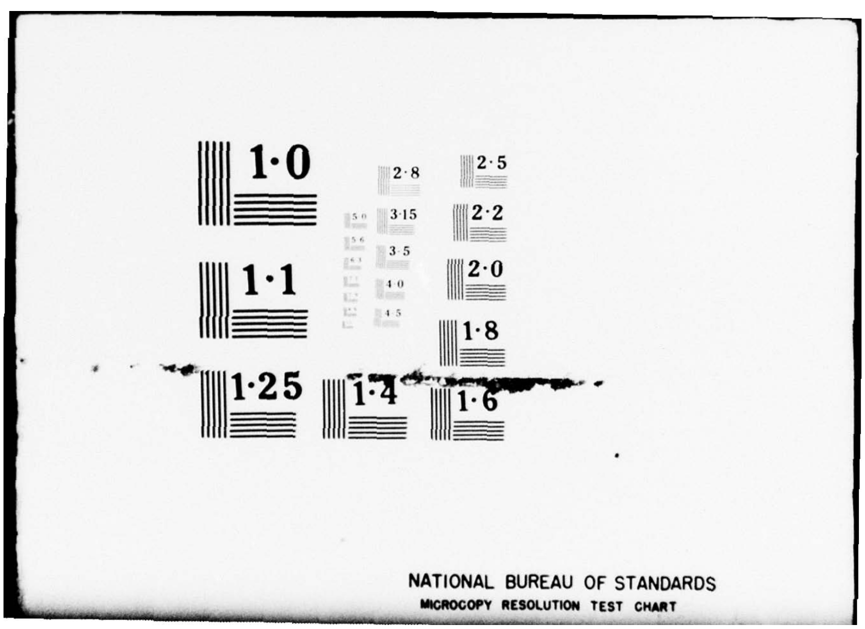
AD-A046 496 NAVAL POSTGRADUATE SCHOOL MONTEREY CALIF F/G 5/1
STUDY OF THE PREDICTION OF MANPOWER CHANGE BEHAVIOR USING REGRE--ETC(U)
JUN 77 R R READ
UNCLASSIFIED NPS55-77-26 NL

| OF |
ADA
046496

END
DATE
FILMED
12-77
DDC


1·0
2·8
2·5
3·15
2·2
3·5
2·0
4·0
1·1
1·8
1·25
1·4
1·6
NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

NPS55-77-26

NAVAL POSTGRADUATE SCHOOL

Monterey, California

Technical rept.,

STUDY OF THE PREDICTION OF MANPOWER CHANGE BEHAVIOR USING REGRESSION METHODS,

by

Robert R. R./Read

June 1977

69 p.

Prepared for:
Naval Personnel Research and Development Center
San Dieo, CA 92152

251 450

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Isham Linder
Superintendent

Jack R. Borsting
Provost

The research being reported on here was supported by the Naval Personnel Research and Development Center, San Diego, California. Reproduction of all or part of this report is authorized.

This report was prepared by:

Robert R. Read
Professor
Department of Operations Research

Review by:

Michael G. Sovereign, Chairman
Department of Operations Research

Released by:

for Robert R. Fossum
Dean of Research

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS55-77-26	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Study of the Prediction of Manpower Change Behavior Using Regression Methods		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Robert R. Read		8. CONTRACT OR GRANT NUMBER(s) N0002276P002611
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE June 1977
		13. NUMBER OF PAGES 63
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Navy Personnel Research and Development Center San Diego, Ca 92152		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Forecasting
Manpower
Ridge Regression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is shown that the use of regression methods in the forecasting of Separations (EAOS), Eligibles (to reenlist) and Non-reenlistments jointly by length of service and pay grade are competitive with the currently used "alpha" method. The question of whether one of the two methods of forecasting is clearly superior could not be addressed with the currently available data. The report describes the data base, presents various general characteristics of the data, summarizes the computational results

OVER →

20. Continued

that lead to the recommended choice of input, and recommends follow-on work to clarify the issues.

ACKNOWLEDGMENTS

This investigation was sponsored by the Naval Personnel Research and Development Center (NPRDC), San Diego, California, and was performed under the direction of Mr. Joe Silverman, Associate Director for Manpower Systems. Access to the data base was arranged by Mr. Robert Bollar of the NPRDC staff.

Professor Richard W. Butterworth of the Naval Postgraduate School has been the principal investigator of the overall project under which this study was performed and may be credited with the original ridge regression program written in APL.

TABLE OF CONTENTS

STUDY OF THE PREDICTION OF MANPOWER CHANGE BEHAVIOR USING REGRESSION METHODS

R. R. Read
Naval Postgraduate School
Monterey, California

ABSTRACT

It is shown that the use of regression methods in the forecasting of Separations (EAOS), Eligibles (to reenlist) and Non-reenlistments jointly by length of service and pay grade are competitive with the currently used "alpha" method. The question of whether one of the two methods of forecasting is clearly superior could not be addressed with the currently available data. The report describes the data base, presents various general characteristics of the data, summarizes the computational results that lead to the recommended choices of input, and recommends follow-on work to clarify the issues.

I. Introduction

Previous work explored the use of regression techniques (especially ridge regression) in the forecasting of contract losses, gains, and attritions by a set of LOS (length of service) categories and (separately) by PG (pay grade) categories. For sake of immediate reference, the results are included as Appendix A of the present report. The reader is referred to its first four pages for the definition of terms whose use continues in the current work.

The current follow-on efforts ignore the gain and attrition quantities in order to focus on three important components of contract losses:

S = Separations (EAOS)

Y = Eligibles

V = Non-Reenlistments

At the same time the forecasts are more refined in that they must be made jointly by 31 LOS cells and 7 PG cells. Pay grades E1 to E3 are lumped together in the first cell and pay grades E4 thru E9 form the remaining six cells.

In addition the current study utilizes an increased and modified data base. Now there are eleven years of data (1966 to 1976 inclusive) and the definition of LOS has changed, the new data reflecting this change and some other changes whose nature is not explicitly known to the author. These changes make obsolete the specific results of the previous work.

The updating of the data base has also had an effect on the currently used method (which we call the alpha method) for forecasting the three object variables S, Y, and V. The results of this method are available only for 1976, making difficult the comparison of the regression forecasts with current method forecasts since no actuals (1977) are yet available. Hence a fair comparison of the two methods cannot be made at this time.

A biased comparison is made instead. All eleven years' data is used to develop the (1976) forecast coefficients for both methods, but are applied to the previous year's data (i.e. 1975 serving as input data) to forecast 1976. Thus the 1976 data can serve as the actuals. The comparison is compromised because the 1976 data was used also to produce the forecast coefficients. Some reasons why this deficiency may favor the alpha method will be suggested further on in the report.

The regression methods involve the use of p input variables p ranging over 2, 3, and 4 and the most favorable set of p variables is always selected. Two measures of comparisons are computed:

MAE = mean absolute error

RMSE = root mean square error

where the errors are the differences between actuals and forecasts, and the means are computed over the 217 (31 × 7) cells. The results are summarized in Table 1.1. It is concluded that the regression method is competitive with the currently used alpha method.

TABLE 1.1

COMPARISON OF SUMMARY MEASURES OF FORECAST ERROR

FOR THE 'ALPHA METHOD' AND THE

BEST MULTIPLE REGRESSION USING p VARIABLES

		alpha	p = 2	p = 3	p = 4
Separations	MAE	123.5	161.5	146.4	118.4
	RMSE	496.0	651.7	576.4	436.7
Eligibles	MAE	104.6	120.2	88.3	75.9
	RMSE	380.6	469.4	292.9	232.8
Non-Reenlistments	MAE	50.6	77.2	54.9	40.3
	RMSE	291.8	338.0	238.7	199.7

MAE = mean absolute error

RMSE = root mean square error.

The report is organized as follows. A description of the new data base and the anomolies remaining in it is contained in Section II along with some general comments about what the author knows about the alpha method. Section III contains some comparisons over time of the macro behavior of the five important variables contained in the new data. Some interpretations and speculations are made. Section IV contains the refined details of the comparison of regression methods and the alpha method. Further interpretations are made. Conclusions and recommendations follow.

As mentioned earlier, the report of the previous work appears in Appendix A. Eleven year means of the three object variables appear in Appendix B. Appendix C contains APL programs pertinent to this report.

II. Description of the Data and the 'Alpha' Method

The new data contains eleven years' (1966 to 1976 inclusive) data for the five variables

V = non-reenlistments

S = separations (EAOS)

Y = eligibles

X = retentions

T = inventory (total)

for each of the original 279 (= 31 × 9) LOS/PG cells. (The telescoping of pay grades E1-E3 is done later when forecasting.)

All of the eleven years' data have been reworked to accommodate a new definition of LOS. The old definition was based upon pay entry base data which included pay credits for other federal service. The new definition refers to TAFMS (total active federal military service). Although the exact meaning is unknown to the author, it is known that reserve duty does not count. (It appears that the data of actual entry is used by no one.) The new definition affects our previous LOS entries quite noticeably.

Previously, the variable V (non-reenlistments) was a derived quantity being the difference of eligibles and retentions. Now it is obtained independently in some way and there is noise in the relationship V = Y-X. This and other data anomolies are described next.

Each of the five variables V, S, T, X, T are recorded in 3069 (= 11 × 31 × 9) cases. Separations exceed Eligibles (i.e. $S \geq Y$) in all of them, as is proper. However eligibles fail to exceed retentions ($Y \not\geq X$) in 97 (of 3069) cases. Of these, 44 failures are in the most recent two years (1975-76) and in PG, E8 and E9. PG E7 contains some concentration of these also, but the remainder appear to be scattered. The non-reenlistment variable fails to be non-negative ($V \not\geq 0$) in 107 cases. Again 44 of these failures are in the last two years and for E8 and E9. In fact these four cells (summed over LOS) have the exact same distribution as the previous analomy ($Y \not\geq X$). Again E7 has a number of negative entries The differences between separations and retentions (contract losses) should be non-negative. This fails ($S \not\geq X$) in 76 cases. The noise is concentrated in PG E7, E8, E9 and years 1975-76 which accounts for 52 of the 76 cases. The relationship $V = Y-X$ (i.e. non-reenlistments form the difference between eligibles and retentions) holds up in 2256 of the 3069 cases (74%). The discrepancies cluster mostly in E4 to E7 and the first 20 LOS categories.

Finally, the new data appear to have larger inventories in 1966-67. Specifically 591×10^3 vs 586×10^3 for 1966 and 662×10^3 vs 653×10^3 for 1967.

The currently used method of forecasting the object variables is a "black box" as far as the user is concerned. It requires the production of "alpha matrices" whose elements serve as multipliers to convert the input into the forecast. Three such 31 by 7 matrices are available to the author, α_S, α_E, α_N and are used as follows: Let $\hat{S}$, $\hat{Y}$, $\hat{V}$ represent the output projections for the next full time period. Then

$$\hat{S} = \alpha_S T_0$$

$$\hat{Y} = \alpha_E S_0$$

$$\hat{V} = \alpha_N Y_0$$

where T_0 is the matrix of most recent inventories, S_0 is the matrix of separations for the most recent period, and Y_0 is the matrix of eligibles for the most recent period. The indicated multiplications of matrices are elementwise.

The alpha matrices are updated each year when the data become available. The technique does not appear to be well documented but is available in the form of computer programs. It is believed to be a version of "Tukey's Smoothing Medians" and Tukey's materials on exploratory data analysis may be useful in tracking it down. (See also: McNeil, Interactive Data Analysis.) Comments concerning the application of it to the problem at hand are contained in the informal papers entitled 'Introduction to Smoothing and Projection of Naval Population Matrix Time Series,' and 'Rifselsm Overview.' These do not appear to be very useful to the analyst.

The three 1976 alpha matrices used in this study appear in Tables 2.1 to 2.3. The values have been converted to percentages and rounded.

TABLE 2.1

ALPHA MATRIX MULTIPLIERS TO FORECAST

SEPARATIONS FROM INVENTORY (includes 1976)

LOS	PG						
	2	5	19	29	58	3	2
	12	27	74	46	45	2	23
	37	28	33	43	38	1	93
	67	73	93	79	29	10	79
	25	25	17	22	36	30	2
	41	41	29	38	31	47	99
	14	18	16	23	27	21	4
	18	22	22	32	32	41	5
	11	22	25	32	52	32	10
	23	29	32	36	44	26	18
	16	21	23	30	43	12	6
	36	32	28	33	44	48	11
	14	18	24	34	38	37	23
	25	24	27	32	29	32	45
	28	22	26	31	29	32	35
	30	35	31	28	29	30	33
	23	16	19	23	25	28	26
	65	11	15	16	17	17	17
	3	7	7	10	16	19	22
	1	1	7	12	19	22	21
	1	26	10	16	18	19	21
	1	46	25	16	18	23	19
	1	2	49	15	22	28	31
	1	1	74	20	21	19	21
	1	1	82	42	17	17	19
	1	4	29	58	22	16	21
	1	1	13	27	17	27	29
	5	1	3	21	15	15	12
	1	1	1	8	7	3	5
	1	1	74	17	5	7	9
	1	2	13	27	8	10	14

TABLE 2.2

ALPHA MATRIX MULTIPLIERS TO FORECAST

ELIGIBLES FROM SEPARATIONS (includes 1976)

PG

LOS

38	79	85	100	100	3	99
48	98	100	100	82	14	1
37	98	99	100	95	3	93
41	98	100	98	100	14	1
53	98	99	100	100	100	2
74	97	99	100	100	100	100
78	96	99	100	95	99	56
56	97	99	100	100	98	14
68	95	99	100	100	100	14
64	96	99	100	100	95	56
50	96	99	100	100	100	96
59	93	99	100	100	100	12
51	94	99	100	100	100	56
68	86	99	100	100	100	100
76	97	99	100	100	100	100
95	97	100	100	100	100	100
43	93	100	100	100	100	100
78	100	100	100	100	100	100
16	100	98	100	100	100	100
1	5	94	100	100	100	100
1	99	100	100	100	100	100
1	95	91	100	100	100	100
1	83	95	100	100	100	100
1	45	85	100	100	100	100
1	1	95	100	100	97	100
0	83	100	100	100	100	100
0	1	17	100	100	100	100
0	0	17	100	100	100	100
0	1	1	100	100	100	100
0	0	87	99	100	99	95
0	2	96	99	97	100	100

TABLE 2.3

ALPHA MATRIX MULTIPLIERS TO FORECAST

NON-REENLISTMENTS FROM ELIGIBLES (includes 1976)

LOS (rows) × PG (columns)

70	43	39	57	62	3	12
89	68	14	48	5	2	1
81	61	36	21	86	4	33
81	75	60	19	54	15	2
56	52	46	40	53	99	3
30	60	68	70	52	10	26
35	37	46	46	19	5	56
18	28	32	33	4	31	5
38	20	26	22	3	5	5
42	20	23	18	7	8	4
2	20	14	14	6	2	7
26	8	15	10	7	14	5
1	12	10	6	4	1	3
16	23	5	4	2	4	5
34	7	5	4	1	1	2
2	9	4	2	1	2	1
45	1	5	1	1	1	1
1	23	2	1	1	1	1
10	15	6	2	1	1	1
1	3	21	3	2	2	2
1	31	46	5	4	1	1
1	46	39	7	4	4	1
1	70	34	6	1	1	1
1	46	1	1	2	1	1
1	1	1	16	5	1	1
1	70	2	19	7	3	1
1	1	1	2	1	1	2
1	1	2	22	5	1	1
1	1	3	4	2	1	1
1	1	1	1	18	1	2
1	1	1	2	8	1	1

III. Plots Illustrating the Macro Behavior of the Data.

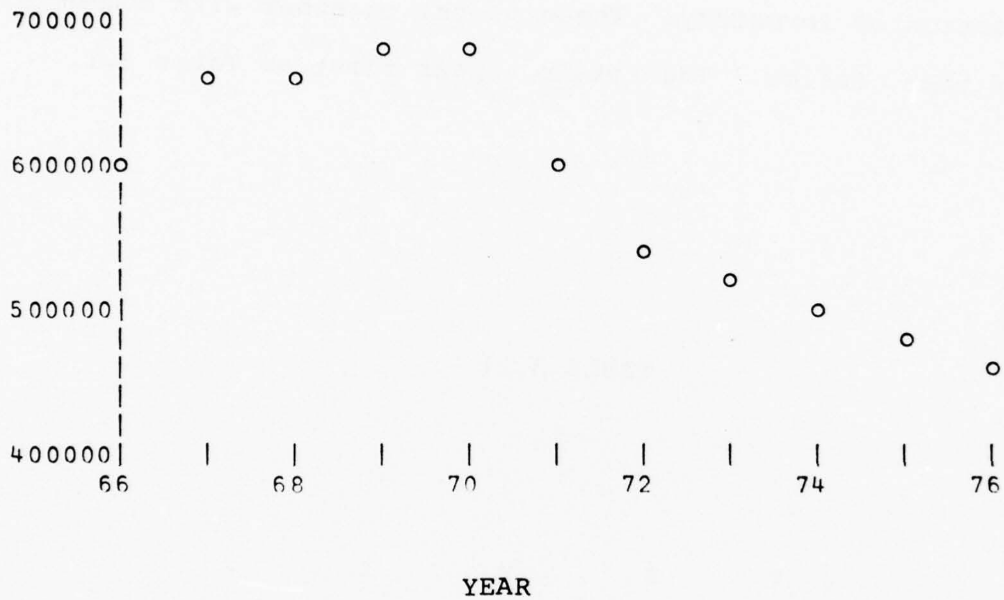
The five variables were summed over LOS and PG, and their behavior has been plotted against time, against each other for selected pairs, and the first four against time as a fraction of Inventory. These plots, together with comments about them, follow. The totals appear first as Table 3.1.

