z_t = a_t$$

while moving average models depend on past values of the noise or error:

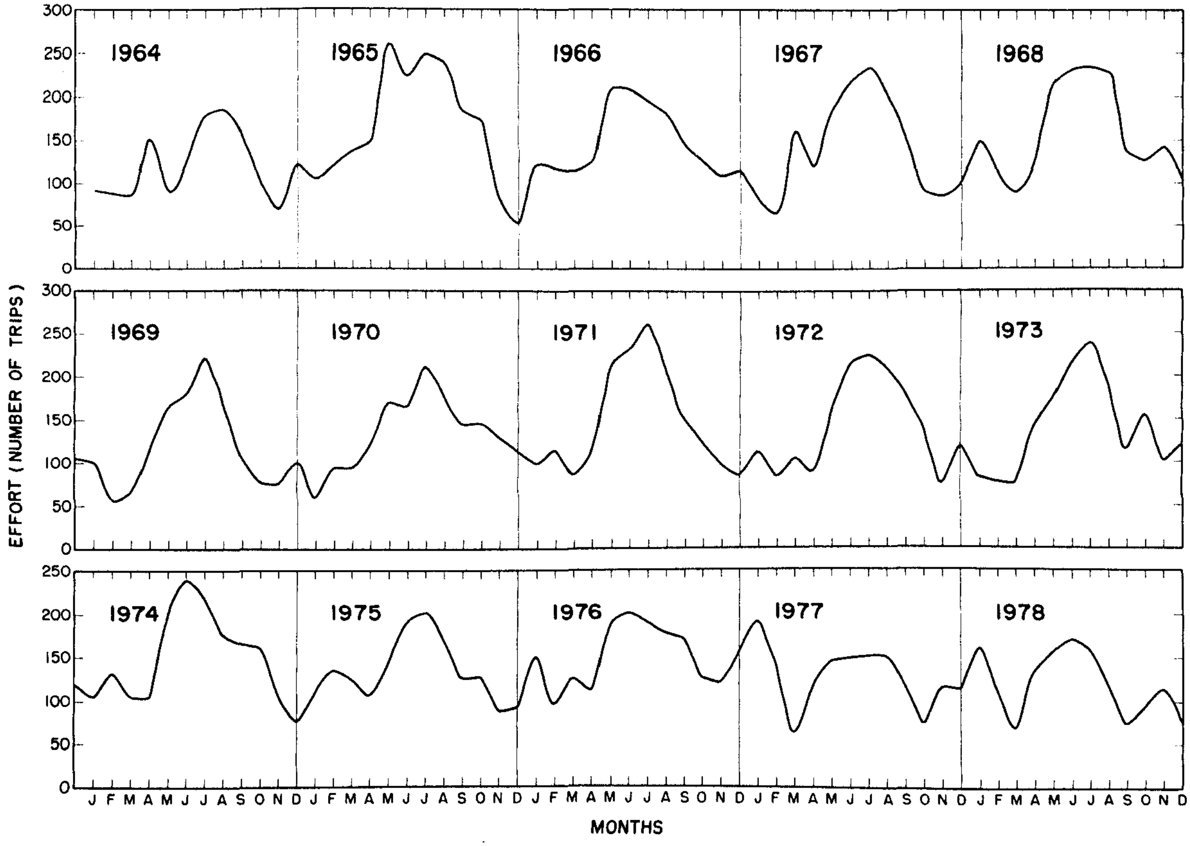

FIGURE 2.—Number of fishing trips per month by the Hawaii skipjack tuna fleet 1964-78, near Oahu, Hawaii.

$$z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

or:

$$z_t = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) a_t.$$

A model that has both moving average and autoregressive parameters is a mixed autoregressive moving average model, whose representation in terms of the backshift operator is:

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)(1 - B)^d x_t$$

$$= (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) a_t.$$

MODEL IDENTIFICATION

The first step in the Box-Jenkins modeling process is to use properties of the data to tentatively identify a model. Even if a multivariate model (i.e., a model based on catch and effort) is the ultimate goal, univariate models of each series are constructed first. Often the univariate model produces forecasts that are almost as accurate as the multivariate model forecast.

