now called Joint Maximum Likelihood Estimation JMLE

VERIFYING THE UNCONDITIONAL ESTIMATION PROCEDURE FOR RASCH ITEM ANALYSIS WITH SIMULATED DATA

by

Benjamin D. Wright

and

Susan R. Bell

The University of Chicago

This document reports a study of the extent to which the unconditional estimation procedure for Rasch item analysis (Wright and Panchapakesan, 1969) used in the Rasch measurement program, BICAL (Wright and Mead, 1977), successfully evaluates fit and recovers parameters for data generated according to the model.

Research Memorandum Number 26

Statistical Laboratory
Department of Education
The University of Chicago
September 1977

Introduction

Andersen (1970) proves that a conditional maximum likelihood procedure for Rasch item analysis produces consistent estimates of item parameters. Since, however, conditional estimation is impractical for tests of more than twenty items (Wright and Douglas, 1977b), Wright and Panchapakesan (1969) developed an "unconditional" maximum likelihood procedure, UCON. Although this method is quite practical even for the largest tests it produces slightly biased estimates. Wright and Douglas (1977a, 1977b) describe a corrected unconditional estimation which should make the bias in UCON negligible. This corrected unconditional estimation procedure is incorporated into a practical and economical FORTRAN program, BICAL (Wright and Mead, 1977), which calibrates items and analyzes fit between model and data according to the Rasch model. The purpose of this report is to document the extent to which the unconditional maximum likelihood procedure used in BICAL produces accurate and consistent estimates, and so, incidentally, to see if there is any practical need to attempt conditional estimation of item difficulty parameters.

To explore this question, a study of data simulated to fit the model was performed. A comparison of estimates with their generating parameters was made for data generated from a variety of typical test and sample shapes.

Definition of Test and Sample

A test or a sample can be described adequately by four basic characteristics (Wright and Douglas, 1975a). The units in which these properties will be expressed are logits. Logits are natural log odds of frequency data such as test and item scores which transform these frequency data into a linear scale. (For a discussion of logits apropos the Rasch model, see Wright, 1977.)

Tests

The first characteristic of a test is its difficulty level or height, H. This is the average difficulty of the test's items. A centered test is one with a height near the mean ability or center of the sample in use.

Width, W, is the second characteristic of a test. This is the range of item difficulties covered by a test, from the easiest item to the hardest.

Length, L, the number of items composing the test is the third characteristic of a test.

Finally, the distribution of the item difficulties must be specified. Typically, either a normal or uniform distribution is adequate to describe test shape. Wright and Douglas (1975) show that uniformly distributed items are the best overall test shape strategy and that in general, "best" tests should be constructed that way.

Samples

The mean ability level, M, of a sample corresponds in interpretation to the height of a test. In the construction of a scale, the origin, in general, is arbitrary. What is determined by the data is $(\beta - \delta)$ or $(M - H)$, the difference between sample and test. For the purposes of these simulations the height of the test was set at zero and the difference (M - H) was varied by varying M. Therefore, a sample less able than the test is difficult would have a mean ability less than zero, while a sample more able than the test is difficult would have a mean greater than zero. Since the response model functions symmetrically, however, we need consider only one side of the scale when simulating persons in order to see how well the estimation algorithm works.

The second characteristic of a sample is its dispersion, or standard deviation, S. Just as the width of a test is the range of item difficulties composing the tests, so S indicates the spread of person abilities in the sample.

Sample size, N, is the third characteristic of a sample.

The last characteristic is sample shape. For most practical purposes a normal distribution of people is an adequate representation of sample shape.

Scope of the Simulations

Fifteen years experience in seven AERA Rasch presessions and innumerable consultations with many different tests used under a wide variety of circumstances, has shown that tests narrower than two logits or wider than six logits are extremely rare. The most frequent test widths encountered have been in the region of three to five logits. Therefore, the simulations were performed for tests of width two, four, and six logits to cover this experience. A few simulations made at one and three logits in order to see the trend around two are shown in the summary tables, but were not included in the graphical analyses.

Test lengths of twenty-one and forty-one items were chosen for the simulations. Few serious tests are shorter than twenty items and, as tests exceed forty or fifty items, the impact of the amount of data becomes sufficient to wipe out the problem of inconsistency in the unconditional estimates. Twenty-one was chosen because it occasionally happens that an in-class test or one subtest of a lengthy battery is as short as twenty items. Forty-one was chosen to show how increasing test length eases estimation problems and improves the estimates. A second reason was that sub-tests in many widely used batteries run about forty to fifty items.

Two levels (means) of samples were chosen for the simulations. The first level was centered on the tests (M = 0), as that is the most desirable set-up. The second level was set at one logit above test center (M = 1), because calibration samples are often selected to be somewhat more able than the test is difficult, in order to diminish the effects of guessing.

A variety of practical experience calibrating school tests has shown that within-grade ability standard deviations tend to run around one logit. In order to cover this, standard deviations of ability were simulated at 0.5, 1.0, and 1.5 logits.

Sometimes attempts have been made to calibrate a test over several grade levels. In those circumstances ability standard deviations can reach 1.5 logits or even somewhat more. However, in the careful development of a calibrated bank of items, it is more efficient and more reliable to calibrate within-grade levels and to equate across grade levels by common item links, rather than to include the possible vicissitudes of across grade level item instability in the initial estimates of item difficulties. Thus, the standard deviations of ability in these simulations did not exceed 1.5 logits.

Sample sizes of four hundred and eight hundred persons were chosen, because in principle, given the model, four hundred suitably chosen persons should be enough to determine item characteristics effectively, and eight hundred persons should be more than sufficient.

All of the simulations made in this study had an average item difficulty of zero, items that were uniformly distributed, and samples that were normally distributed. Two typical combinations of test length and sample size produced two basic set-ups. The first set-up was a test of twenty-one items taken by four hundred persons. The second set-up was a test of forty-one items taken by eight hundred persons. For each of these two set-ups various plausible combinations of test width, sample mean, and sample dispersion were formed. The twenty-three combinations resulting are marked by x in Table 1. Five random replications of each of these combinations were then simulated and provide the data used here to investigate the extent to which BICAL evaluates fit and recovers the item difficulty parameters satisfactorily.

Not all possible combinations of the variables in Table 1 were simulated, because some combinations generate unrealistic or nonsensical situations. For instance, it would be unreasonable to give a narrow test (W = 2) to a sample that was more able than average (M = 1), because too many of the people would fall outside the range of the test.

Recovery of Item Difficulties

The first analysis of the simulation results concerns possible bias in the recovery of item difficulties. An examination of Table 2 suggests that BICAL recovery is excellent at all combinations of the variables. A closer look reveals that for shorter and wider tests (L = 21, W = 6), the recovery of item difficulty dispersion runs one to five percent too large, while for the longer and wider tests (L = 41, W = 6), the recovery runs only one percent too large. Regardless of test width, the average excess in dispersion over all simultions is two percent for the shorter test and one percent for the longer test. This suggests that a slight additional correction for bias based on test length might be useful for shorter tests ($L < 30$).

Further analysis of calibration bias appears in Tables 3 and 4. These tables show the regression of individual item difficulty estimates on their generating parameters, for the tests of twenty-one by four hundred persons and forty-one items by eight hundred persons, respectively. It is hard to see how the BICAL recovery could be any better. Overall, the best estimates are recovered by the longer test. A closer examination of both Tables 3 and 4 reveals again that the greatest recovery inflation (5%) occurs for the shorter wider tests with off-center samples (L = 21, W = 6, M = 1), either narrowly or widely dispersed (S = 0.5 or 1.5) and for average tests which are off-center, but moderately dispersed (L = 21, W = 4, M = 1, S = 1.0), while for the longer tests, the largest inflation (2%) occurs on the narrow test with centered, narrowly dispersed samples (L = 41, W = 2, M = 0, S = 0.5) and on wider tests with off-center, narrowly dispersed samples (L = 41, W = 6, M = 1, S = 0.5).

Figures 1 through 23 are graphs of the five replications of BICAL item estimates plotted against their generating parameters, for each of the twenty-three combinations of test and sample characteristics. It can be seen from these graphs that the item points are well-balanced around the identity line and that nearly all the estimates fall within three standard errors of that line. This documents the consistency of the estimates.

Distribution of Fit Statistics

In order to judge when data are adequately recovered by the model it is necessary to assess quantitatively the agreement between observed data and their expectation under the model. To evaluate the fit of the data to the model it is useful to develop a fit statistic for each item and for each person.

These fit statistics for items and persons are formed by partitioning an approximate χ^2 statistic for the overall fit of the data,

$$\chi^2 = \sum_{v}^{N}\sum_{i}^{L} z_{vi}^2$$

into L parts, one for each item,

$$\chi_i^2 = \sum_{v}^{N} z_{vi}^2$$

and into N parts, one for each person,

$$\chi_v^2 = \sum_{i}^{L} z_{vi}^2$$

where,

$$z_{vi}^2 = \frac{(x_{vi} - p_{vi})^2}{p_{vi}(1 - p_{vi})}$$

and where,

$$p_{vi} = \frac{\exp(b_v - d_i)}{1 + \exp(b_v - d_i)}$$

The degrees of freedom are approximately [(N-1)(L-1)]/L for each item and approximately [(N-1)(L-1)]/N for each person.

It is convenient to express these statistics in the form of mean square residuals. For the overall fit of item i this is

$$V_i = \sum_v^N z_{vi}^2 \left[\frac{L}{(N-1)(L-1)}\right]$$

which has an expected value of one and a variance of 2L/[(N-1)(L-1)].

The corresponding mean square for the fit of person v is

$$V_v = \sum_i^L z_{vi}^2 \left[\frac{N}{(N-1)(L-1)}\right]$$

which has an expected value of one and a variance of 2N/[(N-1)(L-1)].

The fit mean square residual for the entire test is given by

$$V = \sum_v^N \sum_i^L z_{vi}^2 \left[\frac{1}{(N-1)(L-1)}\right]$$

with an expected value of one and a variance of 2/[(N-1)(L-1)].

If we represent the expected value of the standard deviation of the item fit statistic V_i as $\sigma_V = \sqrt{2/df}$, and its observed value over items in any particular calibration as s_V, the ratio s_V/σ_V can become a useful part of the fit analysis, as it standardizes the dispersion of the item mean squares around their expected value of one.

The analysis of the fit statistic is illustrated in Tables 5 and 6. In Table 5 the observed value of the test fit statistic V is compared with its expected value of one. One can see that the fits are closest to expectation when the test is narrow (W = 2). Regardless of test length the fits drift slightly below one as the test becomes wider, although the percent below one is greater for the shorter test (L = 21). Sample properties do not appear to have a significant influence on these fit statistics, as the departure below expectation is about the same at the different levels and dispersions of sample ability. Overall, the longer tests (L = 41) show the least departure below expectation in their fit statistics over the different test widths, two to three percent at all levels versus two to six percent for the shorter test (L = 21).

The standard deviation of the item fit statistics is analyzed in Table 6. These show two interesting trends. There is a distinct inflation above expectation with short, wide tests (L = 21, W = 6). The ratio (s_V/σ_V) becomes as large as two for wider tests (W = 6), regardless of test length. This is twice the expected value of one and indicates that the item mean squares for data simulated to fit the model are substantially more dispersed around one than the elementary theory concerning these residuals implies.

Further examination of Table 6 shows that the expected ratio of one is found on the narrow tests with centered and moderately dispersed samples (W = 2, M = 0, S = 1.0) and on the average width tests with off-center, narrowly dispersed samples (W = 4, M = 1, S = 0.5) for tests of both lengths. But as the theoretical ratio of one is too low compared to the ratios (s_V/σ_V) observed for wide tests (W > 4) taken by dispersed samples (S > 1.0), some adjustments will be required in the reference values used when judging item fit in real data collected under those circumstances.

The examination of the fit statistics for these tests and their items shows how BICAL functions with data known to fit the model. From these simulations it seems desirable to work with longer tests of moderate width well-targeted on their calibration samples (e.g., L = 41, W = 4, M = 0, S = 1.0). Although it is best to develop an item bank using well-matched tests and people, these simulations show that moderate deviations from this ideal can still yield useful results.

Conclusions

The Rasch response model used by BICAL implies ideal consequences for residuals from the model. The residuals actually observed in a calibration are summarized into mean square residual fit statistics for individual items and for the test as a whole. It is unrealistic to expect the results of simulations to match the ideal consequences exactly, but one can ask, "To what extent do the results, and hence the algorithm they document, approximate these ideals?". This is an important question, as the ideals are the frame of reference from which an experimenter must judge the fit of any real data.

The simulations in this study indicate that, when test and sample properties are in approximate rapport, the fit statistics can be used to evaluate fit and, when fit is obtained, that good parameter estimates are also obtained. However, these simulations also show that when sample and/or test properties are pushed to extremes, residuals not fully anticipated by the model can obtain.

As sample and test spread beyond typical values, variation among the item mean squares, for data simulated to fit the model, increases to twice as much as expected from the model. At the same time, the total mean square falls slightly below its expected value of one. Therefore, when judging the fit of real data, for either a wide test ($W > 4$), a wide sample ($S > 1.0$), or both, one might be tolerant of item mean square dispersion larger than expected. Also, one should work toward overall mean squares falling slightly less than one.

In serious practice, however, genuine attempts are ordinarily made to avoid extreme situations. Should extreme conditions nevertheless arise, the experimenter can qualify the interpretation of his results in terms of the trends shown in these simulations.

REFERENCES

Andersen, E.B. Asymptotic properties of conditional maximum likelihood estimators. Journal of the Royal Statistical Society, 1970, 32, 283-301.

Wright, B.D. Solving measurement problems with the Rasch model. Journal of Educational Measurement, 1977, 14, No. 2.

Wright, B.D. and Douglas, G.A. Best procedures for sample-free item analysis. Applied Psychological Measurement, 1977a, 1, No. 2.

Wright, B.D. and Douglas, G.A. Conditional versus unconditional procedures for sample-free item analysis. Educational and Psychological Measurement, 1977b, Autumn.

Wright, B.D. and Douglas, G.A. Best test design and self-tailored testing. Research Memorandum No. 19, Statistical Laboratory, Department of Education, University of Chicago, 1975.

Wright, B.D. and Mead, R.J. BICAL: Calibrating rating scales with the Rasch model. Research Memorandum No. 23, Statistical Laboratory, Department of Education, University of Chicago, 1977.

