

PRE-1994 SEASON PROJECTION OF RUN-TIMING CAPABILITIES USING PIT-TAG DATABASES

Prepared by:

John R. Skalski
Gennady Tartakovsky
Steven G. Smith
Peter Westhagen

Center for Quantitative Science
School of Fisheries
University of Washington
Seattle, WA

Prepared for:

U.S. Department of Energy
Bonneville Power Administration
Environment, Fish and Wildlife
P. O. Box 3621
Portland, OR 9720X-3621

Project Number 9 I-05 1
Task Number DE-AT79-9 1 BP1 6570
Contract Number DE-B179-87BP35885

April 1994

Executive Summary

Regulating the timing and volume of water released from storage reservoirs (often referred to as flow augmentation) has become a central mitigation strategy for improving downstream migration conditions for juvenile salmonids in the Snake River. The success of the flow augmentation, in turn, depends on releasing reservoir waters when and where wild smolt will benefit the most. This requires the ability to predict in real time the status and trend in the outmigration of listed threatened and endangered stocks. This study evaluated the feasibility and the performance of two alternative statistical algorithms to predict outmigration status of Snake River wild spring chinook.

Using historical trends in PIT-tag detections of wild chinook smolt at Lower Granite Dam, pattern recognition techniques were developed to predict the percent of the run-to-date and days to a specific percent of the run. The statistical methods are based on algorithms that smooth historical trends in PIT-tag arrivals and a generalized least squares decision criterion. The methods were evaluated for 16 different river runs of chinook, as well as composites over various river basins. A bootstrapping approach across historical years provided the means to measure the accuracy and precision of predictions and construct approximate interval estimates.

The recommended predictors have an average error rate across stocks of fish and seasons of $\pm 9.6\%$ about the true percent of the run-to-date. The best performance was for Catherine Creek with an average error of $\pm 4.2\%$. The worst performance occurred at Big Creek with an average error of $\pm 19.0\%$. An interactive graphical analysis program written in C-language for an **X-Windows**® environment has been developed to analyze outmigration data for select stocks of Snake River spring chinook.

Table of Contents

Section	page
Introduction	1
Methods	4
Data Sets	4
Analytical Procedures	6
Overview	6
Characterizing Historical Outmigration Patterns	10
Predictions	18
Forecasting	21
Results	21
Alternative Predictive Algorithms	21
Performance of Predictors	21
Discussion	25
Literature Cited	28
Appendix A: Additional Details on Prediction	29
Appendix B: Jackknife Results of Predictive Performance	35

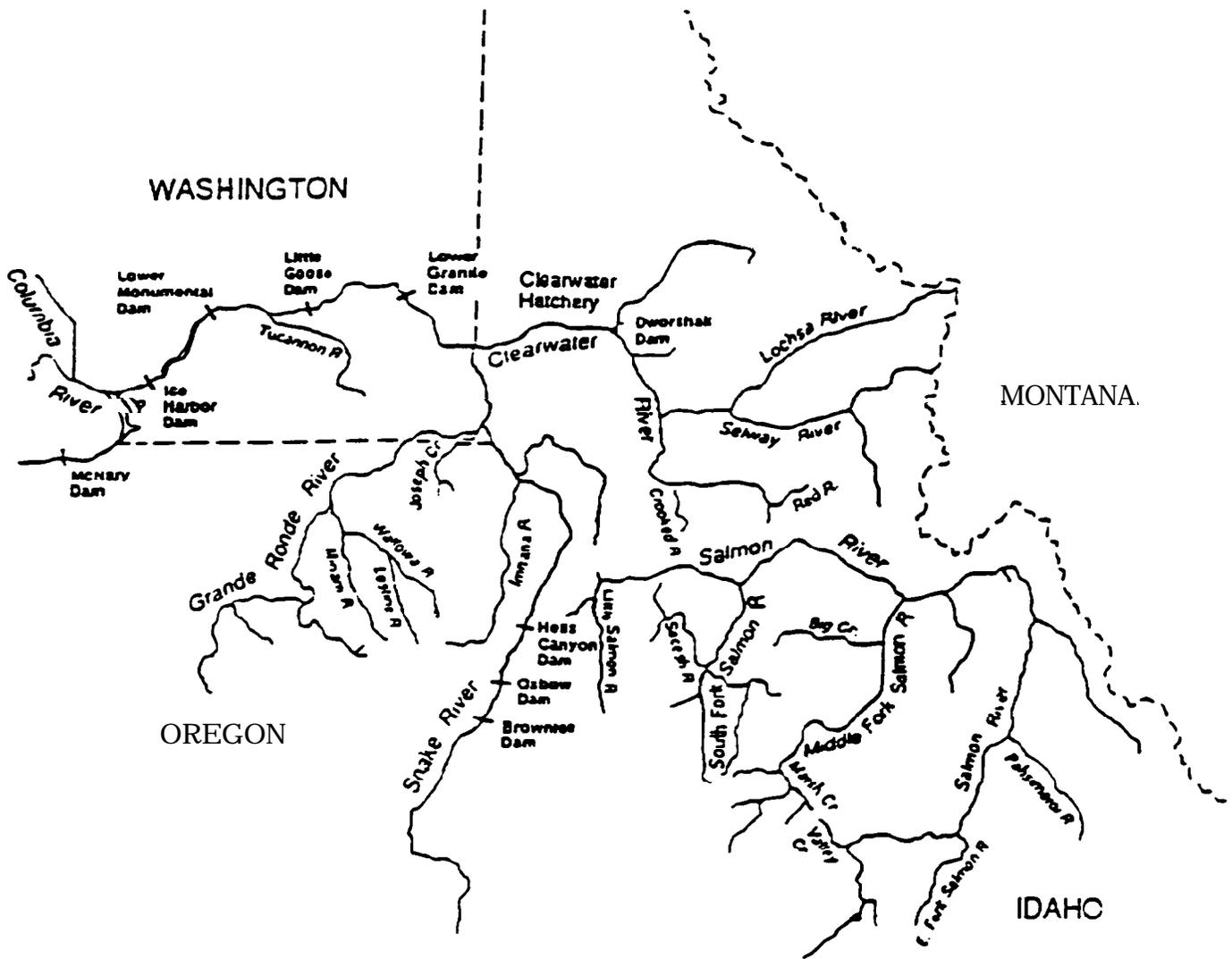
Introduction

Three ecologically significant units (ESU) of Pacific salmon have been designated as either threatened or endangered (T&E) under the Endangered Species Act (ESA) in the Snake River Basin: spring/summer chinook, fall chinook, and sockeye salmon. The tributary populations of spring/summer chinook reside primarily in the Salmon and Grande Ronde drainages, and Imnaha River, all of which are situated upstream of Lower Granite Dam. Additionally, a small population resides in the Tucannon River, which enters the Snake River downstream from Little Goose Dam (Figure 1). Fall chinook reside in the Snake River between Lower Granite and Hell's Canyon Dams. The sockeye reside in the uppermost portion of the main Salmon River in the Stanley Basin. Except for the Tucannon River population, all others reside upstream from Lower Granite Dam on the Snake River.

Regulating the timing and volume of water released from storage reservoirs (often referred to as flow augmentation) has become a central mitigation strategy for improving downstream migration conditions for juvenile salmonids in the Snake River. Threatened and endangered salmon stocks have received increased priority with regard to the timing of flow augmentation. Generally, the strategy is to release water from storage reservoirs at times when the listed stocks are in geographic locations where they encounter the augmented flows.

In the Snake River Basin, regulated water enters the system at two locations, below Hell's Canyon Dam on the Snake River and below Dworshak Dam on the Clearwater River. The preponderance of regulated water available for fish passage is provided by Dworshak Reservoir. None of the listed stocks are located in the Clearwater drainage. Thus, listed stocks must migrate to below the confluence of the Clearwater and Snake Rivers before they are fully exposed to augmented flows. The confluence of the Clearwater and Snake Rivers forms the approximate upper boundary of the Lower Granite Reservoir. Determining when stocks are in the vicinity of Lower Granite Dam and Reservoir is a chief consideration in requesting flow augmentation.

Figure 1. Locations of key tributaries used in the evaluation of real-time predictions of outmigration timing.



It is possible to identify the presence of wild T&E populations in this reach. Since 1988, wild salmon have been PIT-tagged as parr while residing in their natal streams. The majority of PIT-tagging has been conducted by investigators at the National Marine Fisheries Service, Idaho Department of Fish and Game, Oregon Department of Fish and Wildlife, and the U.S. Fish and Wildlife Service. The detection of tagged individuals at Lower Granite Dam provides a measure of the temporal and spatial distribution of each wild population. Investigators have described the migrational timing for spring/summer chinook (Mathews et al. 1992, Kiefer and Forster 1992), fall chinook and sockeye (FPC 1992). However, there have been no quantitative measures or procedures developed that permit fishery and water managers to confidently predict the proportion of any population, or assemblage of tributary populations, that have arrived at the index site in real-time. Nor has it been possible to forecast elapsed time to some future percentile in a migration. There is a critical need for such capability, since stored water volumes are limited and it is advantageous to weigh water allocation to coincide with the presence of an ESU, some population within the ESU, or to equitably allocate water among ESUs.

The objectives of this investigation are as follows:

1. Describe the variability in the shape of the passage distributions at Lower Granite Dam for tributary populations, assemblages of populations, and the composite ESU.
2. Based on historical passage distributions, determine if it is practical to predict in real-time the proportion of a population, or assemblages of populations, that have arrived at the index site.
3. Forecast dates at which specified percentiles of the population distribution will occur.
4. Define the precision associated with the predictions or forecasts.

Methods

Data Sets

We first surveyed all tributary and mainstem populations of wild salmon that have been PIT-tagged in any year. Marking activities were first initiated in 1988, at least for some stream populations of spring/summer chinook. For a few populations, migrational timing data are available for as many as six years, 1988-1993. Populations of fall chinook and sockeye salmon have also been PIT-tagged over this period. However, our analyses treat only spring/summer chinook at this time. Sockeye have simply been tagged in far too few numbers to confidently describe the passage distribution at Lower Granite Dam. To illustrate, in 1991-1993, only 10, 11, and 6 PIT-tagged Snake River sockeye, respectively, were detected passing Lower Granite Dam.

Subyearling fall chinook residing in the Snake River between Hell's Canyon Dam and Lower Granite Dam have been PIT-tagged since 1991. However, at this time, some of the tag/recapture designations in the PITTAGIS system are uncertain (Carter Stein, personal communication, 1993). The database manager is currently resolving and correcting any errors. Once those files are corrected, we will examine those data.

Spring/Summer Chinook: Snake River populations of spring/summer chinook migrate seaward as yearlings. During their first summer, they still reside in natal streams and do not migrate seaward until the following spring. Since 1988, several fisheries investigations have PIT-tagged chinook parr in their natal streams for a variety of purposes. The greatest effort has been expended by the National Marine Fisheries Service (NMFS), Oregon Department of Fish and Wildlife (ODFW), and Idaho Department of Fish and Game (IDFG). Table 1 details the numbers of wild chinook tagged throughout the Snake River drainage since tagging efforts were initiated.

Table 1: Locations, numbers and years of tagging of wild spring chinook in the Snake River drainage.

Tag Site	Migration Year					
	1987	1988	1989	1990	1991	1992
Valley C	?	2248	2512	1031	969	1028
Redfish L C	?	0	0	12	0	0
Huckleberry C	?	0		5	183	0
Fourth of July C	?	0	0	0	16	216
Alturus L C	?	410	1043	407	155	382
Petit L C	?	0	0	0	0	184
Pole C	?	260	0	209	6	0
Smiley C	?	106	0		0	0
Frenchman C	?	310	0	20;	575	561
Pahsimeroi R	?	0	0	0	0	1042
Salmon R E FK	?	745	0	533	559	845
Herd C	?	0	0	0	312	
Salmon R E FK W	?	0	0	0	0	12
Salmon R N FK	?	0	0	0	0	513
Lemhi River	?	0	0	0	807	535
Lead Ore Creek	?	0	0	0	43	0
Loon C	?	0	0	0	0	251
Bear Valley C	?	0	1557	353	1044	1017
CaDehorn C	?	0	0	184	209	206
Marsh C	?	0	2517	861	981	1165
Elk C	?	0	16	247	462	628
Big C	?	0	2035	726	1008	733
Rush C	?	0	0	0	0	25
Camas C	?	0	0	0	0	1013
Sulfur C	?	0	2515	0	210	714
Chamberlain C	?	0	0	0	338	0
Chamberlain C W F	?	0	0	0	1065	500
Salmon R S FK	?	2221	0	992	1031	
Secesh R	?	2182	2371	1018	1013	1E
Lake C	?	658		0	0	0
Johnson C	?	0	0	0	0	640
Salmon R	?	2523	30	533	2182	1385
Imnaha R	?	1211	2004	334	844	1252
Catherine C	?	0	0	1013	940	1107
Lostine R	?	0	84	1017	1107	1015
Grande Ronde R	?	2993	0	0	?	964
Wenaha	?	0	0	0	0	196
Wenaha R S FK	?	0	0	0	0	569
Minam R	?	0	0	0	0	1003

Table 2 summarizes PIT-tag detections at Lower Granite Dam during the spring following tagging. Recall that these populations are tagged as subyearling parr in their first summer post-emergence, and do not migrate seaward until the following spring as yearlings.

In addition to the tagging activities conducted in natal streams, some fish were intercepted, tagged, and released at traps placed below the confluence of several tributaries. We did not consider those fish in this investigation for two reasons. First, the trapped sample reflects a mixture of several individual populations in unknown proportions. Secondly, in some years and locations an unknown number of hatchery fish may have been included in the trapped sample.

Analytical Procedure

Overview

The evaluation of the ability to predict the real-time outmigration timing of wild spring chinook smolt was based on PIT-tag releases from tributaries that resulted in a minimum of 3 years of at least 30 detections per year at Lower Granite Dam (Table 3). Because release and detection numbers were often small or nonexistent at some tributaries for some years (Tables 1 and 2), data from nearby tributaries were pooled to form composites. Table 4 identifies those composites, the tributaries used in their formulation and years of available data.

For each population, we described the empirical cumulative distribution function (cdf) of PIT-tag detections at Lower Granite Dam for each year of available data. The resultant passage patterns for the historical cumulative distributions were used as a template to characterize each population. In real time, a cumulative distribution cannot be constructed until the end of the outmigration. Instead, the real-time distribution is compared to the pattern of historical cdfs by fitting the observed part of the cdf to various proportions of the historical data. The segment of

Table 2: Number of PIT-tags detected at Lower Granite Dam for wild spring chinook listed by location and year of detection.

Tag Site	Migration Year					
	1988	1989	1990	1991	1992	1993
Valley C	0	65	77	47	34	34
Redfish L C	0	0	0	2	0	0
Hucklebery C	0	0	0	0	6	0
Fourth of July C	0	0	0	0	1	8
Alturus L C	4	20	4	11	2	18
Petit L C	0	0	0	0	0	26
Pole C	2	3	0	8	0	0
Smiley C	1	7	0	0	0	0
Frenchman C	0	15	0	5	14	49
Pahsimeroi R	0	1	0	0	0	97
Salmon R E FK	0	57	0	18	33	40
Herd C	0	0	0	0	17	17
Salmon R E FK W	0	0	0	0	0	41
Salmon R N FK	0	0	0	0	0	27
Lemhi River	0	0	0	3	134	92
Lead Ore Creek	0	0	0	0	0	0
Loon C	0	0	0	0	0	24
Bear Valley C	0	0	91	44	69	67
Capehorn C	0	0	0	25	19	22
Marsh C	0	0	179	59	67	125
Elk C	0	0	1	32	36	42
Big C	0	0	145	70	57	65
Rush C	0	0	0	0	0	61
Camas C	0	0	0	0	0	0
Sulfur C	0	:	168	0	24	28
Chamberlain C	0	0	0	0	13	0
Chamberlain C W F	0	0	0	0	47	1
Salmon R S FK	0	85	1	99	82	214
Secesh R	0	190	160	78	42	32
Lake C	0	51	0	1	0	0
Johnson C	0	0	0	0	ii	53
Salmon R	277	75	7	9	66	45
Imnaha R	0	73	161	19	94	173
Catherine C	0	0	0	77	67	102
Lostine R	0	0	8	90	92	130
Grande Ronde R	0	242	2	0	0	93
Wenaha R	0	0	0	0	0	23
Wenaha R S FK	0	0	0	0	0	61
Minam R	0	0	0	0	0	105

Table 3: Individual streams used in the evaluation of the ability to perform real-time predictions of outmigration timing and the years of PIT-tag releases from 1988-1993 used in the analysis.

Tributary	Years of Data
Catherine Creek (CATHEC)	3
Lostine River (LOSTIR)	3
Secesh river (SECESR)	5
Marsh Creek (MARSHC)	4
Valley Creek (VALEYC)	5
Imnaha River (IMNAHR)	4
Bear Valley Creek (BEARVC)	4
Alturus Lake Creek (ALTULC)	6
Big Creek (BIGC)	4
Elk Creek (ELKC)	3
East Fork Salmon River (SALMON R E FK)	3
South Fork Salmon River (SALMON R S FK)	3
Salmon River (SALMON R)	4

Table 4: Composition of streams used in the evaluation of the ability to perform real-time predictions of outmigration timing and the years of PIT-tag releases from 1988- 1993 used in the analysis.

Region	Years of Data
Upper Salmon Valley C	5
Redfish L C	
Huckleberry C	
Fourth of July C	
Alturus L C	
Petit L C	
Pole C	
Smiley C	
Frenchman C	
Beaver C	
Pahsimeroi	3
Pahsimeroi R	
Salmon R E FK	
Herd C	
Upper Middle Fork Salmon	4
Loon C	
Bear Valley C	
Capehorn C	
Marsh C	
Elk C	
South Fork Salmon	5
Salmon R S FK	
Secesh R	
Lake C	
Johnson C	

the template that most closely matches the shape of the current real-time distribution then forms the basis for predictions and forecasts. Based on the similarity between those shapes, we predict the population's current temporal position in the passage distribution, and forecast the expected time to reach some specified future position in the migration. Repeated analysis provides daily predictions of outmigration status and trends throughout the outmigration season.

Characterizing Historical Outmigration Patterns

This section is a synopsis of the statistical and analytical procedures employed in this investigation. Detailed procedures are presented in a technical appendix accompanying this text (Appendix A).

Cumulative Passage Distribution: Treating each population separately, the first task was to construct a cumulative passage distribution for each year of data. The distributions were then normalized by expressing them as the percentage of the total number of tag detections each year (Fig. 2). These distributions often display abrupt changes, and for the purpose of constructing the template, it is analytically advantageous to smooth the distributions.

Smoothing the Passage Distributions: We explored two different smoothing procedures. One was based on running averages passed over the data one to several times (Fig. 3). The second approach involved the use of neural nets, a sophisticated pattern recognition procedure, employing artificial intelligence processes. Each smoothing procedure yields a separate set of smoothed distributions.

Temporal Alignment of Cumulative Passage Distributions: Our interest is the shape of the distributions. However, the date when the smolt migration begins each year is not consistent, and this affects the alignment of the distributions and increases the year-to-year variation in pattern. The initiation of migration is regulated by environmental conditions that influence the onset of

Figure 2. Illustration of (a) cumulative counts and (b) cumulative relative frequency distribution of PIT-tag detections at Lower Granite Dam for Secesh River, 1989- 1993.