My procedure was to identify, estimate, and check a series of models based on the data from January 1964 through July 1977. These models were used to forecast the already observed catch and effort for the period August 1977-December 1978. The models with the best "fit" were then reestimated to make the forecast for 1979. To make clear the feedback nature of identification, estimation, and checking in Box-Jenkins models, results from models fixed to 163 and 180 mo of data are intermingled, but clearly labeled.

A tentative model can be identified by estimating the autocorrelation and partial autocorrelation functions for each series. These are shown in Figures 3 and 4. Significant is the undamped sinusoidal behavior of each, with a period of 12 mo. Failure of both the autocorrelation and partial autocorrelation functions to go to zero is a sign of a nonstationary series, and the need for differencing. The 12-mo period suggested a yearly seasonal model, so that twelfth differences were taken, i.e., $z_t = (1 - B^{12})x_t$.

The estimated autocorrelation and partial autocorrelation functions for the differenced catch and effort series are given in Tables 1 and 2. Following guidelines in appendix 9.1 in Box and Jenkins (1976), seasonal models with period s of the form:

$$z_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta_1 B^s) a_t \quad \text{(1a)}$$

or $$z_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta_1 B^s - \Theta_2 B^{2s}) a_t \quad \text{(1b)}$$

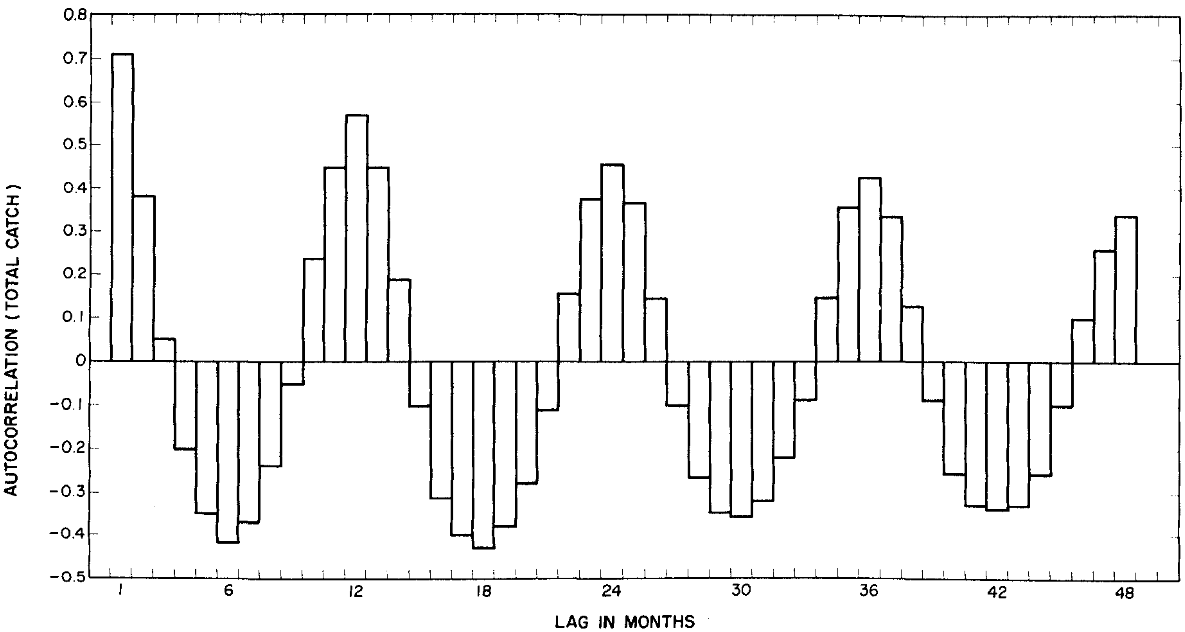

FIGURE 3.—Estimated total catch autocorrelation function for the catch of skipjack tuna near Oahu, Hawaii, 1964-78.

TABLE 1.—Autocorrelation functions for 12th differenced effort series of the Hawaii skipjack tuna fleet, 1964-78.