Wright, B.D. and Panchapakesan, N. A procedure for sample-free item analysis. Educational and Psychological Measurement, 1969, 29, 23-48.

TABLE 1

COMBINATIONS OF TESTS AND SAMPLES WHICH WERE SIMULATED

Test: H=0, W, L=21 Sample: M, S, N=400

	M=0			M=1		
W \ S	0.5	1.0	1.5	0.5	1.0	1.5
1	*	*				
2	x	x	*	*	*	
3		*	*	*	*	*
4	x	x	x	x	x	
6		x	x	x	x	x

Test: H=0, W, L=41 Sample: M, S, N=800

	M=0			M=1		
W \ S	0.5	1.0	1.5	0.5	1.0	1.5
1						
2	x	x				
3						
4	x	x	x	x	x	
6		x	x	x	x	*

x - data based on five replications and reported fully

* - additional data of five or more replications which were generated prior to the final simulations of this report

TABLE 2A

BICAL RECOVERY OF ITEM DIFFICULTY DISPERSION

Test: H=0, W, L=21 Sample: M, S, N=400

Estimated Dispersion s_d

Generated Dispersion σ_d	W \ S	M=0 0.5	M=0 1.0	M=0 1.5	M=1 0.5	M=1 1.0	M=1 1.5
0.31	1	0.31	0.31				
0.62	2	0.63	0.61	0.63	0.62	0.62	
0.93	3		0.94	0.95	0.94	0.93	0.93
1.24	4	1.24	1.24	1.27	1.26	1.30	
1.86	6		1.91	1.92	1.96	1.88	1.95

Ratio s_d/σ_d

W \ S	M=0 0.5	M=0 1.0	M=0 1.5	M=1 0.5	M=1 1.0	M=1 1.5
1	1.00	1.00				
2	1.01	0.99	1.02	1.00	1.00	
3		1.01	1.02	1.01	1.00	1.00
4	1.00	1.00	1.02	1.02	1.05	
6		1.03	1.03	1.05	1.01	1.05

TABLE 2B

BICAL RECOVERY OF ITEM DIFFICULTY DISPERSION

Test: H=0, W, L=41 Sample: M, S, N=800

Estimated Dispersion s_d

Generated Dispersion σ_d	W \ S	M=0 0.5	M=0 1.0	M=0 1.5	M=1 0.5	M=1 1.0	M=1 1.5
0.30	1						
0.60	2	0.61	0.60				
0.90	3						
1.20	4	1.21	1.21	1.20	1.21	1.19	
1.80	6		1.82	1.82	1.84	1.80	1.82

Ratio s_d/σ_d

W \ S	M=0 0.5	M=0 1.0	M=0 1.5	M=1 0.5	M=1 1.0	M=1 1.5
1						
2	1.02	1.00				
3						
4	1.01	1.01	1.00	1.01	0.99	
6		1.01	1.01	1.02	1.00	1.01

TABLE 3

THE REGRESSION OF BICAL ITEM ESTIMATES ON THEIR GENERATING PARAMETERS - 21 ITEMS BY 400 PERSONS

Test H=0 L=21 Width W	Sample N=400 Mean M	Sample N=400 Std.Dev. S	Correlation Between Estimate and Parameter	Standard Deviation[1] of Param.	Standard Deviation[1] of Est.	Regression of Estimate on Parameter Slope	Regression of Estimate on Parameter Mean Square Residual	Regression of Parameter on Estimate Slope	Regression of Parameter on Estimate Mean Square Residual
2	0	0.5	0.985	0.608	0.615	0.996	0.0113	0.974	0.0111
2	0	1.0	0.983	0.608	0.600	0.969	0.0125	0.997	0.0128
4	0	0.5	0.995	1.217	1.215	0.994	0.0142	0.996	0.0142
4	0	1.0	0.996	1.217	1.219	0.998	0.0127	0.994	0.0127
4	0	1.5	0.994	1.217	1.248	1.019	0.0192	0.969	0.0182
6	0	1.0	0.997	1.825	1.871	1.022	0.0233	0.972	0.0221
6	0	1.5	0.997	1.825	1.877	1.025	0.0246	0.969	0.0233
4	1	0.5	0.995	1.217	1.241	1.014	0.0170	0.975	0.0163
4	1	1.0	0.995	1.217	1.274	1.041	0.0179	0.950	0.0163
6	1	0.5	0.992	1.825	1.929	1.048	0.0620	0.939	0.0555
6	1	1.0	0.996	1.825	1.841	1.006	0.0206	0.989	0.0202
6[2]	1	1.5	0.996	1.828	1.912	1.042	0.0317	0.952	0.0290

[1]Because these statistics combined five replications, the degrees of freedom used in the calculation of the standard deviations were 104 instead of 5x20 = 100. As a result, these values are .98 of the ones found in Table 2A. The ratios are comparable.

[2]These calculations were based on four replications rather than the five used in the other combinations.

TABLE 4

THE REGRESSION OF BICAL ITEM ESTIMATES ON THEIR GENERATING PARAMETERS - 41 ITEMS BY 800 PERSONS

Test H=0 L=41 Width W	Sample N=800 Mean M	Sample N=800 Std.Dev. S	Correlation Between Estimate and Parameter	Standard Deviation[1] of Param.	Standard Deviation[1] of Est.	Regression of Estimate on Parameter Slope	Regression of Estimate on Parameter Mean Square Residual	Regression of Parameter on Estimate Slope	Regression of Parameter on Estimate Mean Square Residual
2	0	0.5	0.993	0.593	0.602	1.009	0.0051	0.978	0.0049
2	0	1.0	0.990	0.593	0.600	1.001	0.0071	0.979	0.0070
4	0	0.5	0.997	1.186	1.200	1.009	0.0075	0.986	0.0073
4	0	1.0	0.998	1.186	1.195	1.005	0.0071	0.990	0.0070
4	0	1.5	0.997	1.186	1.191	1.001	0.0080	0.993	0.0079
6	0	1.0	0.999	1.779	1.798	1.009	0.0091	0.988	0.0089
6	0	1.5	0.998	1.779	1.803	1.012	0.0109	0.985	0.0106
4	1	0.5	0.996	1.186	1.202	1.010	0.0105	0.983	0.0102
4	1	1.0	0.997	1.186	1.182	0.993	0.0092	1.000	0.0093
6	1	0.5	0.998	1.779	1.818	1.020	0.0129	0.977	0.0124
6	1	1.0	0.998	1.779	1.787	1.002	0.0121	0.994	0.0120

[1] Because these statistics combine five replications, the degrees of freedom used in the calculation of the standard deviations were 204 instead of 5x40 = 200. As a result, these values are .99 of the ones found in Table 2B. The ratios are comparable.

TABLE 5

AVERAGE MEAN SQUARE FIT STATISTIC OVER ITEMS

Test: H=0, W, L=21 Sample: M, S, N=400

Average Fit Statistic V

	M=0			M=1		
W \ S	0.5	1.0	1.5	0.5	1.0	1.5
1	0.99	0.99				
2	0.99	0.99	0.99	0.99	0.98	
3		0.98	0.98	0.98	0.98	0.97
4	0.97	0.96	0.97	0.96	0.97	
6		0.95	0.94	0.93	0.94	0.94

Test: H=0, W, L=41 Sample: M, S, N=800

Average Fit Statistic V

	M=0			M=1		
W \ S	0.5	1.0	1.5	0.5	1.0	1.5
1						
2	0.99	0.99				
3						
4	0.98	0.98	0.98	0.98	0.98	
6		0.97	0.96	0.97	0.98	0.98

TABLE 6A

STANDARD DEVIATIONS OF MEAN SQUARE
FIT STATISTIC OVER ITEMS

Test: H=0, W, L=21 Sample: M, S, N=400

Standard Deviation of Fit Statistic s_V

σ_V	W	M=0, S=0.5	M=0, S=1.0	M=0, S=1.5	M=1, S=0.5	M=1, S=1.0	M=1, S=1.5
.05	1	.03	.05				
.05	2	.04	.05	.07	.04	.07	
.05	3		.06	.07	.06	.07	.08
.07	4	.06	.07	.09	.07	.10	
.07	6		.14	.16	.12	.13	.17

Ratio s_V/σ_V

W	M=0, S=0.5	M=0, S=1.0	M=0, S=1.5	M=1, S=0.5	M=1, S=1.0	M=1, S=1.5
1	0.6	1.0				
2	0.8	1.0	1.4	0.8	1.4	
3		1.2	1.4	1.2	1.4	1.6
4	0.9	1.0	1.3	1.0	1.4	
6		2.0	2.3	1.7	1.9	2.4

TABLE 6B

STANDARD DEVIATIONS OF MEAN SQUARE
FIT STATISTICS OVER ITEMS

Test: H=0, W, L=41 Sample: M, S, N=800

Standard Deviation of Fit Statistic s_V

		M=0			M=1		
σ_V	W \ S	0.5	1.0	1.5	0.5	1.0	1.5
.05	1						
.05	2	.03	.05				
.05	3						
.05	4	.04	.06	.08	.05	.07	
.05	6		.08	.12	.07	.15	.13

Ratio s_V/σ_V

	M=0			M=1		
W \ S	0.5	1.0	1.5	0.5	1.0	1.5
1						
2	0.6	1.0				
3						
4	0.8	1.2	1.6	1.0	1.4	
6		1.6	2.4	1.4	3.0	2.6

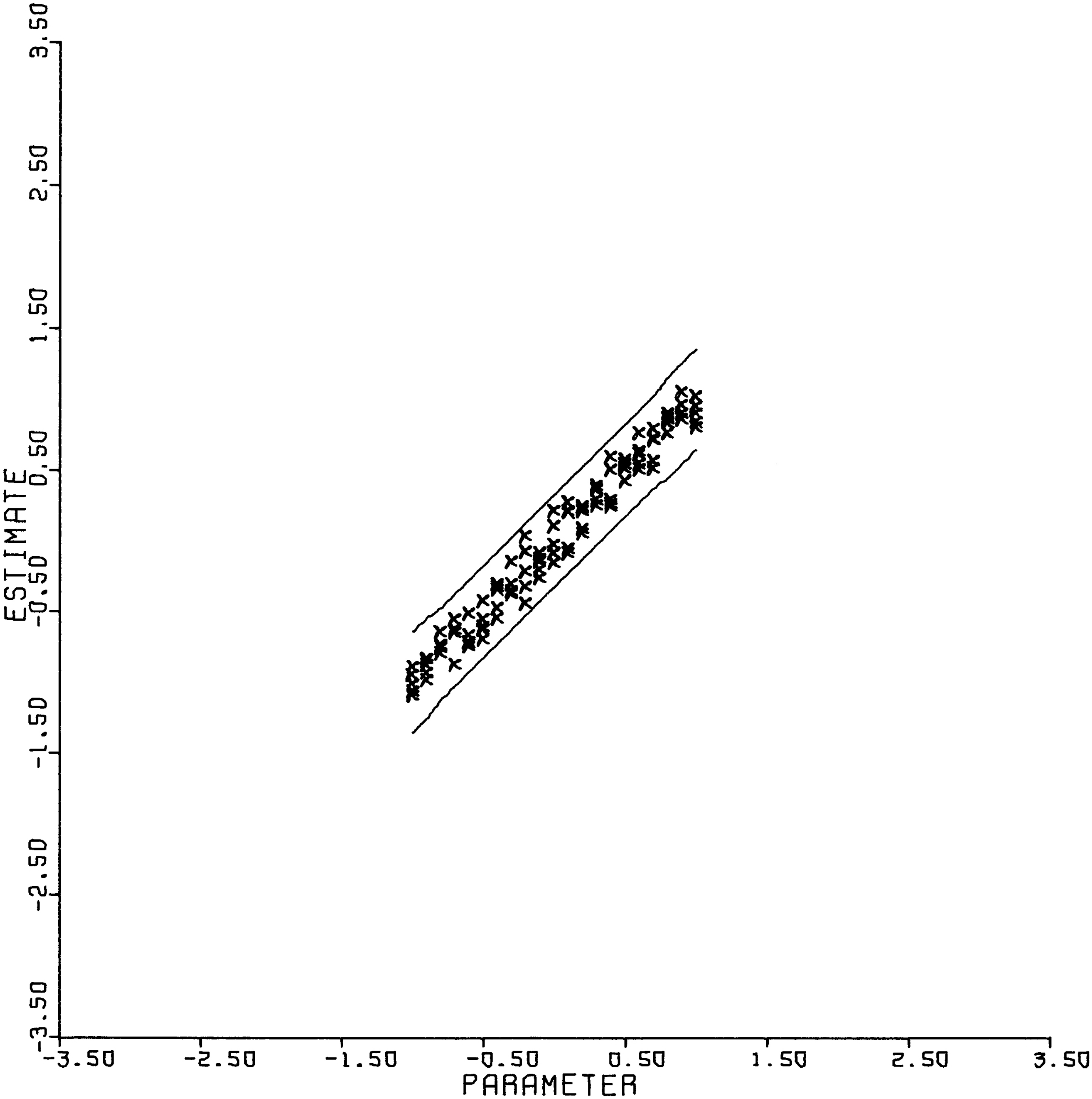

FIGURE 1

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=2 L=21 SAMPLE: M=0 S=0.5 N=400