TABLE 3.11

	V	S	Y	X	T
1966	73319	120368	117929	43409	591463
1967	99598	153964	150426	49604	662056
1968	101281	136697	133677	31348	663779
1969	125236	161051	158008	31897	673589
1970	138025	191950	180070	41599	684109
1971	87690	155077	130599	42573	605898
1972	74181	134110	120458	45865	542298
1973	81670	147908	136894	54383	510669
1974	55365	122347	111399	55538	490009
1975	45815	118115	102735	55625	474735
1976	47471	112983	94730	45639	465748

FIGURE 3.1

INVENTORY VS TIME



Total stocks of Navy enlisted personnel during the Vietnam War period and the decline after that period are illustrated in Figure 3.1. Total Separations (EAOS) are shown in Figure 3.2. The spike for 1970 surely represents an 'end of the war' idiosyncrasy. Separations as a fraction of stocks (Figure 3.3) shows two spikes 1970 and 1973. We do not know how much the latter one affects the alpha matrix coefficients generated to produce a forecast of 1977 EAOS as a multiple of stocks.

FIGURE 3.2

SEPARATIONS VS TIME

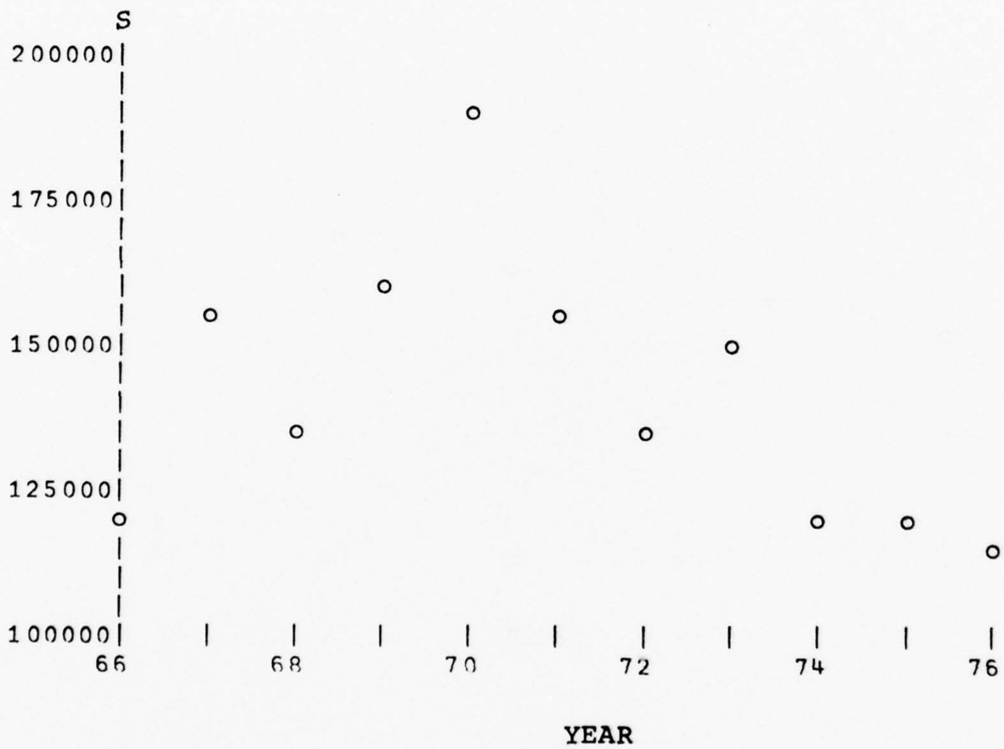
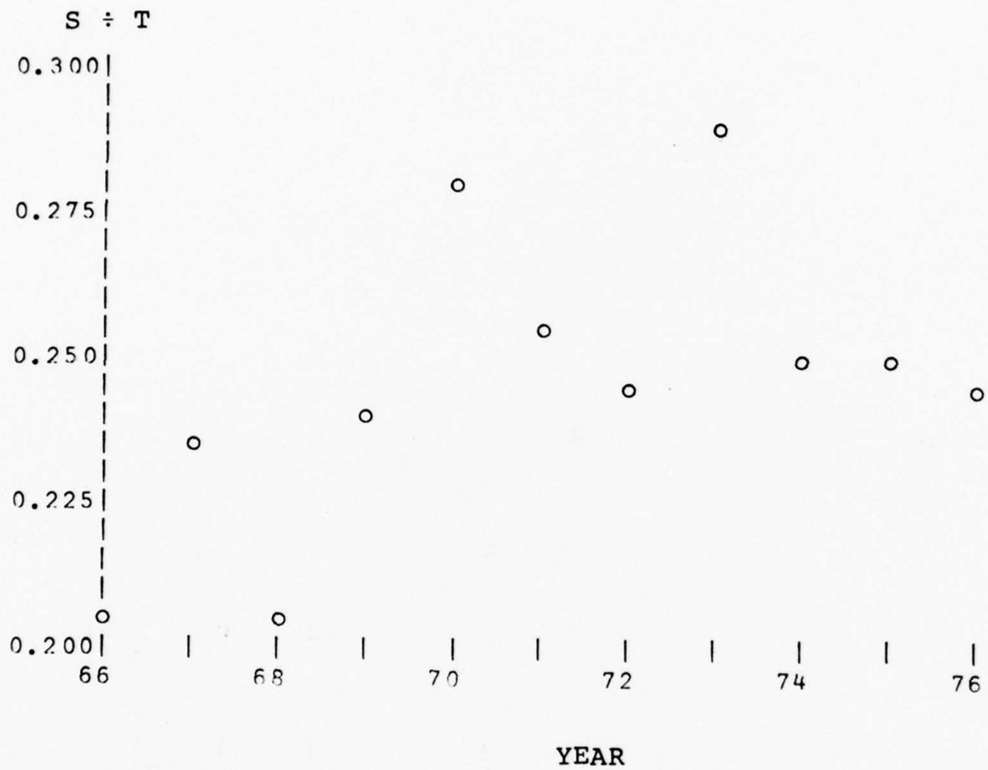


FIGURE 3.3

SEPARATION RATE VS TIME



The next four macro plots are scatter diagrams of separations with each of the other four variables. They show:

1) a moderate correlation of separations with inventory,
2) stronger correlations of separations with non-reenlistments and eligibles,
3) virtually no correlation between separations and retentions.