Item	Lag (mo) 1	2	3	4	5	6	7	8	9	10	11	12	13	14
Regular auto.	0.39	0.17	0.20	0.16	0.10	0.04	0.03	0.07	−0.02	−0.08	−0.20	−0.45	−0.18	−0.08
SE	.08	.09	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.12	.12
Partial auto.	.39	.03	.15	.03	.01	−.04	.00	.00	−.08	−.07	−.19	−.39	.15	.04

Item	Lag (mo) 15	16	17	18	19	20	21	22	23	24	25	26	27
Regular auto.	−0.18	−0.13	−0.12	−0.17	−0.15	−0.17	−0.12	0.05	0.04	0.02	0.00	−0.07	0.03
SE	.12	.12	.12	.12	.12	.12	.13	.13	.13	.13	.13	.13	.13
Partial auto.	−.03	.02	−.07	−.14	−.01	−.02	−.05	.16	−.08	−.19	.05	−.12	.03

TABLE 2.—Autocorrelation functions for 12th differenced catch series of the Hawaii skipjack tuna fleet, 1964-78.

Item	Lag (mo) 1	2	3	4	5	6	7	8	9	10	11	12	13	14
Regular auto.	0.58	0.40	0.33	0.20	0.11	0.05	0.01	0.02	−0.06	−0.12	−0.21	−0.38	−0.21	−0.15
SE	.08	.11	.12	.12	.12	.12	.12	.12	.12	.12	.13	.13	.14	.14
Partial auto.	.58	.09	.10	−.07	−.03	−.04	−.00	.04	−.11	−.07	−.17	−.29	.28	.03

Item	Lag (mo) 15	16	17	18	19	20	21	22	23	24	25	26	27
Regular auto.	−0.16	−0.12	−0.08	−0.09	−0.08	−0.12	−0.10	−0.08	−0.08	−0.09	−0.06	−0.06	−0.05
SE	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14	.14
Partial auto.	−.01	−.06	−.02	−.07	.02	−.05	−.04	−.04	−.11	−.17	−.19	−.01	−.02

were hypothesized as the appropriate univariate models for both the catch and the effort time series.

ESTIMATION AND CHECKING

Given a tentative model, such as Model (1), the next step is a recursive procedure of estimating the parameters of the model, calculating the autocorrelation and partial autocorrelation functions of the residuals from the estimated model, and then testing the residuals for significant departure from the assumption that they are white noise. When a final model has been identi-

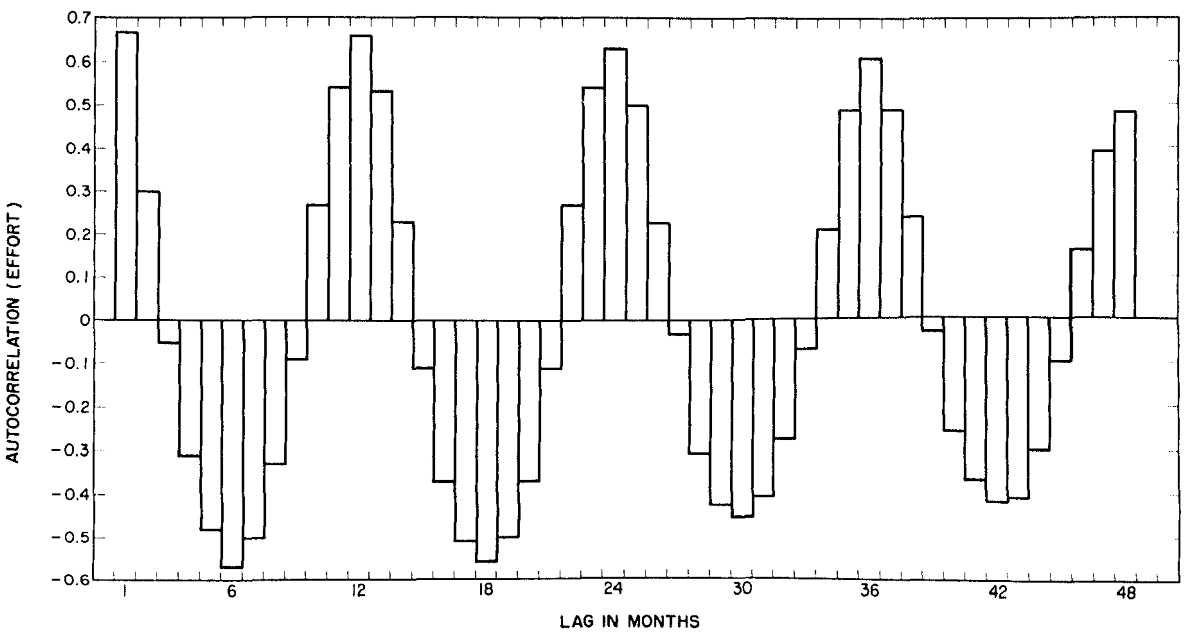

FIGURE 4.—Estimated effort autocorrelation function for the fishing trips by the Hawaii skipjack tuna fleet near Oahu, Hawaii, 1964-78.

fied, overfitting is tried, that is extra parameters are added to see if they are found to be not significantly different from zero.