REGRESSION: Y=1.00X X=0.97Y

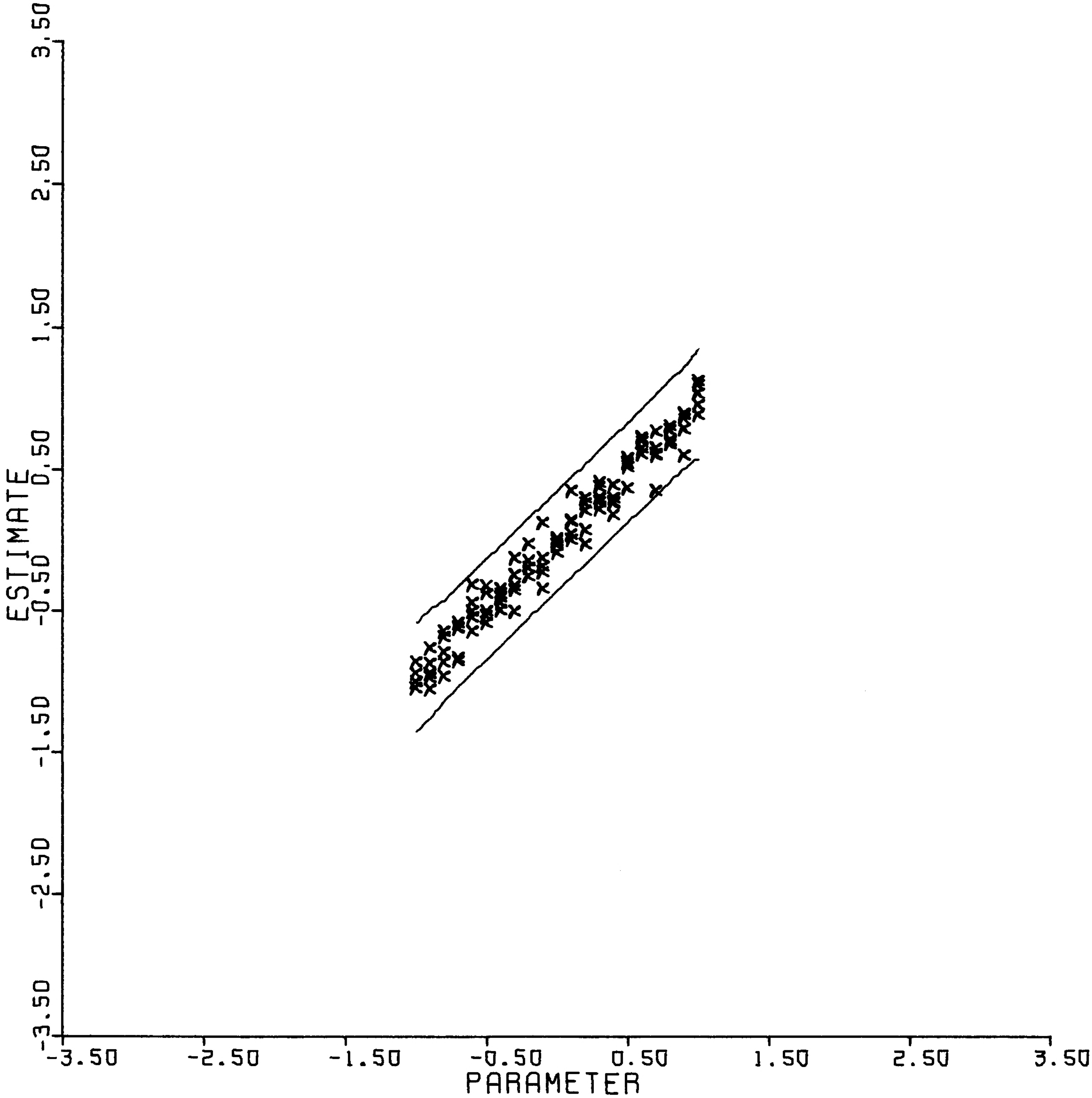

FIGURE 2

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=2 L=21 SAMPLE: M=0 S=1.0 N=400

REGRESSION: Y=0.97X X=1.00Y

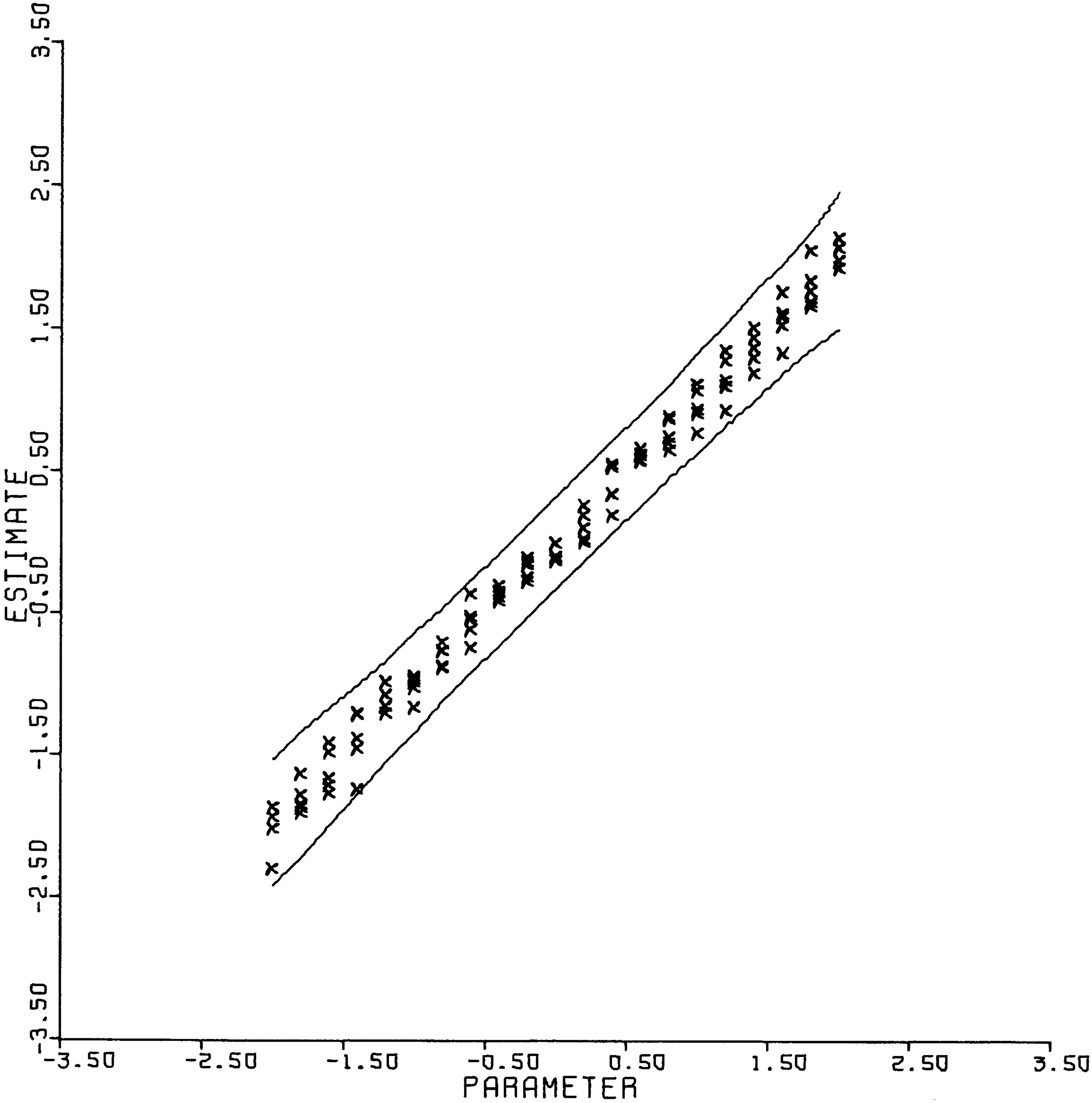

FIGURE 3

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=4 L=21 SAMPLE: M=0 S=0.5 N=400

REGRESSION: Y=0.99X X=1.00Y

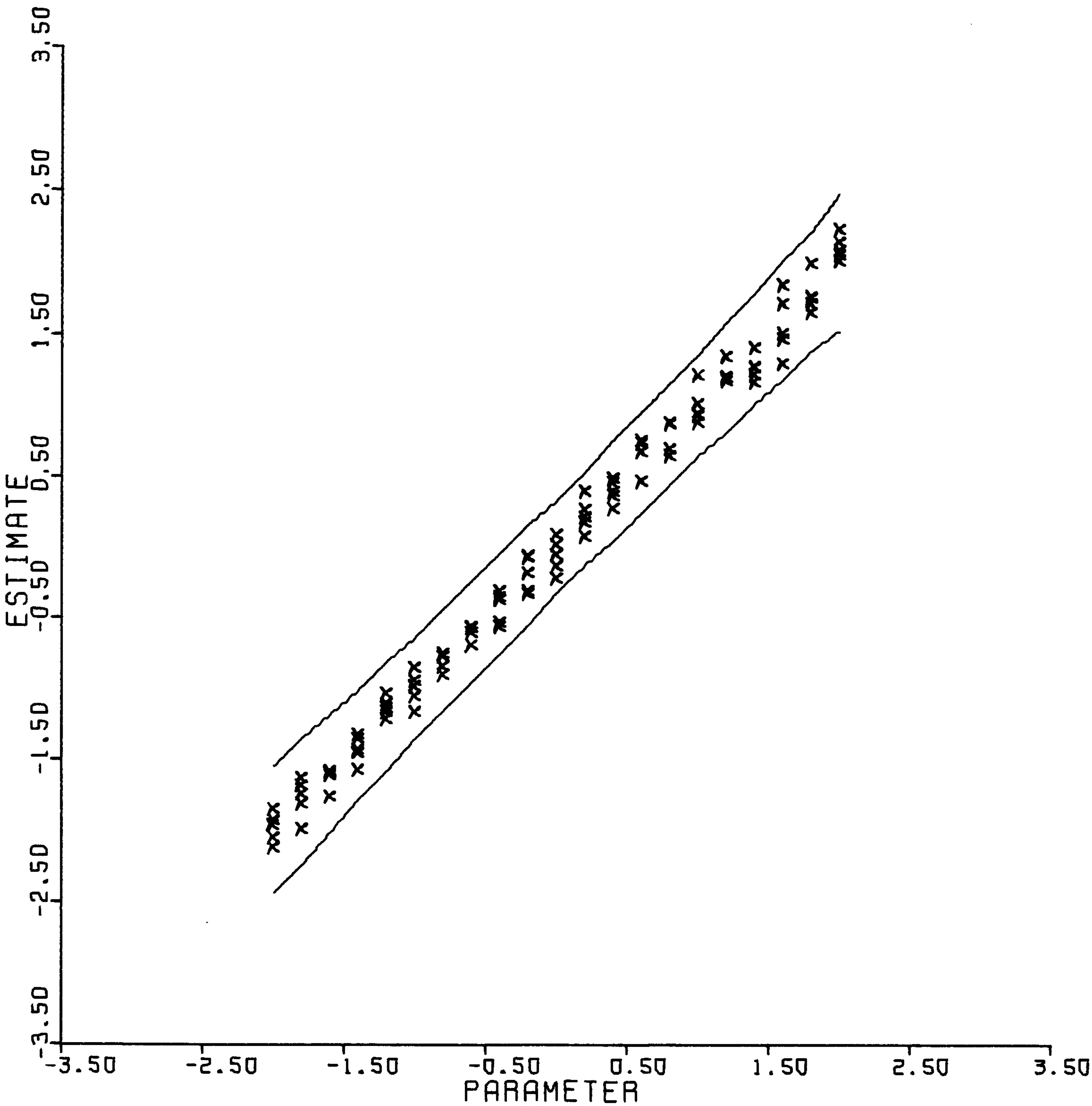

FIGURE 4

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=4 L=21 SAMPLE: M=0 S=1.0 N=400

REGRESSION: Y=1.00X X=0.99Y

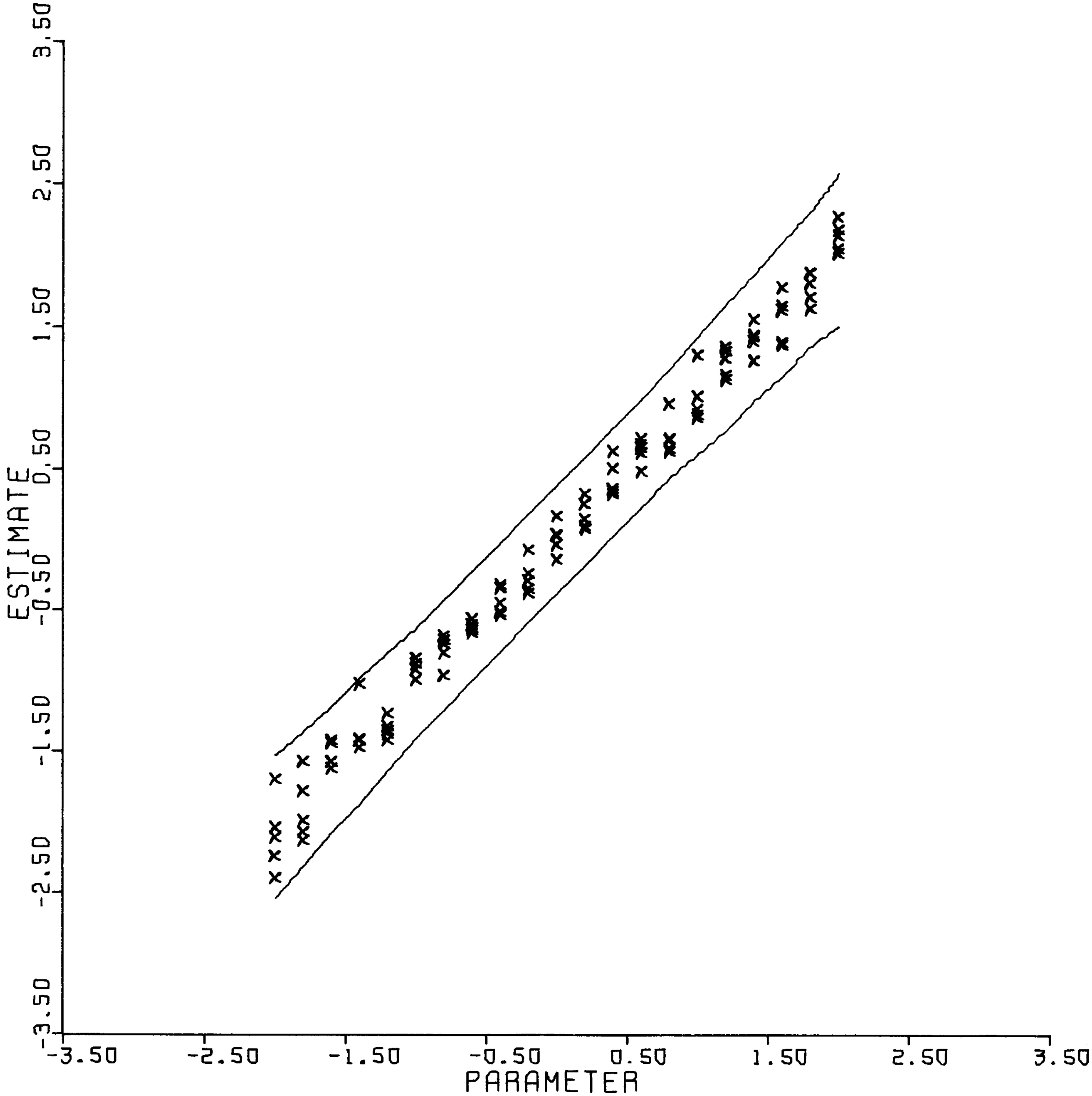

FIGURE 5

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=4 L=21 SAMPLE: M=0 S=1.5 N=400