FIGURE 3.4

SEPARATIONS VS INVENTORY

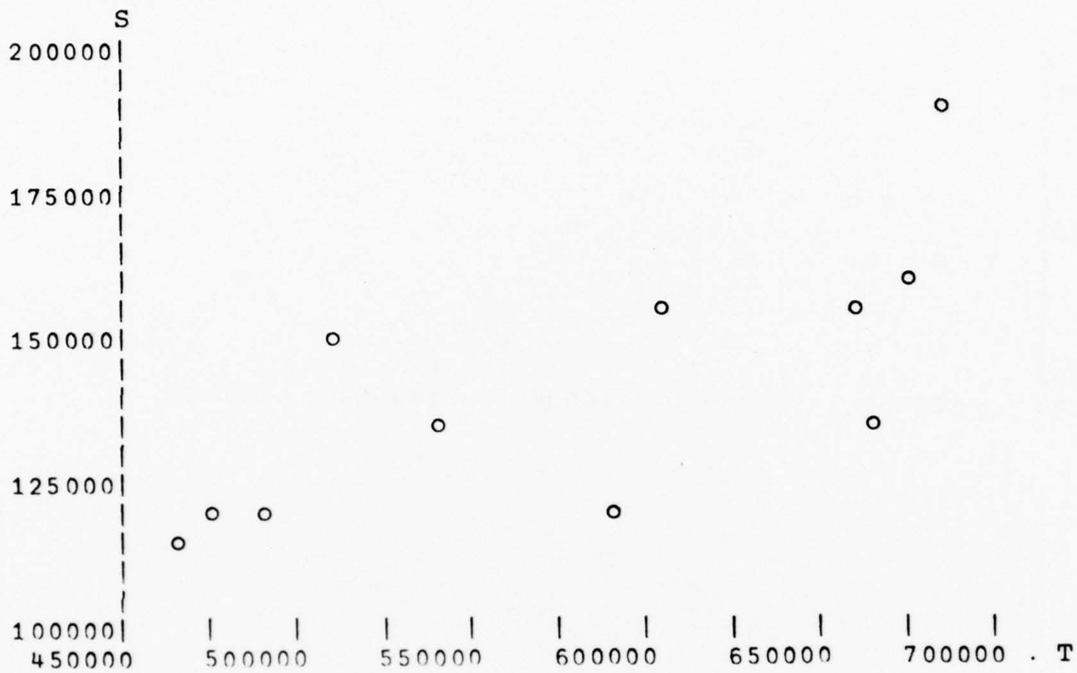


FIGURE 3.5

NON-REENLISTMENTS VS SEPARATIONS

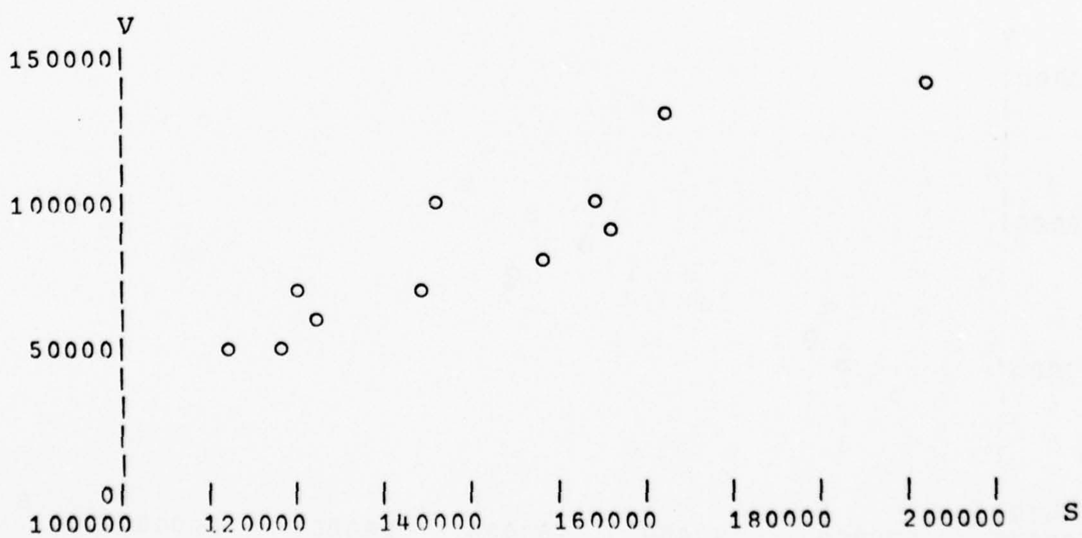


FIGURE 3.6

RETENTIONS VS SEPARATIONS

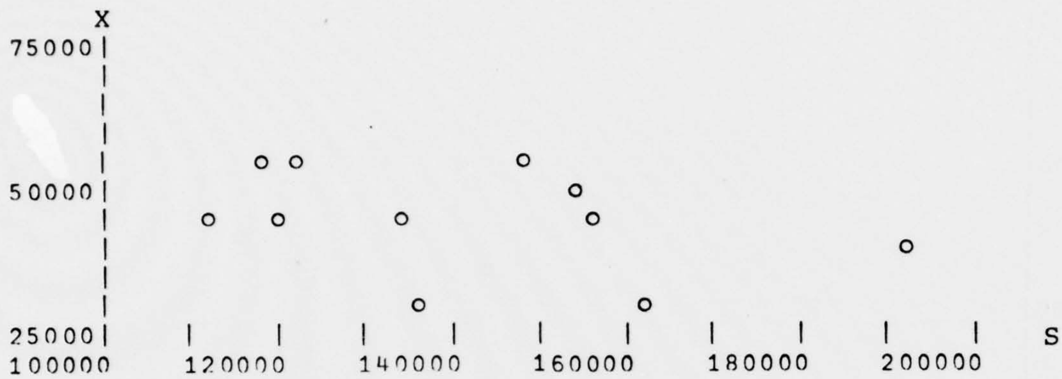


FIGURE 3.7

ELIGIBLES VS SEPARATIONS

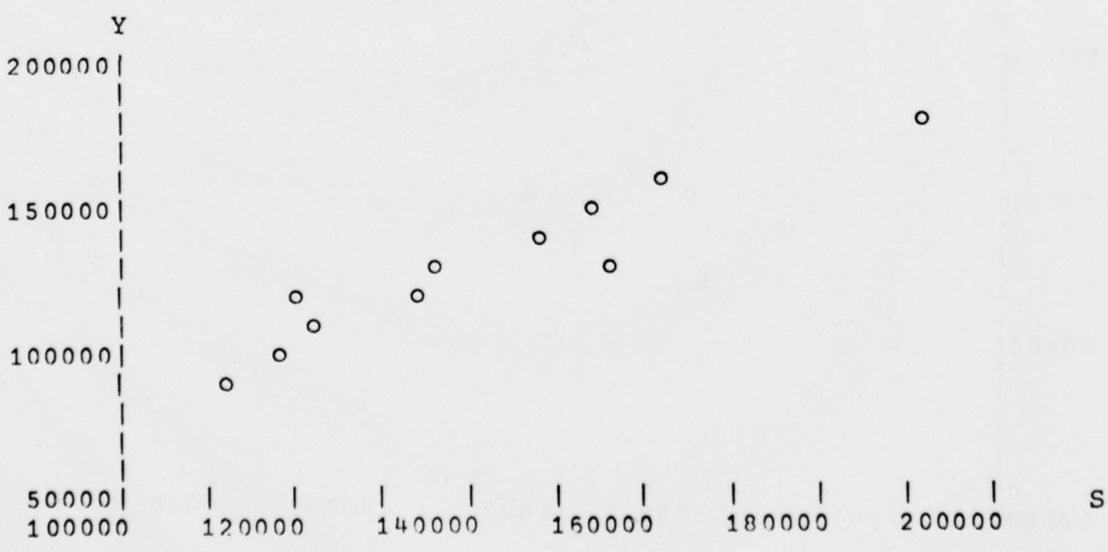


FIGURE 3.8

ELIGIBLES AS A FRACTION OF SEPARATIONS VS TIME

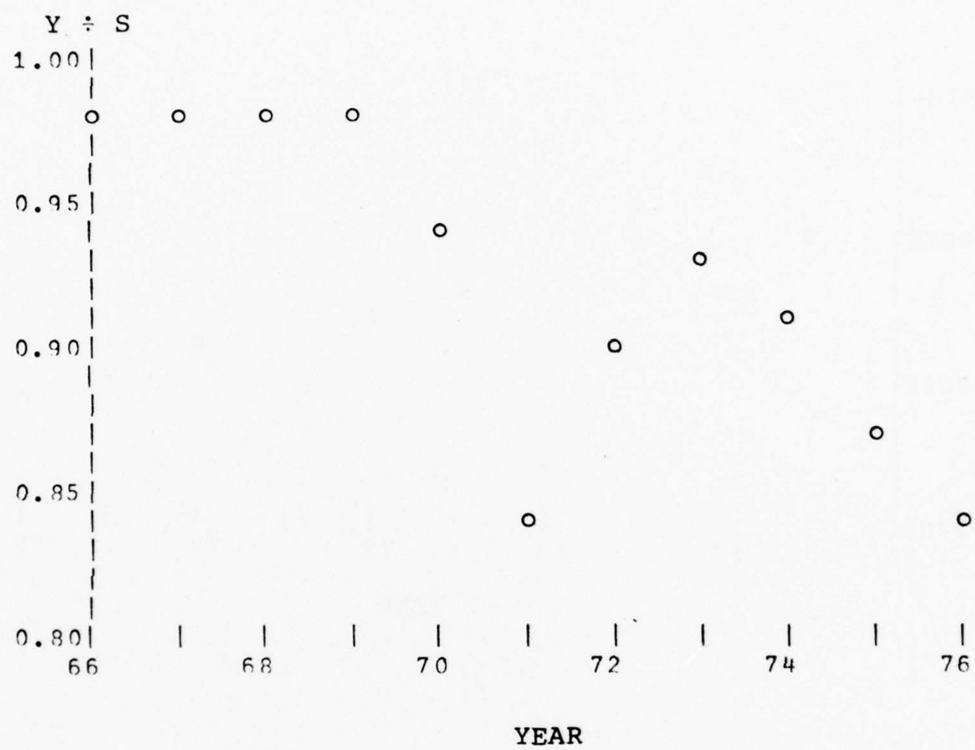


Figure 3.8 has special interest since the 'alpha method' forecasts eligibles as a proportion of separations. Cyclic behavior begins about 1970 (post war). The 1977 forecast is expected to continue the downtrend exhibited above in the more recent years. Indeed, based on this graph the forecast for 1976 is expected to be rather good. Time series smoothing methods have a tendency to do well when the trend continues, but thev are caught when the series either tops or bottoms out.

FIGURE 3.9

ELIGIBLES VS TIME

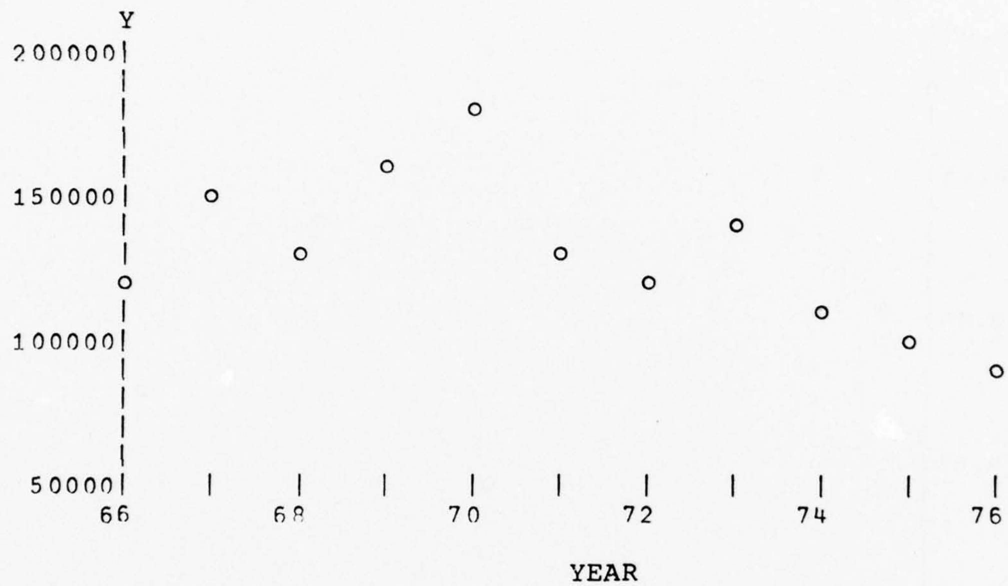


FIGURE 3.10

ELIGIBLES VS INVENTORY

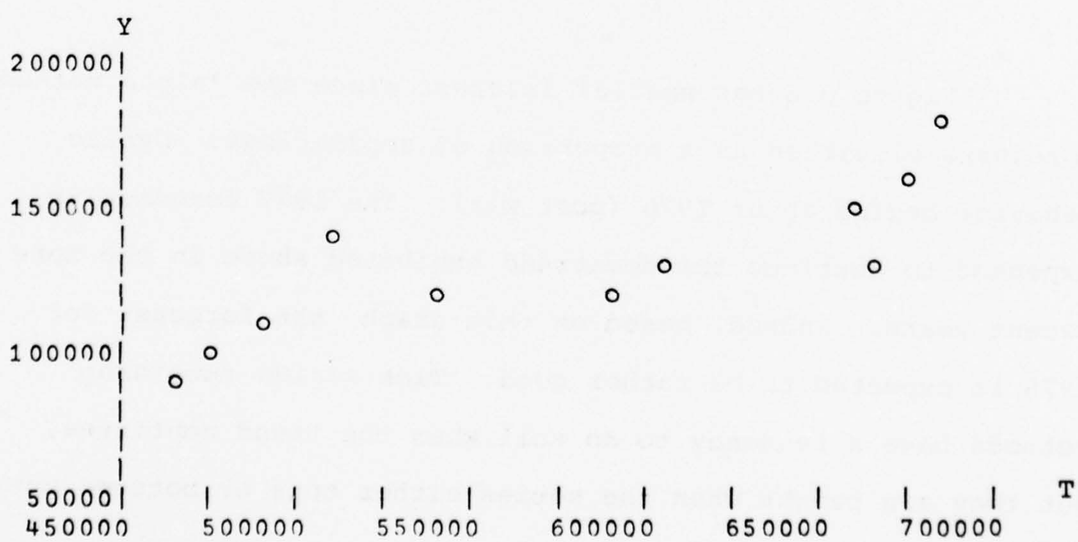


FIGURE 3.11

RETENTIONS VS ELIGIBLES

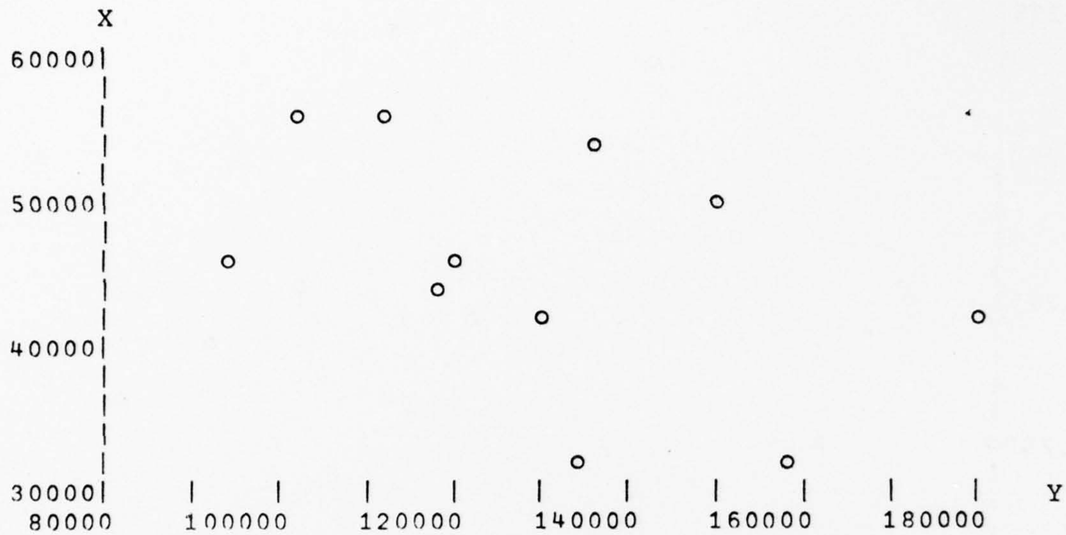


FIGURE 3.12

NON-REENLISTMENTS VS ELIGIBLES

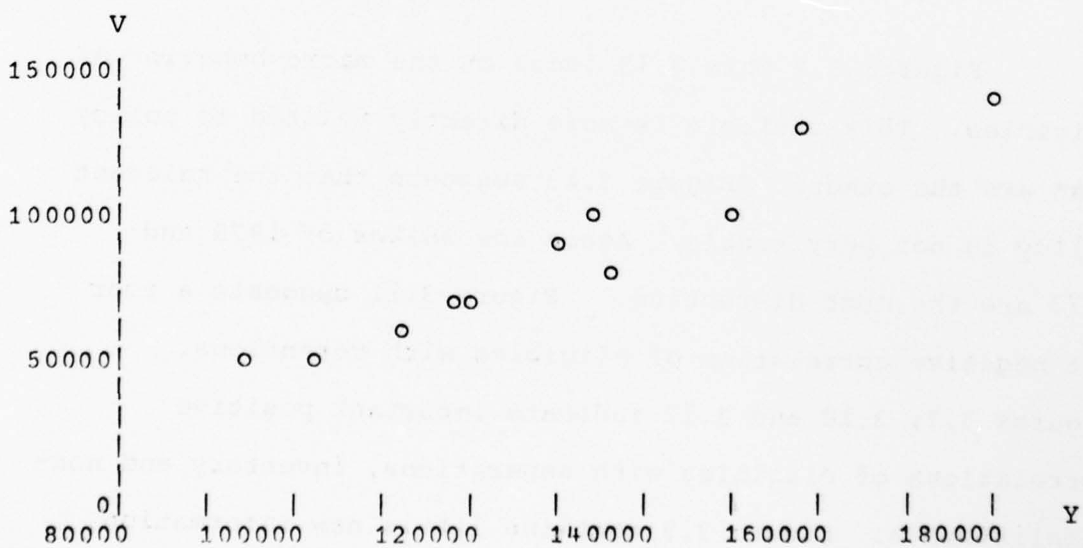
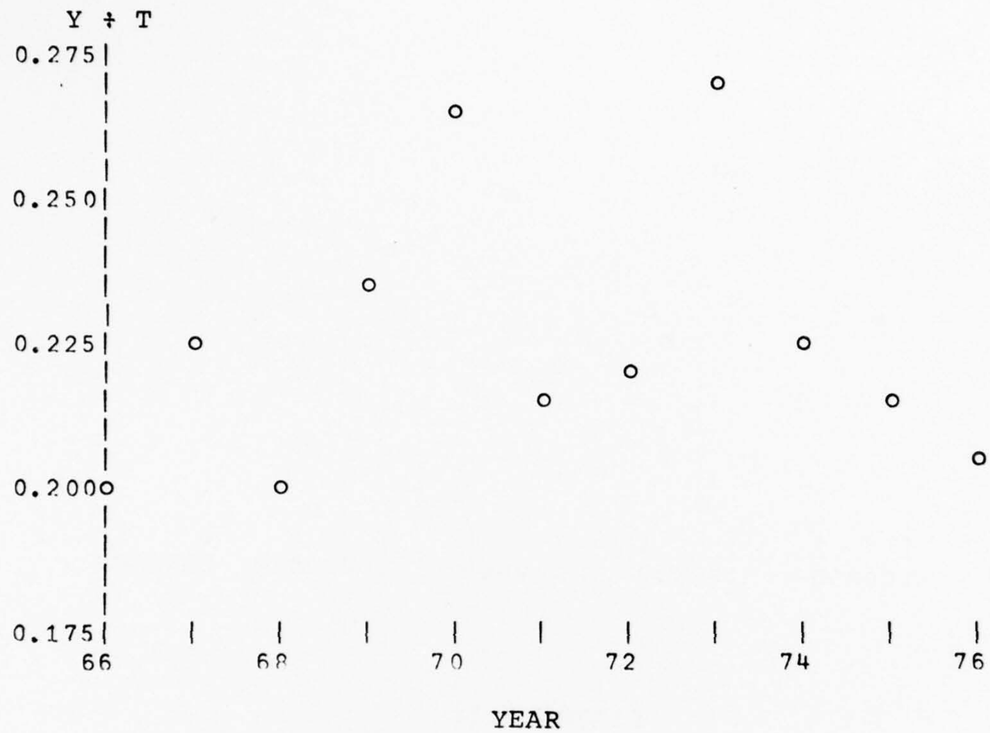


FIGURE 3.13

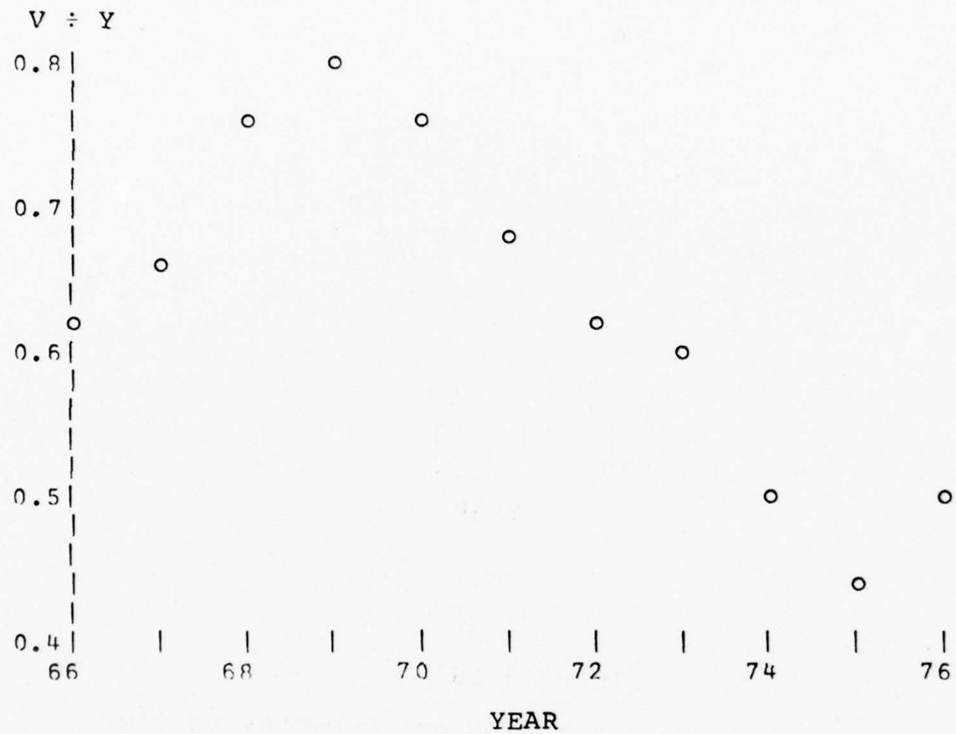
ELIGIBLES AS A FRACTION OF INVENTORY VS TIME



Figures 3.9 thru 3.13 focus on the macro behavior of eligibles. This variable is more directly related to policy than are the others. Figure 3.13 suggests that the relevant policy is not very stable. Again the spikes of 1970 and 1973 are the most disruptive. Figure 3.11 suggests a poor but negative correlation of eligibles with retentions. Figures 3.7, 3.10 and 3.12 indicate important positive correlations of eligibles with separations, inventory and non-reenlistments. Figure 3.9 contains little new information in the light of the others.