To insure that I found the simplest model possible, I fitted first the model $z_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12})a_t$, and then added parameters as seemed necessary based on the diagnostic checking. The estimates for Model (1) for catch and effort are given in Tables 3 and 4. Estimates using two estimation techniques, one using backforecasting and one suppressing it, are presented. Some programs do not have a backforecasting feature; my experience is that the estimated models obtained using backforecasting are far superior, as can be seen in the tables presented.

TABLE 3.—Parameter estimates for effort model, Model (1) (see text). (Based on 180 observations.)

Parameter	Estimate supressing backforecasting	SE	Estimate using backforecasting	SE
θ_1	−0.38349	0.07942	−0.44756	0.07886
θ_2	−.11326	.07996	−.12795	.07911
Θ_1	.5894	.08122	.99493	.00650
Θ_2	.00069	.08609	—	—
χ^2 statistic on residuals	26.894 with 44 df		37.319 with 45 df	
Residual mean square	1,018.60		755.270	
Residual SE	31.915		27.482	
Residual mean	1.629		0.5338	

TABLE 4.—Parameter estimates for catch model, Model (1) (see text). (Based on 163 observations.)

Parameter	Estimate suppressing backforecasting	SE
θ_1	−0.54100	0.08190
θ_2	−.22745	.08235
Θ_1	.75314	.08718
Θ_2	.05184	.09256
χ^2 statistic on residuals	27.470 with 43 df	
Residual mean square	165,410	
Residual SE	406.71	
Residual mean	17.506	

The estimated autocorrelation and partial autocorrelation functions of the residuals from both models are given in Tables 5 and 6. For the effort series, there is no sign of a lack of fit, while for the catch series terms of lag three or four are suggested. An overspecified model:

$$z_t = (1 - \theta_1 B^1 - \theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4)(1 - \Theta_1 B^{12})a_t \quad (2)$$

was estimated for both the catch and effort time series. The results are summarized in Tables 7 and 8. The estimated autocorrelation and partial autocorrelation functions of the residuals (not shown) show no sign of additional lags or trend. The test statistic that the residual series are not significantly different from white noise gave no reason to doubt the models adequacy, and overfitting by including a $\Theta_2 B^{24}$ term found this term to be nonsignificant.

TRANSFER FUNCTION MODELS

If both the catch time series, say y_t, and the effort time series, say x_t, have been suitably transformed so that the resulting series are stationary, a transfer function of the form:

$$(1 - \delta_1 B - \delta_2 B^2 - \ldots - \delta_r B^r)x_t = (\omega_0 - \omega_1 B - \omega_2 B^2 - \ldots - \omega_s B^s)y_{t-b} + \eta_t$$

can be estimated where η_t is not assumed to be white noise, but itself can be modeled as an autoregressive-moving average process of a_t.

The procedures for identifying and estimating a transfer function model are similar to those for the univariate model, except that attention is focused on the estimated cross-correlation function between the "prewhitened" catch and effort series. Series are prewhitened if they are reduced to the residuals left from a given model. In this instance, both series are prewhitened by the univariate model for effort estimated in the preceding section. The estimated correlation function, impulse response function, and residual noise autocorrelation function are given in Table 9. The estimated autocorrelation function for the noise is similar to the original univariate autocorrelations, suggesting a noise model of the form:

$$\eta_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4)(1 - \Theta_1 B^{12})a_t. \quad (3)$$

Based on guidelines in Box and Jenkins (1976:386-388) and knowledge of the fishery, two models were hypothesized:

$$(1 - B^{12})y_t = (\omega_0)(1 - B^{12})x_t + \eta_t \quad (4)$$

and:

$$(1 - \delta_1 B - \delta_2 B^2)(1 - B^{12})y_t = (\omega_0 - \omega_1 B - \omega_2 B^2)(1 - B^{12})x_t + \eta_t. \quad (5)$$

Tables 10 and 11 summarize the estimates when

TABLE 5.—Estimated autocorrelation function for residuals of effort model for the Hawaii skipjack tuna fleet, 1964-78.