REGRESSION: Y=1.02X X=0.97Y

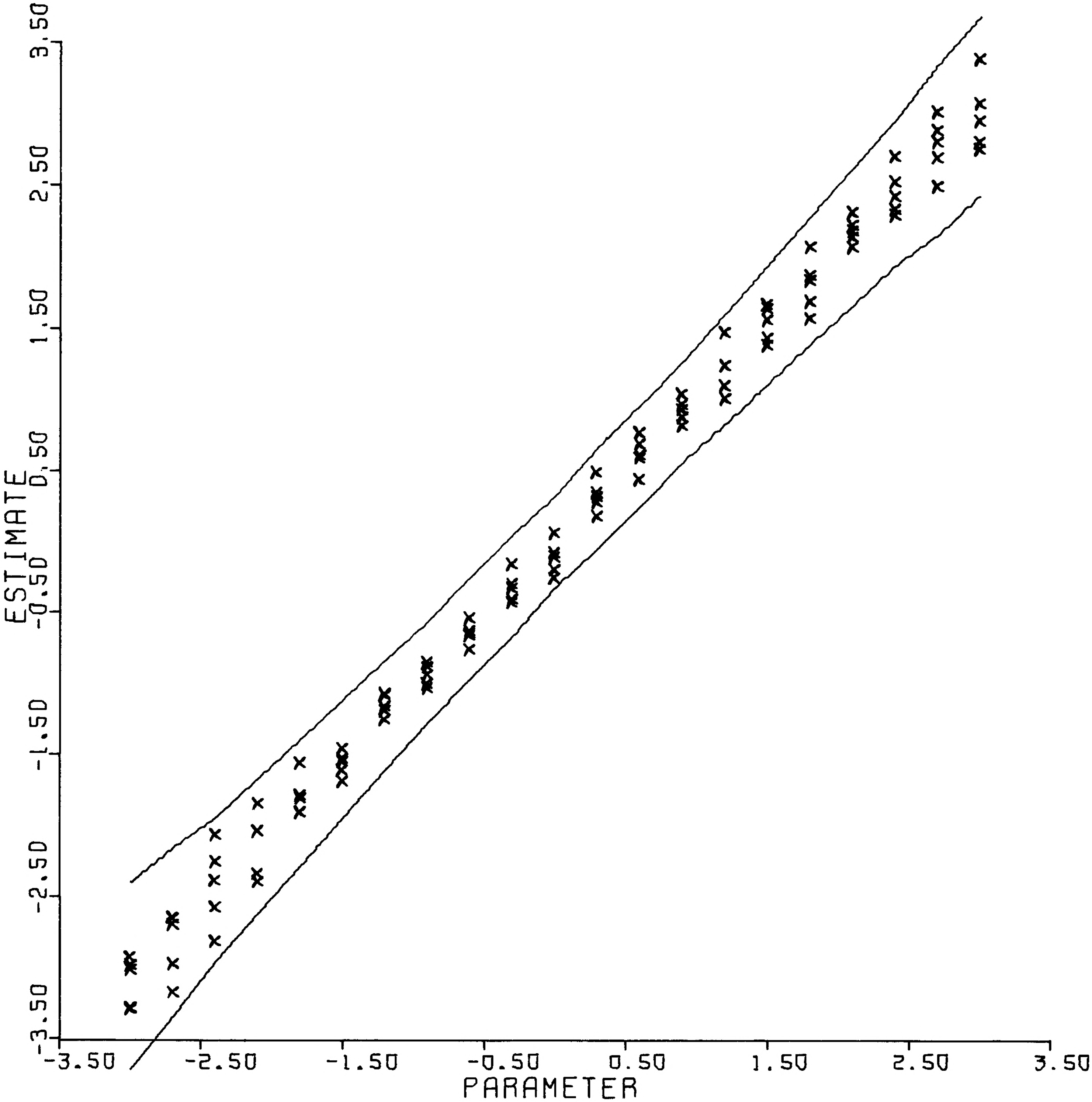

FIGURE 6

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=6 L=21 SAMPLE: M=0 S=1.0 N=400

REGRESSION: Y=1.02X X=0.97Y

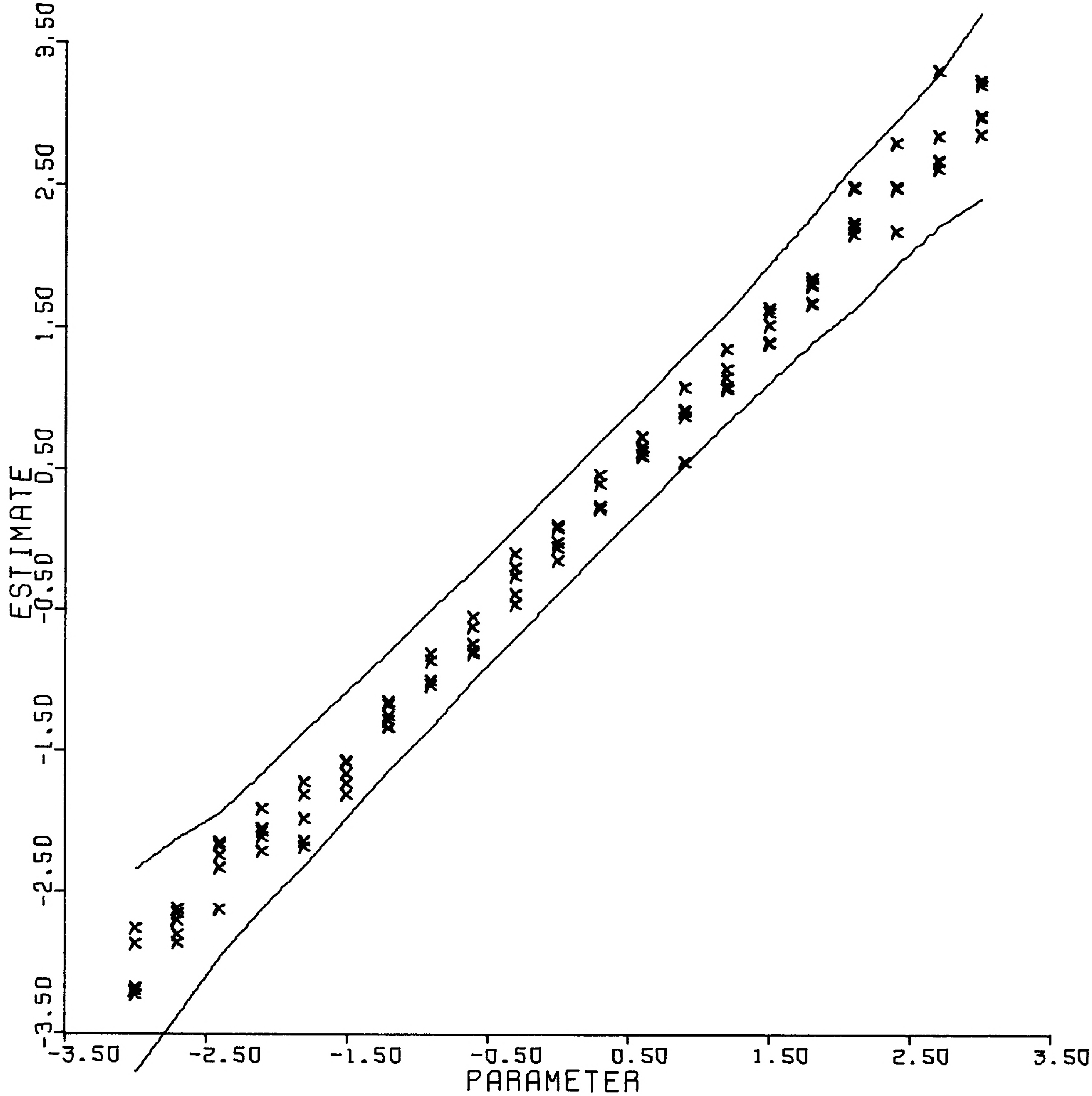

FIGURE 7

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=6 L=21 SAMPLE: M=0 S=1.5 N=400

REGRESSION: Y=1.02X X=0.97Y

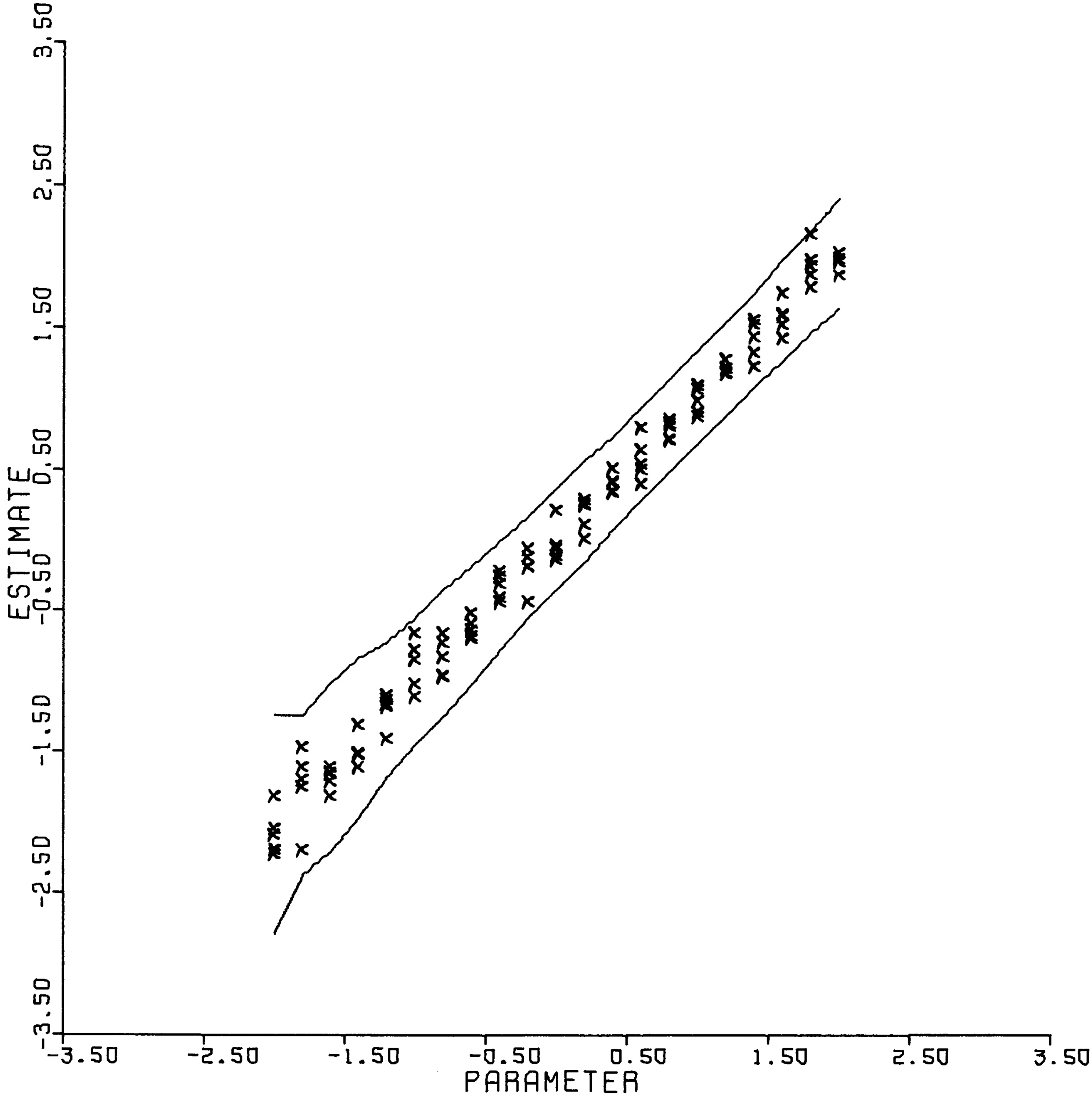

FIGURE 8

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=4 L=21 SAMPLE: M=1 S=0.5 N=400

REGRESSION: Y=1.01X X=0.98Y

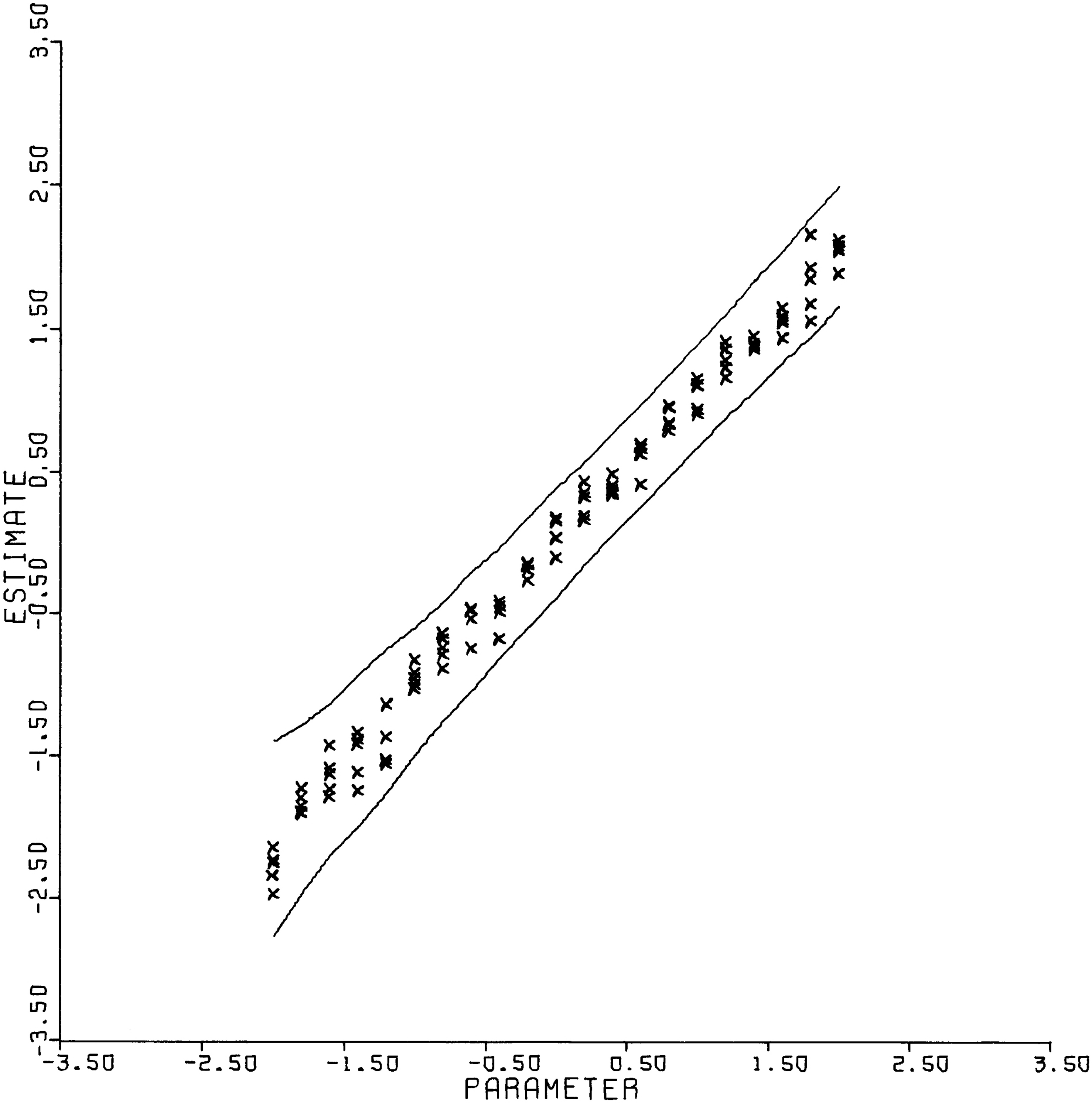

FIGURE 9

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=4 L=21 SAMPLE: M=1 S=1.0 N=400