FIGURE 3.14

NON-REENLISTMENTS AS A FUNCTION OF ELIGIBLES VS TIME



The 'alpha method' forecasts non-reenlistments as a proportion of eligibles. This signal appears to have bottomed out in 1975 and the method is presumed to respond to this. Although we would expect the 1976 forecast (which is based on data thru 1975) to be poor because of the tendency to overswing, forecasts using data thru 1976 should not suffer as much.

FIGURE 3.15

NON-REENLISTMENTS VS TIME

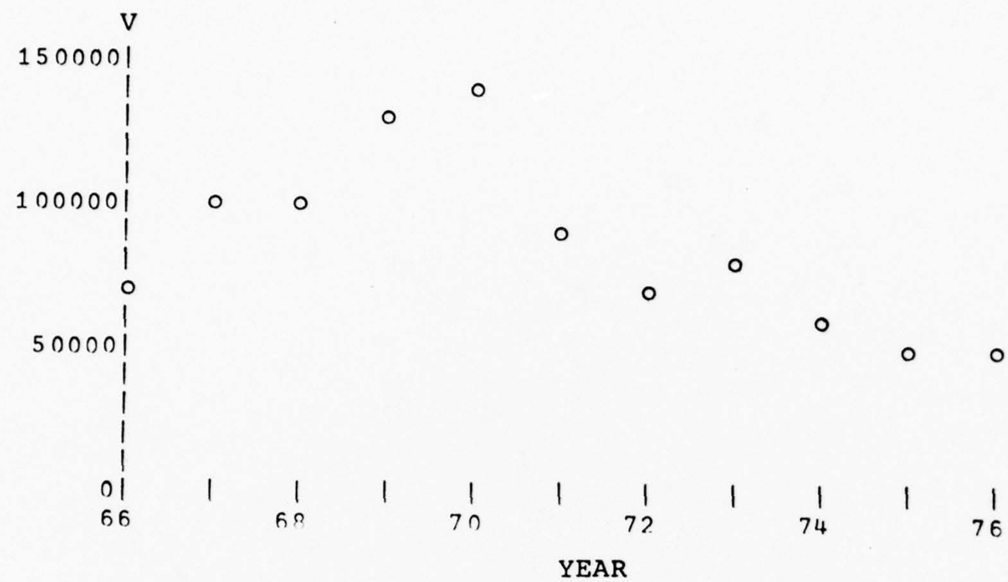


FIGURE 3.16

NON-REENLISTMENTS AS A FRACTION OF INVENTORY VS TIME

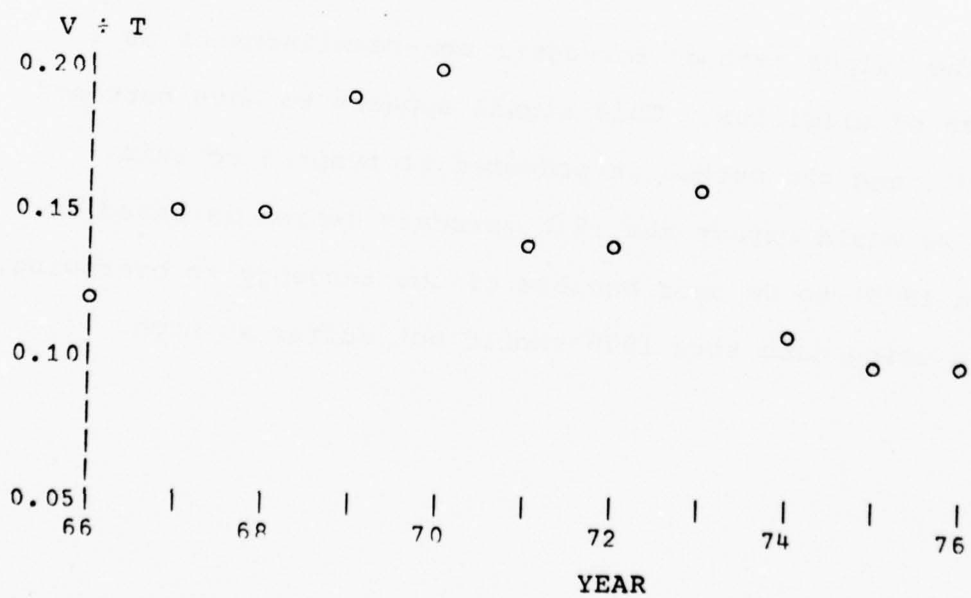


FIGURE 3.17

NON-REENLISTMENTS VS INVENTORY

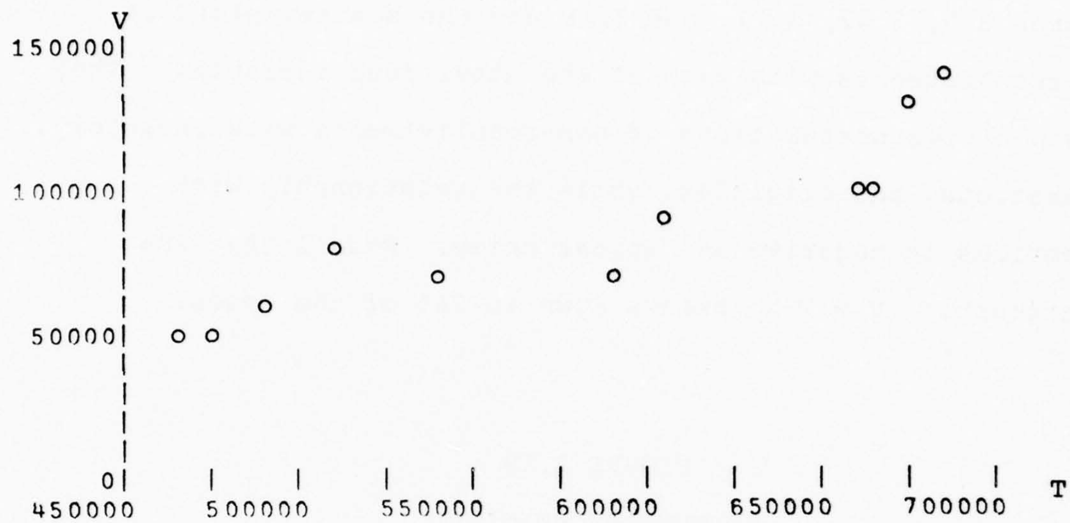
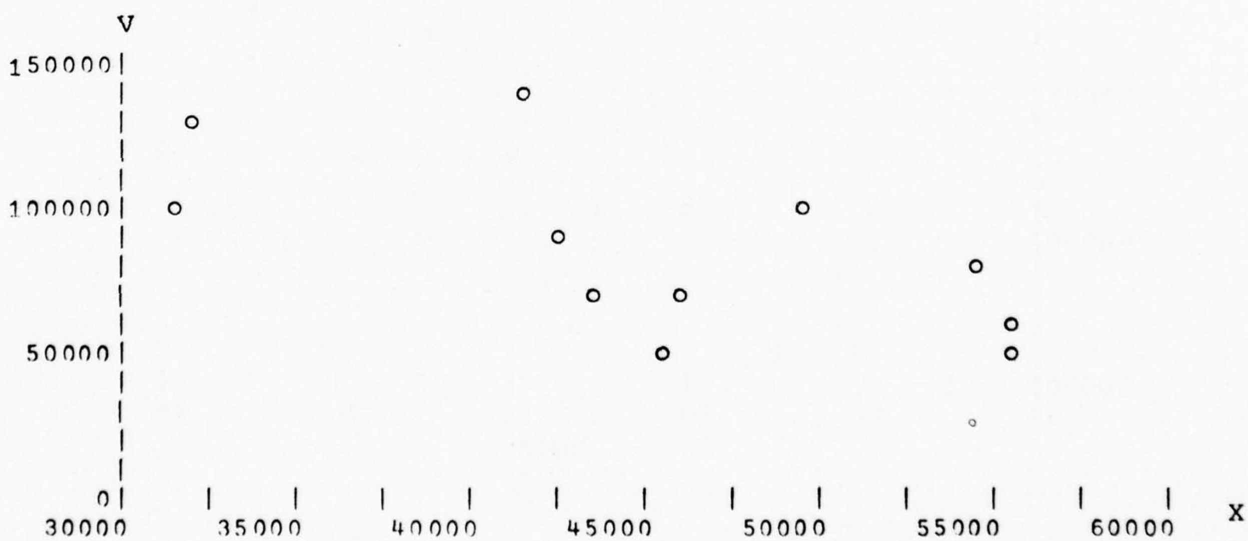


FIGURE 3.18

NON-REENLISTMENTS VS RETENTIONS



Figures 3.15 and 3.16 show macro behavior of non-reenlistments and non-reenlistment rate as a function of time. They are readily interpretable in the light of war and post-war years. Figures 3.5, 3.12, 3.17, and 3.18 are the scatter plots of non-reenlistments with each of the other four variables. They show positive correlations of non-reenlistments with inventory, separations, and eligibles, while the relationship with retentions is negative and appear noisy. Recall that the relationship V = Y-X breaks down in 26% of the cases.

FIGURE 3.19

RETENTIONS VS TIME

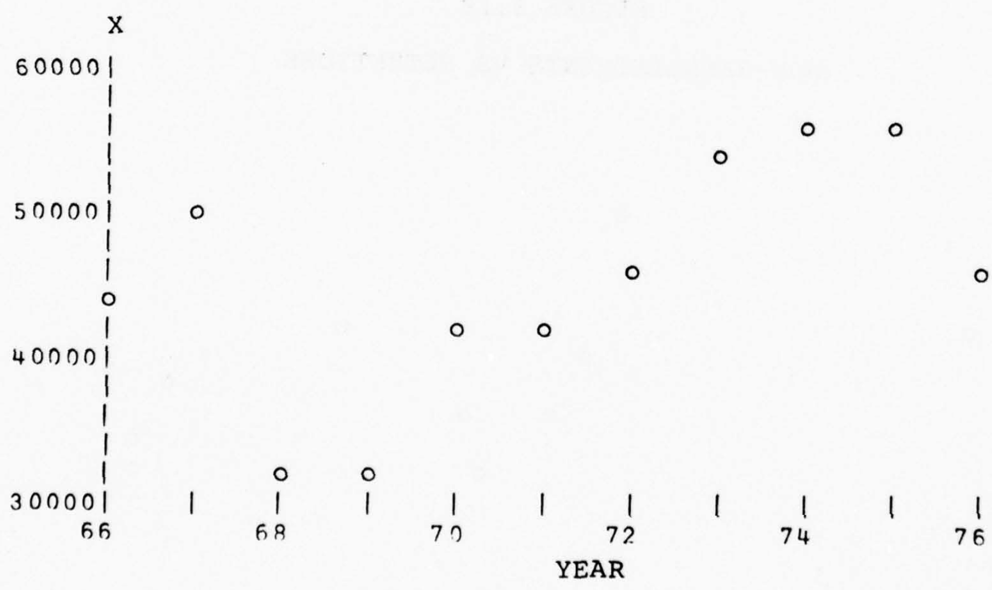


FIGURE 3.20

RETENTIONS AS A FRACTION OF INVENTORY VS TIME

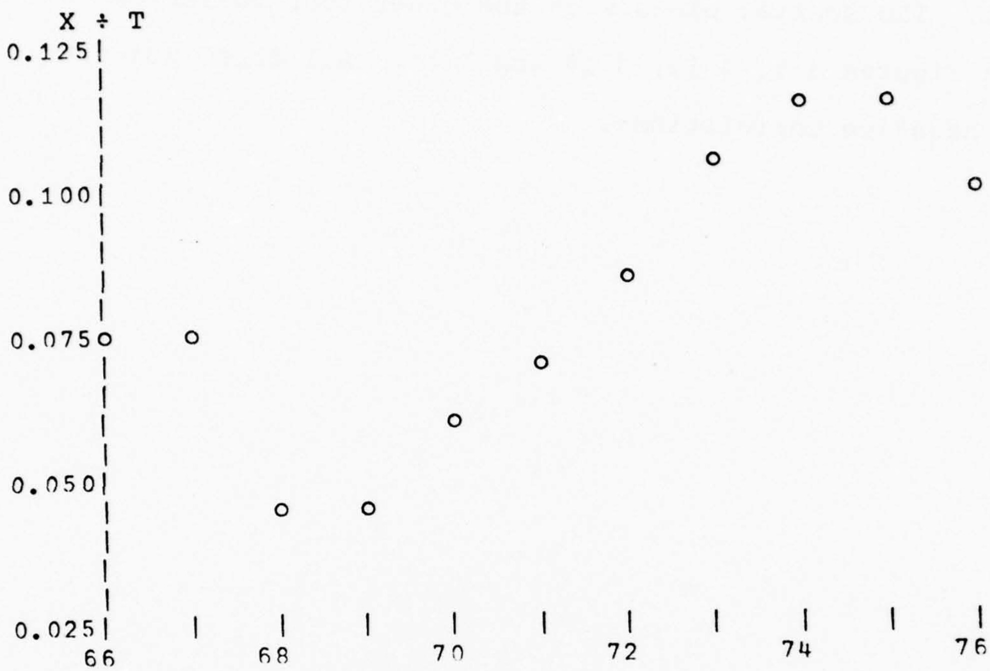
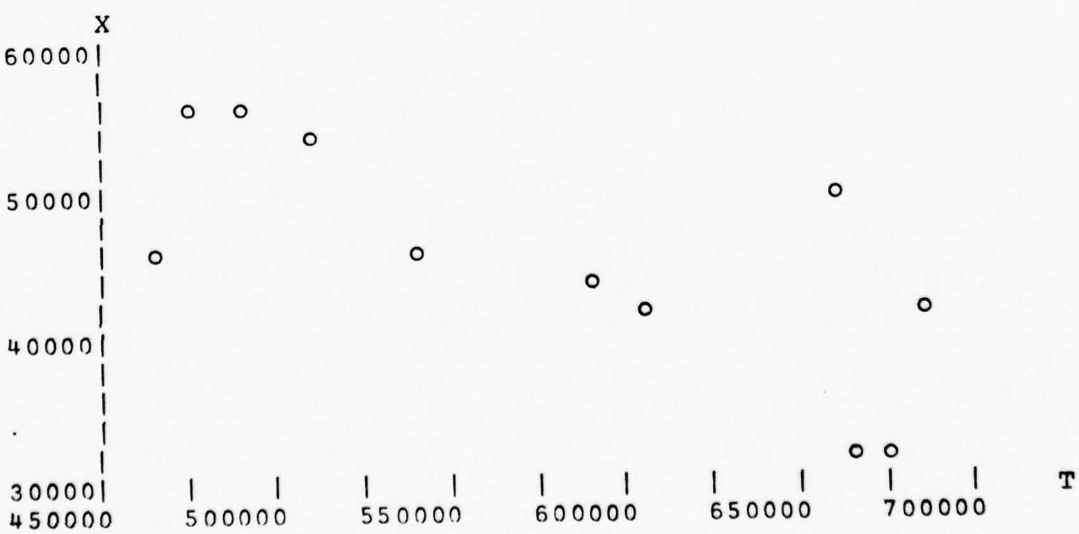


FIGURE 3.21

RETENTIONS VS INVENTORY



For completeness, the macro behavior of retentions is included. Figures 3.19 and 3.20 show somewhat smooth cyclical structure. The scatter plots with the other four variables appear in Figures 3.6, 3.11, 3.16 and 3.21. All four indicate modestly negative correlations.