Item	Lag (mo) 1	2	3	4	5	6	7	8	9	10	11	12
Auto.	0.01	0.03	0.09	0.06	0.04	−0.11	−0.05	0.04	−0.09	0.02	0.03	−0.01
SE	.07	.07	.07	.08	.08	.08	.08	.08	.08	.08	.08	.08
Item	Lag (mo) 13	14	15	16	17	18	19	20	21	22	23	24
Auto.	0.05	−0.01	−0.04	−0.09	−0.08	−0.09	−0.07	−0.08	−0.16	0.10	0.07	−0.02
SE	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08	.08

TABLE 6.—Estimated autocorrelation function for residuals of catch model for the Hawaii skipjack tuna fleet, 1964-78.

Item	Lag (mo) 1	2	3	4	5	6	7	8	9	10	11	12
Auto.	0.04	0.11	0.23	0.06	0.06	−0.05	0.00	0.10	−0.03	0.04	0.01	0.00
SE	.08	.08	.08	.09	.09	.09	.09	.09	.09	.09	.09	.09
Item	Lag (mo) 13	14	15	16	17	18	19	20	21	22	23	24
Auto.	0.05	−0.01	−0.07	−0.01	−0.00	−0.04	0.00	−0.01	−0.06	0.01	−0.04	−0.03
SE	.09	.09	.09	.09	.09	.09	.09	.09	.09	.09	.09	.09

backforecasting is used in estimating the parameters for Models (4) and (5). The chi-square statistics show no reason to suspect model inadequacy. The residuals show no significant cross-correlation with total catch, when $1/\sqrt{180}$ (180 observations in the series) is used as a rough standard error. The residual autocorrelation function shows spikes around lag 15 that are higher than would be desired, but overall the fit is reasonable, and the model residuals could reasonably be modeled as white noise.

DISCUSSION AND FORECASTS

Two transfer function models and one univariate model have been used to forecast the catch and effort in the skipjack tuna fishery during 1979. It is worth emphasizing that the original 12-mo forecasts were made in January 1979 and the updated forecasts were made in May 1979, so the reported results are true forecasts—there was no a priori knowledge of the data to help improve the "fit" of the forecasts. The original catch and effort forecasts are given in Tables 12 and 13 while the updated catch forecasts are given in Table 14.

The models used to produce the forecasts are best understood when written out in difference equation form. The univariate model for catch is:

TABLE 7.—Parameter estimates for effort model, Model (2) (see text). (Based on 180 observations.)

Parameter	Estimate suppressing backforecasting	SE	Estimate using backforecasting	SE
θ_1	−0.36746	0.08004	−0.43862	0.07930
θ_2	−.14976	.08412	−.18144	.08590
θ_3	−.16111	.08458	−.15377	.08617
θ_4	−.17096	.08454	−.16298	.08593
θ_5	−.11547	.08089	−.17291	.07998
Θ_1	.59065	.06431	.99483	.00033
χ^2 statistic on residuals	20.696 with 42 df		27.494 with 42 df	
Residual mean square	1,000.40		752.67	
Residual SE	31.629		27.435	
Residual mean	.82151		.35175	

TABLE 8.—Parameter estimates for catch model, Model (2) (see text). (Based on 180 observations.)

Parameter	Estimate suppressing backforecasting	SE	Estimate using backforecasting	SE
θ_1	−0.55368	0.07972	−0.53771	0.07462
θ_2	−.35882	.08989	−.43825	.07543
θ_3	−.33817	.09056	−.41197	.01144
θ_4	−.24282	.09012	−.30909	.07479
θ_5	−.12294	.07994	−.14974	.07440
Θ_1	.76951	.05062	.99585	.00825
χ^2 statistic on residuals	15.092 with 42 df		20.384 with 42 df	
Residual mean square	143,240		115,170	
Residual SE	378.47		339.37	
Residual mean	2.1150		3.3299	

$$y_t = y_{t-12} + (a_t + 0.538a_{t-1} + 0.438a_{t-2} + 0.412a_{t-3} + 0.309a_{t-4}) \quad (6)$$
$$- (0.996a_{t-12} + 0.535a_{t-13} + 0.436a_{t-14} + 0.410a_{t-15} + 0.308a_{t-16}),$$

i.e., catch this month is equal to the catch during the same month last year, adjusted by a difference of the weighted sums of the forecasting errors over the previous 4 mo. If the forecasts this year have consistently underpredicted compared with last year's forecasts, then the estimated catch is increased, while if the forecasts this year have consistently overpredicted compared with last year's forecasts, then the estimated catch is decreased. The forecast maintains a balance between keeping the catch in equilibrium and keeping the error in equilibrium.

