DOCUMENT RESUME

ED 285 891 TM 870 468

AUTHOR	Ackerman, Terry A.
TITLE	A Comparison Study of the Unidimensional IRT Estimation of Compensatory and Noncompensatory Multidimensional Item Response Data.
INSTITUTION	American Coll. Testing Program, Iowa City, Iowa.
SPONS AGENCY	Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE	Apr 87
CONTRACT	N00014-85-C-0241; NR153-531
NOTE	31p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Ability; Computer Simulation; *Difficulty Level; *Item Analysis; *Latent Trait Theory; *Mathematical Models; Test Items
IDENTIFIERS	BILOG Computer Program; LOGIST Computer Program; *Unidimensionality (Tests)

ABSTRACT

Concern has been expressed over the item response theory (IRT) assumption that a person's ability can be estimated in a unidimensional latent space. To examine whether or not the response to an item requires only a single latent ability, unidimensional ability estimates were compared for data generated from the multidimensional item response theory (MIRT) compensatory and noncompensatory models. It was hypothesized that as the correlation between the two dimensional abilities increased, the response data would essentially become unidimensional and thus the confounding of difficulty and dimensionality would have little effect in either model. This study examined the unidimensional estimates of matched compensatory and noncompensatory data in which difficulty was confounded with dimensionality for different levels of correlation between two dimensional abilities. Eight data sets, four compensatory, and four noncompensatory, were generated. Each set was calibrated using the IRT calibration programs LOGIST and BILOG. BILOG calibration of response vectors generated to the matched MIRT item parameters appeared to be more affected than LOGIST by the confounding of difficulty and dimensionality. As the correlation between the generated two dimensional abilities increased, the response data appeared to become more unidimensional. Six figures are included. (Author/MDE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED285891

A Comparison Study of the Unidimensional IRT Estimation of Compensatory and Noncompensatory Multidimensional Item Response Data

Terry A. Ackerman

The American College Testing Program

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
T.A. Ackerman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

□ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., April 1987. This research was supported by Contract number N00014-85-C-0241, NR153-531 from the Personnel and Training Research Programs of the Office of Naval Research.

Running Head: COMPENSATORY/NONCOMPENSATORY IRT ESTIMATION

BEST COPY AVAILABLE

TM870468

Abstract

The purpose of this study was to compare the characteristics of unidimensional ability estimates obtained from data generated from the multidimensional IRT (MIRT) compensatory and noncompensatory models. Reckase, Carlson, Ackerman and Spray (1986) reported that when the compensatory model is used and item difficulty is confounded with dimensionality, the composition of the unidimensional ability estimates differs for different points along the unidimensional ability scale. Eight data sets (four compensatory, four noncompensatory) were generated for four different levels of correlated two dimensional abilities: $\rho = 0, .3, .6, .9$. In each set difficulty was confounded with dimensionality. Each set was then calibrated using the IRT calibration programs LOGIST and BILOG. BILOG calibration of response vectors generated to the matched MIRT item parameters appeared to be more affected than LOGIST by the confounding of difficulty and dimensionality. As the correlation between the generated two-dimensional abilities increased, the response data appeared to become more unidimensional as evidenced in bivariate plots of $\bar{\theta}_1$ vs. $\bar{\theta}_2$ for specified $\hat{\theta}$ quantiles.

A Comparison Study of the Unidimensional IRT Estimation of Compensatory and Noncompensatory Multidimensional Item Response Data

One of the underlying assumptions of unidimensional item response theory (IRT) models is that a person's ability can be estimated in a unidimensional latent space. However, researchers and educators have expressed concern whether or not the response process to any one item requires only a single latent ability. Traub (1983) suggests that many cognitive variables are brought to the testing task and that the number used varies from person to person. Likewise, the combination of latent abilities required by individuals to obtain a correct response may vary from item to item. Caution over the application of unidimensional IRT estimation of multidimensional response data has been expressed by several researchers including Ansley and Forsyth (1985); Reckase, Carlson, Ackerman, and Spray (1986); and, Yen (1984).