IV. Supporting Details of the Comparison

The data bank for any of the object variables may be viewed as a time ordered set of eleven 31 by 7 matrices. As we move thru time, the numbers in these matrices will roll and twist as in a wave motion that reflects the changing size of the total enlisted force and how those changes affect the various cells. The set of inventory matrices form the base of the stocks of people. Although individual people are not tracked in this set, it is helpful to draw attention to the fact that an individual changes LOS cell each year and PG cell periodically. Thus in these recent years of drawdown (in size) one expects relatively higher exits from the lower LOS cells. Such motion is reflected also in the other four variables since inventory provides the base of stocks on which each of the others draw.

A little more detail concerning the three objective variables can be obtained by studying Tables 4.1 thru 4.3 which compare their changes over the most recent two years for the first 18 LOS cells and all 7 PG cells. The macro behavior of EAOS as shown in Figure 3.2 suggests that recently a smooth decline has taken place. Table 4.1 shows that these changes have been rather drastic in the first two columns and the lower LOS cells. Since the 'alpha' method smooths the time series cell by cell, it is expected to perform well in those cells that sustain a trend and poorly in those that do not.

TABLE 4.1

RECENT BEHAVIOR OF EAOS SEPARATIONS 1975

	E1-3	E4	E5	E6	E7	E8	E9
	2019	700	18	5	10	0	3
	7590	4940	323	30	2	1	0
	7796	7126	2205	25	3	0	3
LOS	5627	17820	12906	115	7	1	0
	551	1666	2038	135	8	1	0
	466	1270	3712	722	10	1	2
	135	209	1216	626	12	1	2
	67	160	1053	1049	18	2	0
	49	189	1094	1505	68	2	0
	37	259	1193	1580	125	1	1
	12	98	600	1379	249	2	0
	9	70	543	1555	408	17	2
	3	38	447	1921	624	36	1
	8	72	465	1833	717	74	5
	4	42	314	1694	908	167	20
	5	48	288	1505	1092	235	42
	0	13	132	1089	1147	322	69
	3	10	91	672	799	242	62

SEPARATIONS 1976

	E1-3	E4	E5	E6	E7	E8	E9
	624	70	18	7	6	0	0
	6017	2942	190	23	5	0	2
LOS	14352	12092	1140	21	8	0	2
	5906	15940	15649	54	2	0	1
	469	1471	1894	103	8	1	1
	221	1280	3379	823	7	1	1
	78	210	1118	620	14	1	0
	37	175	1172	1205	20	4	0
	19	129	750	1168	76	6	0
	35	192	1096	1671	129	2	0
	11	89	631	1093	180	1	1
	7	74	488	1132	264	11	0
	2	20	252	1079	385	17	1
	3	25	310	1219	471	39	5
	3	26	337	1180	615	55	5
	3	27	233	1010	638	94	19
	2	11	115	671	613	132	30
	2	4	60	484	490	124	29

TABLE 4.2

RECENT BEHAVIOR OF ELIGIBLES

Eligibles 1975

	E1-3	E4	E5	E6	E7	E8	E9
	894	685	16	5	10	0	3
	3386	4814	323	30	1	1	0
	2879	6941	2175	25	3	0	3
LOS	2265	17372	12786	112	7	1	0
	284	1623	2003	134	8	1	0
	370	1221	3671	716	10	1	2
	106	202	1203	622	11	1	2
	46	157	1039	1047	18	1	0
	35	178	1084	1497	68	2	0
	25	248	1172	1577	125	0	1
	5	96	594	1371	249	2	0
	4	67	538	1549	408	17	1
	1	37	438	1915	624	36	1
	7	72	461	1828	712	74	5
	3	40	311	1688	907	167	20
	5	45	288	1502	1091	235	42
	0	12	131	1085	1147	322	68
	3	9	91	672	799	242	62

Eligibles 1976

	E1-3	E4	E5	E6	E7	E8	E9
	217	51	15	7	6	0	0
	2739	2876	190	23	4	0	2
LOS	5124	11774	1123	21	7	0	2
	2390	15434	15516	53	2	0	1
	264	1425	1854	103	8	1	1
	131	1233	3338	817	7	1	1
	57	199	1107	618	14	1	0
	18	167	1157	1197	20	4	0
	11	126	739	1163	76	6	0
	21	182	1078	1660	129	2	0
	7	84	620	1090	180	1	1
	4	68	481	1121	263	11	0
	1	18	252	1076	383	17	0
	2	20	305	1216	469	39	5
	3	25	332	1177	611	55	5
	3	27	231	1007	638	94	19
	2	10	115	669	612	132	30
	1	4	60	482	483	124	29

The macro behavior of eligibles as shown in Figure 3.9 also declines smoothly but Table 4.2 shows some sharp drops for LOS = 1 and some sharp increases for LOS = 3. In the former case the advantage is to the alpha method and in the latter it is not. Regression methods must look elsewhere to pick up these signals.

According to Figure 3.15 Non-reenlistments level off in the macro sense. Table 4.3 shows some rather drastic movements for the lower LOS and PG cells. Again LOS = 3 shows some sharp increases.

The forecasting of the object variables using regression methods was performed repeatedly to meet several goals: First, the best set of p variables had to be identified. Second, the influence of the ridge constant (see Appendix A) needed some accounting. Third, the stability of the forecast where 1976 data was not included in the forecast required examination.

Table 4.4 contains a listing of the best set of p variables. The subscripts indicate time and the object variable appearing in the set with subscript t-1 indicates an auto-regressive contribution with a lag of one year. The correspond-MAE and RMSE values appear in Table 1.1.

The resulting forecasting of separations does not depend very much on whether or not 1976 data are included in the forecast coefficients (as measured by the MAE and RMSE values), but the inclusion does make a noticeable difference in fore-casting eligibles and non-reenlistments. This result was not anticipated from study of the macro plots. It may be, in part, a consequence of ill-conditioning (see Appendix A).

TABLE 4.3

RECENT BEHAVIOR OF NON-REENLISTMENTS

Non-Reenlistments 1975

	E1-3	E4	E5	E6	E7	E8	E9
	772	550	6	2	6	0	3
	2976	3125	48	7	-2	0	0
	2401	4103	764	7	3	0	1
	1872	13070	6869	18	4	1	0
LOS	171	877	863	51	2	1	0
	91	736	2352	462	5	0	1
	30	74	536	261	2	0	2
	12	53	277	305	-2	-1	0
	13	37	279	302	3	0	-1
	9	51	261	256	11	-1	0
	0	12	73	156	13	-1	0
	1	6	75	138	13	1	0
	0	9	20	98	12	-6	0
	1	6	24	60	-2	4	-2
	3	2	16	22	-2	-2	-1
	0	1	8	23	1	-1	-2
	0	0	6	6	-10	-5	-2
	-1	0	2	1	-19	-4	0

Non-Reenlistments 1976

	E1-3	E4	E5	E6	E7	E8	E9
	133	14	6	4	5	0	0
	2359	1857	26	10	1	-1	2
	4137	8153	379	2	6	0	1
LOS	1862	11099	8967	18	1	0	0
	152	650	764	47	4	1	1
	52	667	2265	553	4	0	0
	25	72	508	284	3	1	0
	-1	41	364	386	-3	2	0
	4	20	172	242	0	2	0
	10	31	221	287	6	0	-1
	2	19	78	140	13	-1	-1
	1	2	49	94	17	2	0
	0	2	23	47	12	-1	-1
	1	7	8	26	2	-1	1
	1	3	5	29	0	-3	-2
	1	2	6	-1	-6	0	-3
	1	0	1	-3	-9	-1	-1
	0	1	0	-1	-6	-2	-1

TABLE 4.4

VARIABLE SELECTION FOR REGRESSION FORECASTS

Object Variable	p = 2	p = 3	p = 4
Separation S_t	T_t, S_{t-1}	T_t, S_{t-1}, Y_{t-1}	$T_t, S_{t-1}, Y_{t-1}, X_{t-1}$
Eligibles Y_t	T_t, Y_{t-1}	T_t, Y_{t-1}, X_{t-1}	$T_t, Y_{t-1}, X_{t-1}, V_{t-1}$
Non-Reenlistments V_t	T_t, V_{t-1}	T_t, V_{t-1}, Y_{t-1}	$T_t, V_{t-1}, Y_{t-1}, X_{t-1}$

The use of a ridge constant in multiple regression serves to stablilize the regression coefficients when the regression variables are highly correlated (i.e. the problem is ill-conditioned). The picture is cloudy, but generally its use is noticeable in forecasting separations and eligibles but not so much in forecasting non-reenlistments. A high level of ill-conditioning is anticipated whenever V, Y, and X are used in concert because of the logical relationship $V = Y-X$. Also a non-zero ridge constant may be appropriate when $p = 2$ for eligibles and non-reenlistments. The value $RC = .025$ in the space of standardized regression variables, seems to be a reasonable choice for the ridge constant, but overall, the ill-conditioning and stability merits further study.

The remainder of this section is devoted to the numerical comparison of 'alpha' forecasts with regression forecasts for $p = 4$ in the case of separations, and for $p = 3$ for the other two variables. In the latter cases the contribution of the fourth variable is suspect because of the relationship V = Y-X and the improvement indicated in Table 1.1 may be unstable.

The computation of multiple regressions individually for each of the 217 (31 × 7) cells is too time consuming and some grouping was necessary. The grouping chosen was based on the following partitioning of the eleven year averages for each of the object variables. Cells were treated individually as long as the eleven year averages were at least 1000 (1900 in the case of eligibles). Then cells were grouped together in decrements of 100 down to an eleven year average of 100. From then on the groupings were in decrements of 10 with some arbitrary adjustments when zero was approached.

Table 4.5 compares the (rounded) 'alpha' forecasts and regression forecasts (on T_t, S_{t-1}, Y_{t-1}, X_{t-1}) for separations. The ridge constant was .025 and all data were used in developing the regression coefficients. The (rounded) errors of the two forecasts are compared in Table 4.6.

Let us consider Table 4.6 in the light of the recent changes in separation shown in Table 4.1. For LOS = 1 separations experienced a severe drop in 1976 in the first two columns. The "alpha" method picked this up amazingly well. Of course this success is attributed to the use of 1976 data in the development of the coefficients. The same effect is exhibited in LOS = 2 for E4. The first two entries for LOS = 3 in Table 4.1 show a sharp increase. This is a trend reversal that the alpha method does not respond to so rapidly even under these advantageous conditions. It makes its worst showing in these two cells.

TABLE 4.5

FORECASTS OF 1976 SEPARATIONS

	Alpha Forecast								Regression (p = 4) Forecast						
				PG								PG			
	627	137	16	4	3	1	1		1728	1612	53	7	8	2	2
	7188	4661	289	19	9	1	1		5170	6135	464	29	5	2	1
	9614	8116	1922	18	4	1	2		9817	11263	1481	31	6	2	3
	5611	17188	13272	101	5	1	1	LOS	6471	16301	14137	97	13	3	2
	542	1644	1913	134	8	2	1		449	1924	1435	123	19	3	2
LOS	464	1413	3593	683	10	2	3		373	1107	3057	647	20	3	3
	135	212	1176	624	12	2	1		126	271	1391	822	22	3	3
	67	192	1047	932	20	3	1		91	206	1333	1015	37	5	3
	33	193	1048	1322	61	3	2		53	201	1005	1129	81	4	3
	38	266	1251	1485	112	3	2		37	226	1027	1510	155	5	3
	13	98	614	1327	249	3	1		14	91	536	1258	213	5	3
	10	67	498	1539	406	10	2		11	76	471	1368	309	16	3
	6	35	443	1909	582	33	2		6	37	373	1517	429	42	5
	9	72	457	1764	604	61	7		8	40	385	1629	552	86	6
	4	40	310	1598	773	133	16		6	40	312	1653	953	139	16
	4	49	286	1407	971	209	38		6	38	243	1422	1080	184	35
	2	14	132	1087	1004	269	51		4	22	142	970	1078	249	50
	5	7	86	626	656	188	46		3	9	117	649	879	218	70
	1	8	68	436	760	243	85		3	8	68	412	652	245	81
	1	1	36	309	715	260	93		3	6	39	347	602	233	86
	1	6	8	94	243	101	51		1	4	13	125	283	145	60
	1	2	11	40	155	103	46		1	3	8	48	191	71	51
	1	1	13	22	138	107	100		1	3	6	34	140	78	67
	1	1	6	13	63	42	42		1	2	5	22	100	67	66
	1	1	7	12	23	14	17		1	2	3	18	66	35	38
	1	1	3	21	44	29	41		1	2	4	8	46	23	34
	1	1	2	9	41	52	69		1	1	2	7	39	39	46
	1	1	1	4	21	18	20		1	1	2	5	28	29	36
	1	1	1	4	15	5	9		1	1	1	3	13	9	13
	1	1	3	6	5	8	13		1	1	1	3	10	10	15
	1	1	2	8	7	8	15		1	1	2	2	9	11	12

TABLE 4.6

ERRORS OF FORECAST: SEPARATIONS

Alpha Forecast

	PG						
	3	67	-2	-3	-3	1	1
	1171	1719	99	-4	4	1	-1
	-4738	-3976	782	-3	-4	1	0
	-295	1248	-2377	47	3	1	0
	73	173	19	31	0	1	0
LOS	243	133	214	-140	3	1	2
	57	2	58	4	-2	1	1
	30	17	-125	-273	0	-1	1
	14	64	298	154	-15	-3	2
	3	74	155	-186	-17	1	2
	2	9	-17	234	69	2	0
	3	-7	10	407	142	-1	2
	4	15	191	830	197	16	1
	6	47	147	545	133	22	2
	1	14	-27	418	158	78	11
	1	22	53	397	333	115	19
	0	3	17	416	391	137	21
	3	3	26	142	166	64	17
	1	6	38	141	276	106	43
	1	1	-2	13	195	123	31
	1	3	3	3	-47	-17	-4
	1	0	6	4	39	50	21
	1	0	7	4	47	55	64
	1	0	2	5	17	10	6
	1	1	6	5	-8	-7	-8
	1	0	2	16	28	25	28
	1	1	2	4	15	28	35
	1	1	1	2	1	1	3
	1	1	1	3	5	4	4
	1	1	2	3	0	3	5
	1	1	2	5	4	4	3

Regression (p = 4) Forecast

	PG						
	1104	1542	35	0	2	2	2
	-847	3193	274	6	0	2	-1
	-4535	-829	341	10	-2	2	1
	565	361	-1512	43	11	3	1
LOS	-20	453	-459	20	11	2	1
	152	-173	-322	-176	13	2	2
	48	61	273	202	8	2	3
	54	31	161	-190	17	1	3
	34	72	255	-39	5	-2	3
	2	34	-69	-161	26	3	3
	3	2	-95	165	33	4	2
	4	2	-17	236	45	5	3
	4	17	121	438	44	25	4
	5	15	75	410	81	47	1
	3	14	-25	473	338	84	11
	3	11	10	412	442	90	16
	2	11	27	299	465	117	20
	1	5	57	165	389	94	41
	3	6	38	117	168	108	39
	3	6	1	51	82	96	24
	1	1	8	34	-7	27	5
	1	1	3	12	75	18	26
	1	2	0	16	49	26	31
	1	1	1	14	54	35	30
	1	2	2	11	35	14	13
	1	1	3	3	30	19	21
	1	1	2	2	13	15	12
	1	1	2	3	8	12	19
	1	1	1	2	3	8	8
	1	1	0	0	5	5	7
	1	1	2	-1	6	7	0

Tables 4.7 and 4.8 compare the forecasts and errors of forecasts of eligibles for the 'alpha' and regression (on T_t, Y_{t-1}, X_{t-1}) methods. Again when we consider the major changes in 1976 by viewing Table 4.2 we see a sharp drop in columns 1 and 2 for LOS = 1, a substantial drop in these columns for LOS = 2 and an increase for LOS = 3. Again the alpha method picks this up extremely well for LOS = 1, adequately well for LOS = 2, and responds poorly to the turnaround for LOS = 3. The regression method does surprisingly well for LOS = 1 and column 1.