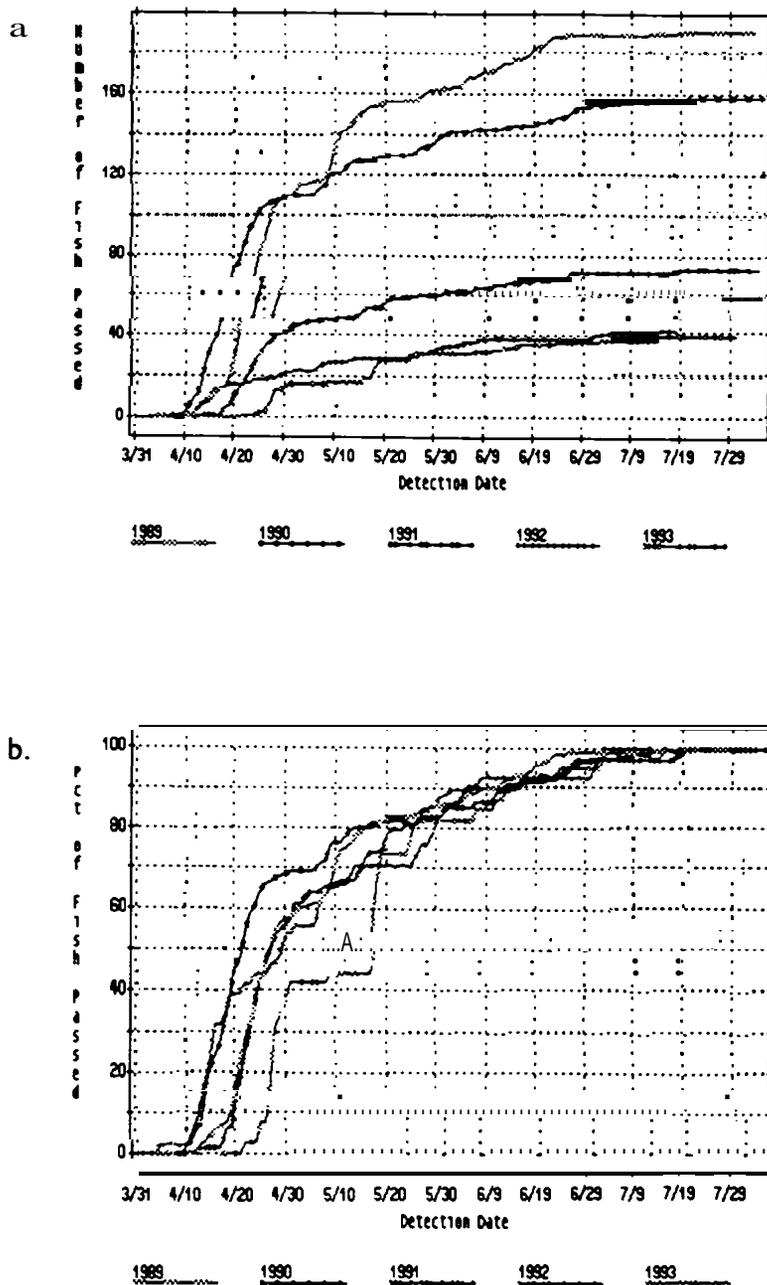
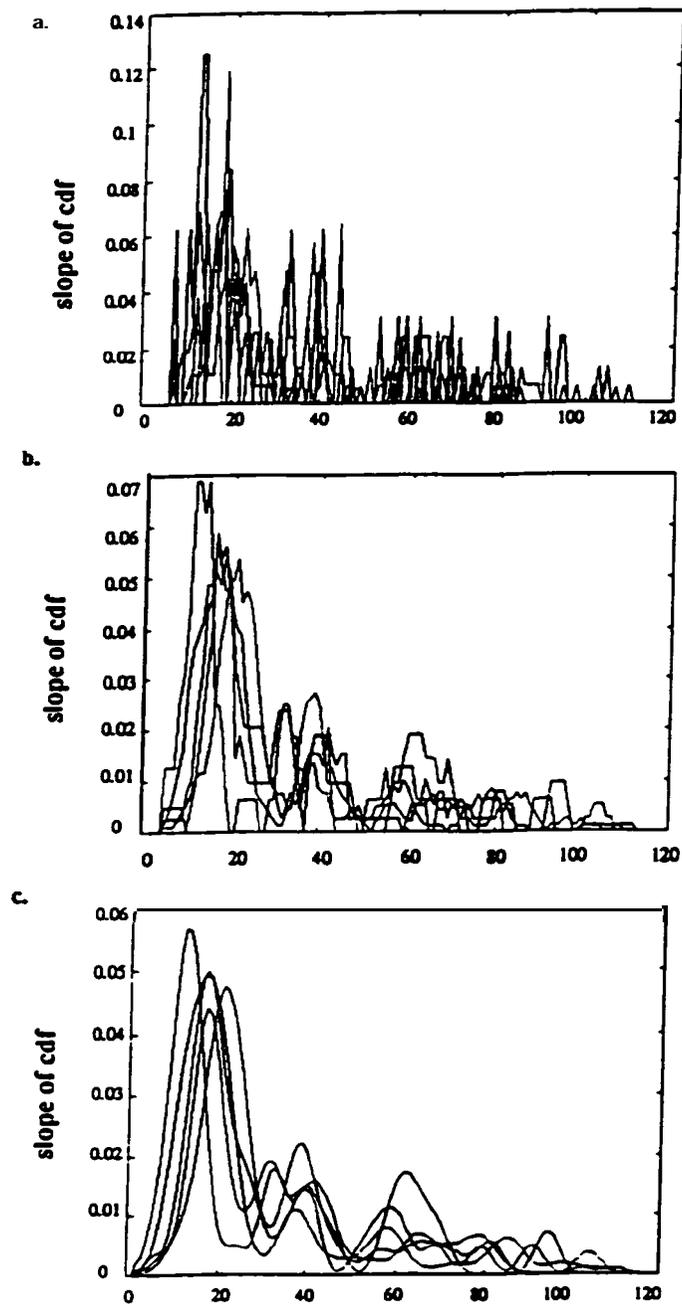


Figure 3. Illustration of the derivatives of cumulative relative frequency distribution of PIT-tag detection at Lower Granite Dam for Secesh River, 1989-1993. Graphs illustrate (a) raw data, (b) 5-day running averages, and (c) 3-pass, j-day running averages. Numbers along x-axes are days since the first fish count.



smolt development (Wedemeyer et al. 1980). The most notable variable in this regard is probably water temperature. The sooner river water warms, the sooner fish will initiate downstream migration. Depending on climatic conditions, this will vary from year to year. This variation can complicate the analysis.

The first prediction algorithm developed does not depend on synchronizing the timing of the outmigration. Instead, the identification is based solely on the shape of the empirical cumulative distribution functions over all years. The second approach uses another procedure which attempts to characterize the average annual pattern (template) for comparison with a current year's data, requires synchronizing the timings of the outmigration. The latter approach uses the peak rate of change in detections (i.e., maximum slope) to synchronize patterns over years. This maximum slope corresponds to synchronizing the maximum value of the first derivative of the cumulative passage distribution curve. Fig. 4 illustrates this synchronization for Secesh River, 1989-1993.

Characterization of Shape: The shapes of the cumulative passage distributions, for the individual years, define the passage pattern for a specific population. The difference in shapes among years defines the variability associated with the distribution. The shape of the arrival distributions can be represented by the sequence of slopes, or tangents, along the distribution (Fig. 5). Any specific segment of the distribution can be represented by a specific sequence of slopes associated with that segment. The sequences of slopes then become the actual measures used in predictions.

Decision Rule: A least-squares criterion was established to identify the most likely estimate for the percent of the run. The objective function can be written as

$$SS = \sum_{i=1}^Y \sum_{j=1}^X (\text{SLOPE}_{i,j} - \text{SLOPE}_{i,j})^2 \quad (1)$$

Figure 4. Synchronized first derivative curves for the cumulative relative frequency distributions for PIT-tag detections at Lower Granite Dam for releases from Secesh River, 1989- 1993.

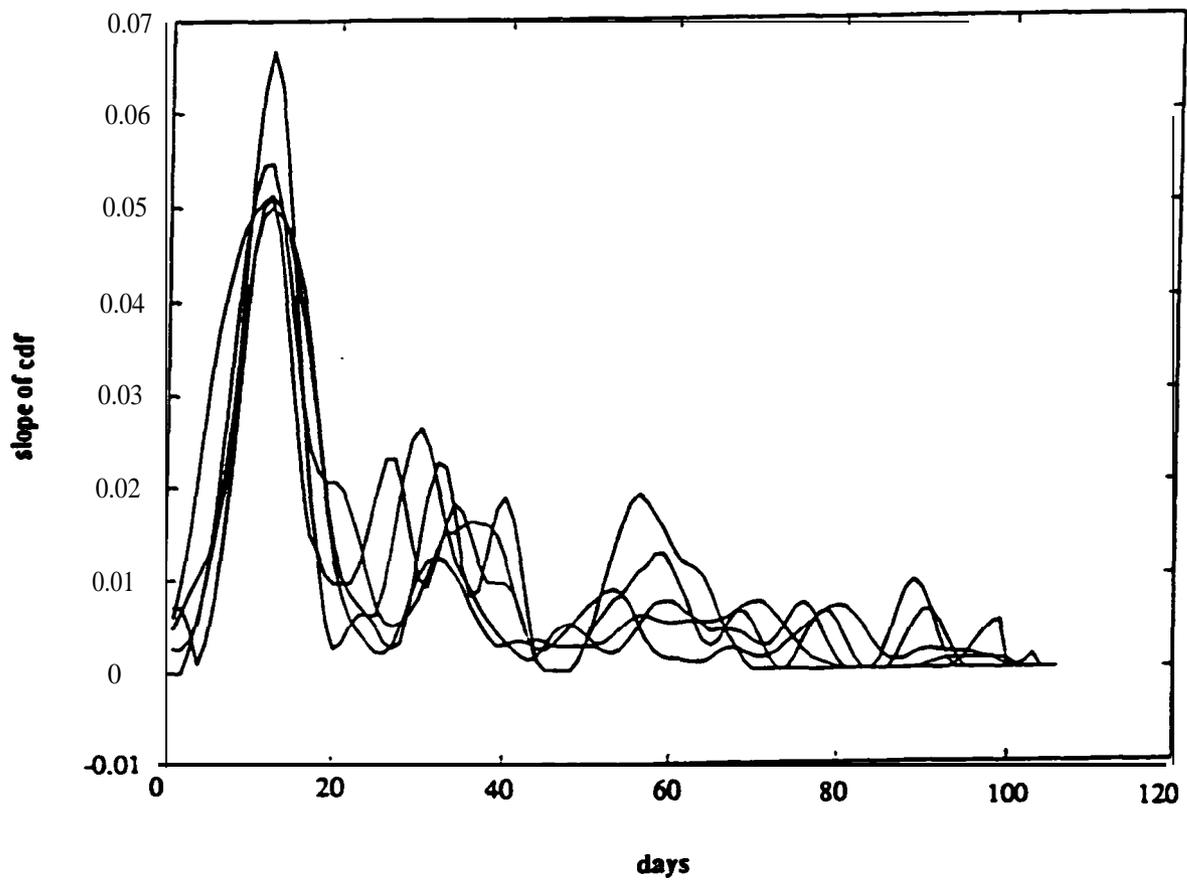
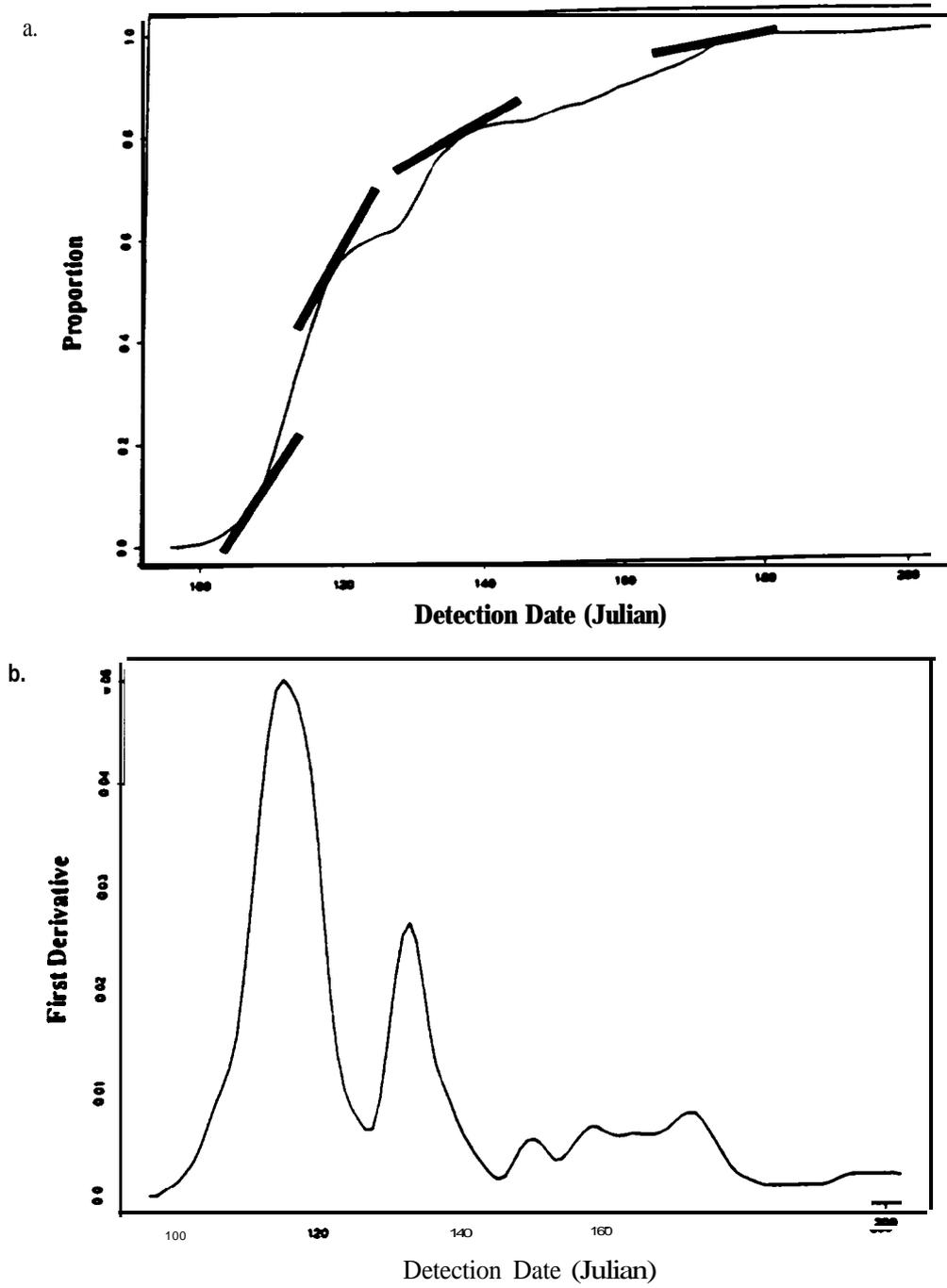


Figure 5. Illustration of characterizing the cumulative relative frequency distribution by (a) slope at strategic percentages, (b) continuously over entire (cdf) based as a curve of first derivatives.



where

SLOPE_{j,i} = slope at the jth percentile ($j=1, \dots, \%$) for the ith historical year ($i=1, \dots, Y$);

SLOPE_j = slope at the jth percentile ($J=1, \dots, \%$) for the current year's data.

The summation used in calculating the sum of squares (Eq. 1) was calculated over all historical years for one predictive algorithm or simply over the average historical cumulative curve for other algorithms. The choice of percentage (%) in the summation (Eq. 1) that minimizes the overall SS is the prediction of the percent of run to date. The summation is calculated over equal intervals of percentage of size ($\%/d$), over the range 0 to %, where d equals the number of data in the run to date for the current year.

Measures of Predictive Performance: A bootstrap (Quenouille 1956) approach was used to measure the precision and accuracy of the predictions of percent of the run to date. Using the historical data, $n - 1$ of the years of data were used to predict the run timing of the remaining n th year. All possible combinations of n choice $n - 1$ year were analyzed. The performance among the historical years provided a measure of the reliability of the prediction algorithms.

Two measures of performance were computed for all analyses. The first measure was the mean squared error (MSE) defined as

$$MSE(\%) = \frac{\sum_{i=1}^Y (\%_i - \%)^2}{(Y - 1)} \quad (2)$$

where

$\%$ = predicted percentage using the computer algorithm for the ith historical year
($i=1, \dots, Y$);

$\%$ - actual percentage of the cumulative relative distribution of PIT-tag detections at Lower Granite Dam.

The MSE can be computed for any specified $\%$ (e.g., $\% = 5, 10, 15, \dots, 95$). The expected value of an MSE is

$$E(MSE) = \sigma^2 + \text{bias}^2$$

where σ^2 is the variance of the prediction method. The units of MSE are in squared $\%$. A limitation of MSE is that it is expressed in squared units, and its square-root is not directly translatable to a standard deviation plus bias. As such, another measure of predictive error that was computed was the mean absolute deviance (MAD) defined as

$$MAD(\%) = \frac{\sum_{i=1}^Y |\%_i - \%|}{Y} \quad (3)$$

The MAD also measures accuracy and precision of the prediction algorithm, but expressed in standard units of percentage. Approximately 95% confidence intervals for a prediction were based on

$$\% \pm 2MAD(\%)$$

The MAD was introduced ad hoc and replaces usual standard deviation (std). It could be proved that **MAD** is close to std. being a slightly weaker criterion.

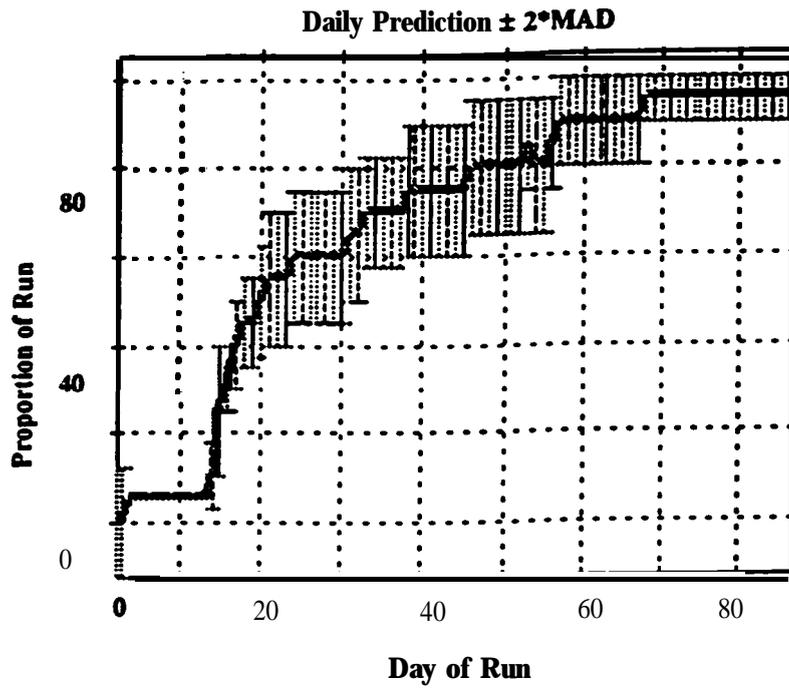
Comparison of the performance of alternative predictors or performance across various tributaries were based on the average value of MSE (%) or MAD (%) across various percentages of the run. Averages were calculated at intervals of 5% from 20% to 95% of the run.

Predictions

A software package has been developed to calculate predictions and forecasts. In real time, as PIT-tagged fish in the population are detected at Lower Granite, a cumulative passage distribution is constructed. As the distribution is updated daily, the shape changes. On any day, the shape of the distribution can be defined by the sequence of slopes accumulated up to that point in time. The predictive program identifies the segment of the historical arrival distribution that most closely matches the shape of the real-time distribution using the least squares criterion (Eq. 1). The computer program also computes the performance of predictive algorithm using the bootstrap approach across historical years. A graphical display of prediction versus expected value is presented (Fig. 6) for a selected tributary across the entire season. In addition, the algorithm reports the MSE and MAD values for the percentage predicted by the program.

The pattern of predictions across the season as the user repeatedly enters the software program can be requested (Fig. 7). A final graphical display superimposes the projected cumulative distribution for the current year over historical patterns including approximately 95% interval estimates (Fig. 8).

Figure 7. Illustration of daily outmigration predictions based on the historical pattern. Bars on the graph show approximate 95% confidence intervals of estimates.



Forecasting

Once the current position in the migration is predicted, it is possible to forecast how many days in the future to some specific percentile (Fig. 8). Forecasts are based on elapsed time between percentiles as calculated from the average cumulative distribution across years. Thus, elapsed times are added to the prediction **of the** percentage of the run to date.

Results

Alternative Predictive Algorithms

Research to develop reliable predictors of outmigration timing had to rely solely on the historic patterns of PIT-tag detections at Lower Granite Dam. Ambient data on tributary flows and temperature patterns have yet to be identified. As such, the predictive algorithms relied on pattern recognition capabilities. Two alternative algorithms were developed, varying in the degree of robustness and specificity incorporated. The algorithms differ in the degree of smoothing employed, whether synchrony of run timing across years was imposed, and whether the decision criterion was applied across all historical years or an average historic pattern. Table 5 summarizes the properties of the two alternative algorithms investigated.

Performance of Predictors

Table 6 summarizes the MAD averaged across percentiles (i.e. 20, 25, 30, . . . , 95%) on a stream and composite basis for two of the alternative algorithms investigated.

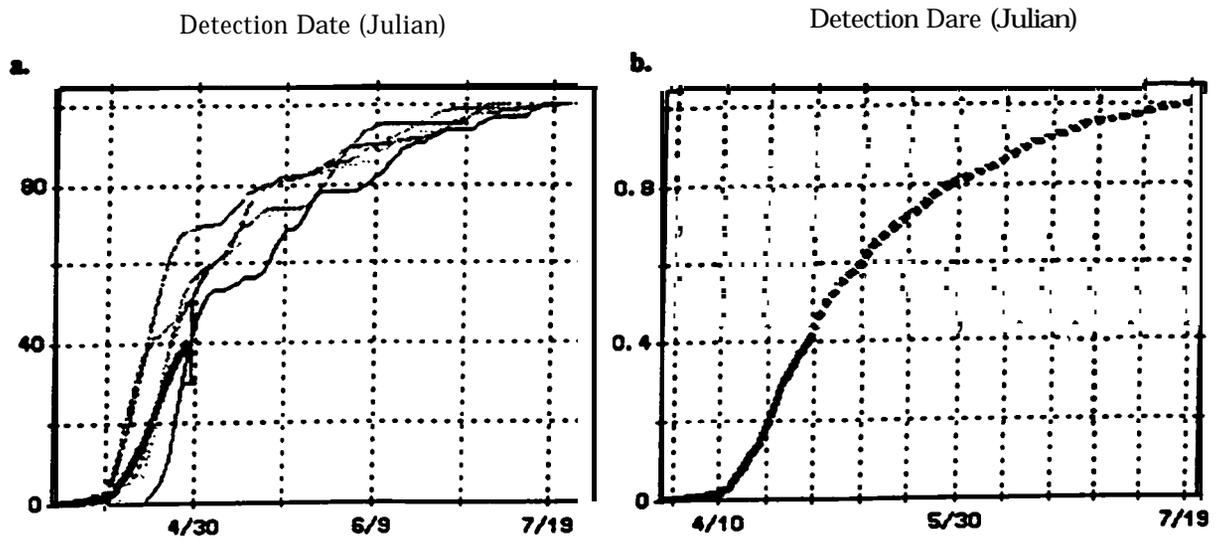
Table 5: Comparison of procedures used in alternative algorithms to predict the timing of spring chinook smolt on the Snake River.

Approach	Method 1 (Least Squares)	Method 2 (Synchronization)
Data Smoothing	5-Day, 3-Pass	5-Day, 3-Pass
Synchronization	None; doesn't make a difference.	1st significant peak of 1st derivative
Across-Year Averaging	Day of the run averaging	Average synchronized curve
Decision Criterion	Least squares; over all years using cumulative %	Least squares; over average year using day of run
Primary Prediction	% with the highest likelihood	% with the highest likelihood

Table 6. Values of the mean average deviance (MAD) averaged across years and within seasons (i.e., 20, 25, 30, . . . , 95%) on a stream or composite basis for two alternative real-time predictors.