REGRESSION: Y=1.04X X=0.95Y

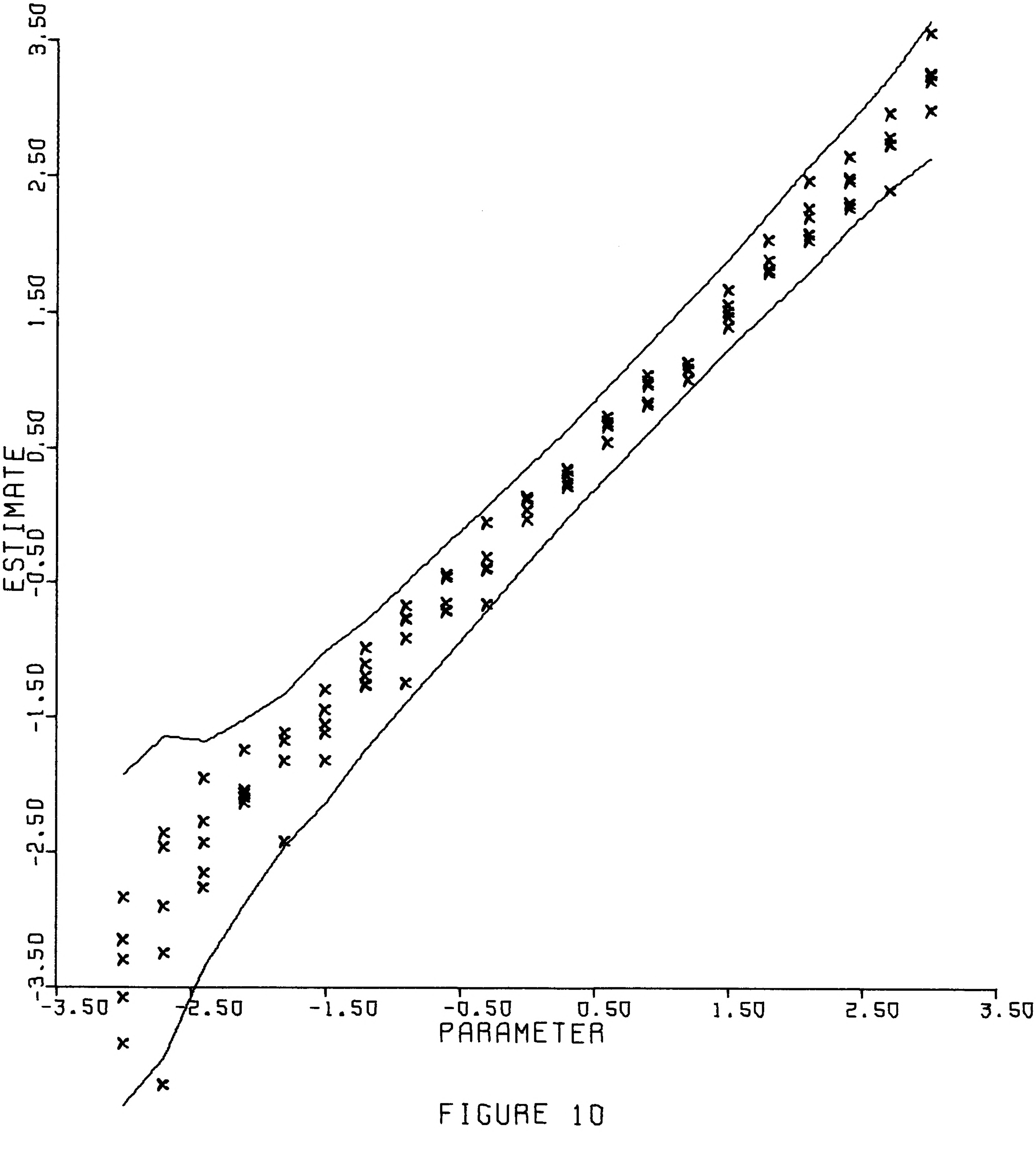

FIGURE 10

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=6 L=21 SAMPLE: M=1 S=0.5 N=400

REGRESSION: Y=1.05X X=0.94Y

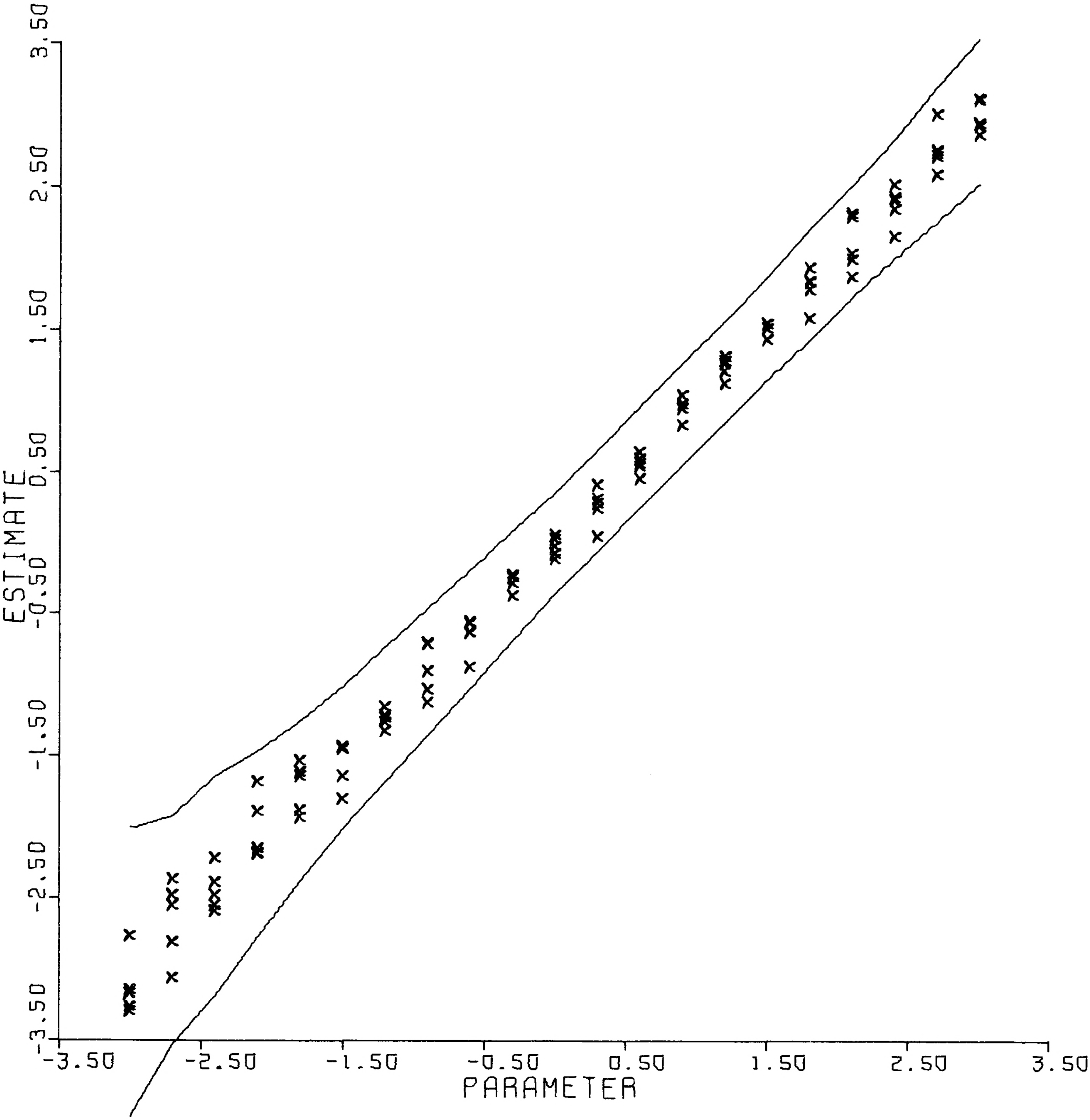

FIGURE 11

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=6 L=21 SAMPLE: M=1 S=1.0 N=400

REGRESSION: Y=1.01X X=0.99Y

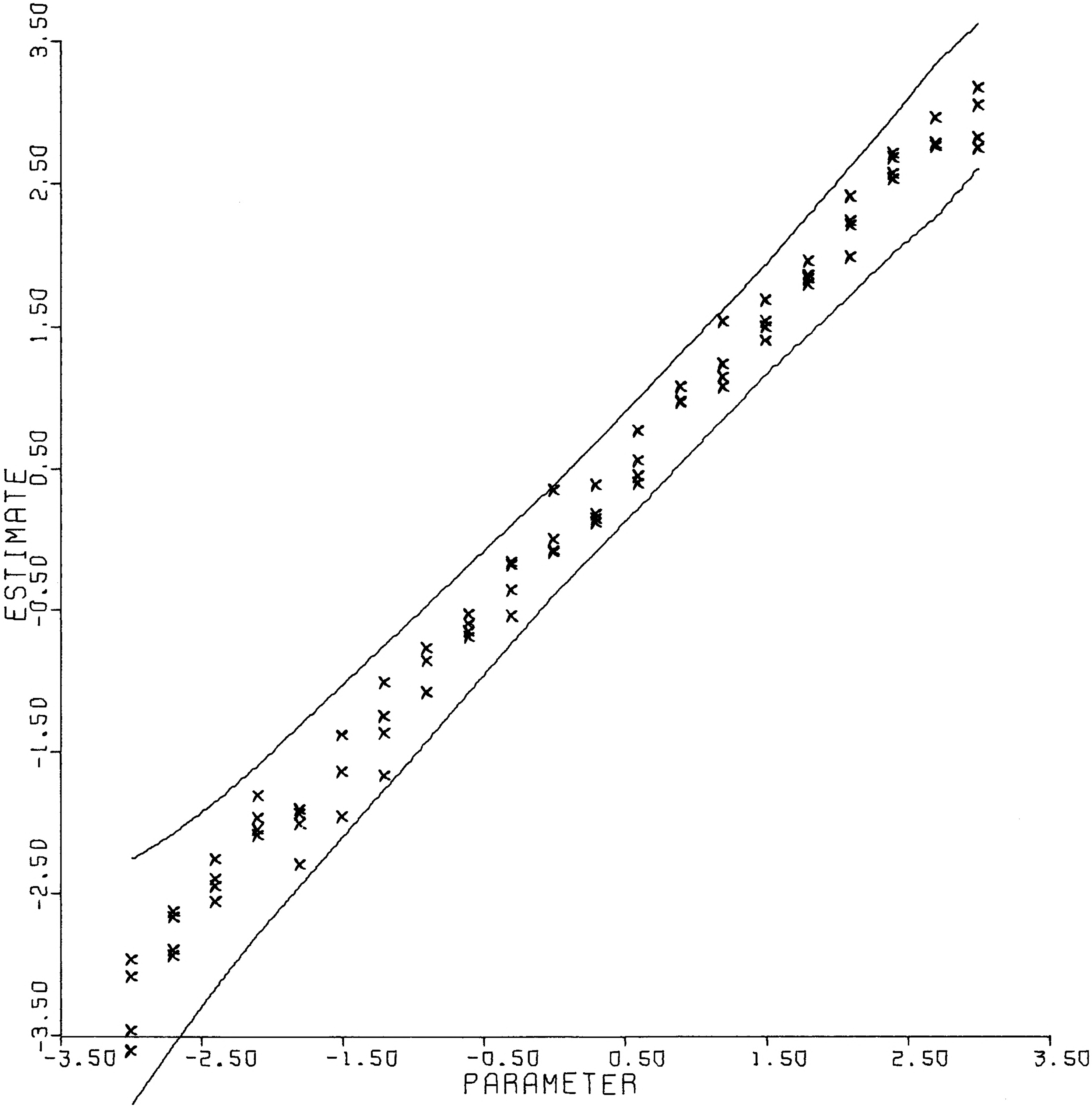

FIGURE 12

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=6 L=21 SAMPLE: M=1 S=1.5 N=400