This same phenomenon continues when comparing the forecasts of non-reenlistments. Table 4.3 shows a sharp drop in LOS = 1 that the alpha method picks up well, a drop in cell(2,2) is not so well forecast. The increases for LOS = 3 and cell (4,3) are again handled much better by the regression methods.

TABLE 4.7

FORECASTS OF 1976 ELIGIBLES

Alpha Forecast

PG

	234	108	13	4	3	1	1
	3423	4544	289	19	7	1	1
	3507	7907	1894	18	4	1	2
LOS	2255	16684	13151	99	5	1	1
	285	1596	1874	133	8	2	1
	343	1367	3552	677	10	2	3
	106	202	1164	621	12	2	1
	38	184	1033	927	20	3	1
	23	182	1034	1315	61	3	1
	25	254	1230	1476	112	3	1
	7	93	605	1320	249	3	1
	6	62	491	1527	405	10	1
	3	32	437	1904	580	33	1
	6	62	451	1760	601	61	7
	4	38	306	1594	769	133	16
	4	47	283	1404	970	209	38
	1	13	131	1084	1003	269	51
	4	7	86	625	655	188	46
	1	8	66	436	759	243	85
	1	1	34	307	713	260	93
	1	6	8	94	243	100	51
	1	2	10	40	154	103	46
	1	1	12	22	138	107	100
	1	1	5	13	63	42	42
	1	1	7	12	23	13	17
	1	1	3	21	44	29	41
	1	1	1	9	41	52	69
	1	1	1	4	21	18	20
	1	1	1	4	15	5	9
	1	1	3	6	5	8	12
	1	1	2	8	6	8	15

Regression (p = 3) Forecasts

PG

	243	1595	54	6	7	1	1
	3908	5988	408	27	3	1	0
	5272	11172	1409	30	5	1	2
	1898	15752	14114	98	12	2	1
	379	1294	1358	122	16	2	1
	292	1197	2990	590	17	1	2
	99	286	1431	826	20	2	2
	70	202	1269	1246	40	4	2
	35	195	904	1095	83	3	2
LOS	27	220	1085	1567	151	3	1
	9	89	546	1258	209	4	1
	7	72	461	1349	306	14	1
	4	35	364	1452	357	29	4
	6	37	375	1577	615	88	5
	5	39	314	1635	903	140	15
	2	35	239	1463	1046	184	33
	3	20	142	1027	1068	245	54
	1	7	117	641	772	217	56
	2	6	63	410	613	243	84
	2	5	36	341	666	232	92
	0	3	12	121	280	145	65
	0	2	7	50	190	76	51
	0	2	5	34	140	66	69
	0	1	4	18	101	69	52
	0	1	2	16	37	31	39
	0	1	2	6	44	43	28
	0	0	1	5	47	34	39
	0	0	1	4	25	27	34
	0	0	0	2	12	8	13
	0	0	0	2	8	9	13
	0	0	0	1	7	9	10

TABLE 4.8

ERRORS OF FORECASTING 1976 ELIGIBLES

	Alpha Forecast								Regression (p = 3) Forecast						
	PG								PG						
	17	57	-2	-3	-3	1	1		26	1544	39	-1	1	1	1
	684	1668	99	-4	3	1	-1		1169	3112	218	4	-1	1	-2
	-1617	-3867	771	-3	-3	1	0		148	-602	286	9	-2	1	0
	-135	1250	-2365	46	3	1	0		-492	318	-1402	45	10	2	0
	21	171	20	30	0	1	0		115	-131	-496	19	8	1	0
LOS	212	134	214	-140	3	1	2	LOS	161	-36	-348	-227	10	0	1
	49	3	57	3	-2	1	1		42	87	324	208	6	1	2
	20	17	-124	-270	0	-1	1		52	35	112	49	20	0	2
	12	56	295	152	-15	-3	1		24	69	165	-68	7	-3	2
	4	72	152	-184	-17	1	1		6	38	7	-93	22	1	1
	0	9	-15	230	69	2	0		2	5	-74	168	29	3	0
	2	-6	10	406	142	-1	1		3	4	-20	228	43	3	1
	2	14	185	828	197	16	1		3	17	112	376	-26	12	4
	4	42	146	544	132	22	2		4	17	70	361	146	49	0
	1	13	-26	417	158	78	11		2	14	-18	458	292	85	10
	1	20	52	397	332	115	19		-1	8	8	456	408	90	14
	-1	3	16	415	391	137	21		1	10	27	358	456	113	24
	3	3	26	143	166	64	17		0	3	57	159	283	93	27
	1	6	37	141	275	106	43		2	4	34	115	129	106	42
	1	1	-4	13	193	123	31		2	5	-2	47	146	95	30
	1	3	3	3	-47	-18	-4		0	0	7	30	-10	27	10
	1	0	5	4	39	50	21		0	0	2	14	75	23	26
	1	0	7	4	47	55	64		0	1	0	16	49	14	33
	1	0	1	5	17	10	6		0	0	0	10	55	37	16
	1	1	6	5	-8	-7	-8		0	1	1	9	6	11	14
	1	0	2	16	28	25	28		0	0	1	1	28	39	15
	1	1	1	4	15	28	35		0	0	1	0	21	10	5
	1	1	1	2	1	1	3		0	0	1	2	5	10	17
	1	1	1	3	5	4	4		0	0	0	1	2	7	8
	1	1	2	3	0	3	5		0	0	-1	-1	3	4	6
	1	1	2	5	3	4	3		0	0	0	-2	4	5	-2

TABLE 4.9

FORECASTS OF NON-REENLISTMENTS 1976

	Alpha Method		PG						Regression (p = 3) Method			PG			
	161	46	6	3	2	1	1		783	1714	49	9	6	0	0
	3000	3059	39	10	1	1	1		3489	4251	91	18	4	0	0
	2802	4752	657	4	3	1	1		4831	8290	69	19	9	1	0
	1800	12335	772[illegible]	19	3	1	1		1465	11508	8364	21	18	2	0
	158	811	843	53	5	2	1		261	148	334	66	17	2	1
LOS	101	813	2397	465	6	1	1		90	559	2647	361	18	1	2
	36	74	530	282	3	1	1	LOS	38	100	747	338	16	3	2
	7	50	324	293	2	2	1		23	84	468	371	15	5	1
	9	35	263	276	2	1	1		13	35	288	281	17	3	1
	11	48	270	254	8	1	1		19	42	217	240	16	2	1
	1	19	82	168	14	1	1		3	20	101	125	18	2	1
	2	5	69	136	26	2	1		5	10	86	107	14	5	1
	1	5	39	91	19	1	1		2	7	25	62	12	1	1
	2	14	19	48	7	3	1		2	9	27	75	13	5	1
	2	3	14	41	4	1	1		2	5	27	56	19	2	1
	1	5	9	12	4	2	1		1	5	18	24	22	3	1
	1	1	6	4	3	1	1		1	2	11	17	9	1	1
	1	2	2	4	2	1	1		0	2	6	16	13	3	2
	1	2	4	5	3	1	1		0	1	4	7	11	3	1
	1	1	7	6	11	2	2		0	1	5	11	12	4	2
	1	2	4	5	8	1	1		0	1	2	7	13	3	2
	1	1	4	3	6	4	1		0	0	2	3	9	2	1
	1	1	5	2	1	1	1		0	0	1	3	8	0	0
	1	1	1	1	1	1	1		0	0	2	3	9	2	2
	1	1	1	3	2	1	1		0	0	0	2	9	2	2
	1	1	1	5	3	2	1		0	0	0	2	5	1	1
	1	1	1	1	1	1	2		0	0	0	2	4	1	0
	1	1	1	2	2	1	1		0	0	0	0	5	1	1
	1	1	1	1	1	1	1		0	0	0	0	3	1	0
	1	1	1	1	2	1	1		0	0	0	0	2	0	0
	1	1	1	1	1	1	1		0	0	0	0	5	2	0

TABLE 4.10

ERRORS OF FORECASTS OF NON-REENLISTMENTS 1976

Alpha Forecast

PG

	27	32	¯1	¯2	¯3	0	0
	640	1202	12	¯1	0	2	¯1
	¯1336	¯3401	277	2	¯3	1	0
	¯63	1235	¯1242	0	1	1	0
	5	161	79	5	0	0	0
	48	145	132	¯88	1	1	1
LOS	11	2	21	¯3	¯1	0	0
	8	8	¯41	¯94	4	¯1	0
	5	15	90	34	2	¯1	1
	0	16	49	¯34	2	1	2
	¯1	¯1	3	28	1	2	2
	1	3	19	42	8	0	1
	1	2	15	44	6	2	2
	0	7	11	22	5	4	0
	0	0	8	11	4	4	3
	0	2	2	12	10	2	4
	0	1	5	7	11	2	2
	1	1	2	4	7	3	2
	1	2	3	1	4	2	1
	1	1	0	5	5	4	2
	1	1	1	2	¯2	3	0
	0	0	3	0	2	2	1
	1	0	2	1	1	1	2
	0	¯1	1	1	1	2	3
	0	0	1	3	0	3	2
	0	0	1	3	2	1	1
	0	0	1	1	2	1	1
	0	0	1	0	1	1	1
	0	0	1	1	¯1	1	0
	0	0	0	0	0	1	1
	0	0	1	1	0	1	1

Regression (p = 3) Forecast

PG

	650	1700	43	5	1	0	0
	1130	2394	65	8	3	1	¯2
	694	137	¯310	17	3	1	¯1
	¯397	409	¯603	3	17	2	0
	109	¯502	¯430	19	13	1	0
LOS	38	¯108	382	¯192	14	1	2
	13	28	239	54	13	2	2
	24	43	104	¯15	18	3	1
	9	15	116	39	17	1	1
	9	11	¯4	¯47	10	2	2
	1	1	23	¯15	5	3	2
	4	8	37	13	¯3	3	1
	2	5	2	15	0	2	2
	1	2	19	49	11	6	0
	1	2	22	27	19	5	3
	0	3	12	25	28	3	4
	0	2	10	20	18	2	2
	0	1	6	17	19	5	3
	0	1	3	4	13	4	1
	0	1	¯2	11	7	6	2
	0	0	¯1	5	3	5	1
	0	¯1	1	0	5	1	1
	0	¯1	¯1	3	8	0	1
	0	¯1	2	3	9	3	4
	0	0	0	3	8	4	3
	0	¯1	0	1	4	1	1
	0	0	0	2	5	1	0
	0	0	0	¯1	5	1	¯1
	0	0	0	0	1	1	¯1
	0	0	0	¯1	1	0	0
	0	0	0	0	4	2	0

V Conclusions and Recommendations

A fair comparison of the two methods could not be made on the one hand because 1977 actuals are not yet available and, on the other hand, because the alpha matrices for 1975 were not available. The alpha method appears to be a time series type method that possesses the typical lag characteristics at peaks and troughs. The regression methods appear to be competitive and have the feature of being more stable.

Some examination of the residuals (over time) of the regression methods took place. Although some firm lag correlations are present, they do not appear to be significant, based on the computation of Durbin-Watson statistics. Thus a hybrid system using both time series and regression methods is not expected to produce highly substantial improvements.

It is recommended that a fair comparison be made and that an appropriate measure of the cost of forecast error be developed for it. The real extent of time series "overswing" would become apparent and its effect would be examined in more realistic terms.

It is desirable to develop the regression approach further. The question of the correct set of ridge constants needs be faced more carefully as well as the question of the grouping of cells. The residuals should be examined more carefully with one eye focused on the time dependence behavior and a second eye looking for suitable transformations to remove skewness and stabilize the variance.

APPENDIX A

FORECASTING MANPOWER CHANGES USING REGRESSION METHODS

Data and Notation:

Ten years data are available, specifically 1966-1976 (fiscal year begins on 1 July of previous year). The first subscript, t, will index the years $(t = 1,\ldots,10)$ with $t = 1$ referring to fiscal 1966 (1 July 1965 to 30 June 1966), etc. The second and third subscripts refer to length of service (LOS) and pay grade (PG), categories respectively, using the indices i and j. Basically then the 31 LOS categories are interpreted as follows: An enlisted man with LOS = i is one who on 1 July of that period had completed at least $i - 1$ years of service but less than i, for $i = 1,\ldots,30$. If $i = 31$ then the number of years completed service is "at least 30." There are nine PG categories referring to the pay grades E1,...,E9.

The resulting data arrays were too cumbersome for our exploratory work and some arbitrary grouping was imposed. Specifically, LOS categories 6 through 17 were grouped together as were categories 18 through 31. Also PG categories E1, E2, E3 were aggregated. Thus the present study treats seven LOS groups and seven PG groups. Notice that the detail lost in aggregating the high LOS cells is partially recovered because the corresponding (high) PG cells are intact. Similarly, the intactness of the low LOS cells retains some of the information lost by aggregating the low PG cells.

The following notation was adopted for the given data.

G gains

A attritions

R retirements

X retentions (re-enlistments)

Y eligibles

S separations (EAOS) $(S > Y > X)$

T inventory (total Navy enlisted)

Some additional derived quantities are useful.

$U = S - Y$ ineligibles

$V = Y - X$ non-reenlistments

$W = U + V = S - X$ contract losses

Some explanation of these quantities is helpful. Gains refer to the number of people from outside the Navy that enter the enlisted Navy during the fiscal year in question, in each LOS, PG category used. New recruits are not included since they need not be forecast. Promotions represent internal movement and are also not reflected in the gains used here. Changes which are included in gains are those persons who reenter the service after having left, under programs called continuous service or broken service reenlistment contracts. A category called miscellaneous gains is also included here, representing gains by various methods, not including the recruits.

