WRINT DWR-206

321

PREDICTIVE ABILITY

OF THE

STRIPED BASS MODEL

July 27, 1992

California Department of Water Resources

PREDICTIVE ABILITY OF THE STRIPED BASS MODEL

The purpose of this exhibit is to investigate the ability of the Striped Bass Model ("Model") to predict the number of adult bass in the Sacramento-San Joaquin Estuary ("Estuary") that would result from a given level of Delta outflow and exports. The Model is summarized from WRINT-DFG-Exhibit 3 in Figure 1.[1] The raw data that the California Department of Fish and Game ("CDFG") reportedly used to develop the Model is presented in Appendix A.[2]

1. VALIDATION

Several reviewers pointed out the need to validate the Model using data that had not been used to develop the Model, including the Department of Water Resources ("DWR"), Hanson, and Jones & Stokes.[3] Regression models are normally validated by comparing predictions to actual values <u>that were not used to develop the model</u>. This can be achieved by collecting fresh data or by

---

[1] David W. Kohlhorst, Donald E. Stevens, and Lee W. Miller, <u>A Model for Evaluating the Impacts of Freshwater Outflow and Export on Striped Bass in the Sacramento-San Joaquin Estuary</u>, WRINT-DFG-Exhibit 3, July 1992.

[2] CDFG's regression equations were checked using the data in Appendix A. All of the equations shown in Figure 1 can be duplicated except those that are based on April through July flows. The equations obtained using the data in Appendix A are:

(2) Residual YOY = 1/[0.00995 + (2.41/Eggs)] - 60

(3) Suisun Index = 46.61 Log(Mean Apr-Jul Outflow) - 158.86

(4) Delta Index = 288.827 Log(Mean Apr-Jul Outflow)
- 34.445(Log(Mean Apr-Jul Outflow))^2
- 0.00560 (Mean Apr-July Diversions)
- 527.400

We suspect that CDFG estimated the above three equations using an early version of DAYFLOW. We used CDFG's data from Appendix A and our refit of Equations 2 through 4 in this footnote in all of the analyses reported herein.

[3] Letter from Edward F. Huntley, DWR, to D. Beringer, SWRCB, November 1, 1992, p. 2, Comment 3; Letter from Charles W. Hanson, Hanson Environmental, to David W. Kohlhorst, CDFG, November 7, 1991, p. 1; Letter from W.J. Shaul, Jones & Stokes, November 1, 1992, p. 1-2.

Figure 1.
DFG STRIPED BASS MODEL

(WRINT - DFG - Exhibit 3)

1 Eggs = 92.25(Adults[inc. hatchery])^2 + 38.58 (p. 27)

2 Residual YOY = [1/(0.0095 + (2.59/Eggs))] - 60 (p. 27)

3 Suisun Index = 46.680 Log(Mean Apr-Jul Outflow) - 159.077 (Fig. 7)

4 Delta Index = 292.332 Log(Mean Apr-Jul Outflow) - 34.866 (Log(Mean Apr-Jul Ouflow))^2 - 0.00561 (Mean Apr-Jul Exports+3108) - 534.5475 (Fig. 7)

5 YOY Index = Residual YOY + Suisun Index + Delta Index

7 Adults = (18940 Weighted* Mean YOY - 446608 Log(Weighted* Mean Loss Rate) + 2960840) x** (p. 14)

6 Log(Loss Rate) = 0.00013593(Mean Aug-Mar Exports) - 0.00001553(Mean Aug-Dec Outflow) + 4.6226 (p. 35)

* When applying Equation 7, multiply Mean YOY and Mean Loss Rate by 0.4238

** x = 1.08 when adults in Equation 1 = 1.7 million; x = 0.936 when adults in Equation 1 = 1 million splitting the data set and using part to fit the model and the rest to validate it.

The CDFG did not validate the version of the Model included in WRINT-DFG-Exhibit 3 and has not adequately validated any version of the Model. Figure 6 in WRINT-DFG-Exhibit 3 is not a validation. Rather, it shows how well one of the six regression equations fits the measured data, not how well the Model will predict future values. The predicted values in CDFG's Figure 6 were apparently estimated using the Adult equation (Eq. 7, Fig. 1) alone with data used to develop the Model in the first place. CDFG also apparently applied the Model in the same way that it was developed, while in practice, the Model is applied very differently.[4]

Basic textbooks on regression[5] routinely recommend validation to assess the predictive accuracy of regression models. For example, Drs. Montgomery and Peck in Introduction to Linear Regression Analysis[6] point out that:

> There is no assurance that the equation that provides the best fit to existing data will be a successful predictor. Influential factors that were unknown during the model-building stage may significantly affect the new observations, rendering the predictions almost useless. Furthermore the correlative structure between the regressors may differ in the model-building

---

[4] In developing the Model, the mean YOY index and loss rate 3 to 7 years earlier are used, while in applying the Model, point estimates of the YOY index and loss rate are used. Similarly, in developing the Model, individual weighting factors for each year are used, while in applying the Model, the mean annual weighting factor of 0.4238 is used. Finally, when applying the Model, adult abundance is multiplied by 1.08 when adults in Equation 1 average 1.7 million (prior to 1976) and by 0.936 when adults in Equation 1 average 1.0 million (after 1976) [WRINT-DFG-Exhibit 3, p. 39]. This introduces significant uncertainty into the predicted results, as discussed in Section 3.3.

[5] See, for example, Dick R. Wittink, The Application of Regression Analysis, Allyn and Bacon, Inc., Boston, 1988, Chapter 10; Douglas C. Montgomery and Elizabeth A. Peck, Introduction to Linear Regression Analysis, 2nd Edition, John Wiley & Sons, Inc., New York, 1991, Chapter 10; Norman Draper and Harry Smith, Applied Regression Analysis, 2nd Edition, John Wiley & Sons, New York, 1981, p. 419-422.

[6] D.C. Montgomery and E.A. Peck, Introduction to Linear Regression Analysis, 2nd Edition, John Wiley & Sons, Inc., New York, 1992, p. 443.

> and prediction data. This may result in poor predictive performance for the model. Proper validation of a model developed to predict new observations should involve testing the model in that environment before it is released to the user.

Dr. Wittink, in The Application of Regression Analysis[7], likewise writes:

> The validity of the final model is usually assessed in terms of predictive accuracy. That is, the estimated equation (from the estimation sample) is used to predict values for the criterion variable in the validation samples. These predicted values are then compared with the actual values. Note that in this case the observations in the validation sample are not used to estimate parameters. Instead, the parameters estimated in the estimation sample are used to obtain the predicted values in the validation sample.

We evaluated the predictive ability of the Model by eliminating 1987 through 1991 from the data sets, refitting the equations, and using the refitted equations to predict adult abundance. The same form of the equations shown in Figure 1 was refit to the shorter data sets. The predictions are based on point estimates rather than 3 to 7 year backaverages and annual weighting to accurately reflect the way the Model is actually applied.[4]

The results of this analysis are shown in Table 1. In three out of the four years, predicted adult abundance is about half of actual abundance, while in one year, predicted abundance is about 20 percent higher than actual. Thus, the predictive ability of the Model is poor.

2. CONFIDENCE INTERVALS

The predictive ability of a regression model is normally evaluated using 95 percent confidence intervals. A confidence interval is a measure of the uncertainty associated with a model. The 95 percent confidence interval, for example, means that there is a 95 percent probability that the true value lies between the upper and lower limits. Although regression equations and their coefficients may be statistically significant (i.e., $p<0.05$), the equations can have wide confidence intervals due to statistical

---

[7] Dick R. Wittink, The Application of Regression Analysis, Allyn and Bacon, Inc., Boston, 1988, p. 267-268.