River System	Average MAD	
	Method #1	Method #2
Catherine C	4.2	5.0
Lostine C	11.6	12.8
Secesh R	5.6	5.3
Marsh C	9.8	9.5
Valley C	9.8	8.8
Imnaha R	12.9	11.6
Bear Valley C		
Big C		19.0
Elk C	29.2	9.2
Salmon R	16.8	15.3
Salmon R E Fk	8.3	7.6
Salmon R S Fk	5.5	6.8
Upper Salmon	9.2	7.3
Pahsimeroi	12.9	11.2
Upper Middle Fork Salmon	10.9	9.1
South Fork Salmon R	5.9	5.7
Average	10.3	9.6

Figure 8. Illustration of graphical display of predictions of real-time status of outmigration. Bold line depicts prediction of cumulative run proportion to date versus historical trends. (a) Dashed line is forecast of remainder of run based on historical time between percentiles. (b) Vertical bars are approximate 95% confidence interval estimate of percent run to date.



Among the two alternative algorithms, the second method, based on identifying an average annual arrival pattern, performed best. This method has an average overall MAD value of 10.3% (Table 6). achieved when additional smoothing has been added to the initial data and decision-making function SS. Both algorithms have almost the same performance, but the second method is still faster and preferable in most cases. In 14 of 16 river systems investigated, the second method had the smaller MAD than the first. We therefore recommend this latter method of predicting the outmigration time of wild Snake River stocks.

For 10 of the 16 river systems evaluated (Table 6). the average MAD value was less than 10% for both methods. The tightened predictions occurred at Catherine Creek, Secesh River, and South Fork Salmon River with average MAD values of 4.2%, 5.3%. and 5.5%, respectively. The worst ability to accurately predict outmigration timing was at Big Creek with an average MAD of 19.0%. Surprisingly, the ability to accurately predict the outmigration timing (as measured by average MAD) was not correlated with either years of data ($r = -0.233$, $p > 0.10$) or mean number of PIT-tag detections ($r = -0.175$, $P > 0.10$) at Lower Granite Dam (Table 7). Instead, the accuracy of predictions seems more related to the consistency of the pattern of the empirical cdf. Plots of predicted versus actual percents of run-to-date for each river system investigated are included in Appendix B. Separate lots are enclosed for both prediction algorithms investigated (Appendix B).

Discussion

This progress report summarizes initial attempts at developing a means to predict in real time the migration status of wild spring chinook smolt in the Snake River. **The approach used** PIT-tag detections at Lower Granite Dam of wild runs of Snake River spring chinook. Pattern recognition procedures were developed, based on historical cumulative relative frequency distributions of detection at Lower Granite Dam.

Table 7. Relationship between average MAD, years of historical data, and mean number of detections at Lower Granite Dam.

River System	Average MAD	Years of Data	Mean No. of Detections
Catherine C	4.2	3	82.00
Secesh R	5.3	5	100.40
Salmon R S Fk	5.5	4	120.00
South Fork Salmon	5.7	5	217.60
Elk C	9.2	3	36.67
Valley C	8.8	5	51.40
Upper Salmon	7.3	5	232.35
Pahsimeroi	11.2	3	101.00
Upper Middle Fork Salmon	9.1	4	142.00
Marsh C	9.5	4	106.75
Salmon R E Fk	7.6	3	43.33
Imnaha R	11.6	4	125.25
Bear Valley C	10.2	4	67.75
Salmon R	15.3	4	115.75
Lostine C	11.6	3	104.00
Big C	19.0	4	84.25

Alternative predictors were developed, based on various procedures to smooth the historical trends, synchronize the onsets of the migration, and the degree of summarization of historical data. Both methods shared the same least squares decision rule that compared slopes of the arrival distributions of past years with real-time data. Method #2 is generally recommended for outmigration predictions for all 12 rivers (Table 3) and 4 composite river systems (Table 4) investigated. The method is anticipated to have an average of 9.6% (MAD) across river stocks and throughout the season. At the end of the spring 1994 season, an evaluation of the performance of the predictions versus actual percentiles should be conducted. Historical error rates should be compared with the error rates observed this first year to determine the reliability of error predictions and associated confidence intervals. The 1994 PIT-tag data should then be incorporated into the historical database for subsequent years.

This investigation also identified a current weakness of existing PIT-tag data (Tables 1 and 2). much more consistent PIT-tagging program should be implemented. Annual tagging on key river stocks should institutionalized with sufficient numbers to adequately characterize outmigration patterns (i.e., detections/year). Furthermore, PIT-tag programs should be initiated at key rivers that historically have received little attention.

Literature Cited

Columbia Basin Fisheries Agencies and Tribes. 1992. Fish Passage Center Weekly Report. Portland, OR: Fish Passage Center.

Kiefer, R. B., and K: A. Forster. 1992. Idaho habitat/natural production monitoring Part 2, Intensive monitoring subproject. Annual Progress Report, 1990. April 1992. Washington, D.C.: Department of Energy.

Matthews, G. M., S. Achord, J. R. Harmon, O. W. Johnson, D. M. Marsh, B. P. Sandford, N. N. Paasch, K. W. McIntyre, and K. L. Thomas. 1992. Evaluation of transportation of juvenile salmonids and related research on the Columbia and Snake Rivers, 1990. Seattle, WA: Army Corps of Engineers. 51 pp.

Quenouille, M. A. 1957. **The analysis of multiple time-series**. New York, NY: Hafner Publishing Company.

Wedemeyer, G. A., R. L. Saunders, and W. C. Clarke. 1980. Environmental factors affecting smoltification and early **marine** survival of anadromous salmonids. **Marine Fisheries Review** 42:1-14.

Appendix A:
Additional Details on Predictors of Outmigration Timing

Forecasting Expert Systems and Artificial Intelligence

During recent years, ideas about flexible decision-making algorithms based on the learning from a given training dataset became more and more popular among professionals in many fields. The applications are widespread, and the fruits of these applications are being reaped by many from diverse fields. This methodology has become an alternative to modeling some physical and nonphysical systems. One of the reasons for this is that absence of full information is not as big a problem in neural networks (NNET) as it is in other methodologies. To think that the modeling of NNET is an attempt to mimic human learning is somewhat exciting. Below we describe how our forecasting algorithm uses a “fuzzy logic” approach to build a real-time analysis and forecasting software. One of the approaches under consideration is to use NNET with the suitable architecture, train it on the sets of historical data, and store the parameters of the NNET. The same NNET applied later to the similar data is able to recognize the memorized pattern and even predict the behavior of the stochastic system of fish outmigration. In a sense, NNET algorithms are closer to an intelligent learning system than to the multivariate regression analysis. Another improvement to the forecasting of the outmigration process could be done by implementing a “fuzzy” mechanism of learning, based on some assumptions or rules. In our case, it corresponds to an intelligent algorithm of the identification of the first significant peak (Fig. 3) in the smoothed count function. The term “fuzzy” means the use of a flexible (changing during the learning stage) criteria used for the final decision. Below we will describe some details of the fuzzy algorithms used in our forecasting software.

Prediction Algorithms Based on a Least Squares Decision-Making Function

To predict the proportion of fish passed during the current run, all of our algorithms have the same general approach. First, we smooth and normalize the historical and current years' cumulative distribution function (cdf) of fish detections. Then, the first derivative of the smoothed and normalized functions are calculated. The smoothing routine and elimination of noise in the

initial data are probably the most crucial part of data preparation. More advanced algorithms, including those that are presently under development, should use NNET and fuzzy logic to combine the years of historical data into one aggregate curve and **take** into account the real-time changes in flows, temperature, and water spills at different dams. We then use a variety of methods to compare the current run to the different portions of the aggregate historical curve (Fig. 5B), or different years' patterns separately, and finally choosing the portion of historical data that most closely (in the least squares sense) resembles the current run. The steps of the algorithm are as follows:

1. **Smoothing.** We found that randomness in the daily count data (Fig. 3a), even after a single 5-point smoothing, adversely affects the performance of the program. **All the algorithms** we developed have been run and tested on thrice (5-point moving average) smoothed data. Using triple smoothing consistently improves results for any tried algorithm, and predictions are usually distinctively better than those for the once and twice smoothed datasets. We gained no perceivable benefits from smoothing the **data** further or applying a sophisticated Backpropagation Neutral Network with self-learning mechanism. In the latter case, NNET "oversmooths" the initial empiric data and starts to obscure actual features of the outmigration (as opposed to noise elimination). The NNET method should be reserved for the most noisy data and will be employed fully in our future development of a forecasting system. Definitely, the same level of prediction could be achieved with the used NNET at the expense of computation time.

2. **Computing Historical Aggregates.** We combined several years of historical data into one historical average pattern that serves as a template for comparison with the real-time migration pattern. The first method that compares each historical year individually to the current year and then computes the most probable percentage of the run based on least squares criterion was described earlier. Substantial improvements in the prediction (significant decrease of MAD) with this method were achieved by thorough data preparation, data interpolation, and averaging over days corresponding to the same percentile in different years.

The second method is based on a heuristic approach, that creates an average (over years of observation) template that later is used for predictions of the current real-time pattern. Simple averaging over the years of data on a day-by-day basis did not produce much of an improvement, as long as runs tend to start at different dates from year to year. The largest improvement was achieved by synchronizing each year according to a chosen paradigm, and then averaging the corresponding synchronized data to form the historical pattern. Our mechanism of synchronization of different years was based on the employment of a fuzzy-logic flexible algorithm that picks the first “significant” peak in first derivatives of the cumulative distribution function (cdf) of fish count.

Other approaches we investigated included synchronizing years by the first day of detection, synchronizing by the date at which 50% of the fish have passed, and synchronizing by the center of mass of the detection graph (very similar to the 50% method). Synchronizing by the center of mass was also a good approach that was robust and stable (producing the least of apparent outliers). One of the problems with synchronization and averaging the data is that peaks in the derivatives of cdf are often obscured or cancelled when the same features occur on different Julian dates in different years. We determined that finding and synchronizing on the “most significant peak is important in the day-by-day averaging approach.

The problem of finding the “most significant” peak is that the peaks can vary in relative amplitude, width, and timing among years. Choosing the overall highest peak does not always identify the first “significant” peaks, because the relative location of the highest peak sometimes varies widely from year to year. Misalignment of the data results in high level of instability in the predictions and large outliers. **We** developed the following criteria for the choice of the right peak:

a. **Timing.** The “significant” peak must occur early in the run. A late peak does not provide coherent data for the predictions at early stages.

b. Size. The “significant” peak must be large enough. Small peaks are usually produced by noise from detections and create large outliers.

c. Consistency. Chosen peaks should occur at similar temporal locations from year to year.

These criteria are definitely a “fuzzy set.” Computer software was developed using these criteria. The best results were obtained by choosing the first peak containing at least 20% of the total fish in the run before the midpoint of the run. In other words, we start with the assumption that the peak sought contains 10% of the total fish in the run and gradually relax this criterion, accepting any available peak by the time that 50% of the fish have passed. This method yielded very strong predictions for the rivers and creeks with reasonably coherent data over the years and improved predictions for most of the others.

Comparison of Methods

Both methods use least squares criterion for the prediction of the proportion of the run. In each prediction method, the decision function SS (sum of the squares of the difference) was calculated with the aid of Eq. 1 or a similar one. The basic difference between these methods was in how the summation over the years Y was performed when SS functions were calculated. The first method used Eq. 1. For the second method of synchronized pattern, the summation over the years was performed before the norm of the difference was calculated, and a synchronization was imposed. That minor change required a sophisticated “fuzzy logic” mechanism of peak “picking” based on the observation that the temporal shape of the year-to-year pattern is relatively stable with respect to the timing of the first “significant peak” in count.

In order to improve the stability of the prediction, particularly near the starting point of the run, an additional bias factor has been incorporated into the algorithm. This factor measures the number of days between the detection of the first fish and the last recorded fish in the historical

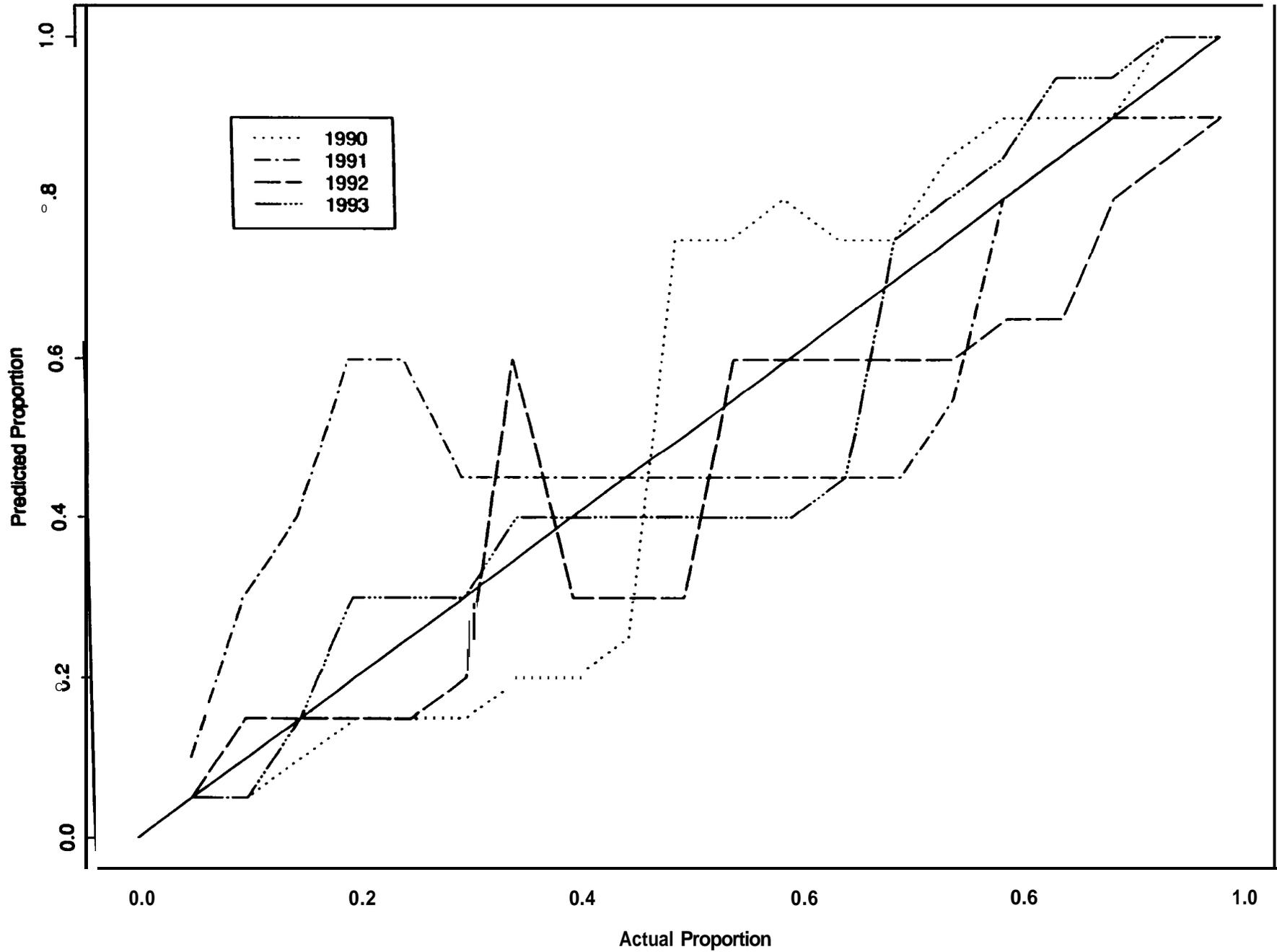
aggregate and in the current run. The decision function SS is multiplied by the ratio of the longer period over the shorter one. This factor makes the minimum in the SS function sharper and improves the stability of the prediction.

During the development and testing, the first (robust) method was substantially improved when some “fuzzy rules” from the second method was employed, and additional smoothing of the SS function was added. The performance of those two different algorithms became fairly close when some notable improvements were added to the robust least squares method.

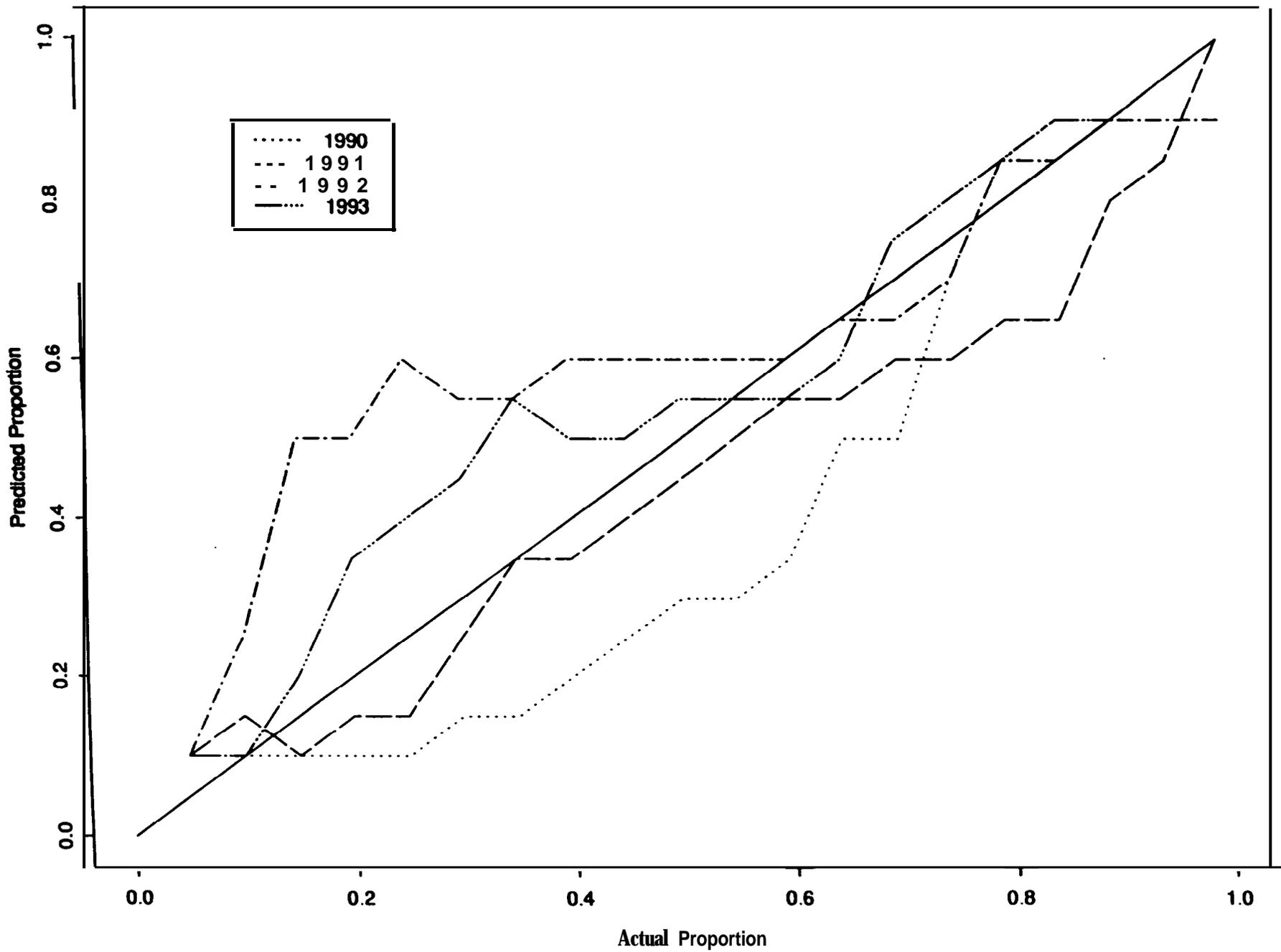
Finally, both methods could be recommended as fairly reliable forecasters for the spring and summer chinook. In the first stages of the run, the second method is preferable in most of the cases.