This impression of a yearly cycle with variability is reinforced when examining the polynomial

TABLE 9.—Estimated cross-correlation function, impulse response function, and noise autocovariance function for a catch-effort transfer model for the Hawaii skipjack tuna fleet, 1964-78.

Lag (mo)	Estimated cross-correlation	Estimated noise autocovariance	SE	Estimated impulse response weights
0	0.651	—		8.409
1	.080	0.49	0.10	1.035
2	.070	.21	.12	.903
3	.086	.16	.12	1.111
4	−.033	.16	.13	−.431
5	.044	.10	.13	.566
6	−.098	.07	.13	−1.269
7	.099	.03	.13	1.276
8	.103	.13	.13	1.334
9	−.017	.14	.13	−.215
10	.043	−.05	.13	.556
11	−.040	−.14	.13	−.517
12	−.20	−.26	.13	−1.555
13	.026	−.05	.14	.338
14	−.109	.05	.14	−1.404
15	.003	−.12	.14	−.038
16	−.098	−.16	.14	−.415
17	.014	−.01	.14	.043
18	−.110	−.05	.14	−1.271
19	−.037	−.12	.14	.185
20	.006	−.12	.14	−1.422
21	−.006	−.16	.14	−.475
22	−.108	−.11	.14	.080
23	.012	−.18	.15	−.075
24	−.108	−.21	.15	−1.393
25	−.001	−.09	.15	.181
26	−.108	−.08	.15	−1.390

TABLE 10.—Parameter estimates for transfer model, Model (4) (see text). (Based on 180 observations.)

Parameter	Estimate suppressing backforecasting	SE	Estimate using backforecasting	SE
ω_0	7.5989	0.69403	8.0003	0.83561
θ_1	−.47621	.07993	−.48894	.07851
θ_2	−.32874	.08734	−.32633	.08541
θ_3	−.17034	.08803	−.14853	.08666
θ_4	−.20033	.07905	−.17506	.07822
Θ_1	.83384	.05271	.99587	.00707
χ^2 statistic on residuals	34.953 with 43 df		32.018 with 43 df	
Residual mean square	83,323		71,300	
Residual SE	288.66		267.02	
Residual mean	−15.152		0.18650	

TABLE 11.—Parameter estimates for transfer model, Model (5) (see text). (Based on 180 observations.)

Parameter	Estimate suppressing backforecasting	SE	Estimate using backforecasting	SE
δ_1	0.01286	0.30389	0.86672	0.22308
δ_2	.88121	.28641	−.70763	.21659
ω_0	7.3488	.73352	8.1855	.82832
ω_1	−1.3011	2.16847	6.7421	1.71214
ω_2	6.8509	2.34577	−7.3133	1.58459
θ_1	−.49924	.08302	−.46980	.08013
θ_2	−.29495	.09102	−.33234	.08870
θ_3	−.16384	.09191	−.17199	.09012
θ_4	−.13639	.08352	−.21746	.08098
Θ_1	.83311	.05511	.99543	.00623
χ^2 statistic on residuals	33.067 with 43 df		38.906 with 43 df	
Residual mean square	85,673		69,066	
Residual SE	292.70		262.80	
Residual mean	−1.9979		−2.4666	

TABLE 12.—Catch forecasts for 1979 for the Hawaii skipjack tuna fleet from Models (1), (4), and (5) (see text).