TABLE 1. VALIDATION OF STRIPED BASS MODEL FOR 1987-1990

Data Used for Model Validation

Year	Adults² (inc. hatchery fish)	Aug-Mar Exports	Aug-Dec Outflow	Log Apr-Jul Outflow	Log Apr-Jul Outflow²	Apr-Jul Diversions	Weighting Factor
1987	1.076649039	8,513	4,435	3.6667	13.4446	9,721	0.4238
1988	0.960533285	8,059	4,402	3.7665	14.1863	10,278	0.4238
1989	0.831278945	10,644	5,208	3.9033	15.2355	10,967	0.4238
1990	0.461551749	5,550	4,337	3.7595	14.1340	8,802	0.4238

Refitted Equations

Equation	Year	Predicted	Actual
1. Eggs = 96.94(Adults)²+25.33	1987	129.7	176.0
	1988	118.4	176.4
	1989	105.9	103.5
	1990	70.1	91.6
2. Correction Factor = [1/(0.0107+(2.14/Eggs))]-60	1987	-23.2	-13.5
	1988	-25.2	-26.3
	1989	-27.6	-31.8
	1990	-35.8	-34.4
3. Suisun Index = 46.61 Log(A-J Outflow)-158.86	1987	12.0	5.3
	1988	16.7	0.7
	1989	23.1	2.0
	1990	16.4	1.5
4. Delta Index = 288.827 Log(A-J Outflow)-34.445(Log(A-J Outflow))²-0.00560(A-J Diversions)-527.4000	1987	14.1	7.3
	1988	14.3	3.9
	1989	13.8	3.1
	1990	22.3	2.8
5. YOY Index = Correction Factor+Suisun Index+Delta Index	1987	2.9	12.6
	1988	5.7	4.6
	1989	9.2	5.1
	1990	2.9	4.3
6. Log Loss Rate = 0.00012476(A-M Exports)-0.00001445(A-D Outflow)+4.6373	1987	5.6353	6.0467
	1988	5.5791	5.6271
	1989	5.8900	6.0772
	1990	5.2671	
7. Adults = 0.936[20336 WMYOY-486819 Log(WMLossRate)+3112988]	1987	539,342	998,349
	1988	587,505	903,899
	1989	473,985	786,380
	1990	707,311	591,241

and structural defects, reducing one's certainty about the magnitude of the predicted value.[8]

Three types of confidence intervals are commonly used in regression analysis. These are confidence intervals on the conditional mean,[9] predicted values, and fitted coefficients.

2.1. <u>CDFG's Confidence Intervals</u>

In WRINT-DFG Exhibit 3, CDFG estimated 95 percent confidence intervals for the Adult equation (Eq. 7, Fig. 1) and reported them in Figure 6 of WRINT-DFG-Exhibit 3. The CDFG confidence intervals are a very narrow band, implying a high degree of confidence in the Model's predictions. Although CDFG did not explain how they estimated their confidence intervals, backcalculation indicates that CDFG's confidence intervals are estimated for the <u>conditional mean</u> for <u>only</u> the Adult equation (Eq. 7, Fig. 1). CDFG's confidence intervals are not relevant for estimating how well the Model predicts because they are on the conditional mean not the predicted value, and they are for one equation only, rather than the set of seven equations that comprise the Model.

2.1.1 <u>Prediction Intervals.</u> <u>Prediction</u> confidence intervals must be used to determine the uncertainty in <u>predicted values</u>. The conditional mean confidence intervals that CDFG used are for the regression line itself, not individual predicted values. Prediction confidence intervals for each of the six regression equations are shown in Figure 2. Prediction intervals for the Model as a whole must include the uncertainity in each of these six equations as well as the uncertainity associated with substituting each result into subsquent equations.

Figure 2f shows confidence intervals on predicted values superimposed over confidence intervals on the conditional mean for the Adult equation. Note that 35 percent or eight out of 23 of the observed values for adult abundance, fall outside of the conditional mean intervals of CDFG's Figure 6. The reason that so many points fall outside of the band is that the conditional mean confidence intervals are not designed to encompass predicted values, only the regression line. The uncertainty in individual predictions is always greater than the uncertainty in the

---

[8] Wittink, 1988, p. 39.

[9] The "conditional mean" is the average predicted value of Y based on the condition of a specific value of X. The conditional mean, for example, might be the average Suisun YOY Index for all years when the April to July Delta outflow was equal to 5,000 cfs. Confidence intervals on the "conditional mean" are the same thing as confidence intervals for the regression line.

Figure 2. 95% Confidence Intervals for the Predicted Values

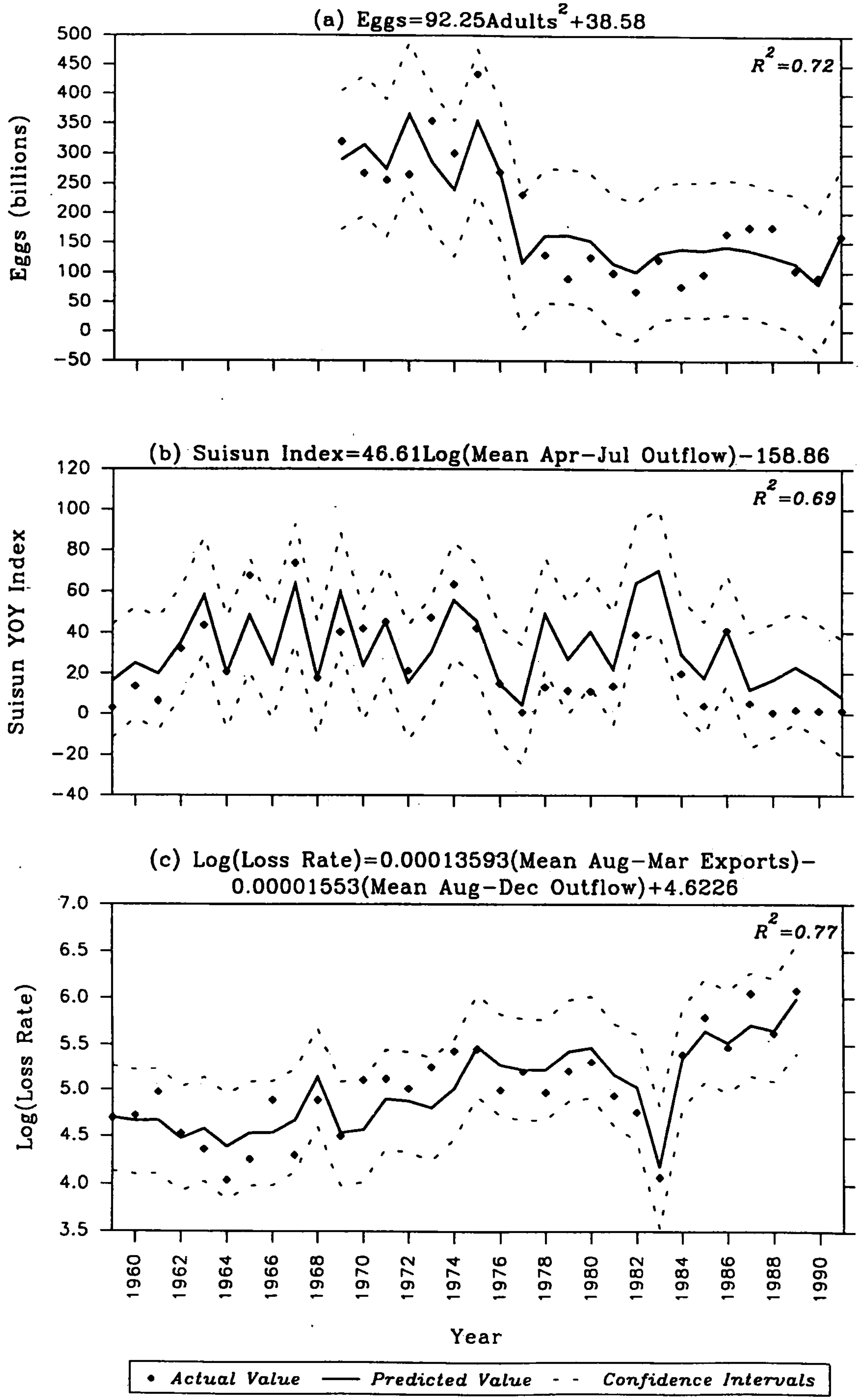
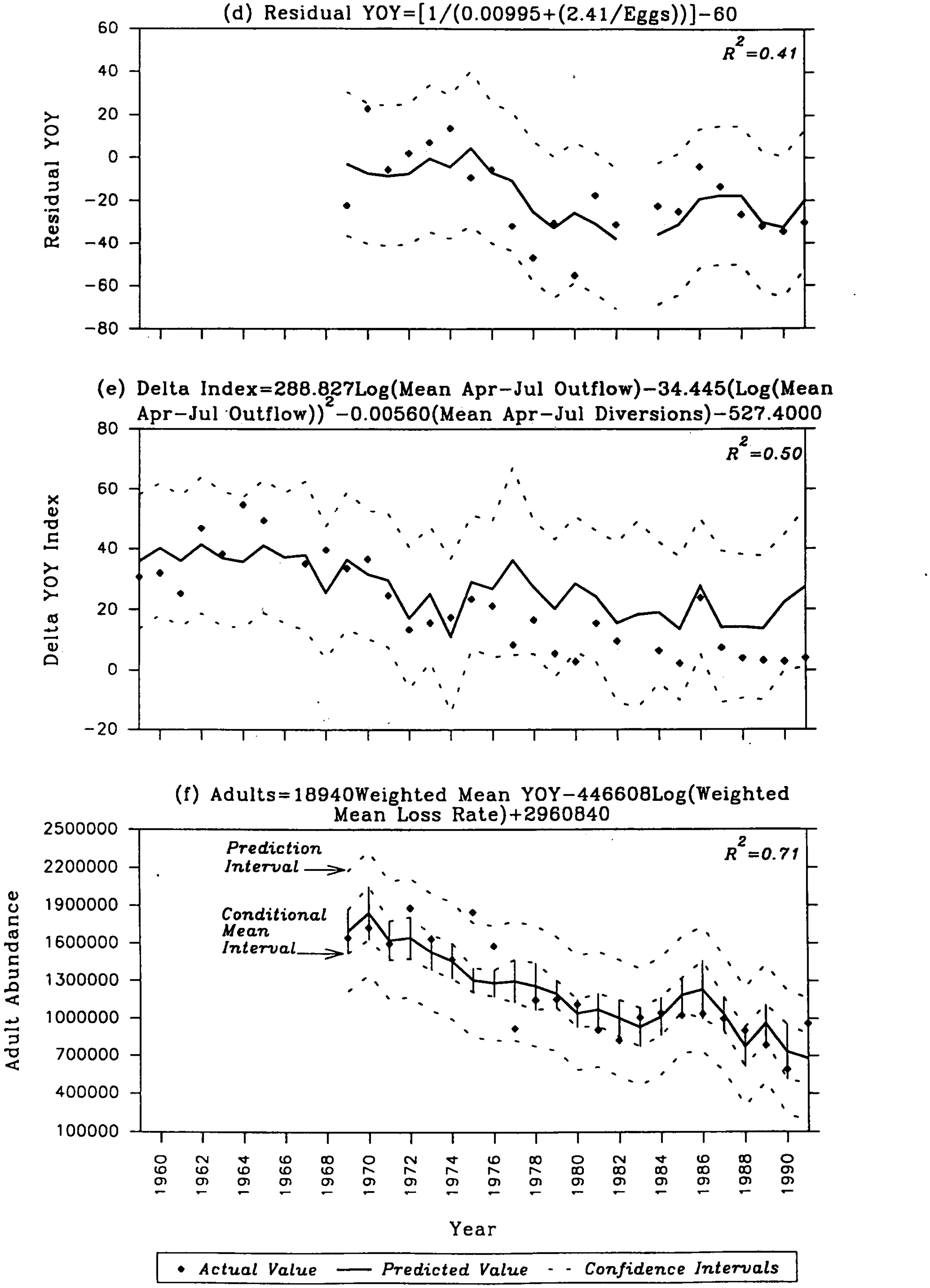