Appendix B:
Jackknife Results on the Performance
of the Predictors of Outmigration Timing
[least squares algorithm (ls), synchronized algorithm (sync)]

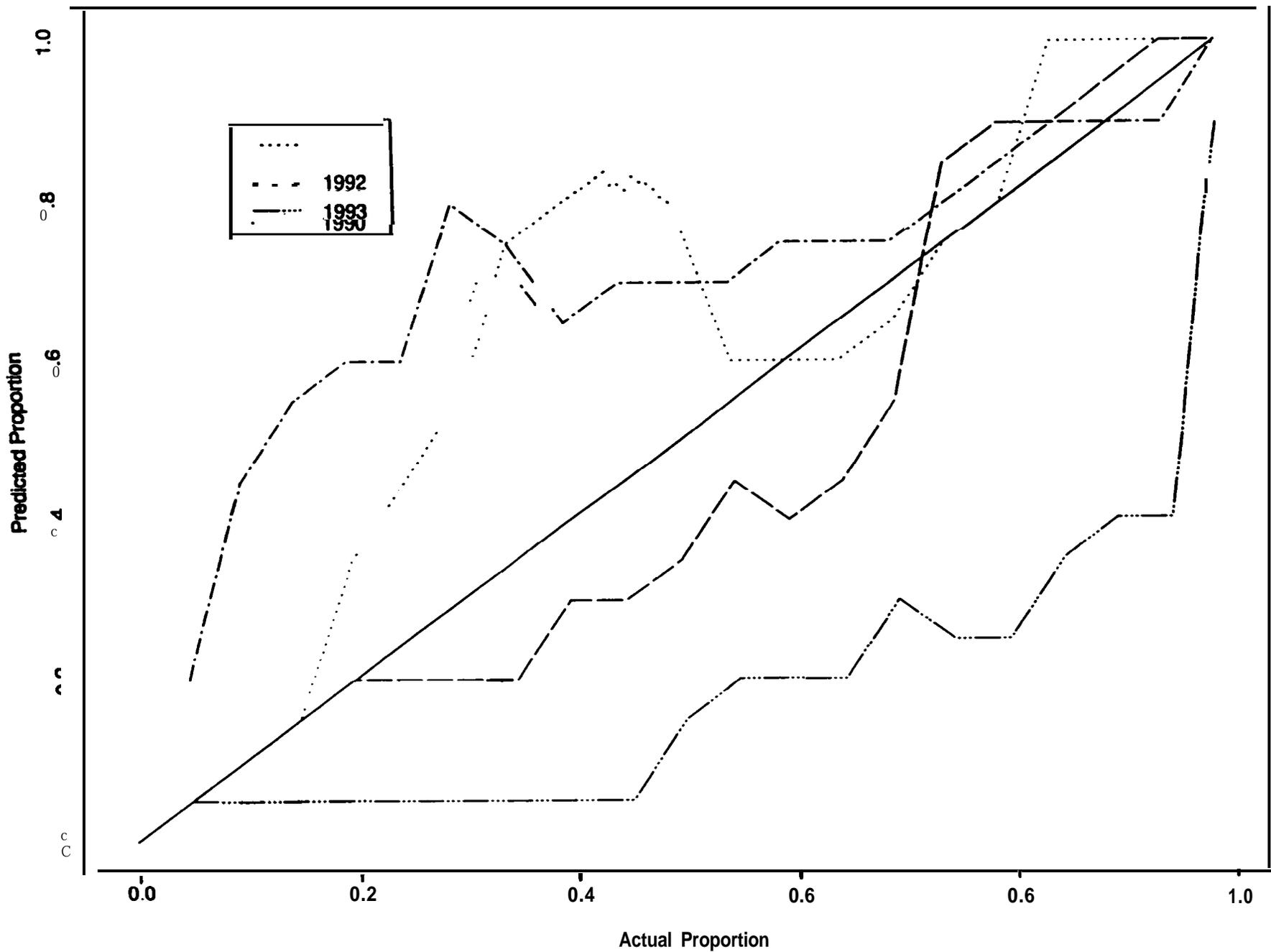
Bootstrap Prediction over Historic Years : Bear Valley Creek (Is)



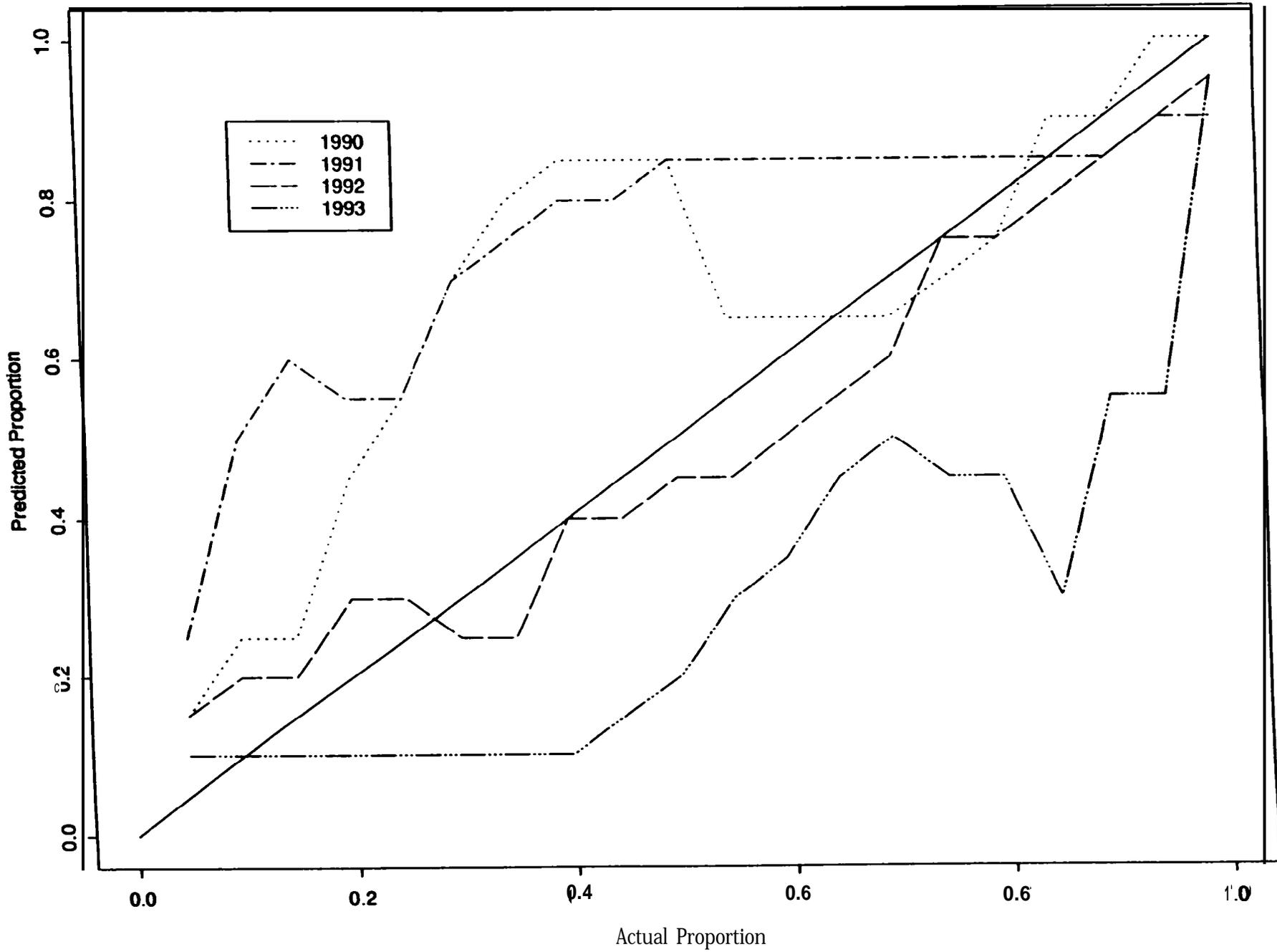
Bootstrap Prediction over Historic Years : Bear Valley Creek (sync)



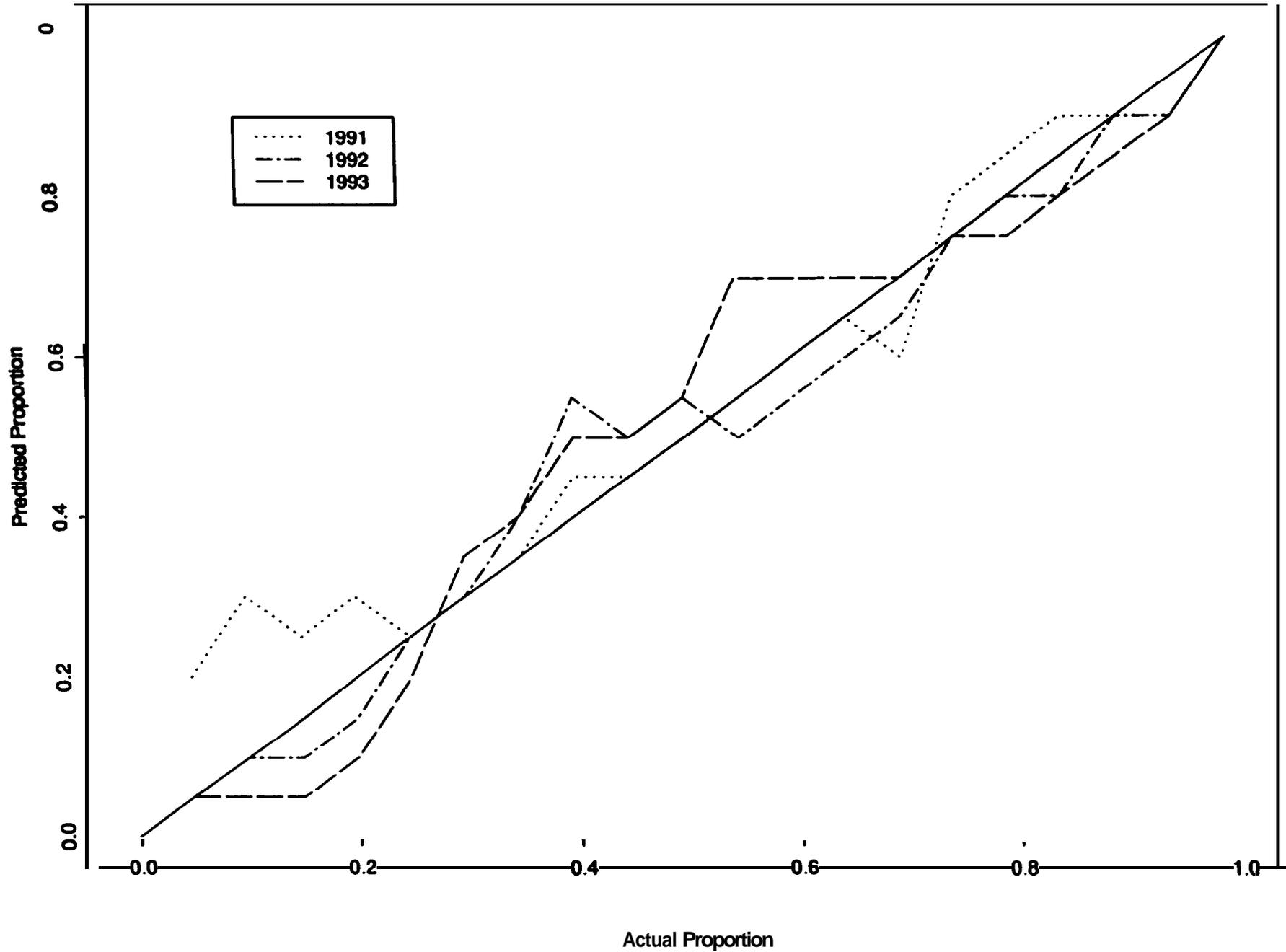
Bootstrap Prediction over Historic Years : Big Creek (Is)



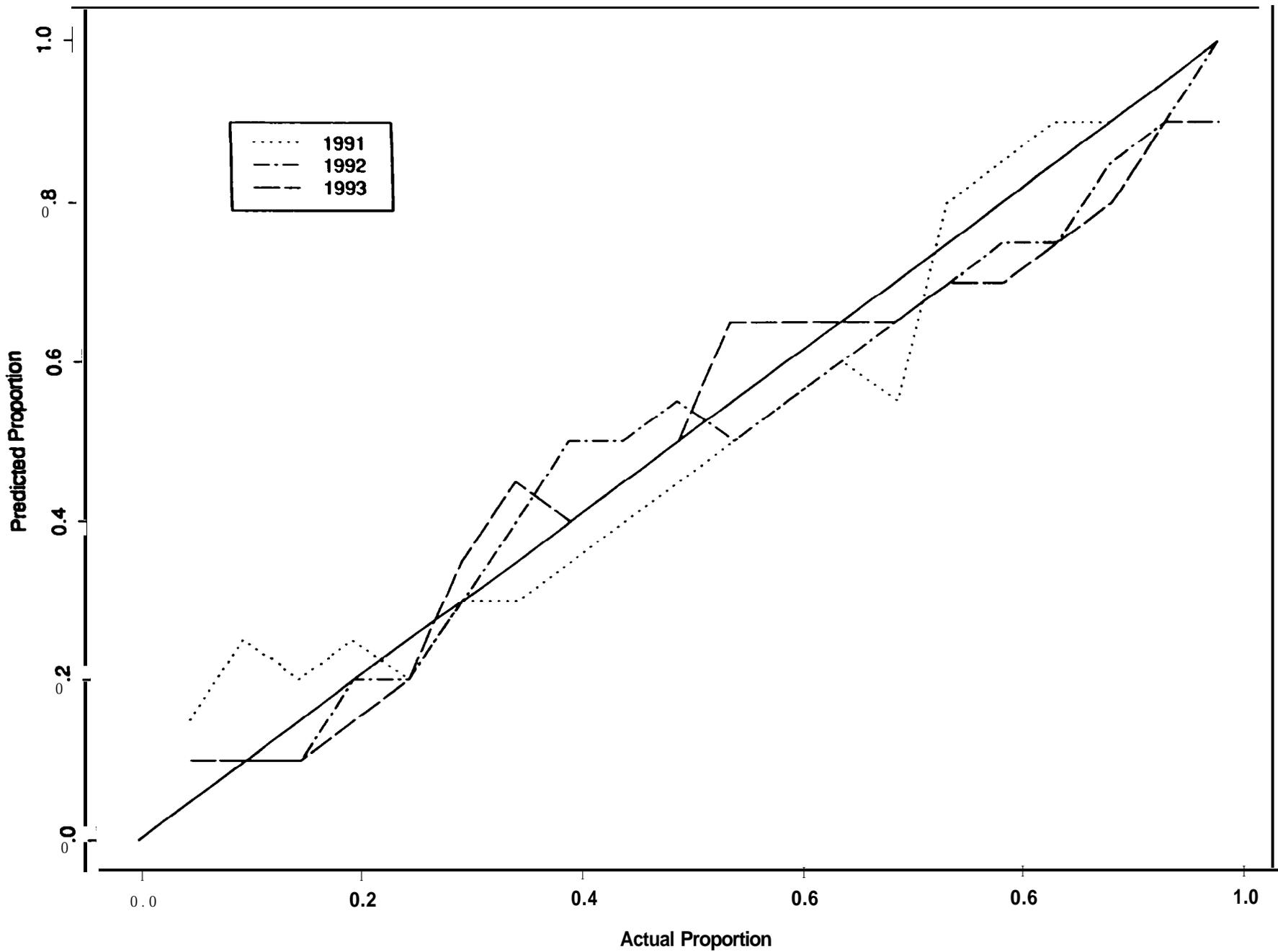
Bootstrap Prediction over Historic Years : Big Creek (sync)



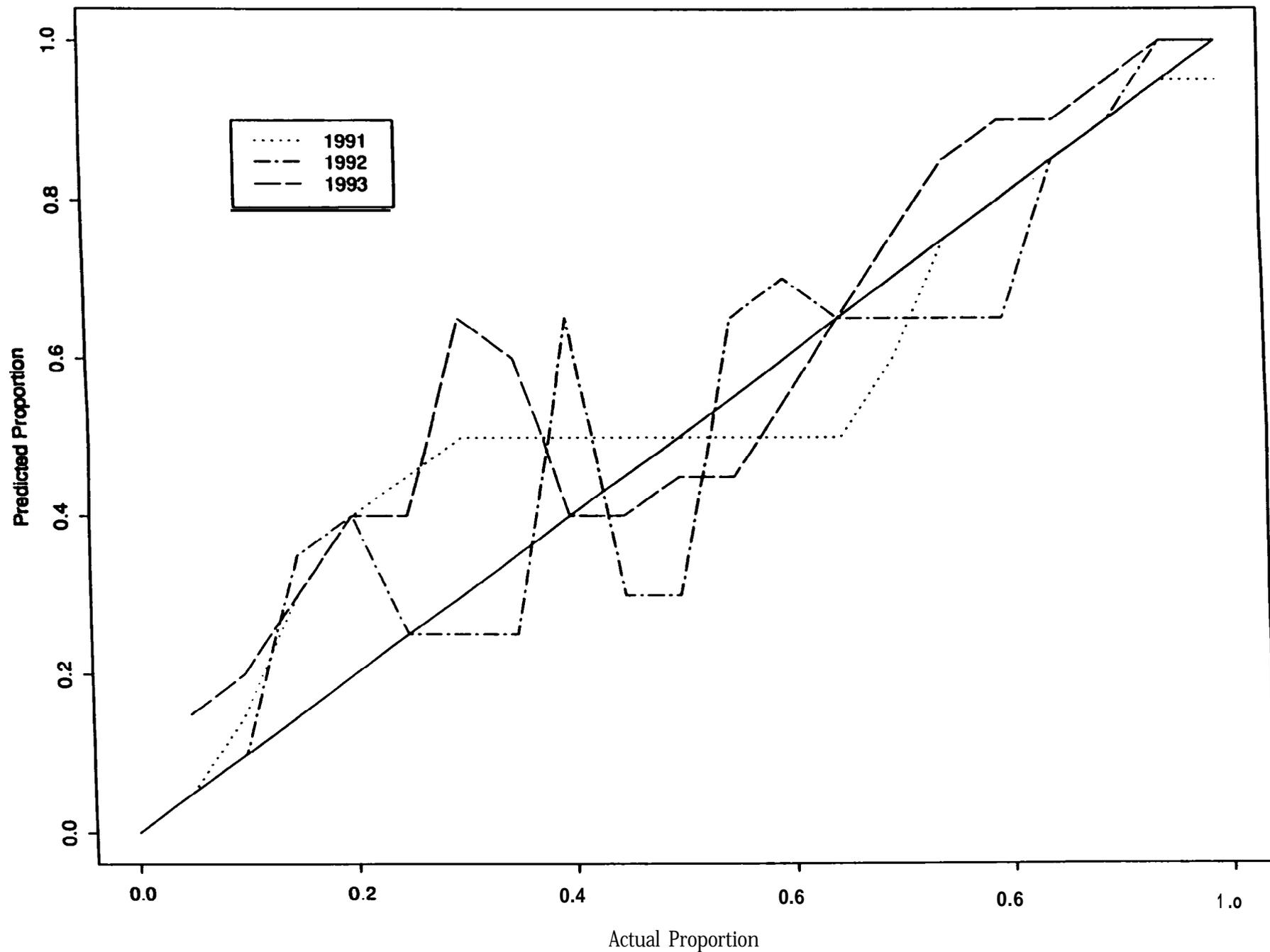
Bootstrap Prediction over Historic Years : Catherine Creek (Is)



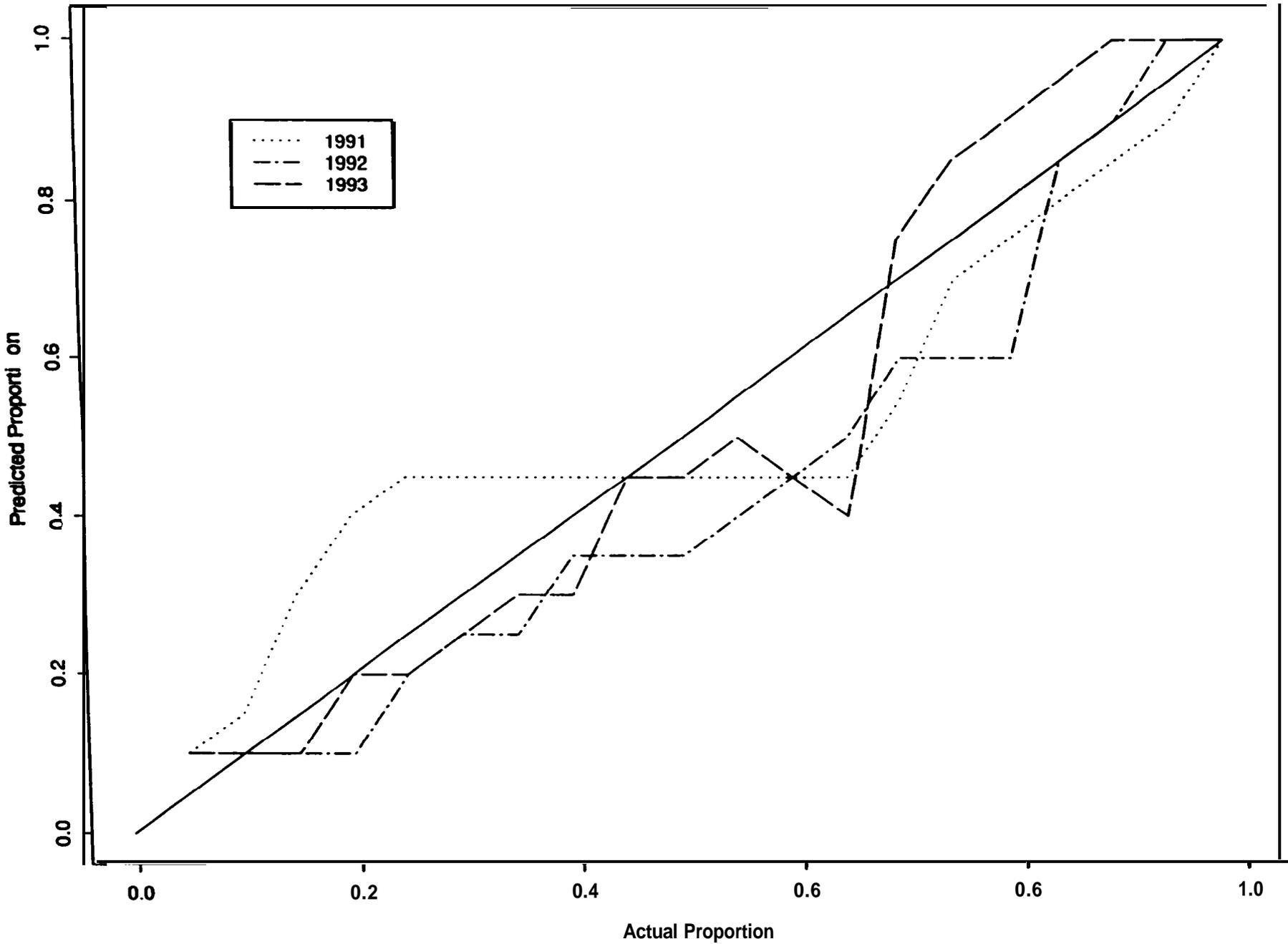
Bootstrap Prediction over Historic Years : Catherine Creek (sync)