Attritions refer mainly to people who are dismissed prior to the expiration of their contract. It also includes deaths, disability discharges, etc. Retirements begin in LOS category 18. The only means of leaving the Navy aside from attrition and retirement is by failure to reenlist at the expiration of the contract, i.e., contract loss. All personnel are separated at the end of their contract. Not all separated personnel are declared eligible for reenlistment and not all eligibles exercise their option to reenlist. Hence the inequality $S > Y > X$. The derived differences ($S - Y$ and $Y - X$) are called ineligibles and non reenlistments, respectively. The sum of these two are the contract losses.

To all the variables may be affixed the subscripts t, i, j which refer to periods and categories already described. All of the variables (except T) are interval functions and refer to the net result of a time period (fiscal year). The variable T referring to the total size of the Navy is a point or "snapshot" variable and refers to the number of personnel on board on the first day of the designated period (left end point or, more specifically, 1 July of the fiscal year).

Scope of Current Study

Regression methods, specifically ridge regression, are applied to the forecasting of contract losses, gains, and attrition by total numbers in each LOS group and in each PG group. All data sets of variables are three dimensional arrays. Since the present work is concerned with exploring the usefulness of a methodology, the dimensions of the data sets were reduced to two in order to obtain simplicity and uniformity. This was done in two different ways since LOS groups and PG groups have separate interest. Thus when studying the predictability of the LOS groups, all data arrays were summed over $j = 1,\ldots,7$ yielding time by LOS group matrices. Similarly when studying the predictability of the PG groups, all data arrays were summed over $i = 1,\ldots,7$ yielding time by PG group matrices. No combining or mixing of the two kinds of groups took place in this study. Thus we attempted to predict the various changes by LOS and by PG marginally, but not jointly.

The choice of regression variables was made on hueristic grounds. The volume of data is quite limited (ten time values) and this limits one to the use of only four or five regression variables so that at least a few degrees of freedom remain for estimating the mean square error. Also any autoregressive feature was limited to the single most recent time period (or time point, in the case of the variable T). Thus the ten time periods still yield nine full sets of observations. The degrees of freedom for estimating error are given by $n - p - 1$ where $n = 9$ and $p =$ the number of regression variables. The choice of variables (same for each of the two kinds of groups LOS and PG) appear below.

Contract Losses regressed on Ineligibles, Reenlistments, Separations, Total Inventory.

Gains regressed on Gains, Attritions, Ineligibles, Separations, Total Inventory.

Attritions regressed on Attritions, Ineligibles, Reenlistments, Total Inventory.

Further exploration could yield a better set of regression variables. The results so far are encouraging as will be seen.

Methodology

The development of ridge regression in the last five years (see Ref. 3, 4, 7) is proving to be an important step in treating the anomalies of regression problems. It's use is especially attractive where the correlation matrix of the regression variable is highly non orthogonal--a condition that is met liberally in the present problem. The key to its successful application is in the selection of the ridge constant k. Current thinking on this question recognizes several competing forces and suggests the selection of a range of values for k in which all the forces are rather stable.

More specifically, these requirements may be summarized as follows:

(i) The variance inflation factor of the estimates of the regression coefficients should be at least one but certainly not as large as ten, (Ref. 6, p. 609ff.; Ref. 7).

(ii) The ridge trace should be stable. This includes the accomplishment of all reversals in sign with respect to the initial signs at $k = 0$ (ordinary least squares regression), (Ref. 4)

(iii) The mean square error (MSE) of forecast should not have increased greatly beyond the initial values at $k = 0$, (Ref. 4, 7).

These three requirements are the author's set of guidelines formed from the materials in the references. The first deals with the variance inflation factors which are defined by Marquardt, as as the diagonal elements of (see Ref. 6 p. 609)

$$[X'X+kI]^{-1}(X'X)[X'X+kI]^{-1}$$

when $X'X$ is in correlation form (i.e., correlation matrix of the regressive variables). For our set of problems, criterion (i) is met uniformly for $k \geq .02$.

The ridge trace is the vector of regressive coefficients viewed as functions of k, the ridge constant (Ref. 3). Typically, they are unstable and change sharply as k moves away from zero, but settle down after that. It is recommended that k be large enough so that any regression coefficient that is going to change its sign from that at $k = 0$, be allowed to do it. Much of the information in the ridge trace is summarized by $L^2(k)$, the squared length of the regression vector of coefficients. The ridge trace is also valuable if one is selecting variables for deletion.

For our set of problems, many of the ridge traces settle down quite quickly by the time k reaches .04. There are some stragglers however, but even so all have stabalized by the time k has reached .2.

Accordingly, the range $.05 \leq k \leq .2$ was chosen for further study. Typically the MSE grows modestly in this range. The movement of the MSE is represented by the movement of R^2, where $1 - R^2$ is the ratio of the sum of squared errors to the sum of squares of the dependent variable. This corresponds to looking at the degradation in the square of the multiple correlation coefficient, the two coinciding when $k = 0$.

Table 1 contains initial information for the application of ridge regression using the chosen variables. On the left are the three kinds of variables W (contract losses), G (gains), and A (attrition). For each then are 7 LOS groups and 7 PG groups. Ridge regressions were performed for k starting at zero, advancing in increments of .005 until 0.1 is achieved and then 0.2, 0.4, 0.6, 0.8, and 1.0.

The ranges of values of k that brought the VIF (variance inflation factor) into the range of one to ten are tabulated next. Values of k this small are expected when all variables have been standardized. (Ref. 6)

The column headed $R^2(0)$ is the square of the multiple correlation coefficient under ordinary least squares $(k = 0)$. It is one minus the ratio of the sum of squared residuals to the sum of squares of the dependent variable. It's use as a measure of the percent of variance accounted for is tenuous in the current application because these are so few degrees of freedom to estimate the variance of residuals. (Ref. 1, 5). Thus, large values are encouraging but not to be depended upon. They measure the level of "explanability" for this particular set of data, but the measure is not reliable for prediction. It can be used to measure the change in the sum of squares of residuals as k increases.

The remaining data in the table give indications of the degree of "ill conditioning" of the problem. The "min eigenvalue" refers to the correlation matrix X'X of the regressor variables. For orthogonal data all eigenvalues are unity. The small values indicate substantial ill conditioning. (Ref. 3, 7). The quantity $L^2(k)$ is the squared length of the regression coefficient vector. It will be greatest when $k = 0$. The L^2 range values correspond to values of k in the preceding k range. These values have stabilized in all cases. The asterisks (*) denote those groups whose regression coefficients have not stabilized in sign within the k range.

Some basic data for our 42 cases are contained in Table 2. Following the designator columns are the means, standard deviations, and coefficients of variation (ratio of standard deviation to mean) of the dependent variables.

For illustrative purposes (and for fun) it was decided to display a complete set of regression predictions for 1976. The value $k = .05$ was chosen arbitrarily and applied uniformly. The values $R^2(.05)$ should be compared with the $R^2(0)$ values of the earlier table to indicate the growth of the sum of squared residuals as k increases. The last two columns contain the 1976 forecasts and their root mean square errors (sum of squared residuals over $n - p - 1$ raised to one-half power). Because of the correction for degrees of freedom the RMSE is actually measurably larger than the dependent variable standard deviation in five of the cases. (W Los 1, W Los 5, G Los 2, A Los 2, A Los 6-17.) In thirteen of the cases it is dramatically smaller--this is especially notable because all of the five figure standard deviations are converted to four figure RMSE values, and several four figure standard deviations are reduced to three figures. The remaining cases are in the range of no change to modest improvement.

Finally the corresponding regression coefficients (converted back to the original dimensions) appear in Table 3. These values need interpreting, i.e., why should contract losses be negatively correlated with separations in some cases and positively in others, etc. These questions may have rational answers, or they may indicate the need for a better selection of variables.

Discussion

The current exploratory work should be continued--seeking better sets of regression variables (not necessarily uniform in kind across the cases), individualized values of the ridge constant (k), and perhaps some refined modeling. Also there may exist some important exogeneous variables (e.g., dates of major policy changes, planning targets for the size of the Navy, the unemployment rate), but they may be hard to identify.

REFERENCES

[1] Goldberger, A. S., Topics in Regression Analysis, Macmillan, 1968.

[2] Hemmerle, W. J., "An Explicit Solution for Generalized Ridge Regression," Technometrics, Vol. 17, No. 3, August 1975.

[3] Hoerl, A. E. and Kennard, R. W., "Ridge Regression: Biased Estimation for Nonorthogonal Problems," Technometrics, Vol. 12, No. 1, February 1970.

[4] Hoerl, A. E. and Kennard, R. W., "Ridge Regression: Applications to Nonorthogonal Problems," Technometrics, Vol. 12, No 1, February 1970.

[5] Huang, D. S., Regression and Econometric Methods, Wiley, 1970.

[6] Marquardt, D. W., "Generalized Inverses Biased Linear Estimation and Nonlinear Estimation," Technometrics, Vol. 12, No. 3, August 1970.

[7] Marquardt, D. W. and Snee, R. D., "Ridge Regression in Practice," The American Statistician, Vol. 29, No. 1, February 1975.

Dep Var	Group	k range (based on VIF)	$R^2(0)$	Min. Eigenvalue	$L^2(0)$	L^2 range	
W	LOS 1	.015, .05	.41	.006	24.6	.55, 2.47	
	2	.02, .03	.94	.002	55.7	.74, .93	*
	3	.015, .03	.98	.005	4.6	.93, 1.11	
	4	.015, .03	.99	.012	1.0	.90, .93	
	5	.015, .03	.36	.005	32.3	.94, 2.3	*
	6-17	0. .1	.91	.104	1.1	.72, 1.12	
	18-31	.005 .1	.55	.058	.9	.53, .89	*
	PG1-3	.01, .03	.84	.002	26.8	.57, 1.10	*
	4	.015, .1	.96	.021	1.1	.73, 1.03	
	5	.015, .05	.74	.010	.8	.70, .77	
	6	0. , .05	.65	.087	1.6	.85, 1.63	*
	7	0. , .05	.81	.060	.5	.42, .49	
	8	0. , .05	.45	.101	.8	.65, .78	
	9	0. , .1	.95	.259	1.6	1.10, 1.64	
G	LOS 1	.01, .1	.99	.035	4.3	.76, 2.86	*
	2	.015, .1	.85	.028	4.7	1.01, 3.47	
	3	0. , .05	.62	.096	1.2	.57, .75	
	4	0. , .05	.79	.138	6.3	.68, 1.57	
	5	0. , .1	.73	.062	4.6	1.02, 4.62	
	6-17	0. , .1	.68	.119	1.2	.89, 1.19	
	18-31	0. , .1	.97	.059	4.0	1.11, 4.03	
	PG1-3	.01, .1	.99	.030	4.2	.78, 3.47	*
	4	.005, .1	.41	.041	.5	.34, .46	
	5	0. , .1	.62	.062	1.4	.90, 1.41	
	6	0. , .1	.73	.041	.8	.56, .78	
	7	0. , .1	.69	.120	1.2	.90, 1.16	
	8	0. , .05	.67	.210	1.3	.96, 1.26	
	9	0. , .1	.71	.265	1.8	.49, 1.83	*
A	LOS 1	0. , .1	.96	.311	1.4	.94, 1.37	
	2	0. , .05	.24	.126	.3	.25, .33	*
	3	0. , .02	.69	.253	1.0	.86, .95	
	4	0. , .1	.78	.295	.5	.44, .51	
	5	0. , .1	.59	.109	1.2	.66, 1.18	
	6-17	0. , .1	.42	.164	.6	.35, .57	
	18-31	0. , .05	.33	.094	2.7	1.22, 2.72	
	PG1-3	.01 , .1	.93	.039	.8	.60, .78	
	4	0. , .1	.94	.110	.8	.53, .79	
	5	0. , .1	.85	.185	1.1	.70, 1.13	*
	6	0. , .02	.59	.119	.9	.75, .90	
	7	0. , .1	.56	.251	1.1	.62, 1.06	
	8	0. , .05	.65	.425	.8	.68, .78	
	9	0. , .1	.55	.277	.9	.62, .92	

TABLE 1

Dep Var	Group	Mean	Std Dev	Coef Var	$R^2(.05)$	Forecast 1976(k=.05)	RMSE
W	LOS 1	889.9	954.6	1.07	.291	569.8	1137.1
	2	10545.6	6410.5	.61	.847	6051.1	3543.6
	3	21957.2	11497.9	.52	.959	3131.8	3298.6
	4	47718.0	16041.8	.34	.988	21117.8	2474.9
	5	8832.9	2833.4	.32	.224	6589.1	3529.7
	6-17	9853.2	2614.7	.26	.901	3712.5	1163.4
	18-31	875.3	283.3	.32	.547	756.0	269.7
	PG1-3	27683.2	8296.4	.30	.791	18604.2	5362.6
	4	44525.9	12773.1	.29	.954	28005.3	3895.3
	5	24146.5	10168.7	.42	.742	10955.0	7301.7
	6	3298.4	747.3	.23	.634	2463.3	639.3
	7	837.1	316.2	.38	.810	249.5	194.7
	8	121.7	32.4	.27	.451	77.7	33.9
	9	59.2	18.7	.32	.948	13.1	6.0
G	LOS 1	3482.3	2476.2	.75	.945	7800.8	1010.6
	2	1978.9	1185.5	.60	.408	5003.2	1490.0
	3	1401.3	508.8	.36	.614	2704.7	516.1
	4	1508.0	483.4	.32	.734	1521.9	406.9
	5	1108.8	207.9	.19	.685	1049.2	190.5
	6-17	5518.1	1242.5	.23	.666	3412.0	1173.4
	18-31	935.6	323.5	.36	.693	929.2	291.2
	PG1-3	8531.9	4707.1	.55	.951	19424.3	1708.9
	4	2675.9	650.4	.24	.791	2803.4	485.4
	5	2508.0	379.4	.15	.445	2229.4	461.2
	6	1233.8	302.4	.25	.669	825.1	284.1
	7	802.0	321.8	.40	.801	756.7	234.5
	8	117.4	38.4	.33	.844	80.0	24.8
	9	64.0	21.7	.34	.623	52.1	21.8
A	LOS 1	20545.1	7329.1	.36	.957	31200.8	2138.9
	2	10131.1	3000.8	.30	.243	9611.2	3691.4
	3	5972.9	1499.3	.25	.686	6606.8	1187.5
	4	3001.3	961.3	.32	.779	1785.7	638.5
	5	1095.2	251.7	.24	.581	721.4	230.5
	6-17	5767.1	1617.6	.28	.419	7608.1	1743.5
	18-31	876.1	105.5	.12	.297	764.0	125.0
	PG1-3	35783.0	8841.4	.25	.924	50379.2	3447.0
	4	5345.8	1839.6	.34	.932	3924.1	678.0
	5	3277.1	1171.3	.36	.846	2583.2	649.2
	6	1851.4	664.6	.36	.587	1074.7	604.1
	7	878.6	377.2	.43	.557	881.9	355.0
	8	189.0	71.8	.38	.653	227.7	59.9
	9	64.0	13.6	.21	.542	72.6	13.1