Figure 2 (Continued). 95% Confidence Intervals for the Predicted Values



conditional mean.[10] The predicted value interval is the appropriate interval to assess predictive ability.

2.1.2 Model Intervals. The second problem with CDFG's confidence intervals is that they are for only the Adult equation, not the six regression equations that comprise the Model (Eqs. 1-4,6,7, Fig. 1). Each of these six equations also has 95 percent confidence intervals, as shown in Figure 2. Therefore, to determine 95 percent confidence intervals for predictions made with the Striped Bass Model, confidence intervals that combine the error in all six equations and simultaneously apply to all six regression equations must be constructed.

CDFG's confidence intervals only apply when the Adult equation (Eq. 7, Fig. 1) is used in isolation to predict the conditional mean. In other words, the CDFG intervals apply only when a measured YOY (rather than one estimated from Eqs. 1 - 5 in Fig. 1) and a measured loss rate (rather than one estimated from Eq. 6) is used to estimate conditional mean adults. This is not how the Model is applied. The Model is applied by using hydrologic variables to estimate the Suisun and Delta indices and the loss rate, which are then plugged into the Adult equation. Thus, the uncertainty in each of the six regression equations must be combined to determine the total uncertainty of the Model.

In sum, CDFG's confidence intervals give the misleading impression that the Model is a precise predictor of adult abundance. As shown below, the Model does not accurately predict adult abundance.

2.2. Confidence Intervals For The Model

The most straightforward way to estimate confidence intervals[11] for the Model would be to combine the seven equations into a single equation and use the raw data (Appx. A) in the usual way to compute confidence intervals. When this is attempted with the Striped Bass Model, it reduces to an equation[12] of the form:

---

[10] Wittink, 1988, p. 47.

[11] Henceforth, when the term "confidence intervals" is used, it refers to confidence intervals on the predicted value unless otherwise stated.

[12] The actual equation reduces to:

$$\text{Adults}_o = [aW/\text{Adults}_h^2 + bW\text{Log}(AJO) + cW(\text{Log}(AJO)^2 + dW(AJD) + e\text{Log}(WxAME) + f\text{Log}(WxADO)]_{\text{lagged 3-7 yrs}}$$

Adults (without hatchery fish) = a function of (1)

[1/**Adults**^{2} (with hatchery fish) + Outflow + Diversions]$_{lag}$

Notice that "adults" is a response variable (Y) and that the reciprocal of lagged adults^2 is one of the predictor variables (X's). The Model, in effect, predicts adults from one over the square of 3 to 7 year lagged adults[12].

We fit this single equation to the data, and the results are shown in Table 2. Examining the probabilities (p), it is obvious that neither the Model as a whole (p=0.07) nor the individual fitted coefficients (p=0.14 to 0.75) with the exceptions of August to December outflow, are significant at the 5 percent level. The p values shown in Table 2 are understated due to violations of the underlying assumptions of regression analysis, including the requirement that the predictor variables are independent and that the residuals have equal variance.[13]

Examining the signs on the various fitted coefficients, intuitively, it appears the Model is not correct. The fitted Model predicts that adult abundance: (1) decreases when adults 3 to 7 years earlier increase; (2) increases when August to March exports increase; and (3) decreases when the square of April to July outflow increases. These results raise significant questions about the validity of the seven-equation Model that the single equation model was derived from.

Clearly, the single-equation model is not useful for prediction because 5 of the 6 fitted constants are not statistically significant and have inconsistent signs. Further, the Model as a whole is not statistically significant at the 5 percent level (r^2=0.45, p=0.07). This is in part due to extreme collinearity and other statistical defects, as discussed in Section 3. Since the single-equation model is useless for prediction, confidence limits have no meaning and are not calculated.

2.2.1 Prediction Error. Because confidence intervals are not strictly defined for the Model, confidence intervals on the

---

where the subscript o indicates adults without hatchery fish, the subscript h indicates adults with hatchery fish, W is a weighting factor, AJO is the April to July outflow, AJD is the April to July diversions, AME is the August to March exports, ADO is the August to December outflow, and "a" through "f" are regression coefficients.

[13] Rupert G. Miller, Jr., Beyond ANOVA. Basics of Applied Statistics, John Wiley & Sons, New York, 1986, Chapter 5.