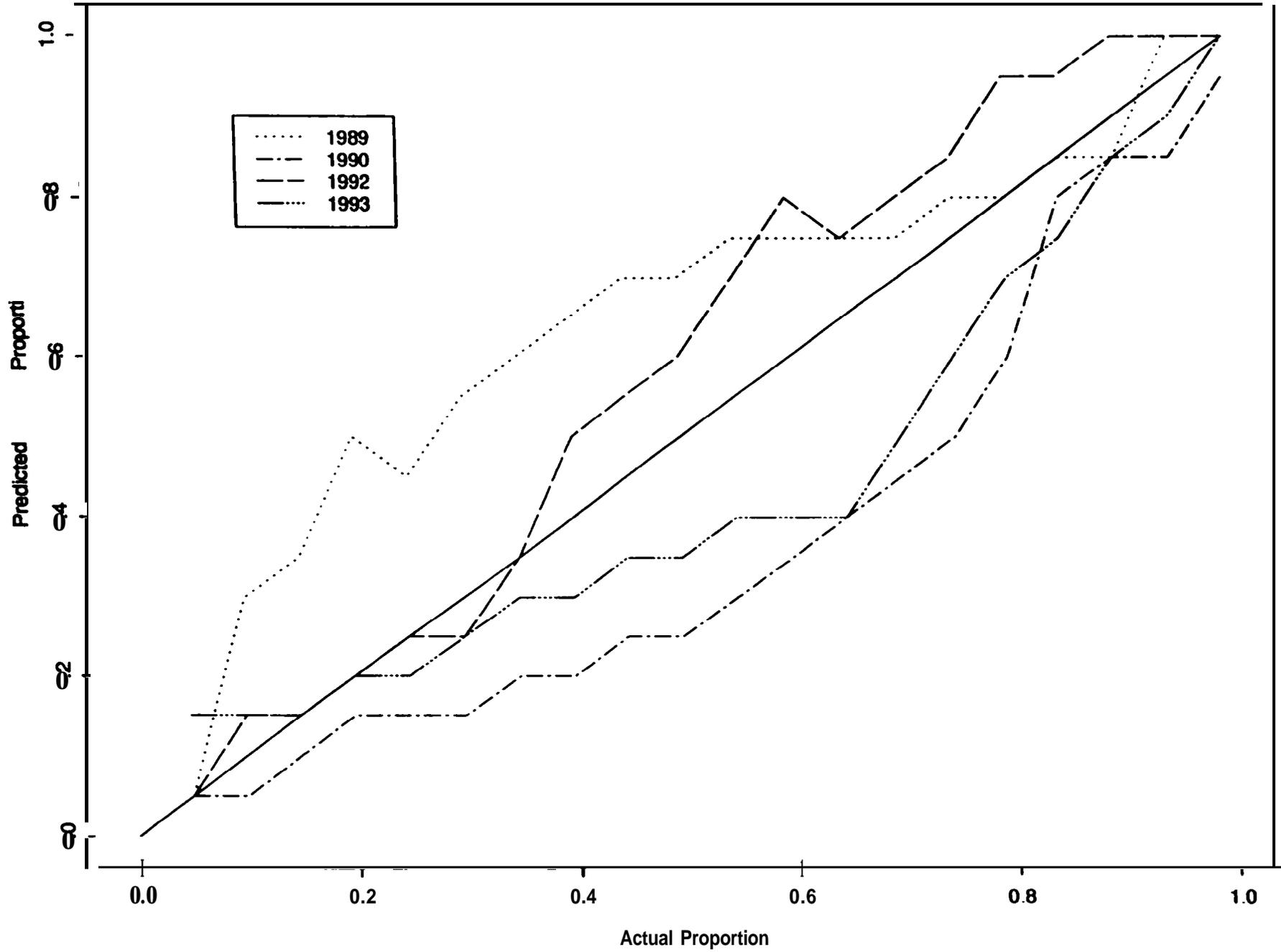
Bootstrap Prediction over Historic Years : Elk Creek (Is)



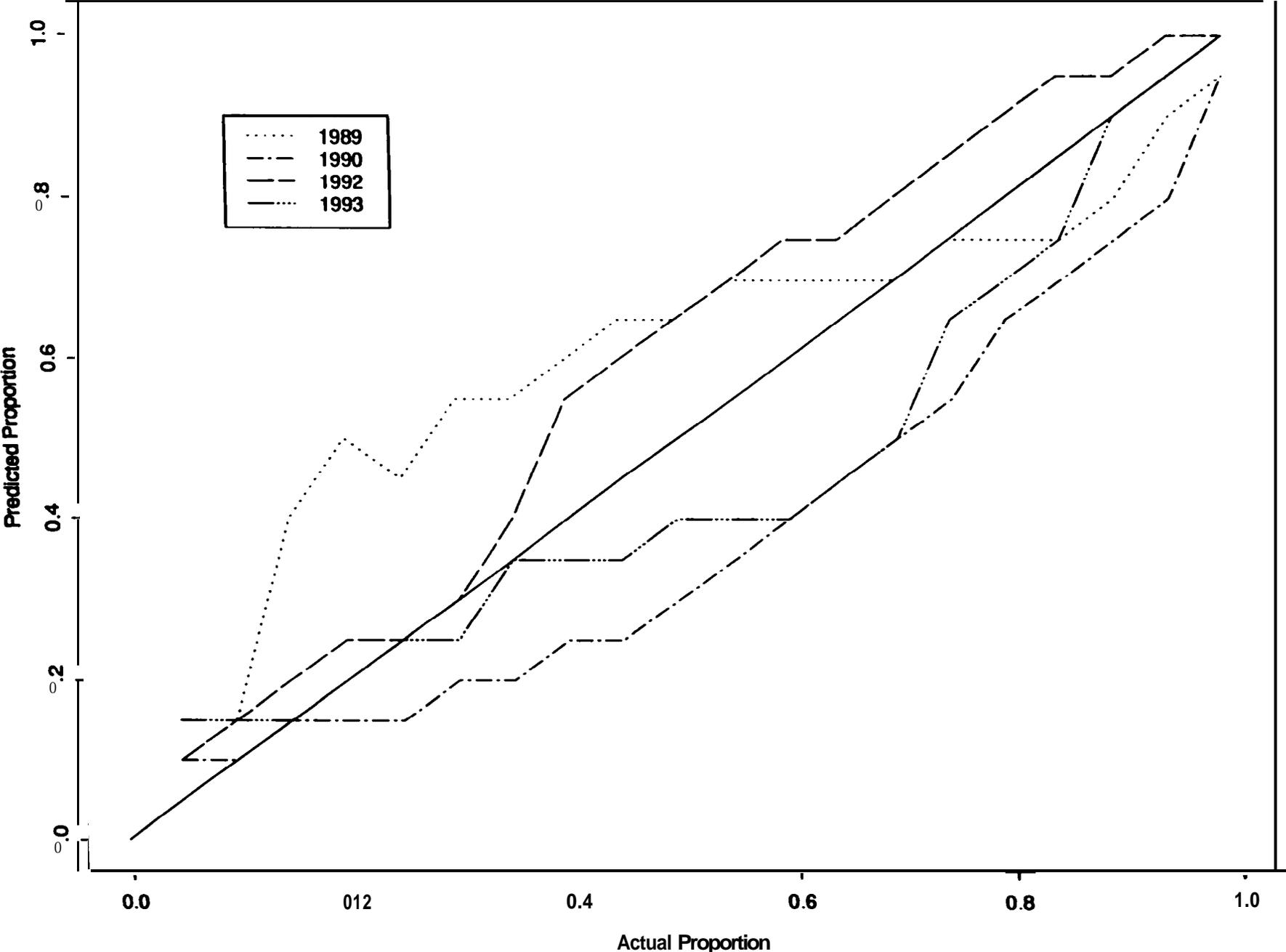
Bootstrap Prediction over Historic Years : Elk Creek (sync)



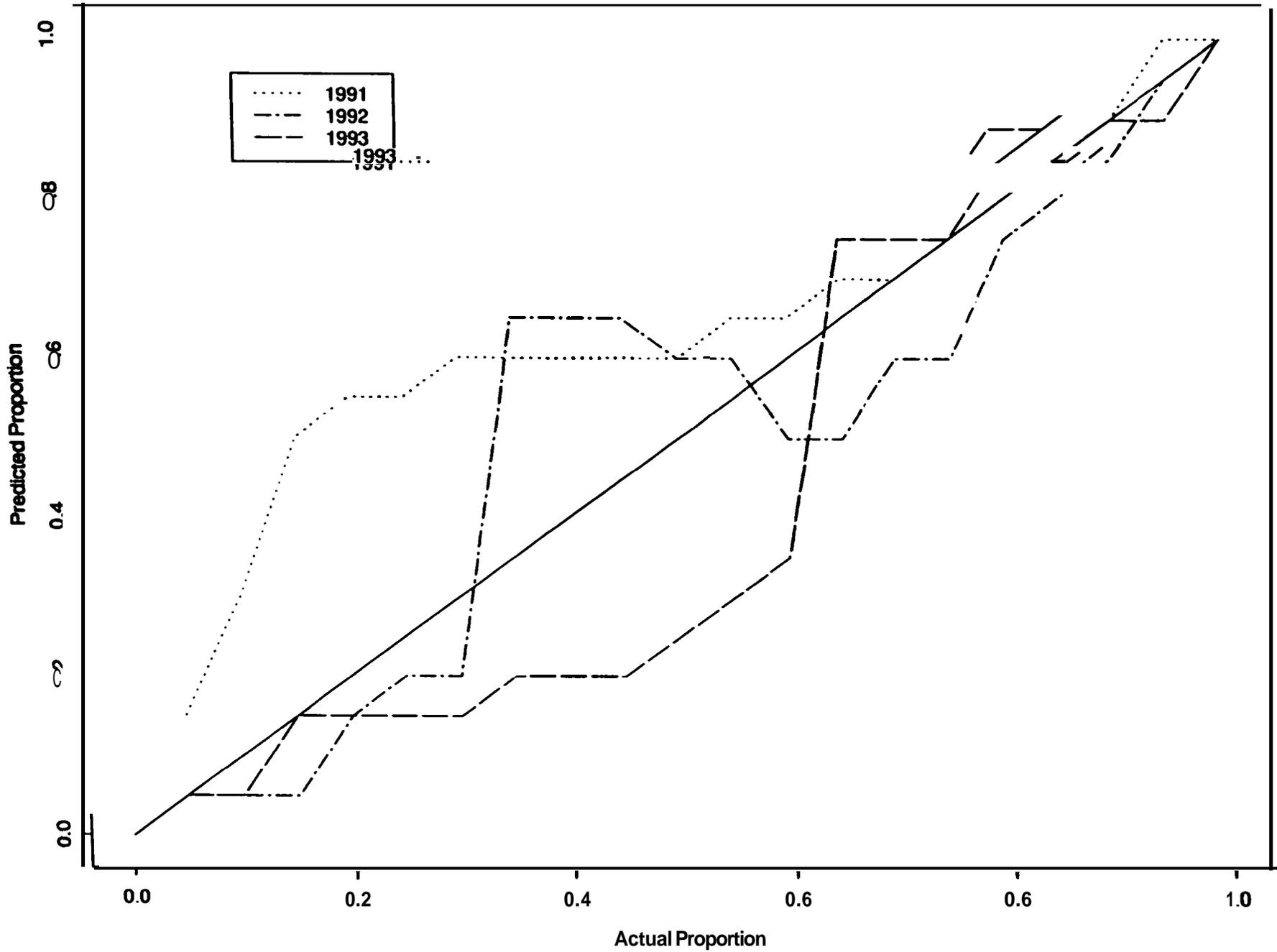
Bootstrap Prediction over Historic Years : Imnaha River (Is)



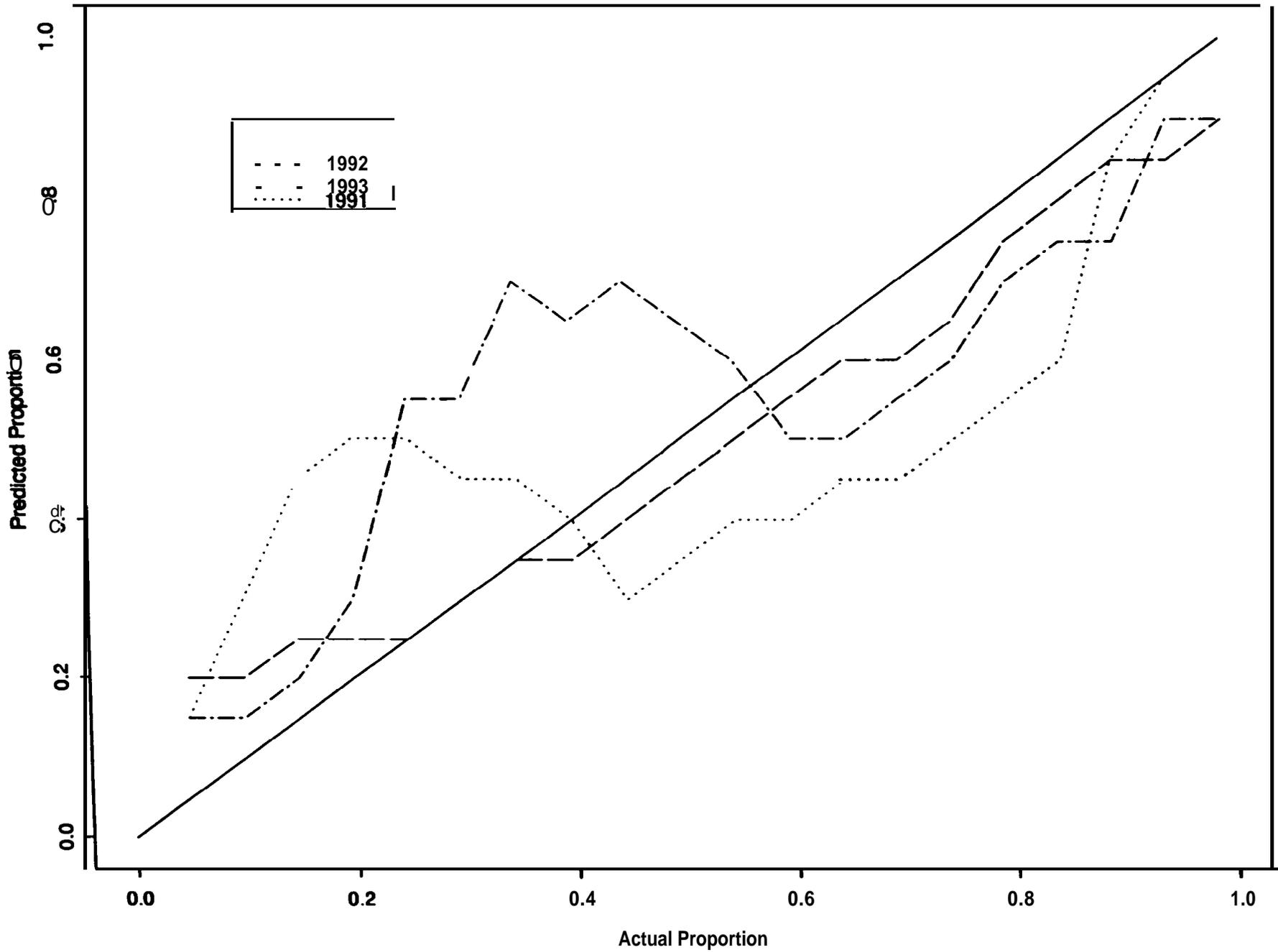
Bootstrap Prediction over Historic Years : Imnaha River (sync)



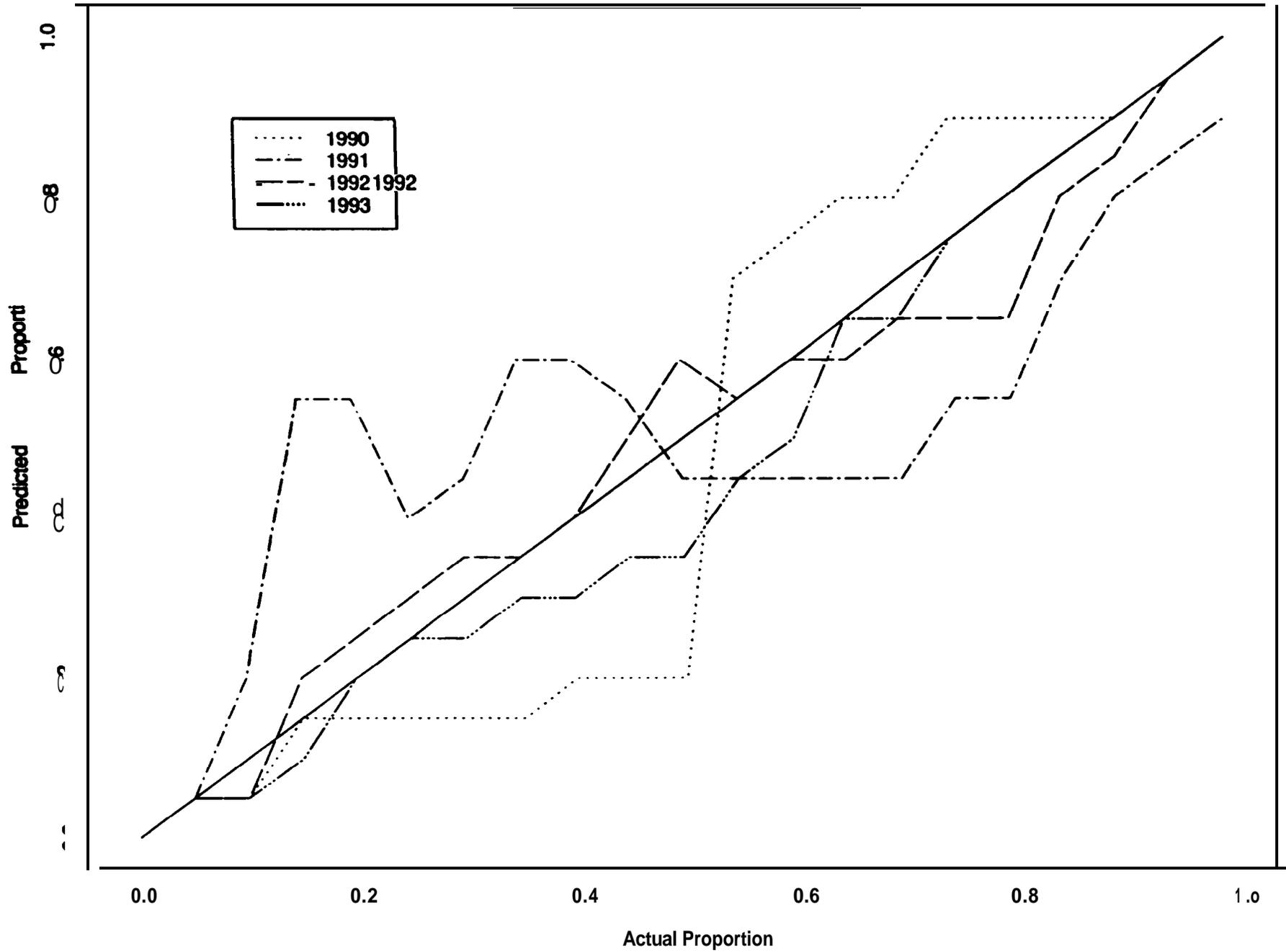
Bootstrap Prediction over Historic Years : Lostine River (Is)



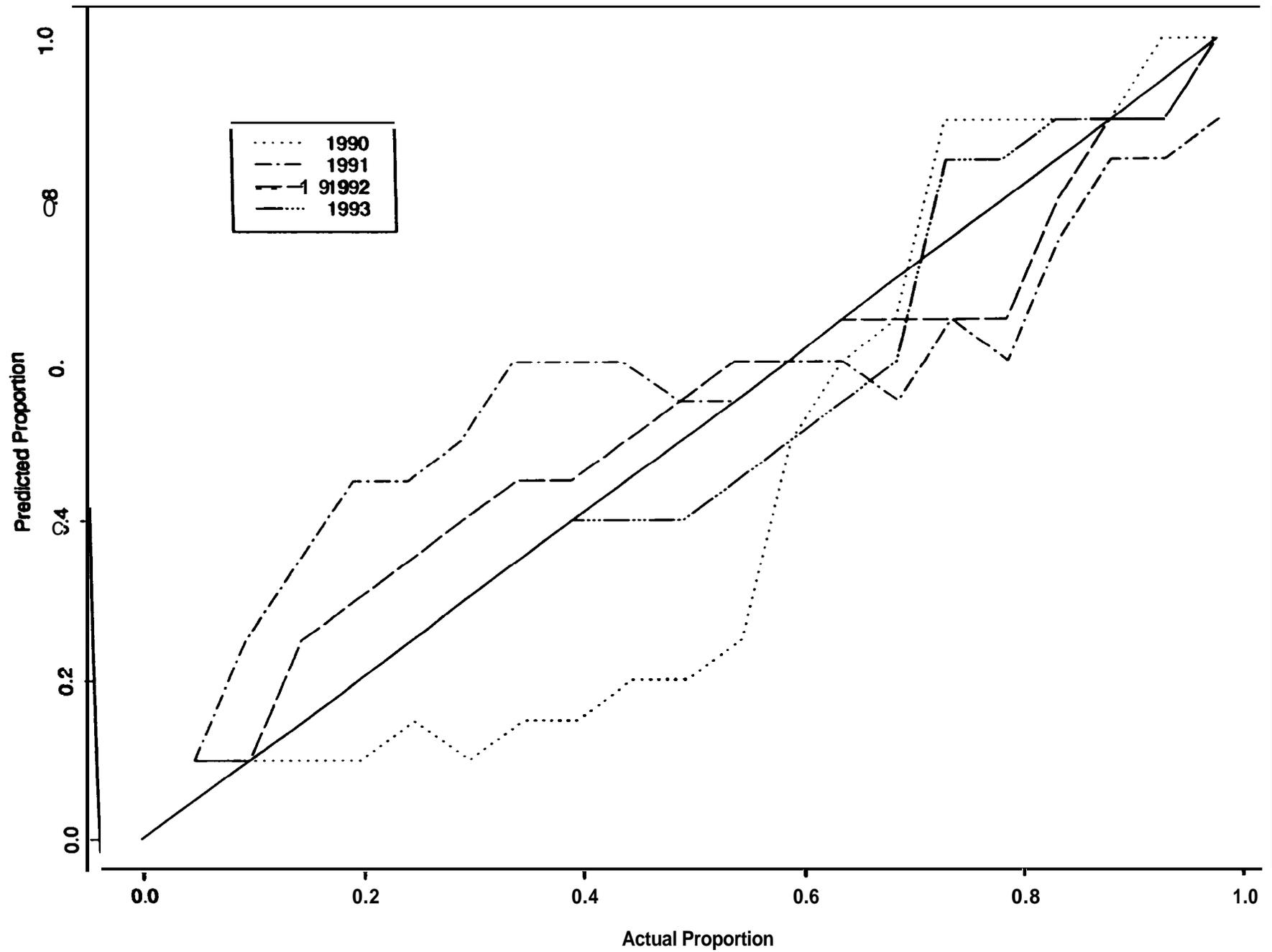
Bootstrap Prediction over Historic Years : Lostine River (sync)