	Model			
Month	(4)	(5)	(1)	Observed catch
Jan.	102.24	157.48	159.97	52.6488
Feb.	78.91	123.32	117.81	74.1184
Mar.	121.86	118.83	108.40	102.4088
Apr.	202.05	169.75	175.82	131.0658
May	423.40	406.87	423.95	470.5450
June	595.39	605.68	598.17	358.5100
July	666.16	684.99	607.07	600.6930
Aug.	528.09	535.73	523.14	600.5200
Sept.	297.96	294.92	291.97	148.3070
Oct.	224.28	216.64	222.96	79.3360
Nov.	173.99	168.83	172.94	27.5084
Dec.	133.22	131.61	132.58	84.7755
Total	3,547.55	3,614.65	3,534.78	2,730.4367

TABLE 13.—Predicted and observed number of fishing trips for the Hawaii skipjack tuna fleet in 1979.

Month	Original prediction	Updated prediction	Observed
Jan.	98.93		53
Feb.	97.50		75
Mar.	101.06		78
Apr.	122.30		118
May	174.00	167.71	173
June	196.16	187.36	182
July	209.37	206.14	200
Aug.	183.04	179.81	174
Sept.	139.18	138.73	84
Oct.	121.20	120.45	84
Nov.	104.23	104.83	51
Dec.	100.92	100.51	109

TABLE 14.—Updated forecasts of total catch for 1979 for the Hawaii skipjack tuna fleet.

	Model			
Month	(4)	(5)	(1)	Observed catch
May	393.214	382.430	401.874	470.545
June	547.014	586.400	589.524	358.510
July	644.638	705.137	668.895	600.693
Aug.	500.151	527.945	521.456	600.520
Sept.	293.130	283.067	289.516	148.307
Oct.	220.557	197.953	222.806	79.336
Nov.	174.567	164.720	173.594	27.5084
Dec.	130.947	136.831	133.148	84.7755
Total	2,904.218	2,984.483	3,000.813	2,370.1949

representation of Model (1). The value of Θ_1 is nearly one. Thus the term $(1 - B^{12})$ appears on both sides of the equation, and can be cancelled. Abraham and Box (1978) showed that this is sufficient reason to suspect a deterministic cosine function trend with a moving average model around the trend. Given the high residual mean square for the model (115, 170), this latter interpretation is consistent with the folklore on the fishery—highly variable but on the average things are similar from year to year.

The first transfer function model is:

$$y_t = 8.003x_t + (y_{t-12} - 8.003x_{t-12}) + (a_t + 0.489a_{t-1} + 0.326a_{t-2} + 0.149a_{t-3} + 0.175a_{t-4}) \quad (7)$$
$$- (0.996a_{t-12} + 0.487a_{t-13} + 0.325a_{t-14} + 0.148a_{t-15} + 0.174a_{t-16}).$$

This model has an interpretation similar to that of the univariate model, except now catch per weighted units of effort are compared between years. The second transfer function model compares lagged values of catch and effort also.

It is difficult to judge the value of a forecast, since this will depend on the use being made of the forecast and the alternatives available. Granger and Newbold (1977) suggested the most appropriate measure of the value of a forecast is a loss function which reflects the loss from inaccurate forecast in the actual application for which the forecasts were developed. For forecasting the skipjack tuna fishery in Hawaii, there were four immediate goals. The first was to give a reasonably accurate estimate of total catch over the year, within a 15-20% error rate. The second was to predict what kind of summer it would be, May through September being the main fishing months. This means predicting what month the fish start running, what month the fish stop running, and whether the catch is high and peaked as in 1979, or flat and low as in 1978. An important concern is the relative size of the drop in catch when it occurs in September or October.

A third concern was an accurate forecast of the catch in December, when the holiday demand for sashimi (a Japanese raw fish delicacy) drives prices very high. And finally, an increased understanding of the dynamics of the fishery was desired.

Based on these criteria, I feel the forecasts have performed well, especially compared with any alternative available. The error in predicting the 1979 total catch is higher than desired. However, for the last 6 mo of 1977 the model forecasted within 8% of the observed total catch, and for the period July 1977-December 1978 the model forecasted within 12% of the observed total catch.

Except for June 1979, the summer months were predicted accurately. Experience with the model on the data from July 1977 suggests that the summer months are almost always predicted within 10% of the observed catch. In fact, in March 1979, an industry representative doubted the high catch forecasted for the summer, due to the low catch in January and February 1979. Similarly, the sharp drop in catch in September was predicted by the model. Again, in August 1979 an industry representative doubted that a sharp decline in catch would occur in September, but said that this could be a useful piece of knowledge since their decisions would change if they knew they could expect the supply to drop sharply.