TABLE 2. LINEAR LEAST SQUARES REGRESSION FOR AN EQUATION OF THE FORM:
Adults[o] = [aW/Adults[h]² + bWLog(AJO) + cW(Log(AJO))² + dW(AJD) + eLog(WxAME) + fLog(WxADO)] [lagged 3-7 years]

where [o] indicates adults without hatchery fish, [h] indicates adults with hatchery fish, W is a weighting factor, AJO is April to July outflow, AJD is April to July diversions, AME is August to March exports, ADO is August to December outflow, and a through f are regression coefficients

Year	Adults[o]	Lag 1/Wght Adults[h]²	Lag Wght Log AJO	Lag Wght Log AJO²	Lag Wght AJD	Lag Wght Log AME	Lag Wght Log ADO
1969	1,646,026		0.7399	3.0615	1188.80	1.2034	1.6707
1970	1,727,394		0.7945	3.5411	1061.82	1.2321	1.6473
1971	1,599,715		0.7414	3.0943	1259.43	1.3189	1.6006
1972	1,882,907		0.7875	3.4853	1179.44	1.2926	1.6405
1973	1,637,159		0.7492	3.1529	1285.55	1.3037	1.6878
1974	1,477,213		0.7668	3.2903	1395.37	1.3351	1.6657
1975	1,849,770		0.7199	2.9051	1566.62	1.3494	1.6470
1976	1,581,076	0.347332	0.7233	2.9230	1626.36	1.3802	1.7004
1977	924,301	0.399867	0.7698	3.3193	1794.81	1.4025	1.6904
1978	1,151,642	0.353697	0.7826	3.4199	1682.04	1.4501	1.6554
1979	1,155,701	0.372179	0.7293	2.9851	1546.41	1.4388	1.5164
1980	1,115,999	0.750805	0.6800	2.5990	1191.19	1.4241	1.4369
1981	911,300	0.779775	0.7334	3.0294	1333.48	1.4243	1.4690
1982	825,126	0.772918	0.7267	2.9590	1553.29	1.4471	1.4966
1983	1,009,748	0.790352	0.7456	3.1065	1583.91	1.4608	1.4980
1984	1,042,668	0.994026	0.7246	2.9321	1603.01	1.4690	1.6051
1985	1,024,188	1.225278	0.7893	3.4974	1621.51	1.4758	1.6987
1986	1,037,127	1.131474	0.8364	3.9219	1488.16	1.4268	1.7913
1987	998,349	1.031355	0.7897	3.5039	1655.50	1.4492	1.7317
1988	903,899	0.990185	0.7350	3.0397	1773.63	1.4730	1.5808
1989	786,380	0.957085	0.7495	3.1462	1704.97	1.4815	1.5358
1990	591,241	0.969013	0.7059	2.7939	1719.77	1.4939	1.4598
1991	961,063	1.090548	0.6872	2.6373	1785.90	1.4983	1.4154

Regression Statistics	
Multiple R	0.817581672
R Square	0.668439791
Adjusted R Sq.	0.447399651
Standard Error	157089.8694
Observations	16

Analysis of Variance

	df	*Sum of Squares*	*Mean Square*	*F*	*Significance F*
Regression	6	4.47753E+11	74625543253	3.024065185	0.066367446
Residual	9	2.22095E+11	24677227072		
Total	15	6.69848E+11			

Variable	*Coefficients*	*Standard Error*	*t Statistic*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-5913771.35	12361974.46	-0.4783840	0.6392747	-33878521.72	22050979.03
1/Adults[h]²	-277734.89	286649.57	-0.9689004	0.3479592	-926181.77	370711.99
Apr-Jul Out	12429856.98	34912934.91	0.3560244	0.7267800	-66548749.00	91408462.96
Apr-Jul Out²	-1995662.99	4114756.94	-0.4850014	0.6346820	-11303896.98	7312571.00
Apr-Jul Diversn	-643.95	417.60	-1.5420213	0.1438984	-1588.64	300.73
Aug-Mar Exp	1038084.05	3152439.04	0.3292955	0.7464854	-6093233.94	8169402.03
Aug-Dec Out	2297247.47	939383.29	2.4454847	0.0272822	172213.22	4422281.72

regression coefficients were used to give an indication of the error in predicted values. Model predictions are no better than the estimated coefficients. The prediction error was computed by first replacing all of the regression coefficients with their upper 95 percent confidence limit and using the Model in the usual way to predict adult abundance. Then, the regression coefficients were replaced with the lower 95 percent limit and adult abundance predicted in the usual way.

The results of this analysis are shown in Figure 3. The intervals shown on Figure 3 are not confidence intervals, but rather a worst-case estimate of the error associated with Model predictions. Additional estimates could be obtained by using different combinations of confidence limits on the coefficients.

Figure 3 shows that errors in the regression coefficients could cause errors of over +/- 2000 percent in predicted adult abundance. The very wide error band is caused by structural and statistical problems such as multicollinearity, which affect the accuracy of the estimated coefficients. The large prediction error of the Model (Fig. 3) suggests that it is not useful for forecasting adult abundance.

3. FACTORS THAT INFLUENCE THE PREDICTIVE ABILITY OF THE STRIPED BASS MODEL

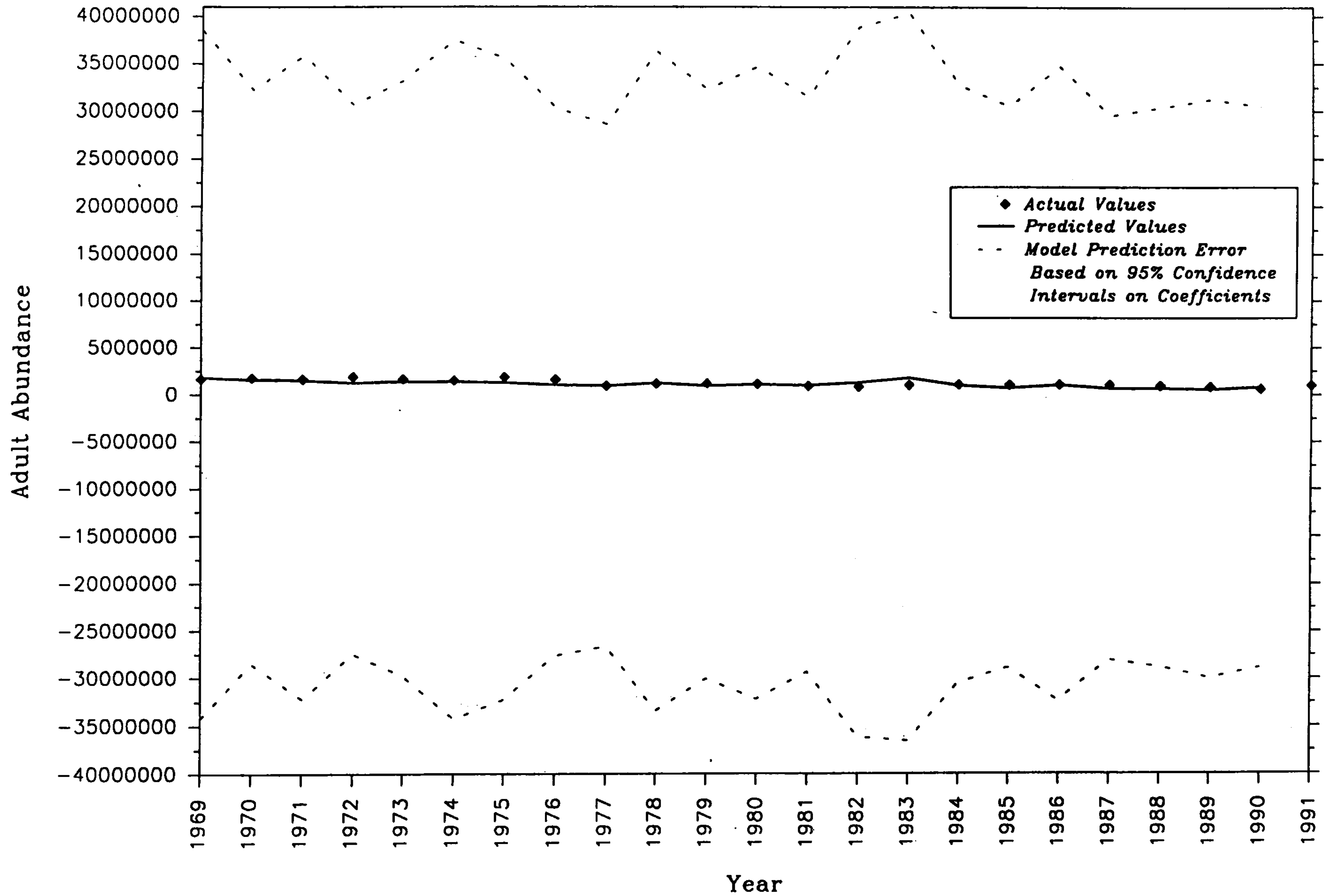
Many reviewers raised statistical criticisms of the Striped Bass Model, including multicollinearity, autocorrelation, averaging, and propagation of errors, among others. CDFG dismissed these statistical concerns, arguing that as long as the equations make biological sense, statistical flaws are unimportant. For example, in their November 21, 1991 response to comments to the State Water Resources Control Board ("SWRCB"), they state: "[w]e also believe that complete statistical rigor is not essential...Nevertheless, the model makes sense conceptually and biologically and does a good job at predicting abundance within the range of past observations."[14]

However, when the errors in CDFG's analysis are corrected, one finds that the Model does not do a good job at predicting adult abundance (Secs. 1,2). Further, some of the structural and statistical defects discussed below suggest that the Model may also have biological flaws. Regardless, a model can make biological sense, but not statistical sense, and thus not be able

---

[14] Letter from Harold K. Chadwick, CDFG, to Leo Winternitz, SWRCB, Re: Department of Fish and Game Responses to Comments on the Striped Bass Model and Additional Analyses, November 21, 1991, p. 3, Comment 6.