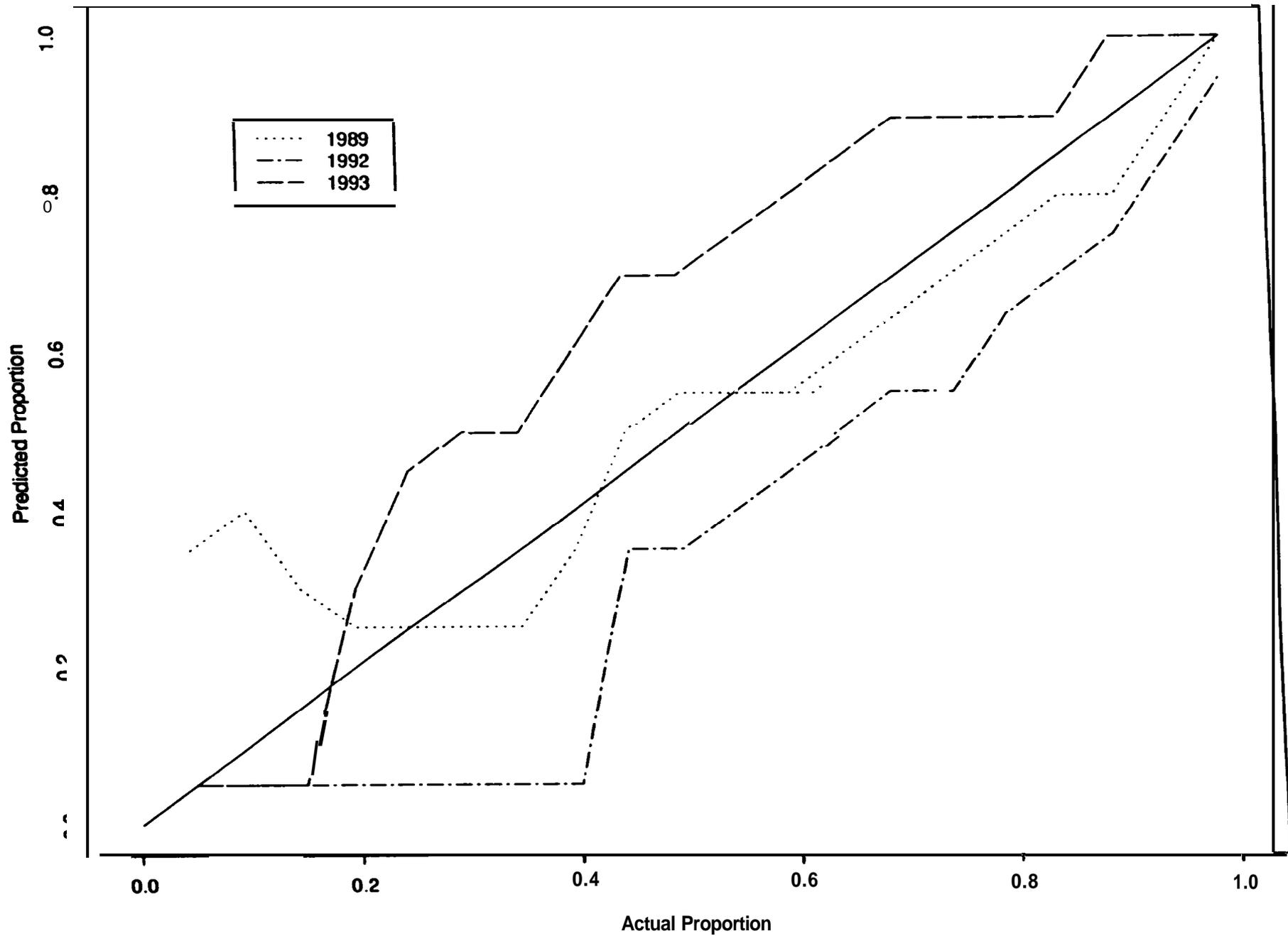
Bootstrap Prediction over Historic Years : Marsh Creek (Is)



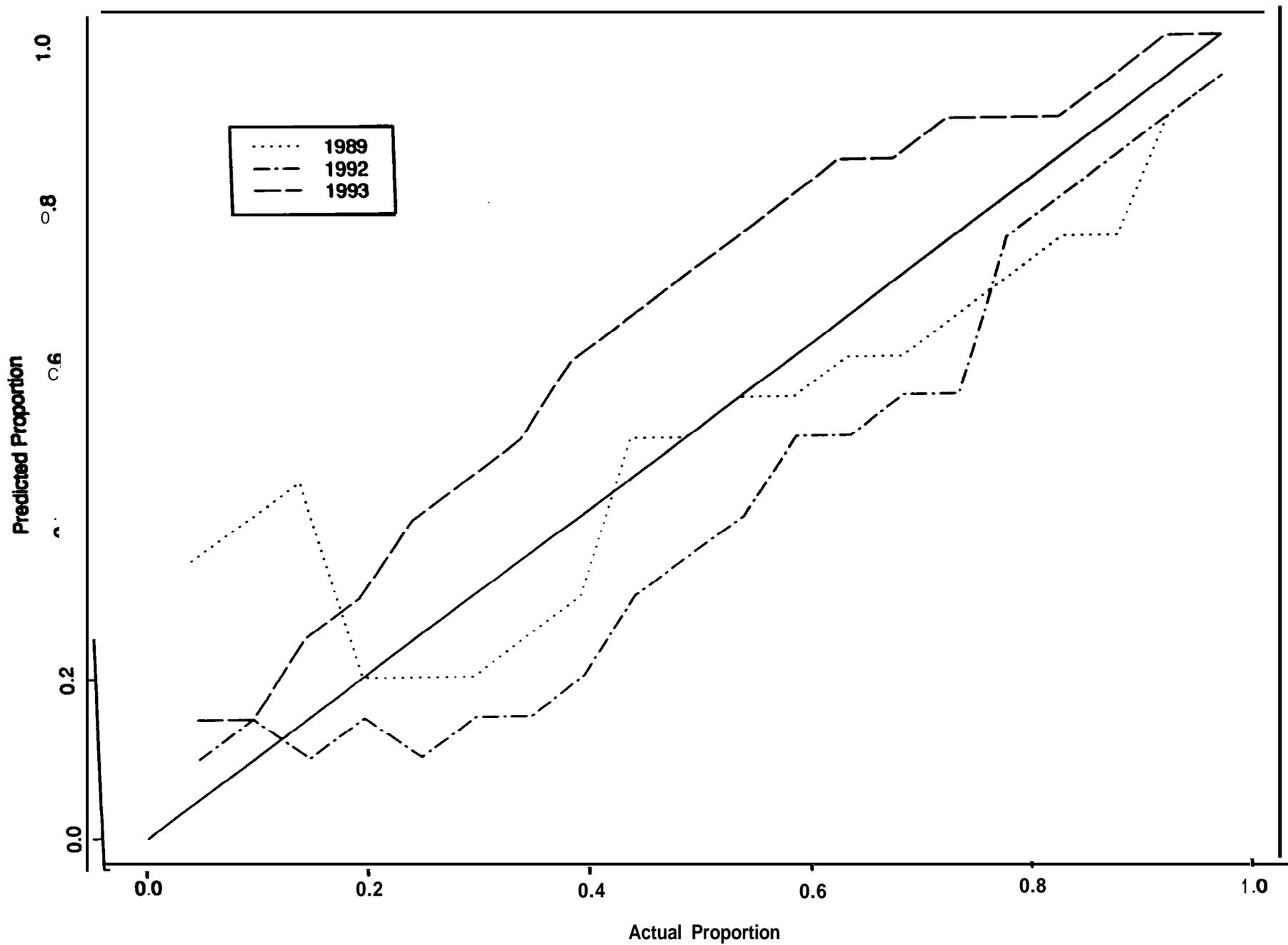
Bootstrap Prediction over Historic Years : Marsh Creek (sync)



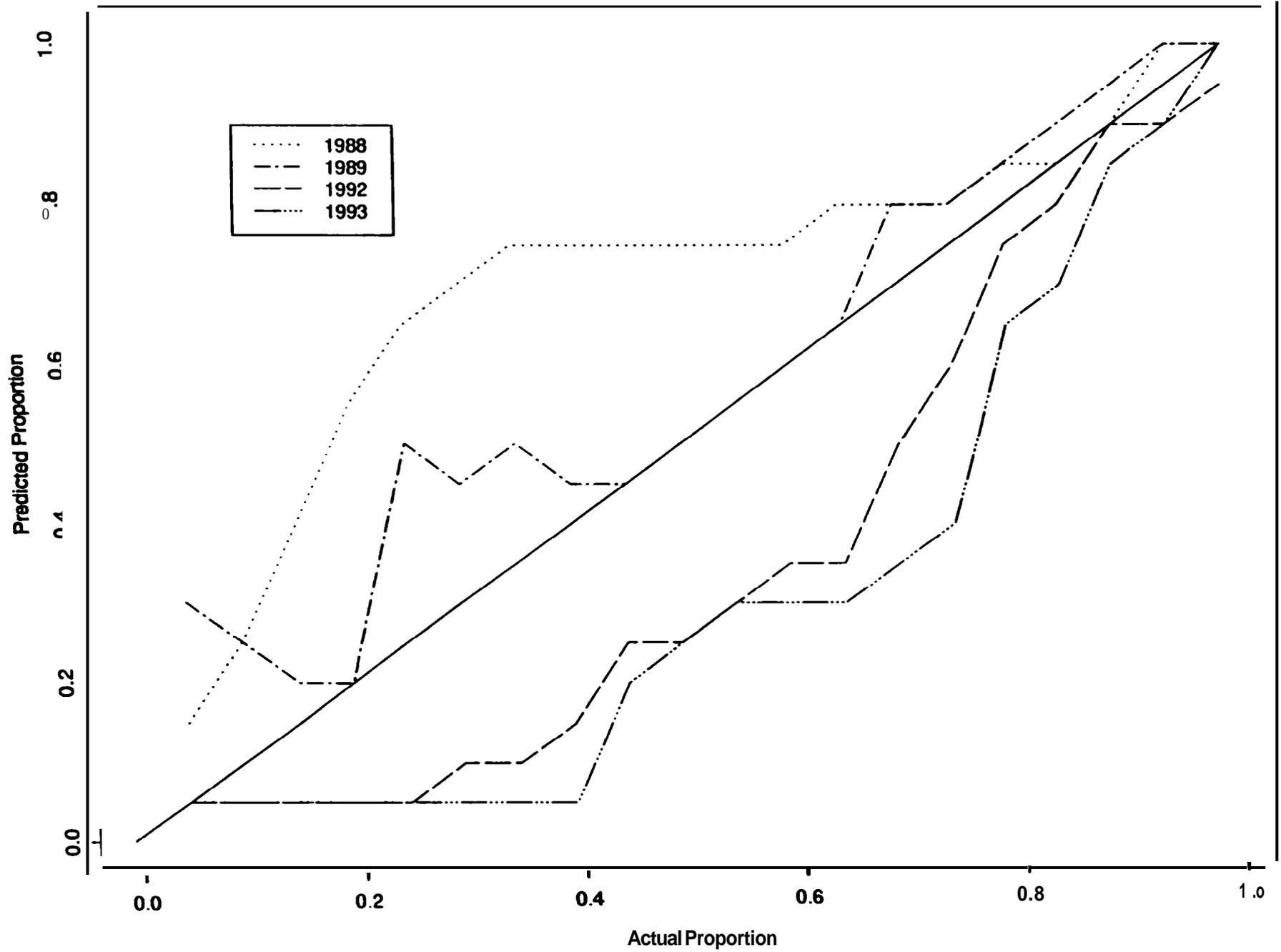
Bootstrap Prediction over Historic Years : Pahsimeroi (Is)



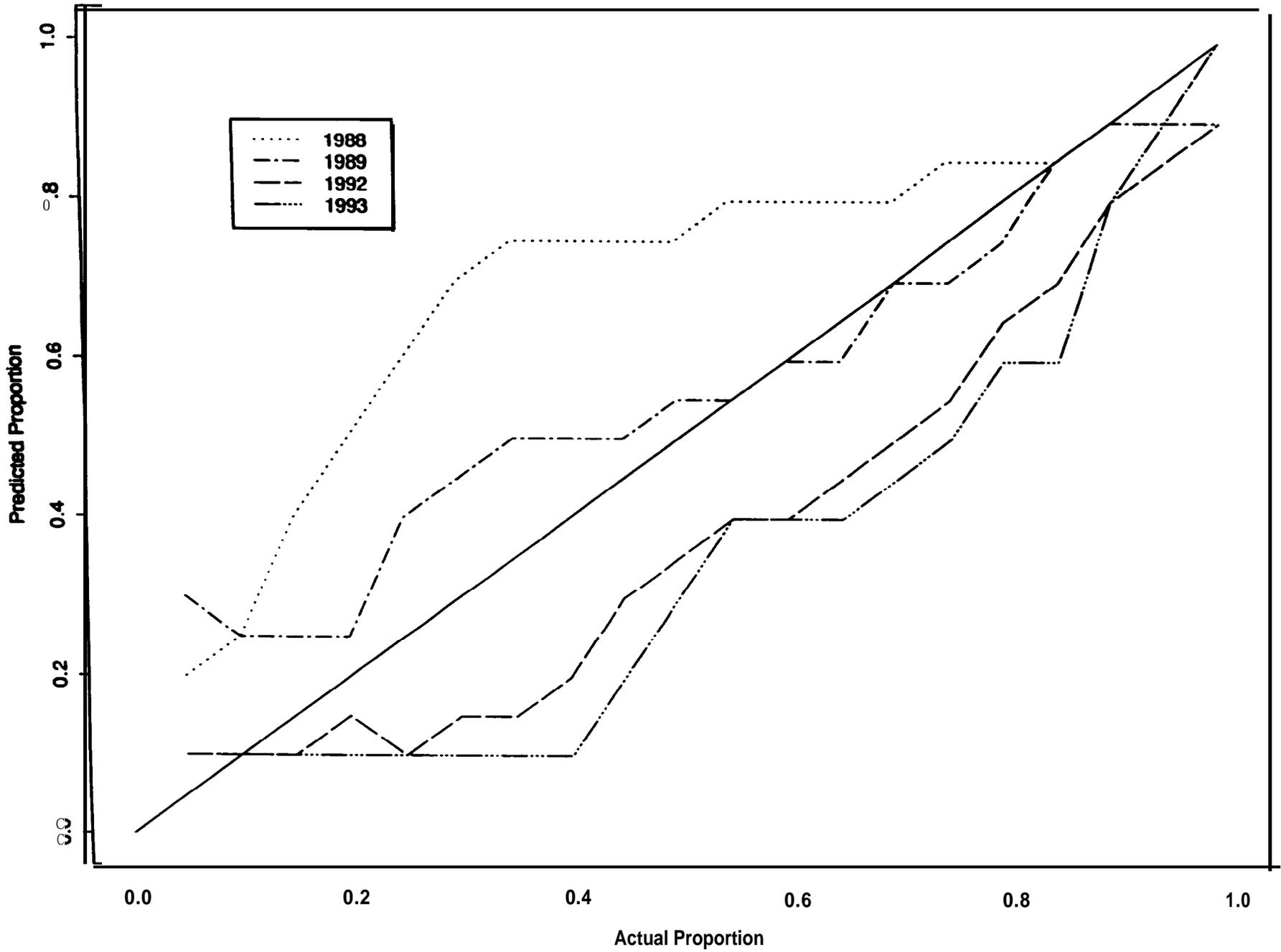
Bootstrap Prediction over Historic Years : Pahsimeroi (sync)



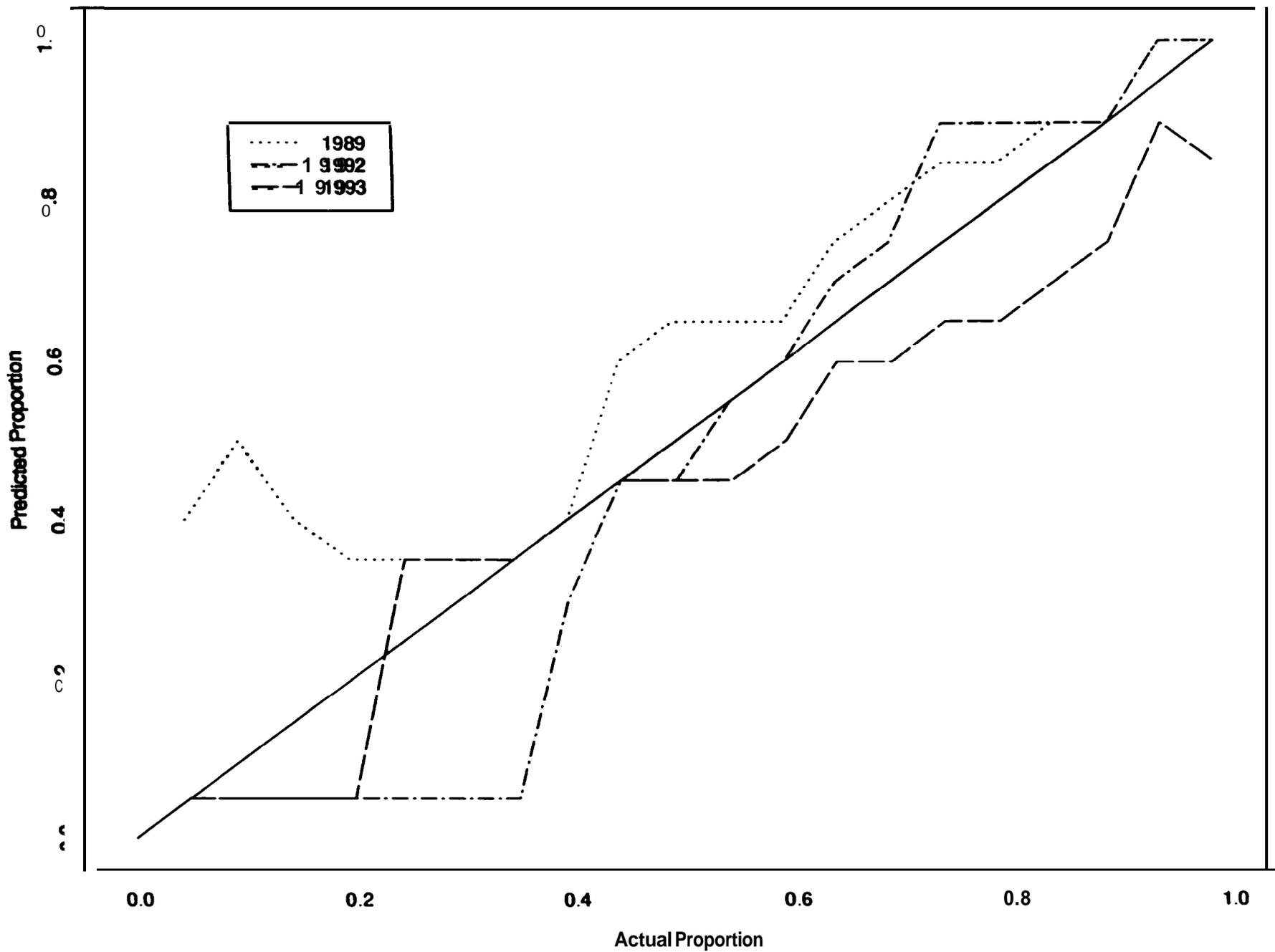
Bootstrap Prediction over Historic Years : Salmon River (Is)



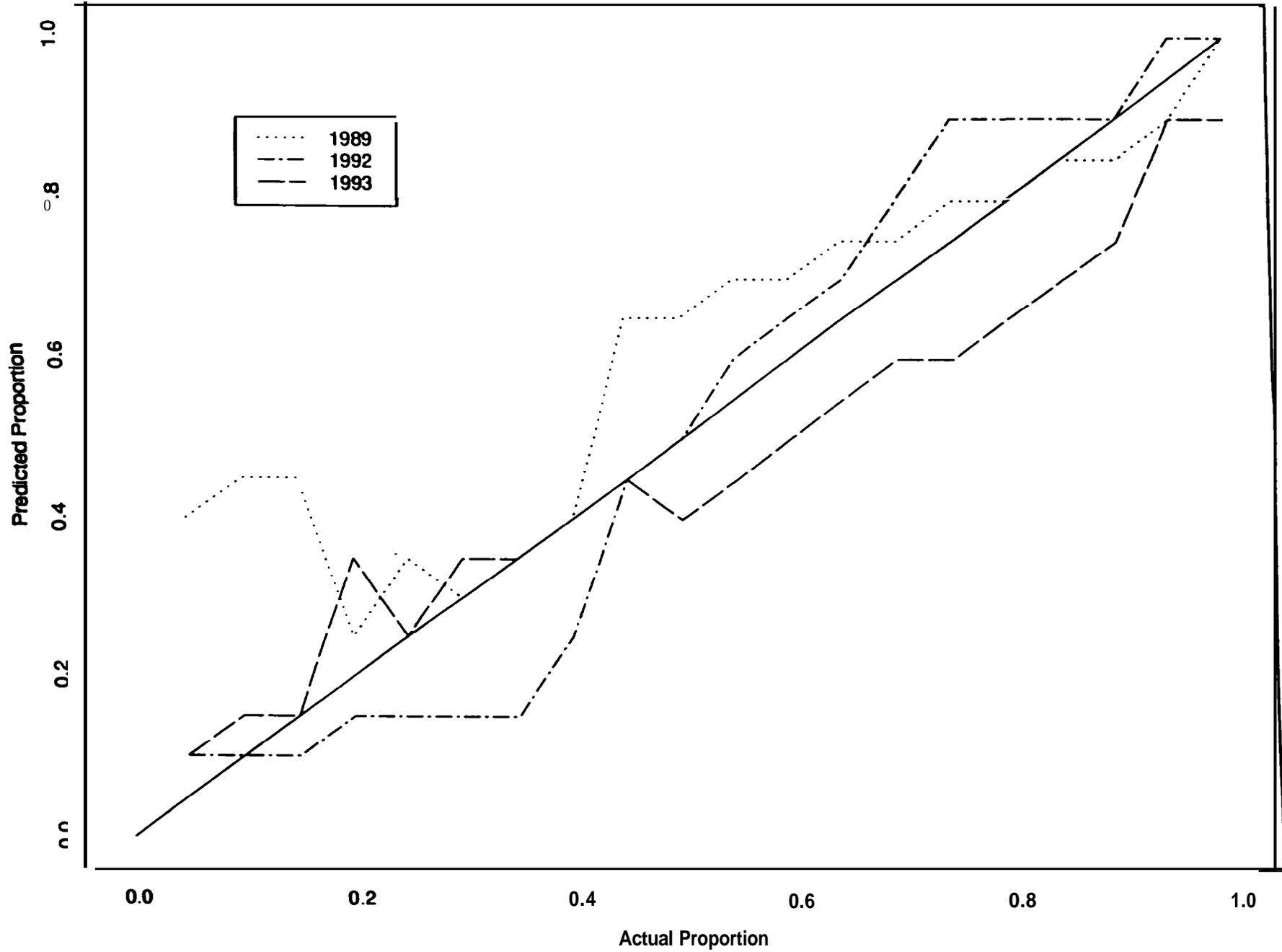
Bootstrap Prediction over Historic Years : Salmon River (sync)