Using a compensatory multidimensional IRT (MIRT) model, Reckase et al. (1986) demonstrated that when dimensionality and difficulty are confounded (i.e., easy items discriminate only on θ_1, , difficult items discriminate cnly on θ_2) the unidimensional ability scale has a different meaning at different points on the scale. Specifically, for their two- dimensional generated data set, upper ability deciles differed mainly on θ_2 while the lower deciles differed mostly on θ_1. These results led the authors to suggest that the univariate calibration of two-dimensional response data can be explained in terms of the interaction between the multidimensional test information and the distribution of the two-dimensional abilities. Reckase et al. (1986) examined the condition in which ability estimates were uncorrelated. Such an approach may not be very realistic, however, since most cognitive abilities tend to be correlated.

Ansley and Forsyth (1985) examined the unidimensional estimates from two-dimensional data generated using a noncompensatory model (Sympson, 1978). Ansley and Forsyth (1985) selected item parameters so that generated response data would match item difficulty parameters as taken from a "real" test. They examined situations in which abilities were correlated .0, .3, .6, .9, and .95. Although the issue of confounding dimensionality with difficulty may have occurred it was not addressed. The researchers found the $\hat{a}$ values were "best considered" as averages of the true a_1 and a_2 values; that the $\hat{b}$ values were "overestimates of b_1", and that θ's were "highly related" to the average of the true θ values.

It was the purpose in this paper to extend this work by examining the unidimensional estimates of matched compensatory and noncompensatory data in which difficulty is confounded with dimensionality for different levels of correlation between two dimensional abilities. Two main issues were examined. The first area of focus was to investigate differences between the two MIRT models when difficulty was confounded with dimensionality. That is, could the results of the Reckase et al. study be replicated for both models. The second issue was to determine if different levels of correlation between the two dimensional abilities had any affect on the confounding of difficulty and dimensionality under each model. It was hypothesized that as the correlation between θ_1 and θ_2 increased, the response data would essentially become unidimensional, and thus the confounding of difficulty and dimensionality would have little effect in either model.

Model Definition

A compensatory model, M2PL, Reckase (1985) was used for specification of compensatory items. The model defines the probability of a correct response as:

$$P(x_{ij} = 1 | \underset{\sim}{a}_i, d_i, \underset{\sim}{\theta}_j) = \frac{1}{1 + \exp\left[- \sum_{k=1}^{n} a_{ik}\theta_{jk} + d_i \right]}$$

where x_{ij} is the response to item i by person j,

θ_{jk} is the ability parameter for person j on dimension k,

a_{ik} is the discrimination parameter for item i on dimension k,

d_i is the difficulty parameter for item i.

The probability of a correct response for the noncompensatory model proposed by Sympson (1978) is:

$$P_{ij}\,(X_j = 1 | \theta_{in}) = c_j + \frac{1 - c_j}{\prod_{n=1}^{n} \left(1 + \exp\left[-1.7a_{jn}\,\{\theta_{in} - b_{jn}\}\right]\right)}$$

where b_{jn} is the difficulty of item j in dimension n. For this study, c_j, the guessing parameter, was set to zero.

Method

To test the effects of correlated ability dimensions, four levels of correlation were selected ρ = .0, .3, .6, and .9).

Parameters for a set of 40 two-dimensional compensatory items were selected with difficulty and dimensionality confounded. Discrimination parameters ranged from $a_1 = 1.8$, $a_2 = .2$ to $a_1 = .2$, $a_2 = 1.8$. Difficulty was confounded with dimensionality such that the difficulty parameters ranged from $d = -2.4$ (for $a_1 = 1.8$, $a_2 = .2$) to $d = 2.4$ (for $a_1 = .2$ and $a_2 = 1.8$). Thus as the items became more difficult, they discriminated less along θ_2 and more along θ_1. The guessing parameter was set to zero because there was concern over how much "noise" would be added to the multidimensional data with a nonzero guessing parameter.

An item vector plot (See Reckase, 1985) representing the distance and direction from the origin to the point of maximum slope (discrimination) is shown in Figure 1. The longer a vector is in the third quadrant the easier the item, and the longer a vector in the first quadrant, the more difficult the item.