TABLE 2

Dep Var	Group	Regression Coefficients of the Indicated Variables (k=.05)					
		Const.	U	V	S	T	
W	LOS 1	¯854.697	0.330288	0.693534	¯0.353212	0.017917	
	2	¯3323.404	¯0.25541	0.952801	0.04428	0.065939	
	3	¯3583.668	¯2.480848	¯0.12472	0.078764	0.315323	
	4	¯15586.46	¯0.748573	0.148314	¯0.014002	0.793914	
	5	¯169.258	¯2.282954	0.467961	¯0.168073	0.318064	
	6-17	¯4943.373	¯11.169272	0.130415	0.021375	0.109561	
	18-31	1437.579	0.007831	¯0.058284	¯0.101243	0.006547	
W	PG1-3	11797.36	¯0.193937	0.568101	0.229846	¯0.004464	
	4	¯2303.351	¯14.6953	¯0.086057	¯0.109196	0.529989	
	5	¯45096.391	12.058667	0.253253	0.07591	0.590159	
	6	4465.6	¯5.851896	0.4956	¯0.157238	0.002881	
	7	¯934.238	¯0.899586	0.371512	¯0.06209	0.053288	
	8	3.105	¯1.013038	0.411105	¯0.034354	0.016494	
	9	59.102	0.842668	1.052868	¯0.048284	¯0.007198	
		Const.	G	A	U	S	T
G	LOS 1	¯4275.732	0.039999	0.393833	¯2.790196	0.462755	0.006997
	2	693.392	0.895267	0.055717	0.00732	¯0.089016	0.002754
	3	1891.658	0.779972	¯0.150336	0.074671	¯0.022291	¯0.002087
	4	1765.656	0.322559	¯0.304602	¯0.016921	0.001838	0.00108
	5	428.317	0.638435	¯0.419748	¯0.282194	0.011759	0.017218
	6-17	10478.53	¯0.667479	¯0.489492	¯2.190218	0.140414	¯0.019836
	18-31	¯341.382	¯0.055081	1.402199	0.860208	¯0.094002	0.016774
G	PG1-3	6663.93	0.453469	0.289034	¯0.243511	¯0.152454	¯0.019333
	4	6354.993	¯0.404924	¯0.285919	0.633956	¯0.004342	¯0.010881
	5	2478.81	¯0.231998	¯0.156558	¯0.258073	0.001918	0.010872
	6	¯1655.742	¯0.207115	¯0.228523	1.245263	¯0.003333	0.046304
	7	867.984	¯0.209065	0.418086	¯1.634225	¯0.099786	0.018155
	8	¯227.022	¯0.167405	0.176314	0.320814	¯0.049615	0.04825
	9	¯68.438	¯0.643818	0.901776	3.739978	¯0.038501	0.039221
		Const.	A	U	X	T	
A	LOS 1	32096.51	0.754409	¯0.401486	¯0.826655	¯0.069612	
	2	¯1762.357	0.439853	2.202321	¯0.162965	0.037994	
	3	¯4252.39	0.460742	3.664622	¯0.013841	0.05637	
	4	¯5700.411	0.1195	3.283925	0.046024	0.085348	
	5	2597.368	¯0.246084	0.234793	¯0.159988	¯0.006043	
	6-17	1115.538	¯0.38336	0.232208	¯0.054455	¯0.085045	
	18-31	26.042	0.245201	3.8736	0.026573	¯0.001588	
A	PG1-3	¯2625.808	0.916001	¯2.507623	¯29.055833	0.104343	
	4	6853.783	0.298235	0.002521	¯1.592162	0.022486	
	5	5675.728	¯0.207994	0.497185	¯0.51989	0.024126	
	6	867.841	0.300298	0.000693	¯0.118133	0.027116	
	7	423.653	0.674542	¯0.619689	¯0.005627	0.007372	
	8	7794.112	0.151657	6.408782	¯0.172355	¯0.00384	
	9	429.847	¯0.196835	0.779111	¯0.021464	0.01474	

TABLE 3

APPENDIX B

ELEVEN YEAR AVERAGES

TABLE B.1

Average Separations

	E1-3	E4	E5	E6	E7	E8	E9
	2057	1254	119	15	11	2	2
	9720	7688	969	41	12	2	1
	7183	11837	4210	43	18	3	2
	8535	24107	17792	146	25	4	2
	740	2014	2512	159	32	4	2
LOS	404	917	2069	560	40	4	3
	111	349	1197	915	48	6	3
	79	296	1282	1433	80	7	4
	42	249	1278	1880	197	6	4
	49	229	1078	1712	242	6	4
	18	97	488	1066	278	10	4
	14	87	443	1136	398	25	4
	7	45	359	1299	675	66	6
	8	51	384	1434	844	103	9
	7	48	374	1548	1145	182	25
	6	40	279	1175	1003	188	32
	5	24	140	709	817	213	52
	3	15	108	521	681	188	55
	2	11	61	280	543	185	68
	2	8	36	189	426	162	68
	1	6	19	95	280	128	65
	1	4	14	64	192	94	54
	1	3	11	45	148	81	60
	1	2	9	33	128	77	57
	1	2	5	28	102	66	59
	1	2	5	15	76	54	48
	1	1	2	9	57	42	47
	1	1	2	6	36	30	29
	1	1	1	3	17	9	11
	1	1	1	3	10	7	9
	1	1	2	2	11	7	6

TABLE B.2

AVERAGE ELIGIBLES

	E1-3	E4	E5	E6	E7	E8	E9
	1307	1234	116	14	9	1	1
	6104	7573	960	40	10	1	0
	5038	11699	4187	41	16	1	1
	5666	23770	17690	144	24	3	1
	579	1982	2488	157	31	3	1
LOS	351	898	2050	556	38	3	2
	91	339	1189	911	47	5	2
	62	288	1271	1428	79	6	3
	33	240	1266	1873	196	5	3
	39	221	1065	1706	241	5	3
	14	94	483	1061	276	9	3
	11	84	440	1132	397	24	3
	5	43	355	1296	673	65	5
	5	49	380	1429	842	102	8
	6	45	370	1544	1143	181	23
	5	38	275	1172	1001	187	31
	4	22	138	707	816	212	51
	2	14	107	519	679	187	54
	1	10	59	278	541	184	67
	1	7	35	187	423	161	67
	0	5	18	93	278	127	64
	0	3	13	61	189	93	53
	0	2	10	42	144	79	59
	0	1	8	30	124	76	55
	0	1	4	24	97	65	58
	0	1	3	12	71	52	47
	0	0	1	8	53	40	46
	0	0	1	4	32	28	28
	0	0	0	2	14	8	10
	0	0	0	2	8	6	8
	0	0	1	1	9	6	5

TABLE B.3

AVERAGE NON-REENLISTMENTS

	E1-3	E4	E5	E6	E7	E8	E9
	1190	1134	102	11	9	2	1
	5841	6343	647	30	10	2	1
	4639	9600	2594	30	16	2	2
	5260	21297	13434	52	22	4	2
	432	1536	1711	79	28	4	2
	127	525	1361	354	34	3	3
	56	190	566	404	37	6	3
LOS	43	154	511	460	37	6	4
	19	91	396	428	38	5	4
	24	83	347	397	50	5	3
	6	27	107	182	55	5	3
	6	21	82	142	52	7	4
	3	11	55	104	47	5	3
	4	13	45	99	46	7	3
	3	9	35	78	40	6	3
	2	6	21	53	35	7	2
	2	3	11	24	23	5	2
	1	2	7	19	21	6	3
	1	2	4	10	19	4	3
	1	2	5	9	19	5	3
	1	2	3	6	13	3	2
	1	1	2	5	11	3	2
	1	1	2	4	14	3	1
	1	1	2	5	16	3	3
	1	1	1	4	13	3	3
	1	1	1	3	11	3	2
	1	1	1	2	9	3	2
	1	1	1	1	6	2	2
	1	1	1	1	4	2	1
	1	1	1	1	3	1	1
	1	1	1	1	7	3	1

APPENDIX C

APL PROGRAMS

The main programs are REGR and RESID which perform the ridge regression computations and develops the residuals and their properties (resp.). These are used in FCAST which compute the regression forecast for the entire 31 by 7 set. Individual cell forecasts are computed by PRED. The other programs prepare the data.

More explicitly, the raw data consist of five 11 by 31 by 9 arrays D1, D2, ... , D5 which carry the eleven year values of non-reenlistments, separations, eligibles, retentions, inventory, (resp.). The program COMPRESS merely telescopes pay grades E1, E2, E3 together producing 11 × 31 × 7 arrays. This must be done separately.

The function FCAST requires an explicit input vector CR which is the set of partition boundaries for grouping the range space of the 217 eleven year means of the object variable. It also uses the data array DD implicitly.

```
        ∇COMPRESS[⎕]∇
      ∇ Z←COMPRESS D;V;Z1
[1]     V←(ρD)ρ0
[2]     V[;; 1 2 3]←1
[3]     Z1←+/V×D
[4]     Z← 0 0 3 ↓Z←(~V)×D
[5]     Z←Z1,Z
      ∇
```

```
      ∇FCAST[⎕]∇
    ∇ PP←FCAST CR;J;P
[1]   V←((ρCR),31,7)ρ0
[2]   PP← 31 7 ρ0
[3]   J←0
[4]  L1:J←J+1
[5]   V[J;;]←(DD≥CR[J+1])-DD≥CR[J]
[6]   P←PRED V[J;;]
[7]   PP←PP+P
[8]   →L1×ιJ<¯1+ρCR
    ∇
```



```
      ∇PRED[⎕]∇
    ∇ P←PRED V;PRP
[1]   D←PREP V
[2]   R←RC REGR D
[3]   E←D RESID R
[4]   P← 31 7 ρPRP←(,V)\PR
    ∇
```

The function FCAST creates a 31 by 7 screening matrix of zeros and ones which is used by the data shaping functions PREP, BUILD, and SHAPE. The latter, SHAPE, merely shaves off the first or last face (according to whether the variable Q is minus one or plus one) of the individual data arrays. The function BUILD assembles the object variable in its first face and the p regression variables in the remaining faces. This is the only function that needs to be changed with each application. The version shown is for regressing non-reenlistments. on inventory, previous non-reenlistments, and previous retentions. The output of PREP prepares the assembled data for REGR.

```
      ∇PREP[⎕]∇
    ∇ D←PREP V
      D←BUILD V
      X←D[(ρD)[1];;1+ι(ρD)[3]-1]
      D←(1,(10×(ρD)[2]),(ρD)[3])ρD
    ∇
```

```
      ∇BUILD[⎕]∇
    ∇ D←BUILD V;S1;S2;T;E;Q
      Q←1
      S1←V SHAPE DD1
      E←V SHAPE DD4
      Q←¯1
      S2←V SHAPE DD1
      T←V SHAPE DD5
      D←S2,[2.5] S1
      D←(D,E),T
    ∇
```



```
      ∇SHAPE[⎕]∇
    ∇ S←V SHAPE D;N;K
      N←1↑ρD
      K←+/+/V
      V←(N,ρV)ρV
      D←(N,K)ρ(,V)/,D
      S←((Q×N-1),K)↑D
    ∇
```

The inputs to REGR are the data matrix D and the vector of ridge constants RC. The rank of D is three; the first dimension (faces) indexing the problems (separate predictions), the second (rows) the observations (years), and the third (columns) the variables--the object variable coming first followed by the regression variations. The output R also has three dimensions. Again the faces are the prediction problems. For any face, the first ρRC element of column one

```
      ∇REGR[⎕]∇
    ∇ R←RC REGR D;MU;SIG;X;Y;C;B;R2;A;FAC;RT;N;P;ID;K;CT
[1]     RC←,RC
[2]     →(2+I26)×ι3=ρρD
[3]     D←(1,ρD)ρD
[4]     P←(ρD)[3]-1
[5]     ID←(P,P)ρ1,Pρ0
[6]     N←(ρD)[2]
[7]     R←((1↑ρD),(ρRC),(P+1))ρ0
[8]     MU←+/[2] D÷N
[9]     A←D- 2 1 3 ⍉(N,ρMU)ρMU
[10]    SIG←+/[2](A*2)÷N-1
[11]    SIG←SIG*0.5
[12]    D←A÷ 2 1 3 ⍉(N,ρSIG)ρSIG
[13]    A←0
[14]    ⍝ MU,SIG ARE MATRICS OF MEANS AND STD DEVS.
[15]    ⍝ THE DATA D HAS BEEN STANDARDIZED.
[16]    ⍝ A WAS A TEMPORARY VALUE.
[17]    ⍝ BEGIN OUTER LOOP ON FACES.
[18]  L3:FAC←0
[19]  OUTER:FAC←FAC+1
[20]    X← 0 1 ↓D[FAC;;]
[21]    Y←D[FAC;;1]
[22]    C←(÷N-1)×(⍉X)+.×X
[23]    ⍝ BEGIN INNER LOOP ON K (RIDGE TRIAL.)
[24]    RT←0
[25]  INNER:RT←RT+1
[26]    K←RC[RT]
[27]    A←⌹C+K×(P,P)ρ1,Pρ0
[28]    B←A+.×(⍉X)+.×Y÷N-1
[29]    R2←1-(÷N-1)×+/(Y-X+.×B)*2
[30]    R[FAC;RT;]←R2,B
[31]    →INNER×ιRT<ρRC
[32]    →OUTER×ιFAC<1↑ρD
[33]    ⍝ ALL RIDGE REGRESSIONS COMPLETE.
[34]    ⍝ FINISH OUTPUT ARRAY
[35]    R←R,[2] MU,[1.5] SIG
[36]    R←R,(¯1↓ρR)ρRC, 0 0
    ∇
```


are the coefficients of determination (i.e. square of mult. correl.) and the first ρRC elements of the last column are the values of the ridge constant. All intermediate columns contain the standardized regression coefficients in the same order as they appear in D. The last two rows contain the mean and standard deviation (resp.) of the object and regression variables. Zeros are used to square up the array.

The function RESID takes as input the data matrix D and the output R of REGR. Its explicit output E is the set of residuals. Implicit output includes EB, the mean residual; SEE, the root mean square error of forecast; and BB, the regression coefficients converted back to dimensional form. The array C is the constant term of the regression equations. Line 15 shows that only the positive parts of the predictions are used.

```
      ∇RESID[⎕]∇
    ∇ E←D RESID R;M;S;B;M1;FB1;SS;S1;D1;D2;PR1;C;CC;KK
[1]    RC←,RC
[2]    P←(¯1↑ρR)-2
[3]    M←R[;1+ρRC;⍳P+1]
[4]    S←R[;2+ρRC;⍳P+1]
[5]    B←R[;K;1+⍳P]
[6]    S1←⍉(P,N←1↑ρR)ρS[;;1]
[7]    BB←B×S1÷(S←(N,P+1)ρS)[;1+⍳P]
[8]    M1←M[;;1]
[9]    C←M1-+/M[;;1+⍳P]×(N,1,P)ρBB
[10]   D1← 0 0 1 ↓D
[11]   D2← 0 0 ¯2 ↓D
[12]   B← 2 1 3 ⍉((KK←(ρD)[2]),N,P)ρ,BB
[13]   CC←⍉(KK,N)ρC
[14]   PR1←0⌈CC++/B×D1
[15]   PR←0⌈(1,(ρX)[1])ρC++/X×(ρX)ρBB
[16]   E←(D2←(N,KK)ρD2)-PR1
[17]   SEE←S[;1]×((1-R[;K;1])×(KK÷KK-P))*0.5
[18]   EB←(÷KK)×+/E
[19]   SE←÷(KK-P)÷+/(E-⍉(KK,N)ρEB)*2
[20]   SE←SE*0.5
[21]   BB← 11 6 DFT BB
[22]   C← 11 3 DFT C
[23]   BB←C,BB, 11 3 DFT SE←((1↑ρR),1)ρSE
    ∇
```