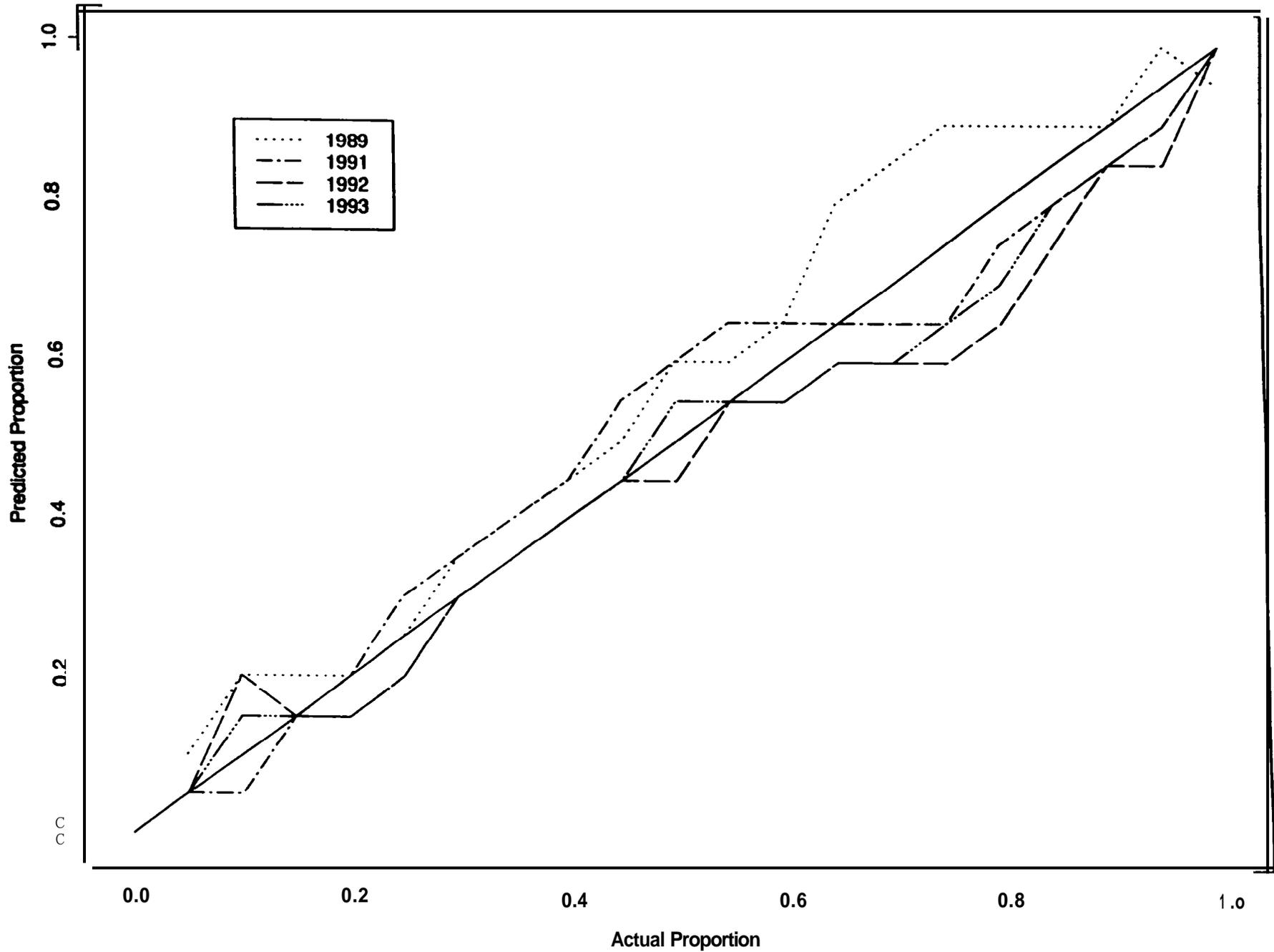
Bootstrap Prediction over Historic Years : Salmon River East Fork (Is)



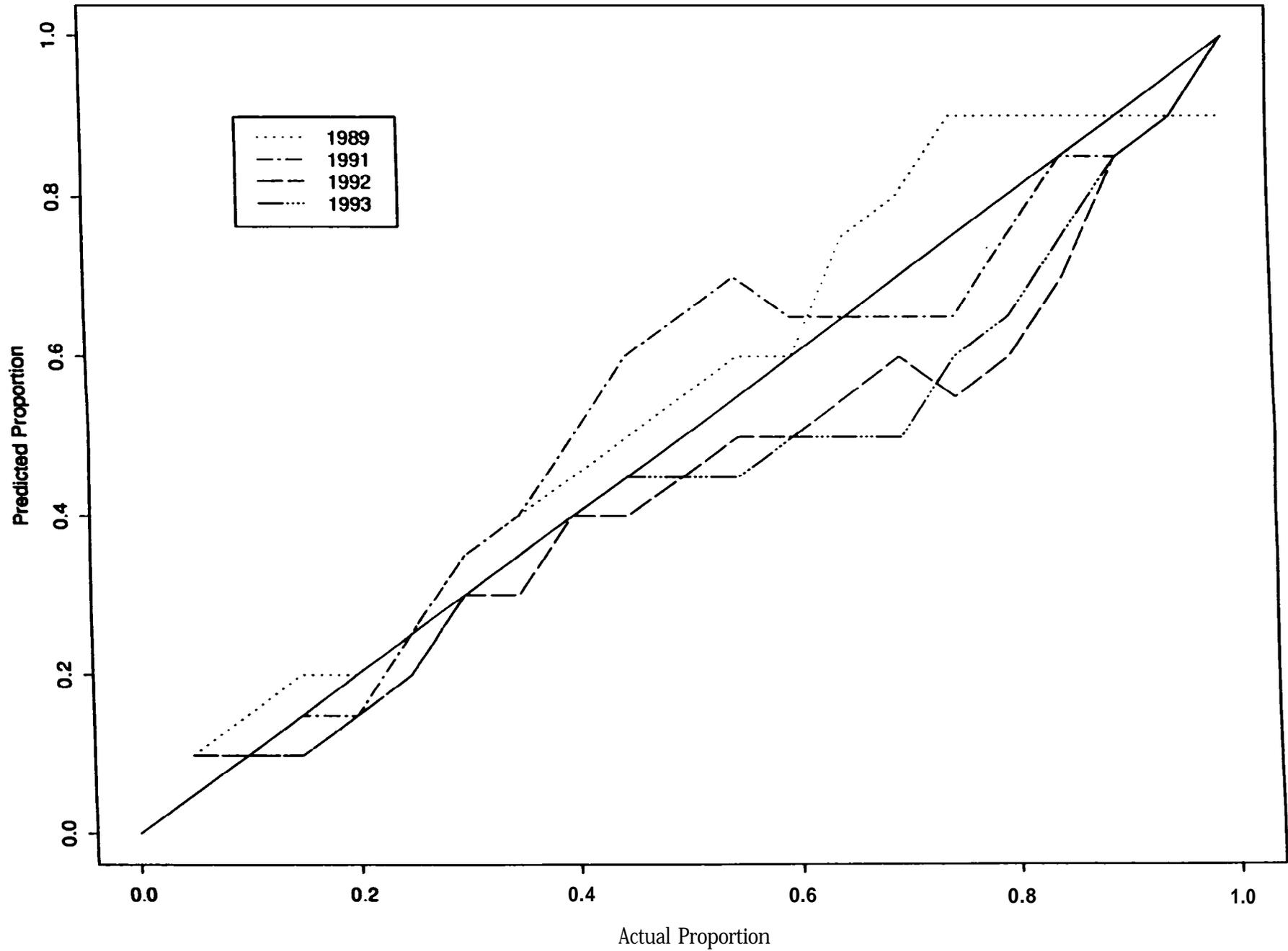
Bootstrap Prediction over Historic Years : Salmon River East Fork (sync)



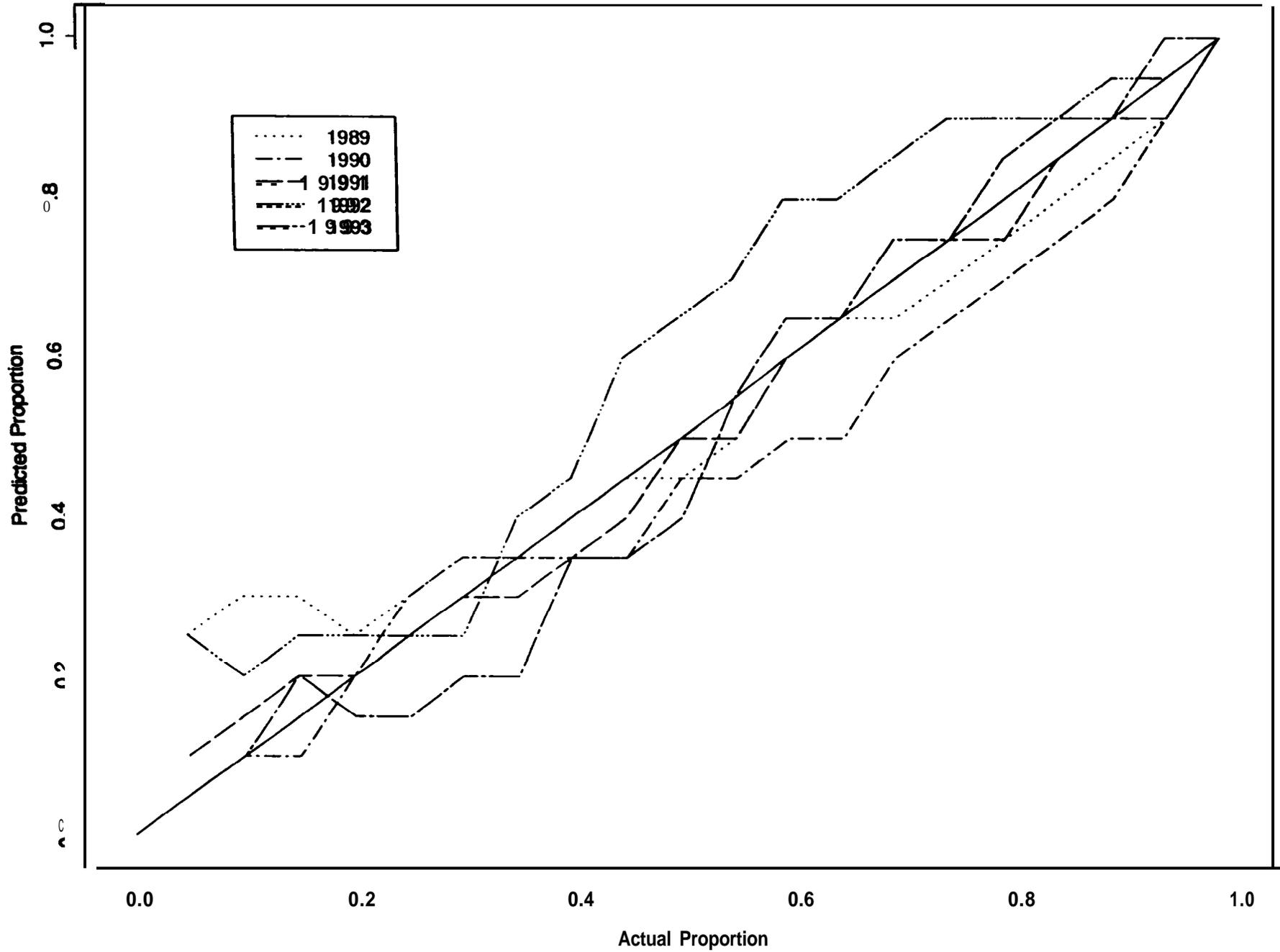
Bootstrap Prediction over Historic Years : Salmon River South Fork (Is)



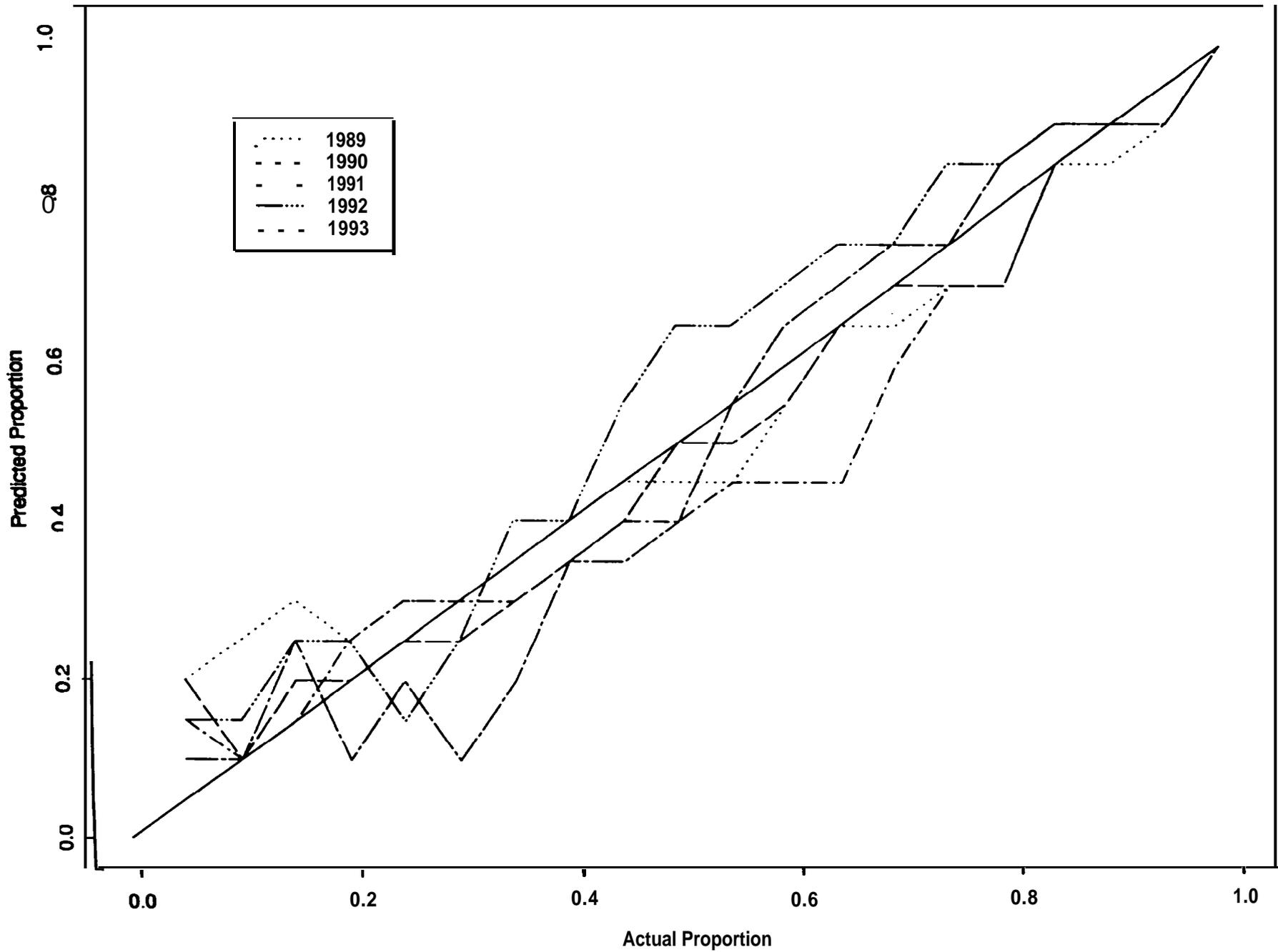
Bootstrap Prediction over Historic Years : Salmon River South Fork (sync)



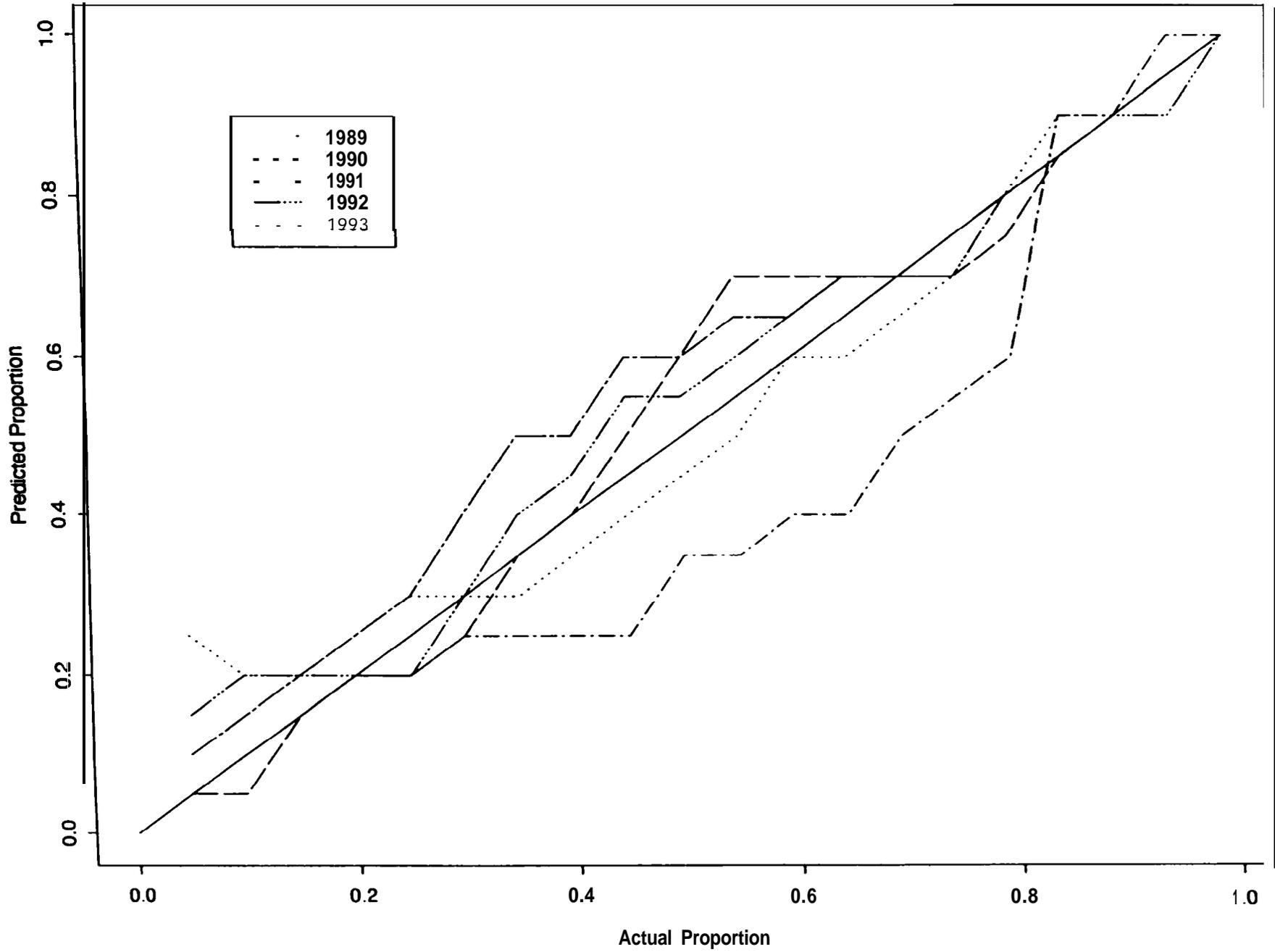
Bootstrap Prediction over Historic Years : Secesh River (Is)