REGRESSION: Y=1.04X X=0.95Y

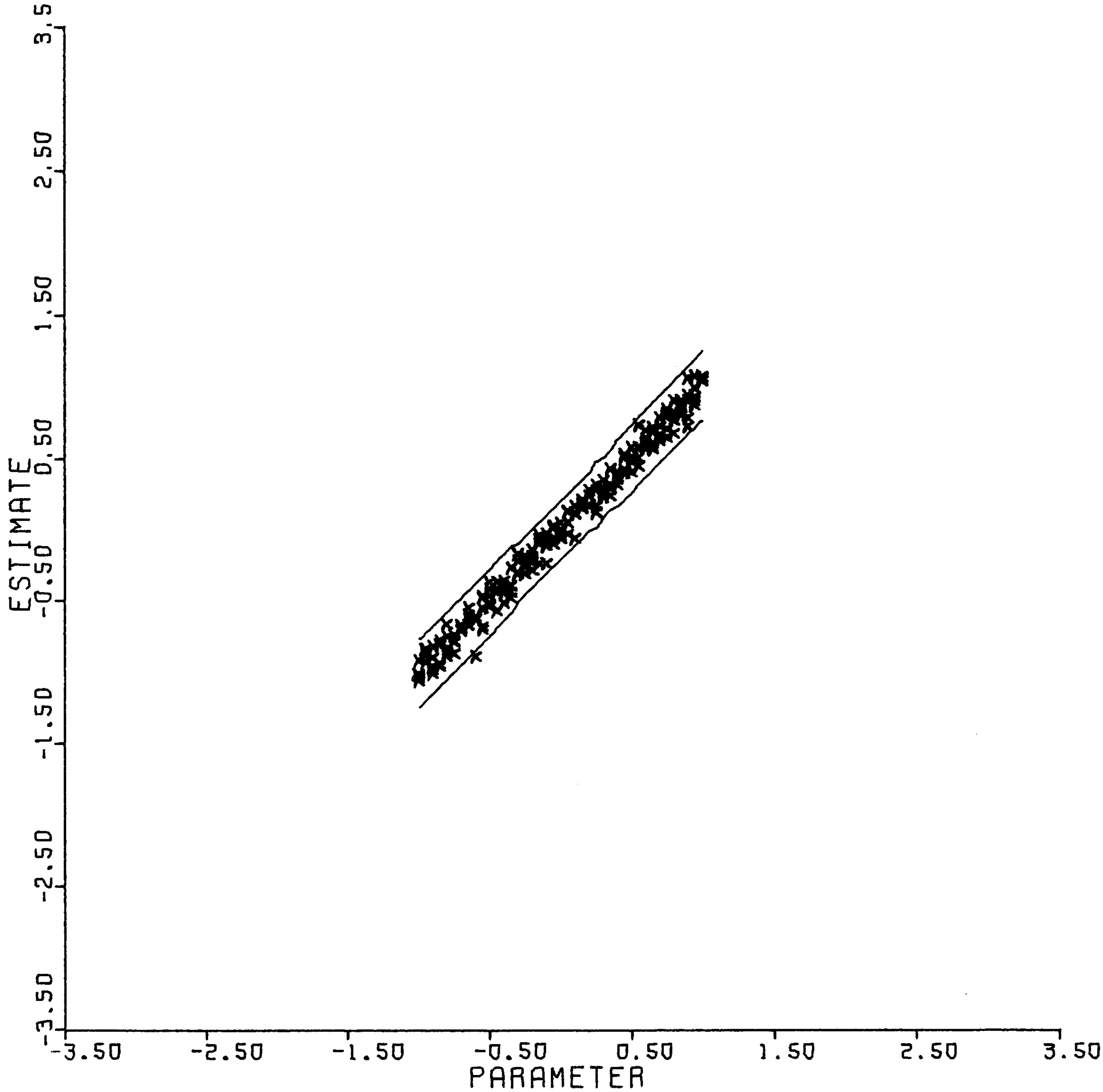

FIGURE 13

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=2 L=41 SAMPLE: M=0 S=0.5 N=800

REGRESSION: Y=1.01X X=0.98Y

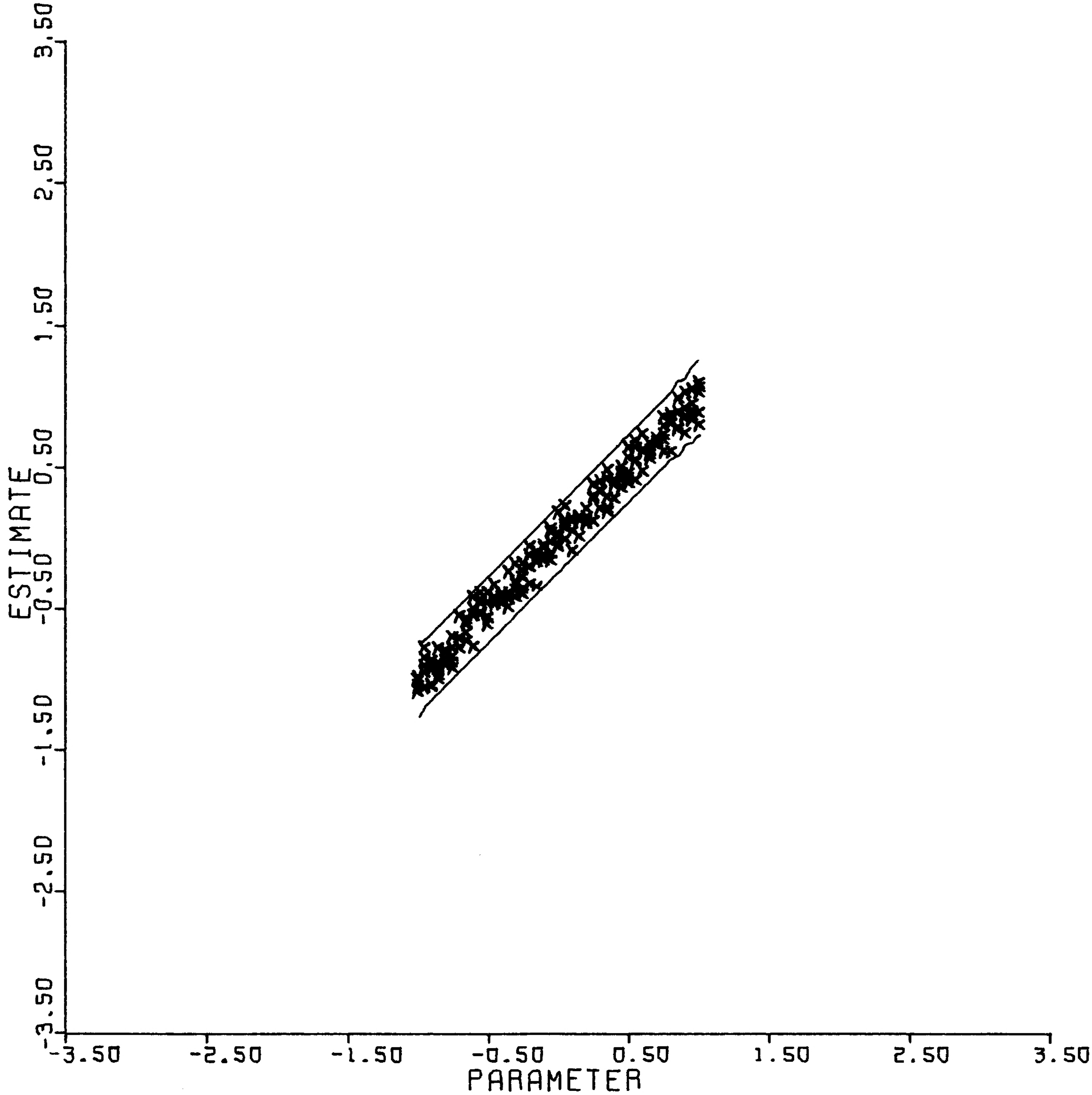

FIGURE 14

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=2 L=41 SAMPLE: M=0 S=1.0 N=800

REGRESSION: Y=1.00X X=0.98Y

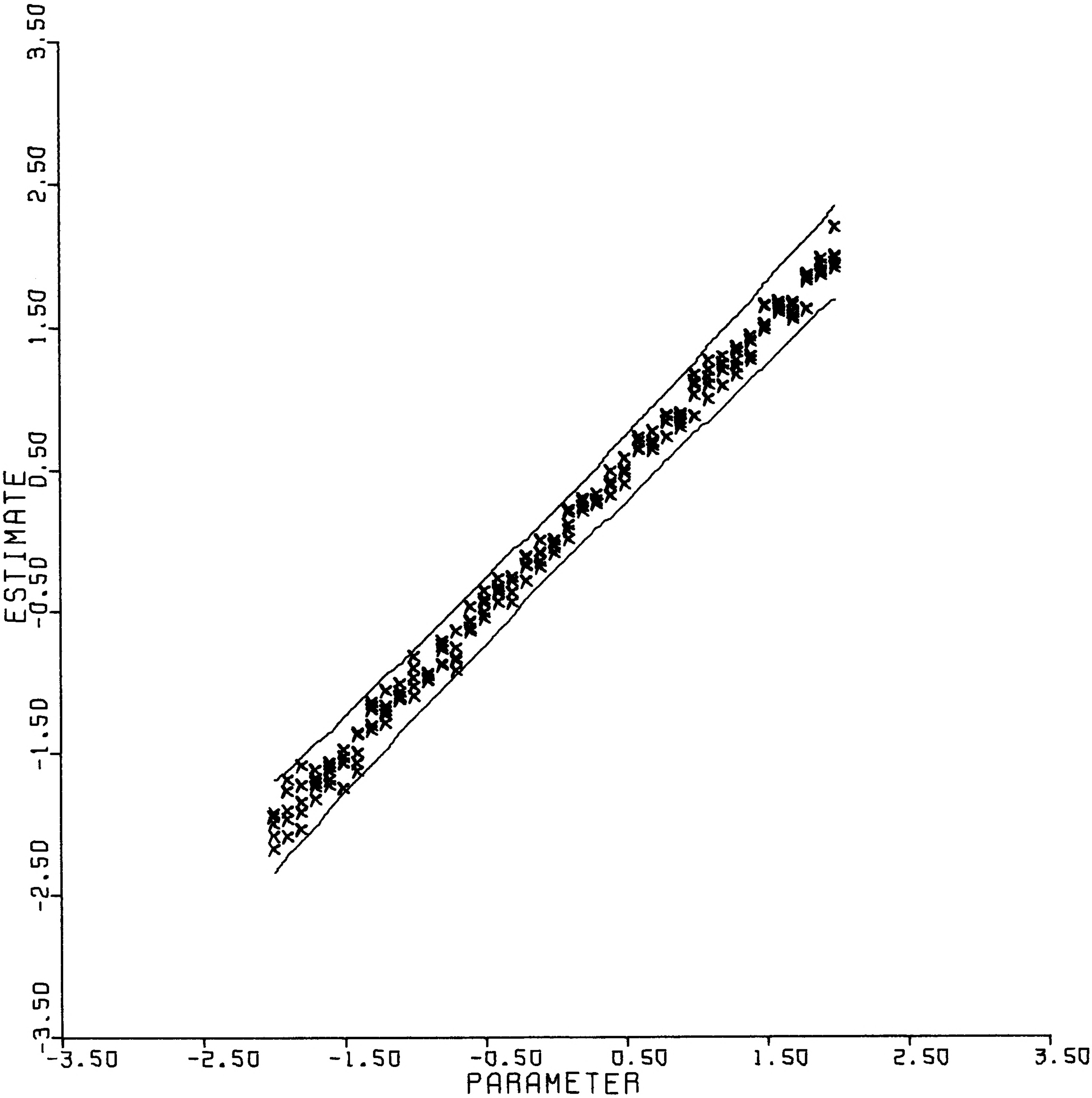

FIGURE 15

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=4 L=41 SAMPLE: M=0 S=0.5 N=800

REGRESSION: Y=1.01X X=0.99Y

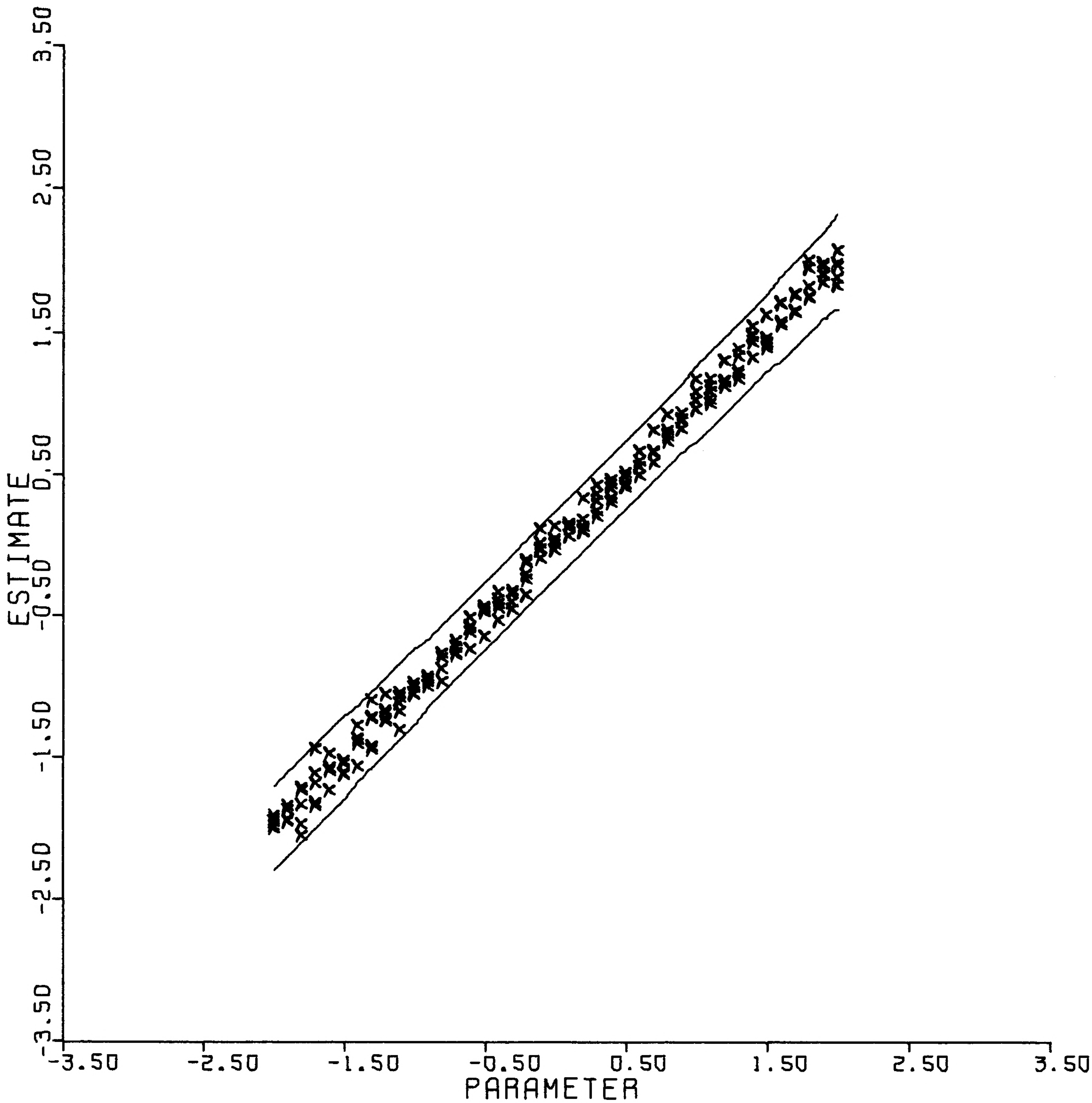

FIGURE 16

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=4 L=41 SAMPLE: M=0 S=1.0 N=800