Figure 3.
Model Prediction Error Based on
95% Confidence Limits for the Coefficients



to make accurate predictions. In order to be useful as a management tool, a model must be able to make reasonably accurate predictions. The statistical problems raised by the reviewers are valid concerns that strongly affect the ability of the Model to make accurate predictions. Further, because the inter-relationships among variables are not accurately modelled due to multicollinearity and autocorrelation, the Model also is not useful for studying relative changes and differences among alternatives because management options necessarily involve changing the inter-relationships among variables.

The large uncertainty in the Model's predictions are due to statistical flaws, measurement errors, and model specification errors, as discussed below.

3.1. Multicollinearity

One of the principal assumptions of regression analysis is that the predictor or explanatory variables (the X's) are independent (i.e., the value of one predictor variable does not depend on the value of any of the other predictor variables). Multicollinearity or simply "collinearity," refers to the situation when the predictor variables are related to each other. Severe collinearity indicates that a substantial part of the information in one or more of the predictor variables is redundant, which makes it difficult to separate the effects of the different predictor variables on the response variable (the Y).

When collinearity is present, the standard errors associated with the regression coefficients are inflated, the coefficients themselves may be unreasonable, and confidence intervals are larger than they otherwise would be. In fact, the coefficient of determination (r^2) can be high and the regression equation can be statistically significant while the individual fitted parameters are not statistically significant and the prediction intervals are large.[15] Chatterjee and Price warned that "one should be extremely cautious about any and all substantive conclusions based on a regression analysis in the presence of multicollinearity"[16] and that multicollinearity "...can seriously limit the use of regression analysis for inference and forecasting. Extreme care is required when attempting to

---

[15] Wittink, 1988, p. 83-90.

[16] Samprit Chatterjee and Bertram Price, Regression Analysis By Example, 2nd Edition, John Wiley & Sons, Inc., New York, 1991, p. 174.

interpret regression results when multicollinearity is suspected."[17]

The six individual regression equations were tested for the presence of collinearity using principal component analysis. The analyses indicate that collinearity is a severe problem for the Delta Index equation (Eq. 4, Fig. 1) and for the Adult equation (Eq. 7, Fig. 1) and is also present to a lesser extent in the Loss Rate equation (Eq. 6, Fig. 1). This indicates that the regression coefficients in these equations may have been degraded and that confidence intervals and predictions could be improved by using better conditioned data.[18] Collinearity appears to be one of the reasons that the error in predicted values (Fig. 3) is so large.

In the Adult equation, collinearity is present because the YOY index and loss rate are correlated with each other (r^2 = 0.79, $p<0.01$). Most reviewers[19] correctly noted that in the Adult equation, the YOY index is correlated with loss rate because loss rate is the ratio of export losses to the YOY index. When the log loss rate equation is refitted using export losses rather than loss rate, the r^2 drops from 0.77 to 0.17, demonstrating that August to March exports and outflow explain very little of the variability in export losses. In contrast, replacing loss rate with export losses in the Adult equation does not significantly affect the r^2, indicating that adult abundance

[17] Chatterjee and Price, 1991, p. 184.

[18] David A. Belsley, Edwin Kuh, and Roy E. Welsch, Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, John Wiley & Sons, New York, 1980, Chapter 3.

[19] Letter from Warren J. Shaul, Jones & Stokes, to Pete Chadwick, CDFG, Re: Comments on California Department of Fish and Game's Proposed Striped Bass Impact Methodologies for the Bay-Delta Water Right Environmental Impact Report (EIR), November 1, 1992, pp. 3-4; Letter from Roger K. Patterson, Bureau of Reclamation, to Pete Chadwick, CDFG, Re: Comments on Draft Striped Bass Paper of Department of Fish and Game, Presented at State Water Resources Control Board Workshop on October 25, 1991 (Bay-Delta Hearings) (Water Rights), November 18, 1991, p. 2, Comment 5a; Letter from Bruce K. Orr, EA Engineering, Science, and Technology, to William Johnston, MID, Re: Comments on the CDFG report by Kohlhorst, Stevens, and Miller, November 1, 1991, p. 3; Letter from James H. Cowan, Jr., University of South Alabama, to Randy Brown, DWR, Re: Comments on Striped Bass Model, January 28, 1992, p. 1, Comment 3; Joseph G. Loesch, Critique of a Means of Evaluating Impacts of Alternative Outflow and Export Criteria on Striped Bass in the Sacramento-San Joaquin Estuary, December 2, 1991, p. 2.

is not significantly affected by August to March exports and outflow.

In the Delta index equation (Eq. 4, Fig. 1), collinearity is present because April to July outflow and April to July outflow squared are highly correlated ($r^2 = 0.99$, $p<0.01$). As a result, at the 5 percent significance level, the fitted coefficients for the intercept (p=0.27), log April to July outflow (p=0.21), and squared log April to July outflow (p=0.20) are not statistically significant, even though the regression equation as a whole is statistically significant (p=0.026). Further, the negative coefficient for squared log April to July outflow is not consistent with the positive coefficient for log April to July outflow. In fitting polynomials, collinearity can be eliminated by centering the measured predictor variables on their mean values before computing the power (e.g., squared) and fitting the equation.

In the loss rate equation, collinearity is present because August to March exports are correlated with August to December outflow. As discussed in Section 3.2 and pointed out by Willits, "the predictors that are of central interest (water exports) are correlated with time (they increase over time) so this creates collinearity between exports and time. Moreover, most of the additional predictors that have been suggested also vary with time, and so it's rather difficult to separate between an effect due to water exports and due to other variables that vary similarly..."[20]

In some cases, collinearity may not adversely affect the ability of a model to predict, so long as the relationship between the predictor variables does not change. For example, if the relationship between YOY index and loss rate remained constant in the future, the mis-specified Model could still be used to make reasonable predictions if there were no other problems. However, if the relationships between these correlated variables change, the Model would not be able to make accurate predictions. For example, if export losses were reduced by structural changes in Clifton Court Forebay, the relationship between loss rate (export losses/YOY index) and YOY index would change, invalidating predictions with the Model. Similarly, if reservoir operations were changed, the relationship between outflow and exports could change.

Because the Model would be used to explore exactly these types of options, it is of questionable utility as a management tool because it does not model the actual inter-relationships among variables. Thus, regardless of whether the Model makes biological since, collinearity is a serious concern because it

---

[20] Willits, March 29, 1992, p. 2-3.

limits the utility of the Model for predicting and for exploring management options.

3.2. Autocorrelation

For time series data, autocorrelation or "self-correlation" means that individual values in a time series, such as Delta outflow or exports, are correlated with past or future values in the series. In other words, for example, April Delta outflow may depend on March Delta outflow, and so forth. The presence of autocorrelation in time series data is usually tested by using the first-order autocorrelation coefficient, which is nothing more than the correlation coefficient (r) obtained by regressing the time series with the same series shifted by one observation (e.g., $Y_2,Y_1;Y_3,Y_2;Y_n,Y_{n-1}$).

First-order autocorrelation coefficients were estimated for all data sets used to developed the Striped Bass Model. The results, shown in Table 3, indicate that strong autocorrelation is present in all individual time series, except log April to July outflow. Autocorrelation in the response and predictor variables can cause badly misleading results when estimating regression models due to time-lagged relationships, feedback from predictor to response variables, and correlated regression residuals, among others. These problems, which are discussed below, can be corrected by using dynamic regression which takes into account the time-lagged relationships between response and predictor variables.[21]

3.2.1 Time-Lagged Relationships. The response variable Y_t may be related to the predictor variable X_t with a time lag; that is, Y_t may be related to X_{t-1}, X_{t-2}, X_{t-3}, and so forth, in addition to (or instead of) being related to X_t. For example, instead of the usual first order regression equation

$$Y_t = mX + b \tag{2}$$

which is the basis of the Striped Bass Model, the correct relationship may be

$$Y_t = m_oX_t + m_1X_{t-1} + m_2X_{t-2} + b \tag{3}$$

where b is a generic constant that would have different values in equations (1) and (2). If equation (3) is the correct formulation, useful information is lost, for example, about the roles of X_{t-1} and X_{t-2} in explaining the change in Y_t. Thus, the estimate of the residual variance would be larger than necessary, and predictions would likely be less accurate than otherwise.