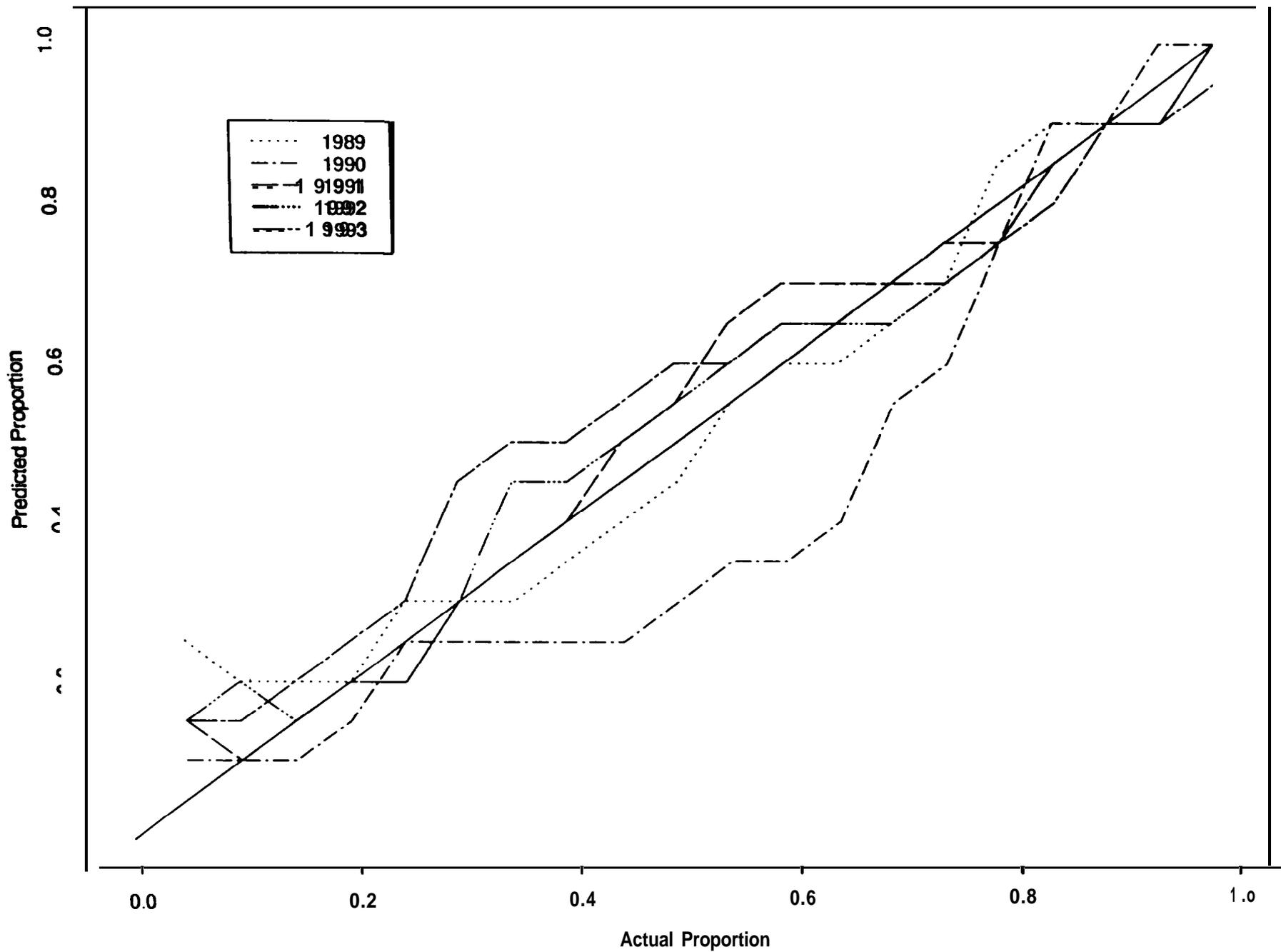
Bootstrap Prediction over Historic Years : Secesh River (sync)



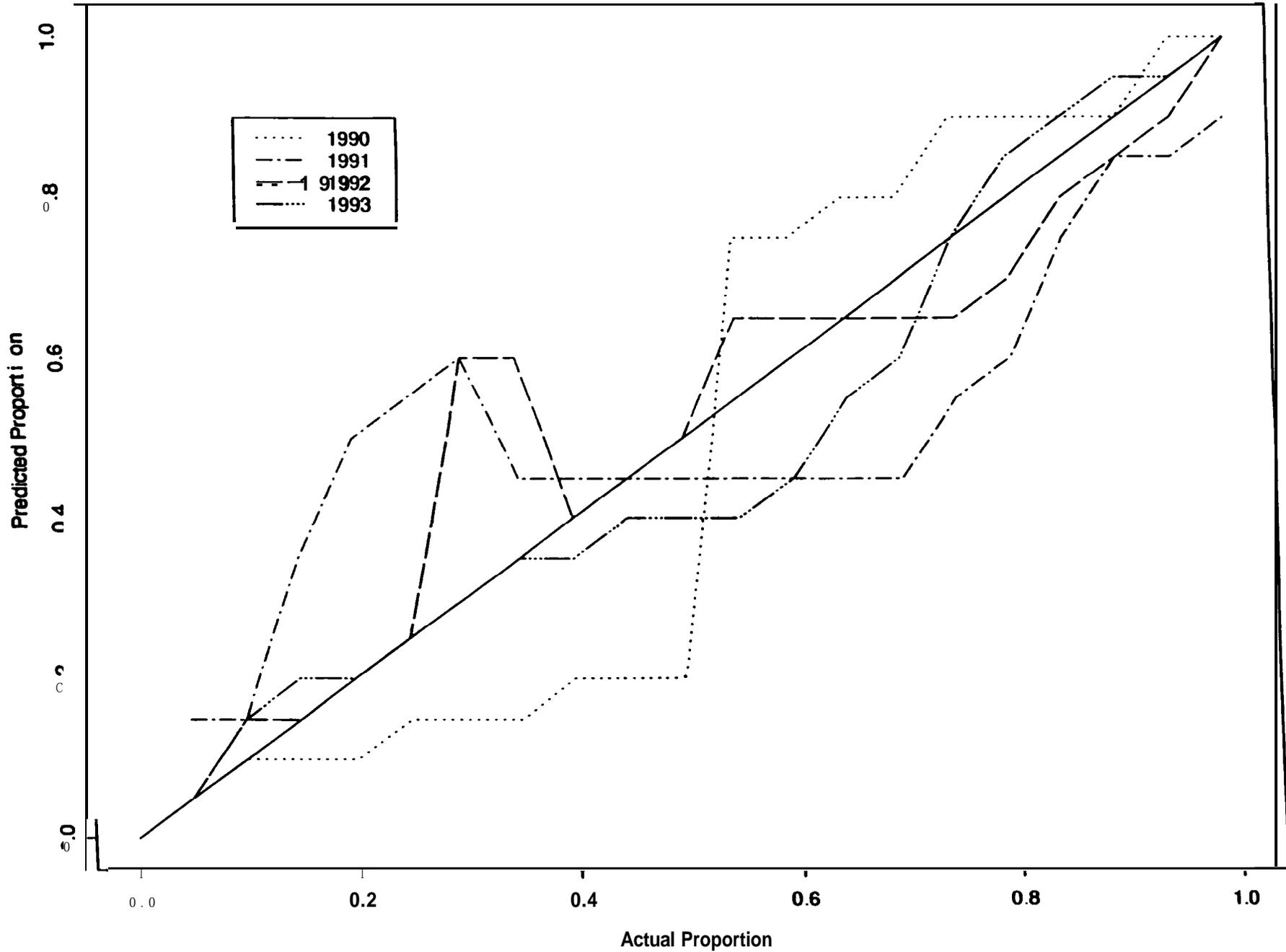
Bootstrap Prediction over Historic Years : South Fork Salmon (Is)



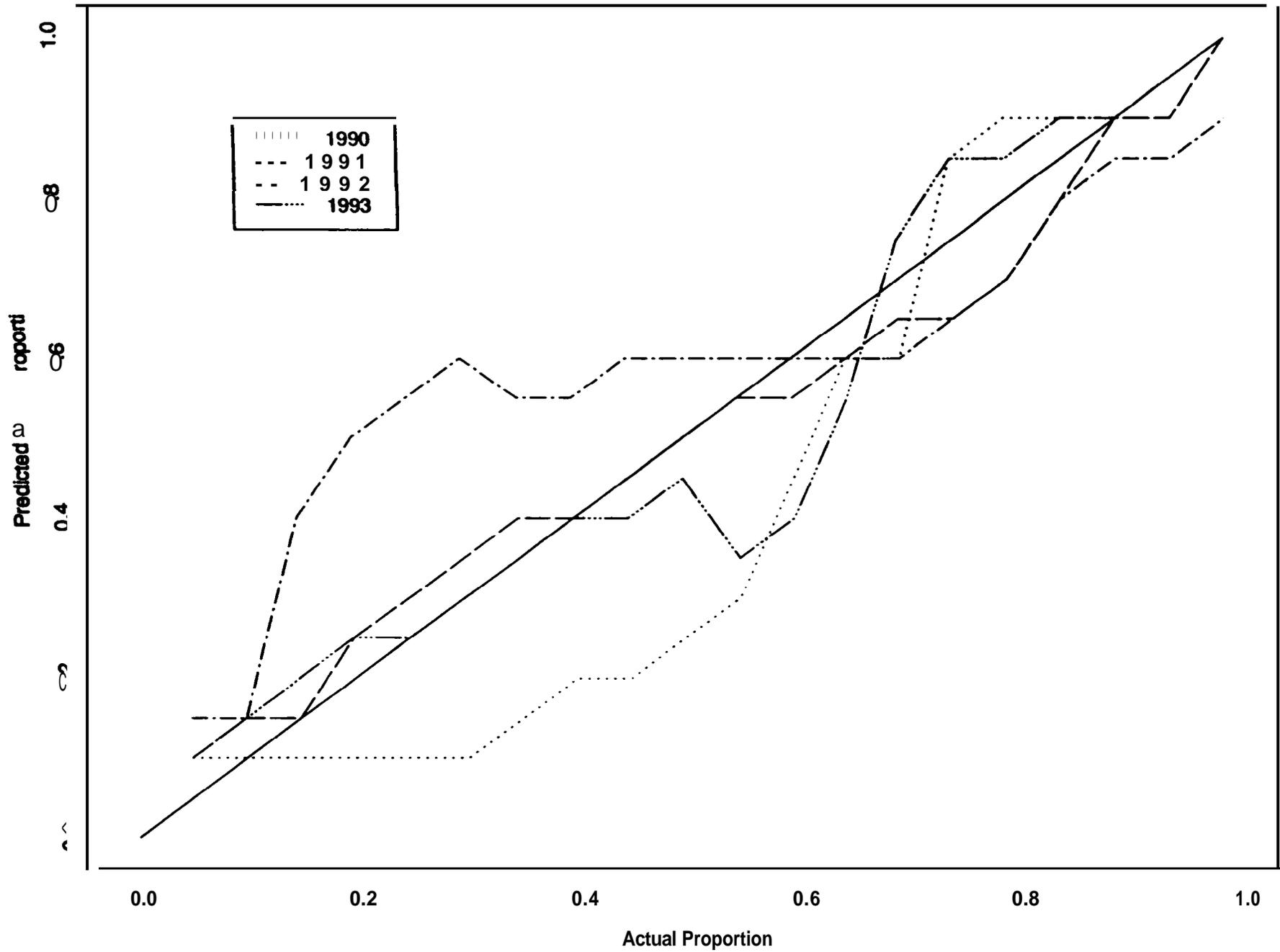
Bootstrap Prediction over Historic Years : South Fork Salmon (sync)



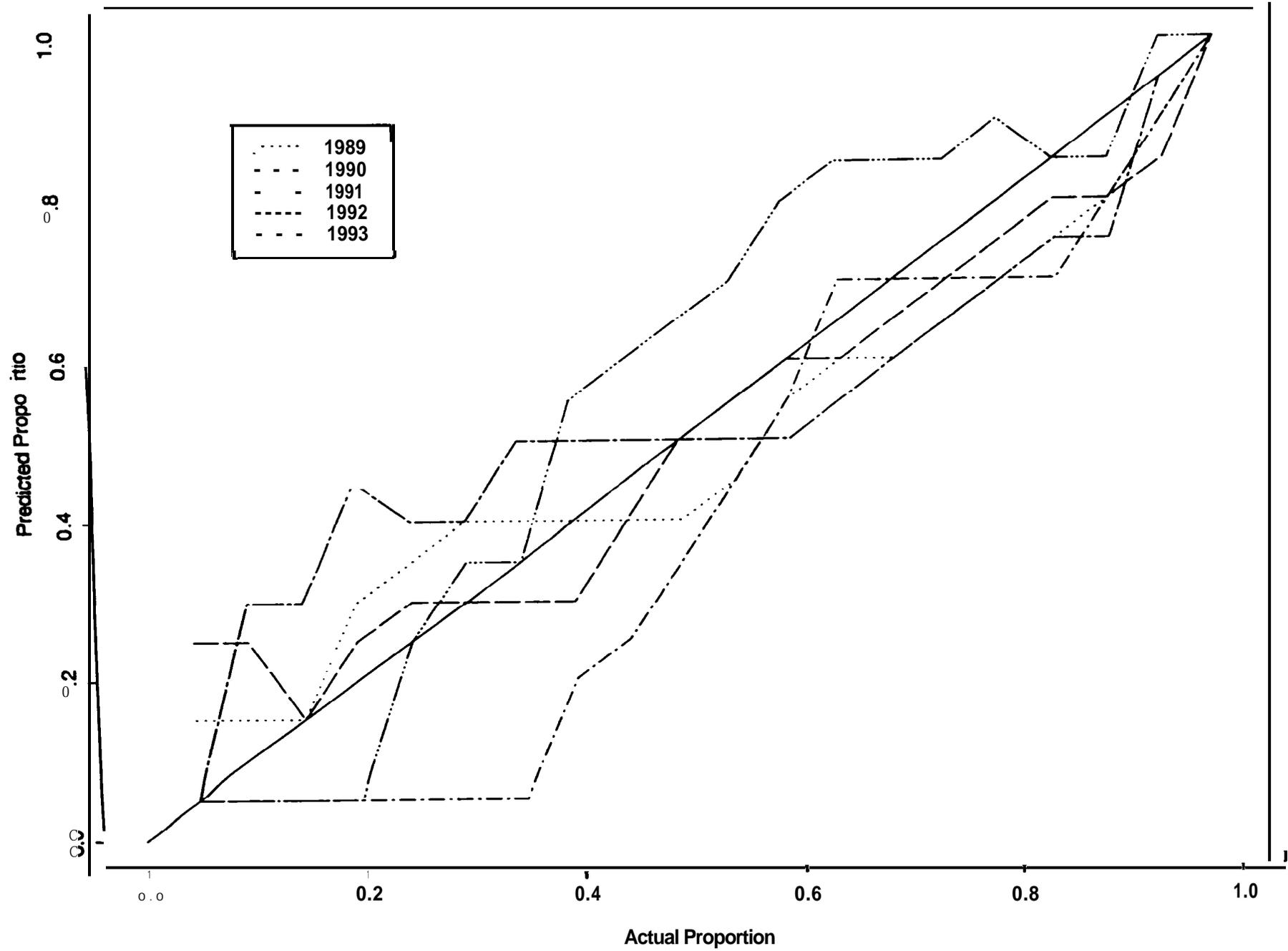
Bootstrap Prediction over Historic Years : Upper Middle Fork Salmon (Is)



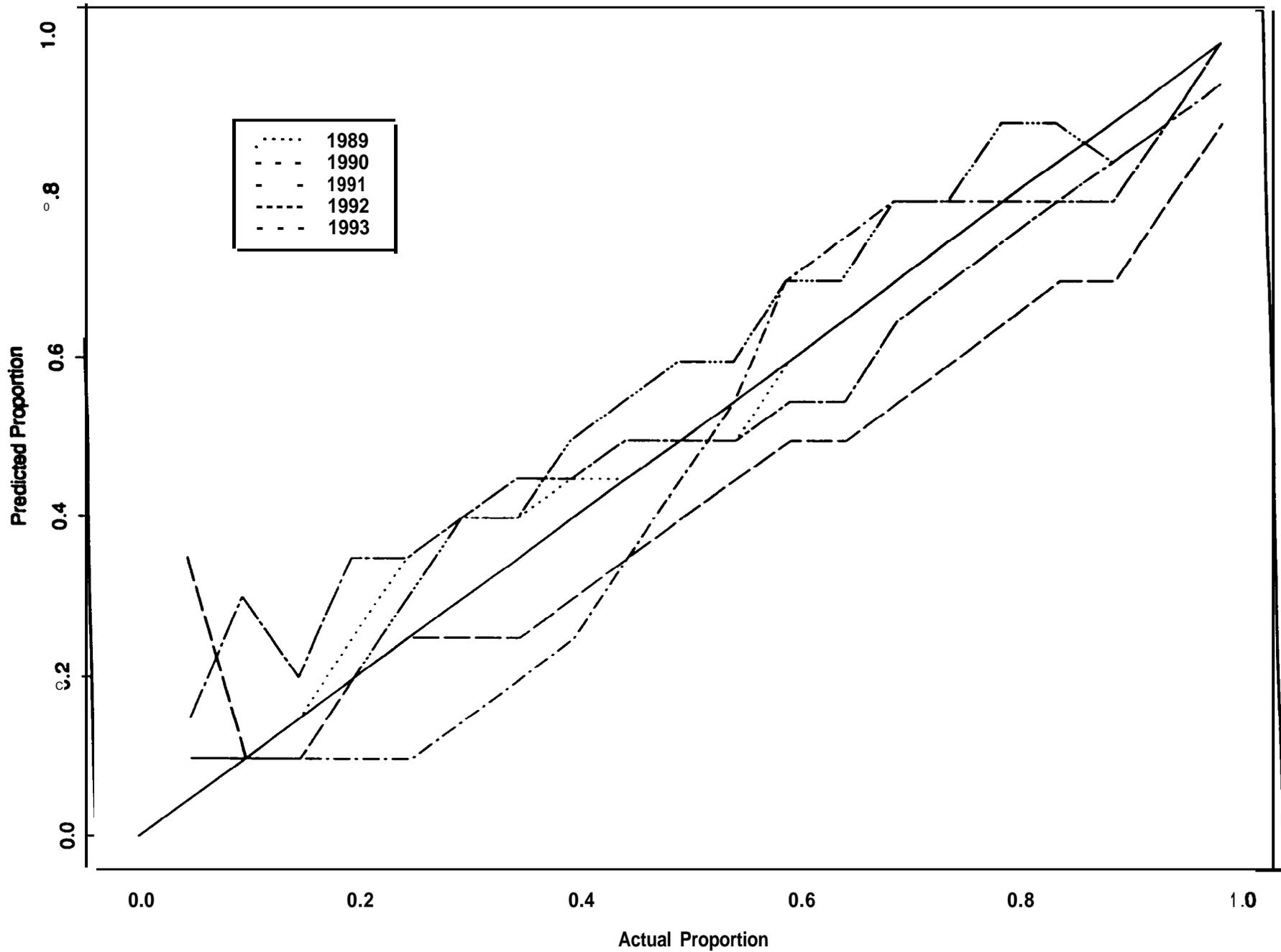
Bootstrap Prediction over Historic Years : Upper Middle Fork Salmon (sync)



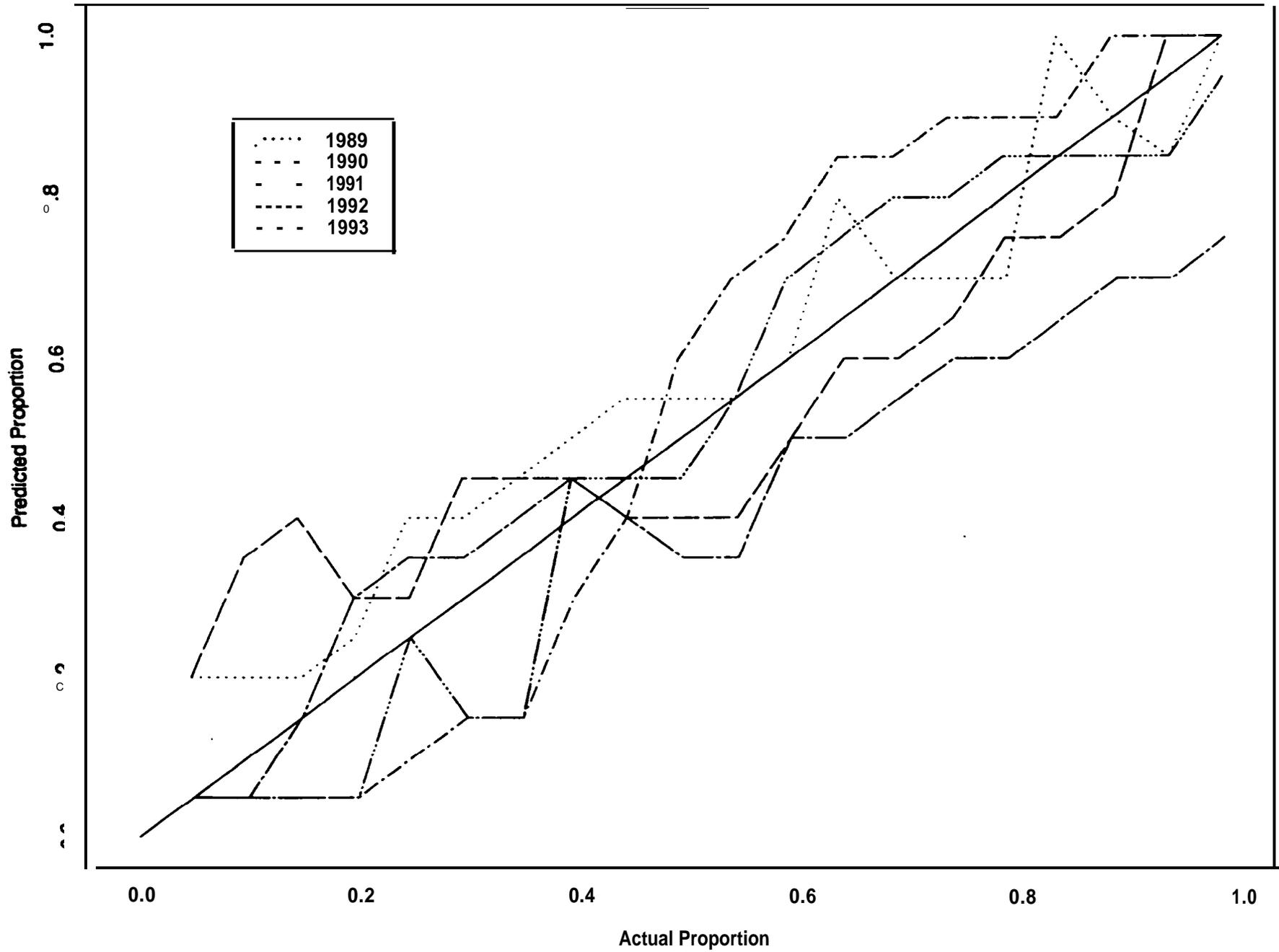
Bootstrap Prediction over Historic Years : Upper Salmon (Is)



Bootstrap Prediction over Historic Years : Upper Salmon (sync)



Bootstrap Prediction over Historic Years : Valley Creek (Is)



Bootstrap Prediction over Historic Years : Valley Creek (sync)