REGRESSION: Y=1.00X X=0.99Y

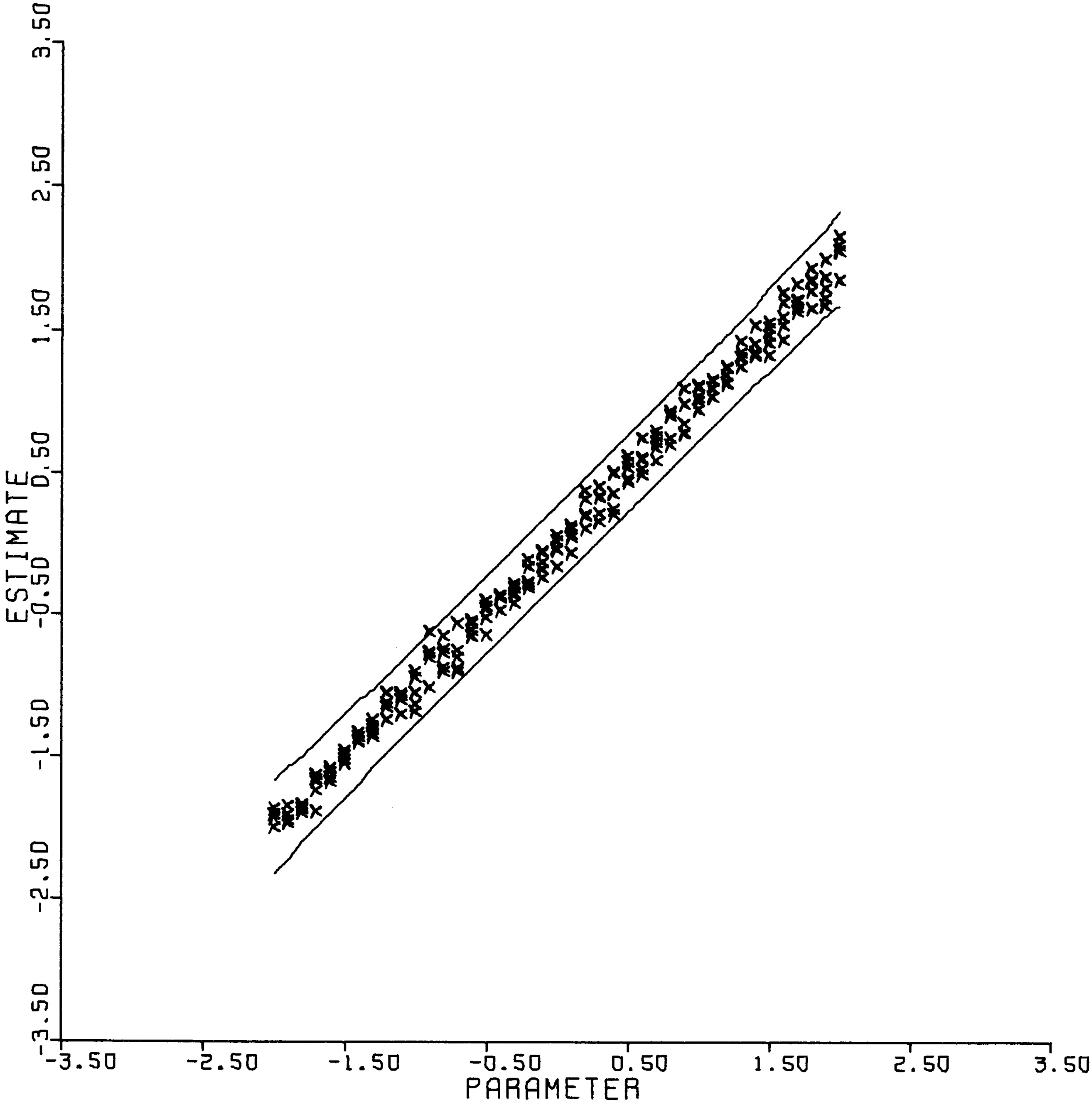

FIGURE 17

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=4 L=41 SAMPLE: M=0 S=1.5 N=800

REGRESSION: Y=1.00X X=0.99Y

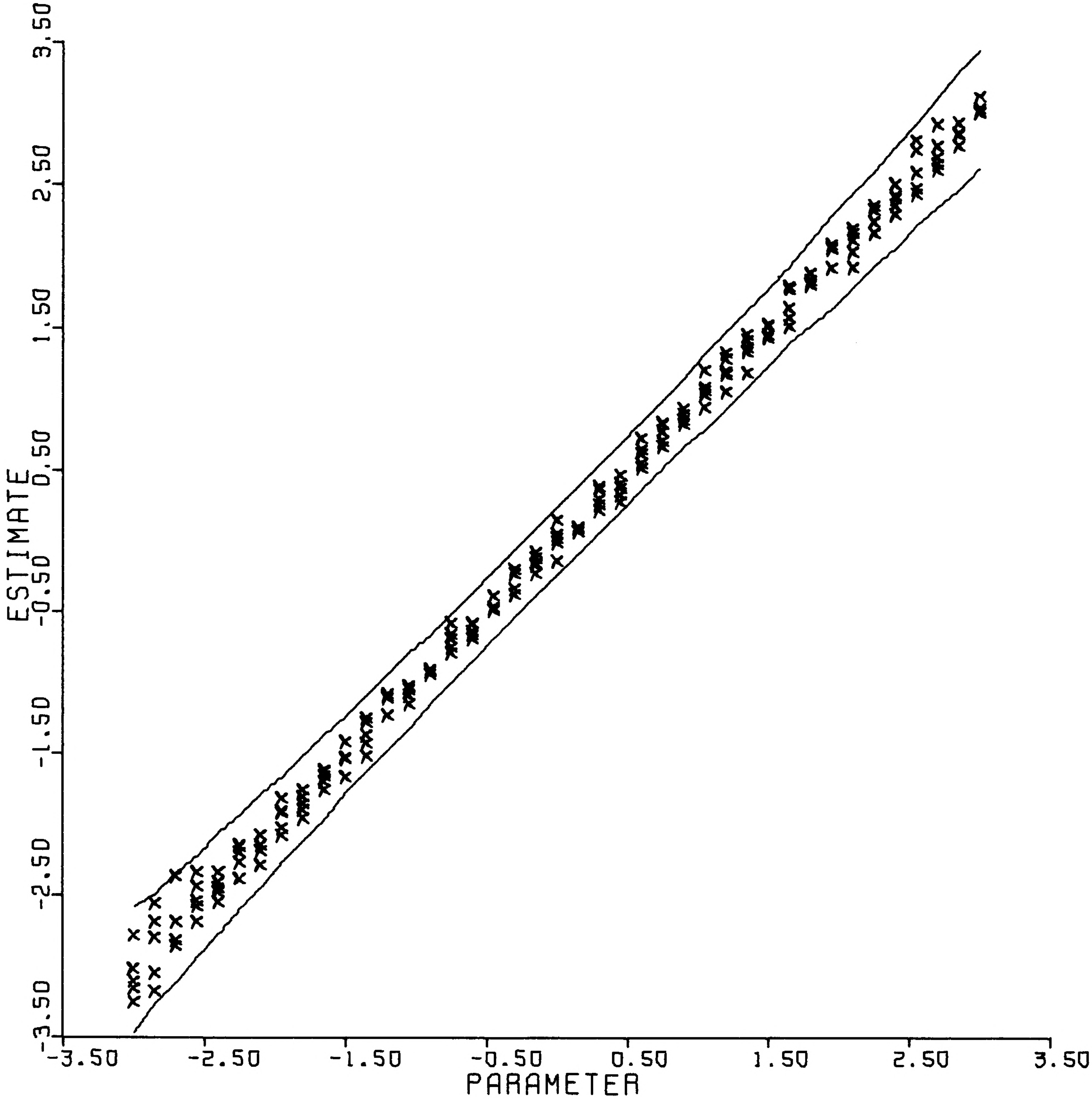

FIGURE 18

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=6 L=41 SAMPLE: M=0 S=1.0 N=800

REGRESSION: Y=1.01X X=0.99Y

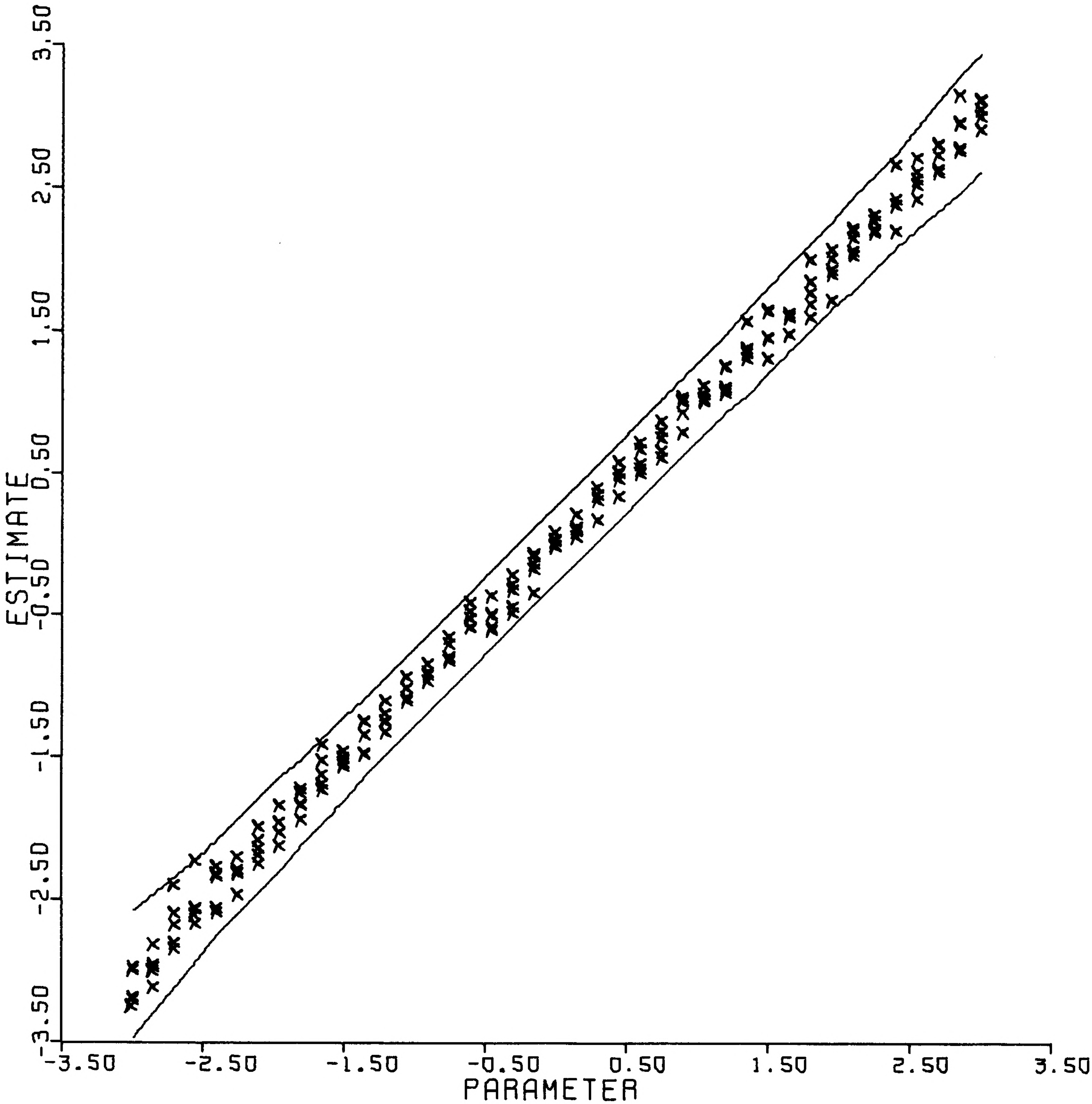

FIGURE 19

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=6 L=41 SAMPLE: M=0 S=1.5 N=800