---

[21] Alan Pankratz, Forecasting with Dynamic Regression Models, John Wiley & Sons, Inc., New York, 1991, p. 8-15.

TABLE 3. AUTOCORRELATION IN DATA SETS USED TO DEVELOP STRIPED BASS MODELS

Variable	*n*	Lag-1 Autocorrelation Coefficient	Standard Error	Bartlett's[1] t	Student's t ($\alpha = 0.05$)	Significant[2]
Adults	23	0.770	0.209	3.68	2.07	yes
Log Apr-Jul Exports	33	0.371	0.174	2.13	2.04	yes
Log Aug-Dec Outflow	33	0.631	0.174	3.63	2.04	yes
Aug-Mar Exports	32	0.758	0.174	4.36	2.04	yes
Delta YOY Index	31	0.776	0.180	4.31	2.04	yes
Eggs	23	0.750	0.209	3.59	2.07	yes
Log Weighted Mean Loss Rate	23	0.731	0.209	3.50	2.07	yes
Log Apr-Jul Outflow	33	-0.065	0.174	-0.37	2.04	no
Log Loss Rate	31	0.569	0.180	3.16	2.04	yes
Export Losses	31	0.476	0.180	2.64	2.04	yes
Suisun YOY Index	31	0.437	0.180	2.43	2.04	yes
Weighted Mean YOY	23	0.820	0.209	3.92	2.07	yes

[1] Bartlett's t is the ratio of the Lag-1 autocorrelation coefficient to the standard error.

[2] If the absolute value of Bartlett's t is greater than Student's t, the Lag-1 autocorrelation coefficient is statistically significant at the 5 percent level in a two-tailed test. This means that the data set is autocorrelated.

Further, the estimate of the fitted regression coefficients associated with X_t will be biased and inconsistent if the excluded variables (X_{t-1} and X_{t-2}) are correlated with the included variable X_t. These consequences of autocorrelation affect the ability of the Model to make accurate predictions and the confidence that one can place in predicted values.

The Adult equation contains time-lagged relationships, which CDFG attempted to address by using the weighted mean YOY index and weighted mean loss rate 3 to 7 years earlier. As pointed out correctly by several reviewers[22], this is an inaccurate way to model time-lagged relationships because useful information on the correct relationships is lost, reducing the accuracy of predictions. Because time averaging is not used when the Striped Bass Model is applied, inaccuracies are compounded.

Although no tests were performed, it is likely that time lagged relationships also exist in the Suisun index, Delta index, and loss rate equations. All of these equations should be refit using dynamic regression to improve their predictive ability.

3.2.2 Misleading Correlations. As explained by Dr. Pankratz in Forecasting with Dynamic Regression Models, "if there are common elements in the autocorrelation structure of Y_t and X_t, then an ordinary regression equation can show a strong relationship between these variables when, in fact, X_t has no real explanatory power."[23] Several reviewers correctly pointed out that the apparent good correlations for some regression equations in the Striped Bass Model are due to the fact that the predictor variables are related to time in the same way as the response variable.

For example, Jones & Stokes noted that both YOY, exports, and outflow are related to time in the same way, suggesting that the correlation between exports and the Delta YOY equation is due to their common relationship to time.[24] Common time relationships are present in the Suisun Index, Delta Index, Adult, and Loss Rate equations. As a result of these common time relationships, correlation coefficients mistakenly show strong relationships among variables that may not be otherwise related, adversely affecting the ability of the Model to make accurate predictions.

---

[22] Letter from David G. Hankin, Humboldt State University, to Randall L. Brown, DWR, January 7, 1992, p. 2, Comment 1; Hanson, November 7, 1991, p. 4; Willits, March 29, 1992, p. 2.

[23] Pankratz, 1991, p. 12.

[24] Shaul, November 1, 1991, p. 3.

The CDFG investigated the effect of common time relationships by detrending adult abundance, weighted mean YOY index, export losses, and loss rate. They found that when the time trend is removed, correlations of adult abundance with weighted mean YOY index and weighted mean loss rate disappear.[25] This suggest that the good correlation for the Adult equation ($r^2 = 0.71$, $p<0.01$) is largely due to time trends, as suggested by the reviewers, rather than any fundamental relationship between variables.

The effect of time trends was further explored by regressing adult abundance on time alone. The resulting regression equation[26] explains more of the variance and has lower p values for the regression as a whole and individual coefficients than the Adult equation. Thus, time alone is a better predictor of adult abundance then loss rate and the YOY index.

Although the CDFG believes that this does not mean that the strong relationships are "...spurious, only that they are mostly due to simultaneous major changes in yoy striped bass abundance, entrainment losses, and loss rate that have occurred over time,"[27] they are incorrect. They have demonstrated that the good correlation obtained for the Adult equation is largely due to common time relationships, rather than to a fundamental relationship between adults and loss rate and YOY index. Even though, theoretically, there may be a biological relationship between adults, the YOY index and loss rate, it is obscured by common time trends and is not modelled correctly. For example, if the time trend associated with Delta outflow shifted relative to the time trend in adults due to, for example, a reduction in export losses in Clifton Court Forebay, the Model would no longer make accurate predictions.

3.3. Differences In Development And Application

In developing the Model, the mean YOY index and loss rate 3 to 7 years earlier are used to estimate adult abundance, while in applying the Model, point estimates of the YOY index and loss rate are used. Similarly, in developing the Model, individual

---

[25] CDFG, Exhibit WRINT-DFG-Exhibit 3, Table 11 and p. 21 and 24.

[26] Regressing adult abundance on year yields the following equation:

$$\text{Adults} = -47{,}513(\text{Year}) + 95{,}285{,}912$$

where $r2 = 0.74$ ($p<0.01$) and the fitted coefficients are statistically significant ($p<0.01$).

[27] Id., p. 25.

weighting factors for each year are used, whereas in applying the Model, the mean annual weighting factor of 0.4238 is used. Finally, in applying the Model, adult abundance is multiplied by 1.08 when adults in Equation 1 average 1.7 million (prior to 1976) and by 0.936 when adults in Equation 1 average 1.0 million (after 1976).[28] As pointed out by Hanson,[29] developing a model using a five year average and applying it for single years can introduce considerable error into Model predictions.

The effect of differences in Model development and application was tested by repeating the validation shown in Table 1, using five-year back averages for the YOY index and loss rate. The predictions for the backaveraged case (comparable to Model development) are 949,027 bass in 1987; 681,645 in 1988; 869,404 in 1989; and 637,221 in 1990. On average, the ratio of the actual to predicted adults is 1.05 for the development case and 1.50 for the application case (predicted using point estimates). This demonstrates that significant errors are introduced into Model predictions when the Model is applied without regard for time lagged dependencies (Sec. 3.2.1). Flow standards set using this Model would have to be met for five consecutive years to be consistent with Model assumptions. This problem can be cured by using dynamic regression to accurately model the time-lagged relationships of the Model, as discussed in Section 3.2.

3.4. Data Inconsistencies

The periods of record vary from data set to set. The only period common to all data sets is 1969 to 1976. Thus, the regressions for the six equations may capture different relationships due to fundamental changes in the Estuary over time. For example, regressions based on data sets that extend back to 1959 (Delta and Suisun index, loss rate) capture conditions present in the Estuary before the State Water Project pumps came on line in 1968, while those based on post-1969 data (eggs, residual YOY, adult abundance) would not reflect pre-SWP conditions. Likewise, regressions based on data sets that end in 1976 (Delta and Suisun index) fail to capture changes that occurred after the 1977 drought.

Data points for 1966 and 1983 appear to have been inconsistently included or excluded from the various data sets, which could affect the accuracy of predictions made using the Model. For example, 1983 was excluded in developing the Residual YOY equation (Eq. 2, Fig. 1), while 1983 was apparently included in developing the Egg, Loss Rate, and Adult equations (Eqs. 1,6,7, Fig. 1). As another example, the 1966 YOY index was not

---

[28] WRINT-DFG Exhibit 3, p. 39, 41.

[29] Hanson, November 7, 1991, p. 4.

used to develop Equations (2) through (4) (Fig. 1), while an estimate of the 1966 YOY index was used in developing the Adult equation (Eq. 7, Fig. 1).