The forecasts have provided insight into the fishery. The major failures of the forecasts were January 1979 and October-December 1979. January 1979 was a period of unusually bad storms, so that few fishing trips were made. However, the observed catch per trip was 0.993 metric tons (t), while Model (4) predicted a catch per trip of 1.033 t. The main source of the error in the forecast was the predicted number of trips to be made.

Similarly, the high summer catches, coupled with very high catches of yellowfin tuna, drove the price for skipjack tuna to very low levels. At the end of September, most of the boats went into drydock because of the prevailing low prices. The few boats that remained tended not to be the industry leaders (i.e., boats with a proven record of higher catch rates), and made only short forays rather than their usual fishing trips.

The point of these explanations is that the causes of the poor forecasts appear to be related not to the behavior of the fish stocks but rather to the behavior of the fishermen. Therefore, the effort to improve the forecasts needs to be directed at understanding the fishery, rather than the fish. (An economic study of the industry is near completion.)

Finally, water temperature and salinity data for one location off Oahu were included in the transfer function models. These variables added little to the forecasts, and since there is no mechanistic explanation as to why these variables should affect the catch and effort, they are not being used at this time in the forecasts. (However, the ability to include random environmental factors into the forecasting model is an advantage when using stochastic models as compared with the normal deterministic production models.) Disaggregating by size class might also improve the forecasts. Prior to 1973, the catch of the large skipjack tuna and the total catch were highly correlated. Since 1973, this has not been true and there has been a definite change in the size composition of the catch. A disaggregated intervention model may be able to explain this change.

SUMMARY

Box-Jenkins models have been proposed as an alternate model for forecasting fishery data. ARIMA models provide maximum likelihood estimators that are not biased when the data is seasonal and autocorrelated, and when a variable is lagged on itself. Techniques are explored which allow the model to be constructed from the data up, rather than from theoretical models that may not be supported by the data. The procedure is illustrated on skipjack tuna catches in Hawaii, which traditionally has been considered too variable to forecast on a monthly basis in a reasonable manner.

ACKNOWLEDGMENTS

I am indebted to Lisa Katekaru for her invaluable assistance in performing most of the programming for this paper. Robin Allen and two anonymous referees provided several comments that have improved the presentation and discussion of this paper.

LITERATURE CITED

ABRAHAM, B., AND G. E. P. BOX.
1978. Deterministic and forecast-adaptive time-dependent models. Appl. Stat. 27:120-130.

ANDERSON, O. D.
1975. Time series analysis and forecasting: The Box-Jenkins approach. Butterworths, Lond., 182 p.

ATCHLEY, W. R., C. T. GASKINS, AND D. ANDERSON.
1976. Statistical properties of ratios: I. Empirical results. Syst. Zool. 25:137-148.

BOX, G. E. P., AND G. M. JENKINS.
1976. Time series analysis: forecasting and control. Rev. ed. Holden-Day, San Franc., 575 p.

CHAYES, F.
1949. On ratio correlation in petrography. J. Geol. 57: 239-254.

EBERHARDT, L. L.
1970. Correlation, regression, and density dependence. Ecology 51:306-310.

FOX, W. W., JR.
1970. An exponential surplus-yield model for optimizing exploited fish populations. Trans. Am. Fish. Soc. 99: 80-88.
1971. Random variability and parameter estimation for the generalized production model. Fish. Bull., U.S. 69: 569-580.
1975. Fitting the generalized stock production model by least-squares and equilibrium approximation. Fish. Bull., U.S. 73:23-37.

GRANGER, C. W. J., AND P. NEWBOLD.
1977. Forecasting economic time series. Acad. Press, N.Y., 333 p.

JOHNSTON, J.
1972. Econometric methods. 2d ed. McGraw-Hill, N.Y., 437 p.

NEWBOLD, P., AND N. DAVIES.
1978. Error mis-specification and spurious regressions. Int. Econ. Rev. 19:513-519.

PARRISH, R. H., AND A. D. MACCALL.
1978. Climatic variation and exploitation in the Pacific mackerel fishery. Calif. Dep. Fish Game, Fish Bull. 167, 110 p.

PELLA, J. J., AND P. K. TOMLINSON.
1969. A generalized stock production model. Inter-Am. Trop. Tuna Comm., Bull. 13:419-496.