REGRESSION: Y=1.01X X=0.99Y

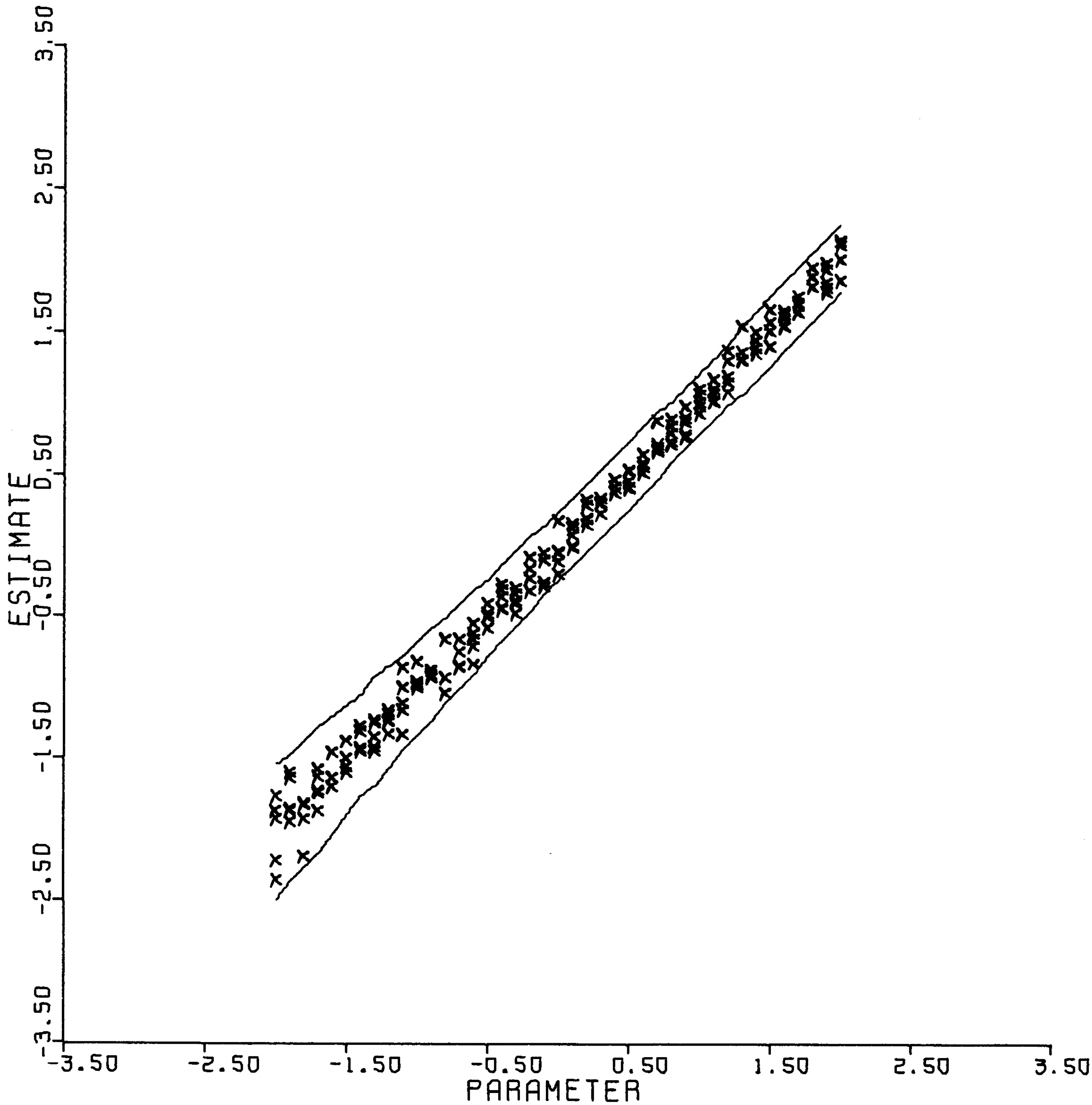

FIGURE 20

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=4 L=41 SAMPLE: M=1 S=0.5 N=800

REGRESSION: Y=1.01X X=0.98Y

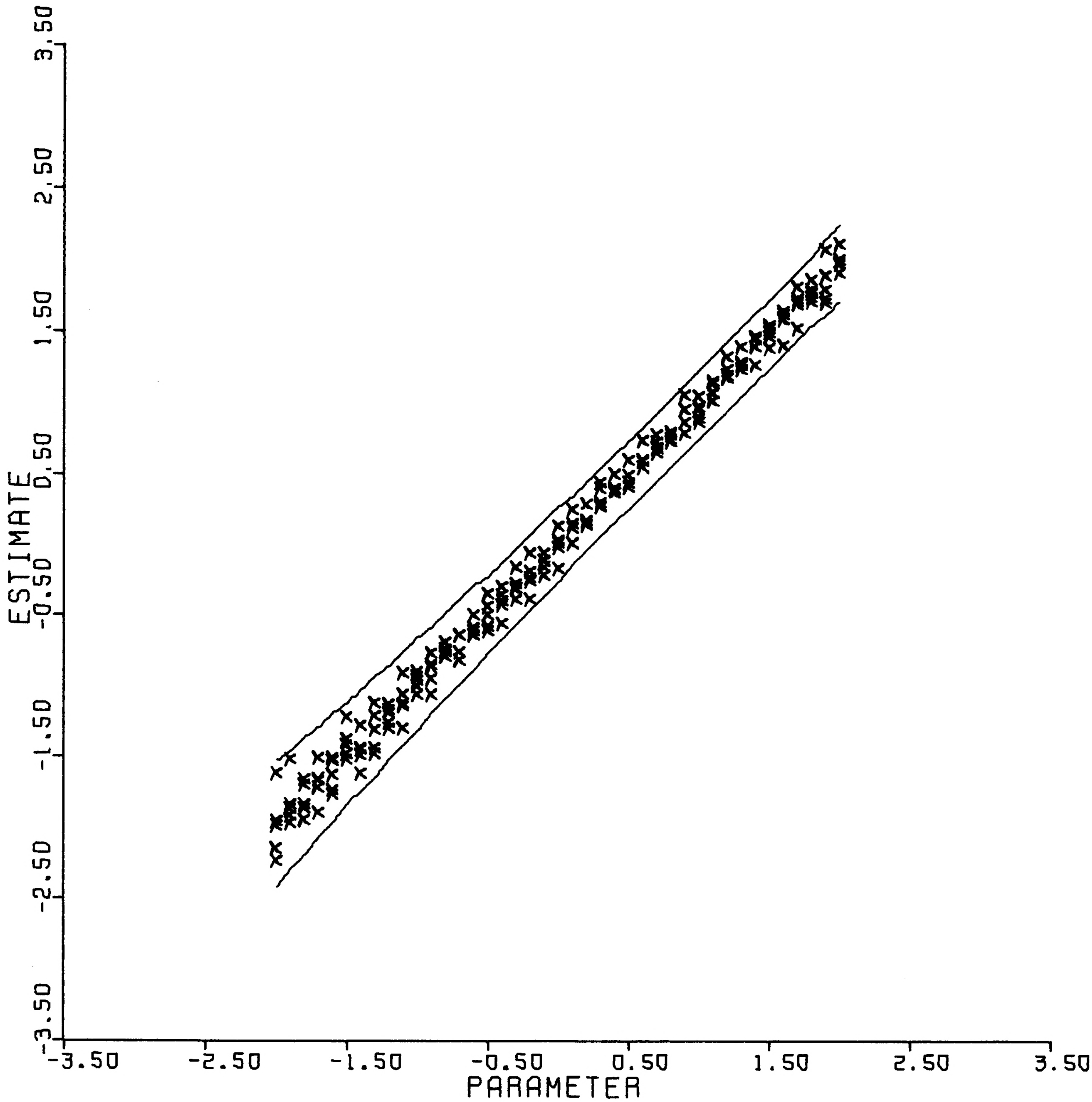

FIGURE 21

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=4 L=41 SAMPLE: M=1 S=1.0 N=800

REGRESSION: Y=0.99X X=1.00Y

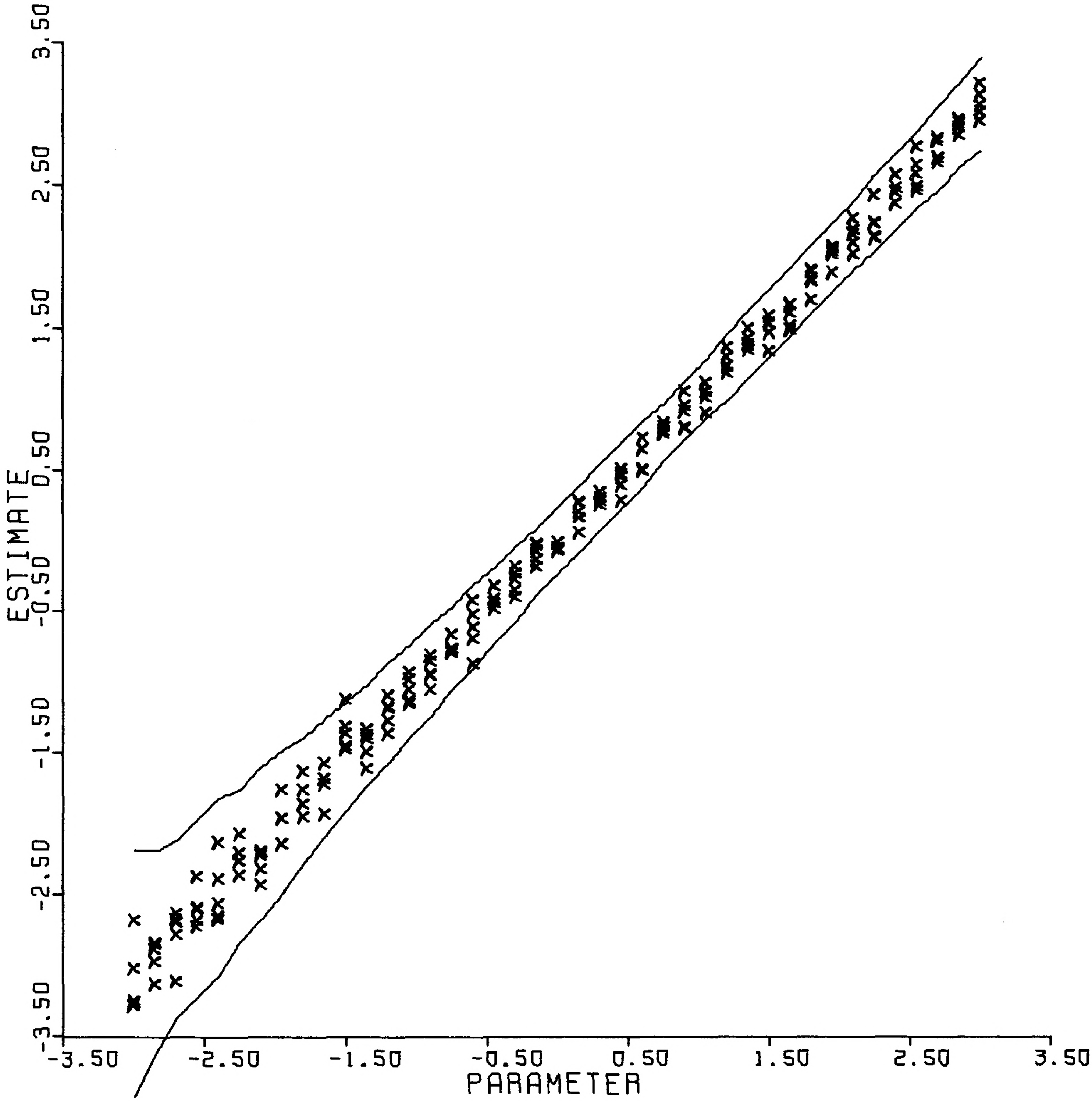

FIGURE 22

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=6 L=41 SAMPLE: M=1 S=0.5 N=800

REGRESSION: Y=1.02X X=0.98Y

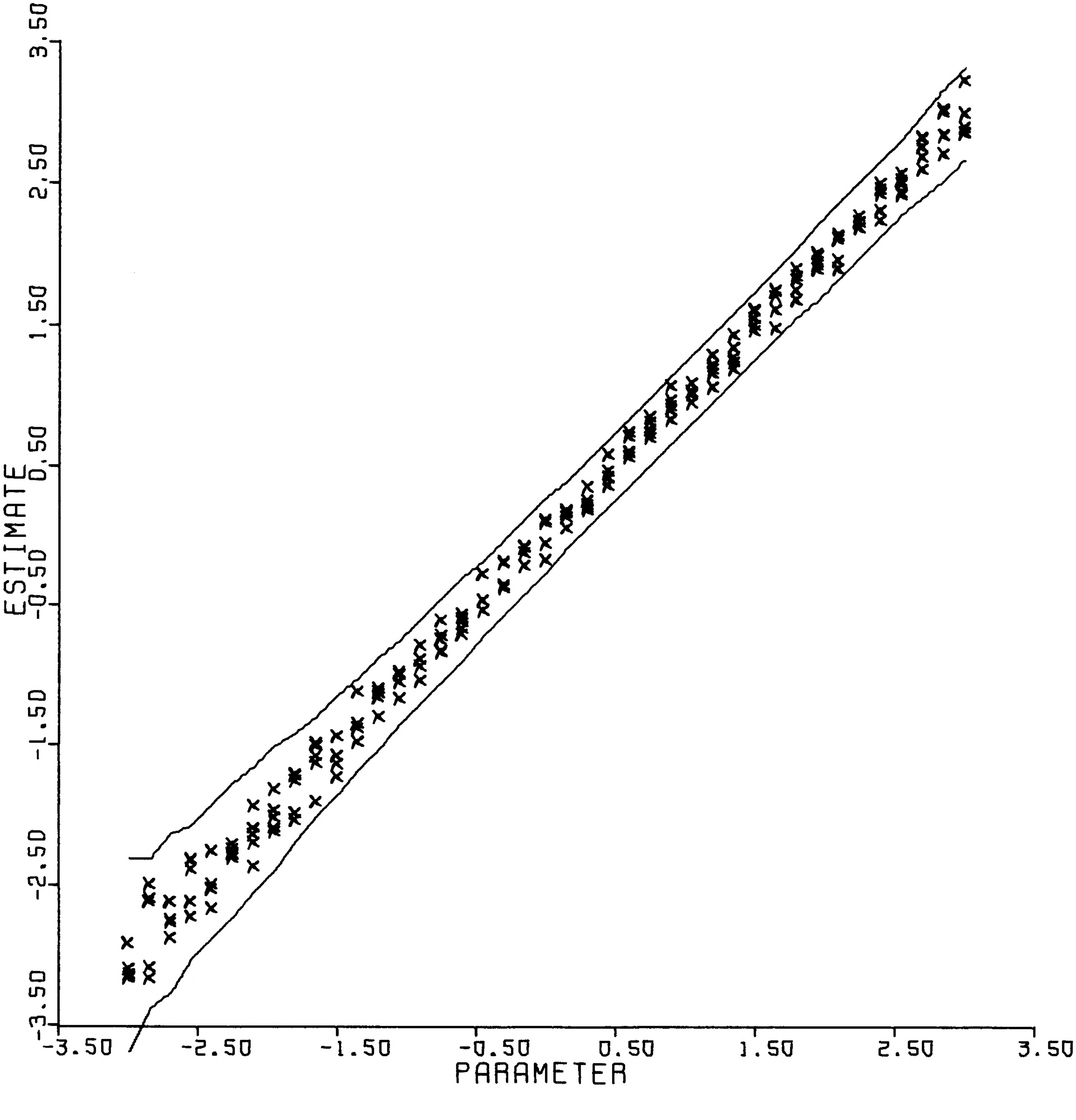

FIGURE 23

REGRESSION OF BICAL ESTIMATES ON THEIR PARAMETERS

TEST: H=0 W=6 L=41 SAMPLE: M=1 S=1.0 N=800

REGRESSION: Y=1.00X X=0.99Y