The use of data sets for different time periods and the use of varying assumptions about 1966 and 1983 raise doubts about the adequacy of the Model for prediction. The relationships among the various variables and equations have shifted since 1959 and the six equations reflect different realities. As discussed in Section 3.5.2, this is compensated for by lumping the majority of the error into a fudge factor referred to as the "residual YOY."

3.5. Specification Errors

In addition to the purely statistical flaws that lead to inaccurate predictions, mis-specifying the Model can also cause inaccurate predictions. Specification errors present in the Model include omission of variables, the use of fudge factors, and using the same variable as both the response and predictor variable. Sometimes these problems actually cause collinearity and autocorrelation.

3.5.1 Missing Variables. The Striped Bass Model assumes that the only significant variables affecting the abundance of striped bass are Delta outflow and exports. In WRINT-DWR Exhibit 30, a number of other factors were identified that potentially affect striped bass, including weather phenomena, over-fishing, poaching, pollutants, introduction of exotic species, and agricultural diversions.[30] One reviewer of the Model also noted that "[w]hile certainly outflow and exports are major factors, they are not the only factors and the model does not appear to recognize anything else, as for example, pesticides and other pollutants, introduced species, changes in environmental conditions (for example, water clarity), or plantings of hatchery raised striped bass."[31] The CDFG admitted that other factors may also affect bass abundance but felt that "...their impact must be minor for adults to be so well predicted by the model."[32]

However, Sections 1 and 2 above demonstrate that the Model does not accurately predict bass abundance when CDFG's analysis are corrected. It is possible that the poor predictions are due,

---

[30] Randall Brown, Bay/Delta Fish Resources, WRINT-DWR Exhibit 30, July 1992.

[31] Patterson, November 21, 1991, p. 2, Comment 6.

[32] Letter from H.K. Chadwick, CDFG, to Roger K. Patterson, Bureau of Reclamation, Re: Responses to Comments on Striped Bass Model, November 22, 1991, p. 2, Comment 6.

in part, to the failure to specify one or more important predictor variables. Additional work is required to explore the influence of non-flow predictor variables on adult abundance.

3.5.2 Fudge Factors. The Delta and Suisun Index equations (Eqs. 3,4, Fig. 1) overpredict the YOY index after 1976 (Fig. 2b, 2e).[33] To correct for overprediction, CDFG fit a Beverton-Holt relationship to the residuals (actual YOY-predicted YOY) from the Suisun and Delta index regression equations and used it in the Model to predict the YOY index (Fig. 1, Eqs. 2, 4). Of the six regression equations comprising the Striped Bass Model, the residual YOY equation has the poorest coefficient of determination ($r^2 = 0.41$) and the widest confidence intervals (Fig. 2d), and accounts for much of the uncertainty in Model predictions. Absent compelling biological evidence that a stock recruitment or linear relationship between residual YOY and eggs is actually present in the Estuary, the residual YOY equation is nothing more than a fudge factor that contains all of the error associated with the Egg, Delta Index, and Suisun Index equations (Eqs. 1,3,4, Fig. 1).

Several reviewers were not persuaded that a stock recruitment relationship makes sense. Jones & Stokes pointed out that a stock-recruitment type relationship (e.g., one expressed by an equation of the form $Y = 1/(a + b/X)$ could be due to other changes in the Estuary (e.g., food supply, introduced species) after 1977 and that at any rate, the stock recruitment relationship likely would not have been constant since 1969.[34] Jones & Stokes' concerns are supported by Cowan[35] and EA Engineering, Science, and Technology.[36] Cowan suggests that slight, undetectable changes in survival could have a big impact on the population. EA Engineering points to the need to use age-structured population estimates and age-specific fecundities to directly examine the egg-to-YOY relationship to verify that a stock recruitment relationship actually exists and to determine what portion of the variability is explained by stock recruitment versus environmental variables (rather than assigning 100 percent to stock recruitment). If stock recruitment or some other relationship cannot be verified, the residual YOY should be abandoned since it would have no predictive ability and would be no more than a fudge factor.

---

[33] WRINT-DFG-Exhibit 3, p. 27 and Figure 7.

[34] Shaul, November 1, 1991, p. 1.

[35] Cowan, January 7, 1992, p.3.

[36] Orr, November 1, 1991, p. 2, Comment 1.

A fudge factor is also applied to the Adult equation "to make predicted values better mimic observed values..."[37] The factor is 1.08 for an average adult abundance of 1.7 million and 0.936 for an average adult abundance of 1 million. The only justification given for this adjustment is to make the predicted values better fit the measured data. It is well known that a model that is tightly fit using fudge factors is not robust and has limited predictive ability.

3.6. Measurement Errors

The difficulty in collecting precise fishery data is well known. Some of these problems are discussed elsewhere.[38] As several reviewers have pointed out, measurement errors associated with variables in each equation are propagated from one component of the Model to another, contributing to inaccurate predictions.[39] These are valid concerns, and error propagation is treated in most basic statistic texts on data reduction.[40] Part of the uncertainty represented by the wide confidence bands shown in Figure 2 and large prediction error shown in Figure 3 is due to measurement errors that have been propagated through the Model.

---

[37] WRINT-DFG-Exhibit 3, p. 39.

[38] Donald E. Stevens, Striped Bass (Morone saxatilis) Monitoring Techniques in the Sacramento-San Joaquin Estuary, In: Proceedings of the Conference on Assessing the Effects of Power-Plant-Induced Mortality on Fish Populations, Webster Van Winkle (Ed.), Pergamon Press, New York, 1977, pp. 91-109; CDFG, Factors Affecting Striped Bass Abundance in the Sacramento-San Joaquin River System, Exhibit 25, 1987; CDFG, Striped Bass Egg and Larval Monitoring, and Effects of Flow Regulation on the Larval Striped Bass Food Chain, in the Sacramento-San Joaquin Estuary, Final Report to SWRCB, May 1988.

[39] Orr, November 1, 1991, p. 3, Comment 3; Shaul, November 1, 1991, p. 1.

[40] See, for example, Philip R. Bevington, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill Book co., New York, 1969, Chapter 4.

TABLE A-2 (Continued). DAYFLOW

SEP		OCT		NOV		DEC	
Exp	Out	Exp	Out	Exp	Out	Exp	Out
1937	9958	1288	5683	679	6016	258	6938
2015	5500	1573	5013	595	13543	54	19090
2165	5649	1440	4260	684	8251	253	16140
2141	8515	1431	42900	804	16351	57	35013
2149	13480	2081	14978	545	27945	164	22825
2434	9442	2269	8118	746	17243	61	108447
2072	12917	1806	15091	664	27350	57	30136
2315	6905	1904	6610	1026	21505	501	60456
2634	16556	1795	16749	1132	16202	675	20498
5603	6004	6249	5453	5043	11120	3765	25682
2514	20188	2006	19484	1072	19964	821	46190
3089	14587	2585	13423	2031	26117	1915	85369
3907	19659	3812	13957	3047	13743	2435	23967
6106	10476	5847	11919	3309	25943	3432	27133
5768	11153	5927	14071	4903	59945	3338	76406
5055	20981	4595	18529	1949	23991	2814	28017
7799	13419	7561	16900	8010	17921	7820	19953
8331	3670	4606	3623	4244	3644	2784	4213
1959	2791	854	2075	2757	4004	5983	8488
7487	11793	5124	9633	5572	10928	6051	8779
9262	5058	7730	7821	5857	12176	5973	19029
7683	9902	6694	7368	6456	6670	6763	12488
6797	4690	5930	5218	4717	35971	5166	86579
5284	25926	5285	22986	6073	39152	8428	88937
4203	31501	2496	32293	1754	74138	2142	155458
5498	13650	5605	11916	7996	25953	8464	31067
8719	3211	7703	3378	7329	6891	9858	9431
10490	10778	7566	10628	6860	7732	7260	8987
9070	1790	5908	3789	5460	4291	8986	9455
8119	2391	5631	3226	6087	6660	7184	7259
10753	6555	10529	4926	10379	5503	10442	4422
5918	2594	3549	3498	3857	4558	5205	6425
4235	3884						

TABLE A-2. DAYFLOW

Delta Outflow and Exports -- Monthly Means (cfs)

	JAN		FEB		MAR		APR		MAY		JUN		JUL		AUG	
Year	Exp	Out	Exp	Out	Exp	Out	Exp	Out	Exp	Out	Exp	Out	Exp	Out	Exp	Out
1959	303	32890	630	58739	2023	27692	2757	11607	2661	7303	3564	1322	4005	2561	3435	5194
1960	255	14231	630	49316	2280	33273	2605	16878	2688	12407	3825	3847	4095	2244	3556	2731
1961	307	15580	813	41997	2057	28425	2900	13397	2837	8580	3992	3541	4656	1672	3923	4007
1962	404	11132	257	74766	905	47503	2761	27385	2963	18173	3799	10317	4229	2795	3722	5028
1963	507	20995	815	103173	1823	29180	1231	102776	2774	53124	3543	19180	4198	5639	3865	5038
1964	577	29970	1588	20915	2172	13073	3065	9187	3261	9784	3795	5302	4619	3185	4242	4704
1965	224	135616	1561	56241	2218	27860	1204	56912	3193	32370	3694	16190	4361	5865	3904	8467
1966	115	43464	914	36316	2484	24328	3108	18946	3381	9835	4075	2460	4597	3155	4280	4846
1967	816	62522	765	84142	2002	56325	1207	77685	1921	74550	2162	61265	2697	23864	4292	9827
1968	1168	24257	1828	52061	4487	40314	5380	9932	5611	6737	4708	3666	5168	3684	4868	5264
1969	7212	123140	4588	159046	3123	93506	2932	69375	2990	64564	2214	46596	3252	13143	5018	12458
1970	1116	193121	1950	111326	2265	55986	4653	11027	4012	10761	4997	6214	5227	5256	4608	7947
1971	1905	64152	3139	34211	4702	32069	4431	36983	4549	26406	5768	21218	6509	11654	6701	12988
1972	1615	21339	3731	21968	6682	18078	6356	7542	6495	5140	8111	2891	5001	6211	6072	6470
1973	2962	101685	1178	102165	1282	76907	3352	22191	6501	11699	7355	7211	7693	4599	7774	5963
1974	1974	138699	5455	59178	6275	77575	4203	109547	7130	25544	9130	16943	10691	9365	9474	12783
1975	5472	17489	6717	57330	6079	66834	6304	34519	5583	28796	4520	22508	5184	11129	8988	9523
1976	8260	9346	7793	7495	8351	7858	5037	8833	5488	4066	4152	3915	4109	4343	6836	4509
1977	7042	4365	4335	4924	3813	3070	1295	3083	2987	3999	739	2521	845	3212	1529	2514
1978	9886	66171	10309	56159	5919	85544	3271	61276	3058	40874	7621	9086	8088	3974	8425	5927
1979	4104	30522	2939	46341	4348	38086	5882	14485	6245	13435	6341	5326	9339	5384	10362	3475
1980	6377	118212	6185	121653	4341	99171	5343	28689	4630	20912	5961	14870	6869	11191	9212	4253
1981	8264	18326	7239	21174	4834	26467	8090	11653	4478	9143	4032	4596	7046	5296	9315	3161
1982	5176	97706	9452	88727	10417	80089	9603	142203	5994	57876	3935	28515	4032	16849	8096	13438
1983	10085	89755	10246	175757	5371	266688	3814	118109	3293	98707	5010	71038	5207	43860	7190	24567
1984	1720	100906	5767	41515	6916	34929	7685	14732	5929	11204	6165	8038	9457	10252	9515	8272
1985	5835	15120	7614	15590	8616	10432	7342	6913	6215	7378	6530	5215	9465	4934	10111	2325
1986	9070	15209	9821	205414	2231	169448	3273	46572	5361	15911	6076	9322	8607	7384	9951	5135
1987	6251	10819	6845	16859	5600	22916	7021	6291	5313	4951	5183	3496	8952	3829	9800	2851
1988	10418	19593	10023	3045	8441	4542	8570	11499	6263	4747	5900	3197	7967	3920	8794	2472
1989	10194	3635	8201	6405	10261	38951	10447	11808	6220	7531	5271	6317	9515	6357	11319	4634
1990	10621	9913	10553	6815	10558	3906	9667	6041	3391	7837	3492	4999	6245	4115	6676	4612
1991	4912	4013	4521	7420	9763	24626	7499	3787	2685	3998	1925	4169	2574	3479	3818	2696

TABLE A-1. DATA USED TO DEVELOP STRIPED BASS MODEL

Year	Eggs	Adults (wild)	Adults (w/ hatchery)	Suisun	Delta	YOY	Export Losses	Apr-Jul Outflow	Apr-Jul Exports	Apr-Jul Diversn	Aug-Mar Exports	Aug-Dec Outflow
1959				3	30.7	33.7	1,668,181	5,698	3,247	6,350	1,345	6,758
1960				13.6	32.0	45.6	2,414,048	8,844	3,303	6,407	1,371	9,175
1961				6.4	25.2	31.6	2,989,860	6,798	3,596	6,700	1,254	7,661
1962				32.1	46.8	78.9	2,694,309	14,668	3,438	6,541	1,413	21,561
1963				43.5	38.2	81.7	1,909,488	45,180	2,937	6,040	1,643	16,853
1964				20.7	54.7	75.4	834,137	6,865	3,685	6,788	1,719	29,591
1965				67.8	49.4	117.2	2,149,306	27,834	3,113	6,216	1,502	18,792
1966							4,811,106	8,599	3,790	6,894	1,701	20,064
1967				73.6	35.1	108.7	2,220,430	59,341	1,997	5,100	2,251	15,966
1968				17.7	39.6	57.3	4,473,927	6,005	5,217	8,320	5,056	10,705
1969	319.1	1,646,026	1,646,026	40.2	33.6	73.8	2,360,665	48,420	2,847	5,950	2,095	23,657
1970	267.2	1,727,394	1,727,394	41.9	36.6	78.5	10,053,728	8,315	4,722	7,826	2,997	29,489
1971	255.7	1,599,715	1,599,715	45	24.6	69.6	9,185,289	24,065	5,314	8,418	3,991	16,863
1972	265.3	1,882,907	1,882,907	21.1	13.4	34.5	3,568,058	5,446	6,491	9,594	3,774	16,388
1973	354.2	1,637,159	1,637,159	47.1	15.6	62.7	11,176,662	11,425	6,225	9,329	5,177	33,508
1974	300.5	1,477,213	1,477,213	63.4	17.4	80.8	21,807,197	40,350	7,789	10,892	5,269	20,860
1975	434.2	1,849,770	1,849,770	42.1	23.4	65.5	18,620,410	24,238	5,398	8,501	8,073	15,543
1976	269.0	1,581,076	1,581,076	14.8	21.1	35.9	3,615,048	5,289	4,697	7,800	5,249	3,932
1977	231.4	924,301	924,301	0.7	8.3	9	1,428,579	3,204	1,467	4,570	4,900	3,974
1978	129.6	1,151,642	1,151,642	13.1	16.5	29.6	2,798,063	28,803	5,510	8,613	5,506	9,412
1979	90.3	1,155,701	1,155,701	11.5	5.4	16.9	2,746,237	9,658	6,952	10,055	7,011	9,512
1980	125.7	1,115,999	1,115,999	11.2	2.8	14	2,853,551	18,916	5,701	8,804	7,143	8,136
1981	99.9	911,300	911,300	13.7	15.4	29.1	2,536,186	7,672	5,912	9,015	7,121	27,124
1982	68.5	825,126	825,126	39.2	9.5	48.7	2,820,971	61,361	5,891	8,994	7,359	38,088
1983	121.8	1,009,748	1,009,748			15.4	182,129	82,929	4,331	7,434	4,024	63,591
1984	77.1	1,042,668	1,048,244	20	6.3	26.3	6,486,207	11,057	7,309	10,412	7,393	18,172
1985	97.9	1,024,188	1,038,126	4.1	2.2	6.3	3,909,444	6,110	7,388	10,491	8,105	5,047
1986	165.2	1,037,127	1,064,142	41.1	23.8	64.9	19,221,607	19,797	5,829	8,933	7,603	8,652
1987	176.1	998,349	1,037,617	5.3	7.3	12.6	14,031,947	4,642	6,617	9,721	8,513	4,435
1988	176.4	903,899	980,068	0.7	3.9	4.6	1,949,138	5,841	7,175	10,278	8,059	4,402
1989	103.5	786,380	911,745	2.0	3.1	5.1	6,092,415	8,003	7,863	10,967	10,644	5,208
1990	91.6	591,241	679,376	1.5	2.8	4.3		5,748	5,699	8,802	5,550	4,337
1991	160.2	961,063	1,161,071	1.6	3.9	5.5		3,858	3,671	6,774		

APPENDIX A

DATA USED TO DEVELOP THE STRIPED BASS MODEL