Corresponding noncompensatory items (same probability of a correct response) were created using a least squares approach to minimize the difference.

Insert Figure 1 about here

$$\sum((P_C|\underset{\sim}{\theta}, \underset{\sim}{a}, d) - (P_{NC}|\underset{\sim}{\theta}\ \underset{\sim}{a}\ \underset{\sim}{b}))^2$$

where P_C is the compensatory model's probability of correct response;
P_{NC} is the noncompensatory model's probability of correct response.

and $\underset{\sim}{\theta}$ is a vector of two dimensional abilities generated from a bivariate normal distribution.

Four noncompensatory item sets corresponding to the four levels of correlation among the two ability dimensions ($\rho_{\theta_1\theta_2}$ = .0 .3, .6 and .9) were created.

Item difficulties for each item (d_j for the 40 compensatory items and b_1, b_2 for the 40 noncompensatory items) for the $\rho_{\theta_1\theta_2}$ = 0.0 case are plotted in Figure 2a. It is interesting to compare the two sets. The selected d_j values are positively related to the item number. In the noncompensatory items, b_2 is highest for item 1 and decreases steadily as the item number increases. Difficulties for dimension 1 do not vary greatly over the item set.

Insert Figure 2a, b about here

The discrimination parameters for both the compensatory and noncompensatory items for the $\rho_{\theta_1\theta_2}$ = 0.0 case are displayed for each item in Figure 2b. The a_1 parameters for each model are greatest in item 1 and decrease with item number. The a_1 parameter is greater for the noncompensatory model for all items and decreases at a slower rate than its compensatory counterpart. The a_2 parameters for each model are lowest for the first item and greatest for the last item. The a_2 parameters constantly increase with item number.

To help understand how the probability of a correct response differs in each model, several item response surfaces (IRS) and corresponding contour plots for matched items are presented in Figure 3. The IRS and its corresponding contour plot are shown for items 1, 20, and 40 for both the compensatory and matched noncompensatory models. Little difference exists between the IRS for each model when the item discriminates only along θ_2 (Figure 3a) and only along θ_1 (Figure 3c). However, when both discriminate equally along θ_1 and θ_2 (Figure 3b) the noncompensatory equiprobability curves contrast the parallel lines of equiprobability of the compensatory item.

Insert Figures 3a, b, c about here

Multidimensional test information plots (INFLINE, see Reckase, 1985) for two sets of match item parameters are shown in Figures 4a and 4b. For both sets little information is provided for examinees with extremely high or extremely low ability on both dimensions. In general, more information is provided by the set of compensatory items. However, there are some isolated areas where this is not true.

Insert Figures 4a and b about here

Eight response data sets were then produced. Using the compensatory item parameters, 1,000 response vectors were simulated for each of four correlational values ($\rho_{\theta_1\theta_2}$ = 0, .3, .6, .9) from a bivariate normal. For each set of noncompensatory item parameters, 1,000 response vectors were generated using the same (θ_1, θ_2) combinations as produced the compensatory response data sets.

Descriptive statistics were then obtained for each of the eight data sets. This was done to validate the similarities in item difficulty and to show the dimensionality of the data. These results are displayed in Table 1.

Insert Table 1 about here

The eight item response sets have the same mean difficulty, with the range of p values also similar. The mean biserials for compensatory and noncompensatory item sets appear to be more similar as the correlation between abilities increase. As the mean biserials increase, the KR-20 reliability coefficient also increase. Eigenvalues of the principal component analysis of the inter-item tetrachoric matrix were computed. Evidence of multidimensionality can be seen by forming a ratio of the first to the second eigenvalue, $\lambda_1|\lambda_2$ (See Hambleton & Murray, 1983). As the correlation between the abilities increases, the ratio increases suggesting more dominant first principal component and that at $\rho_{\theta_1\theta_2}$ = .9 the data are almost unidimensional.

Each dataset was then calibrated twice, once using LOGIST (Wingersky, Barton, & Lord, 1982), and again using BILOG (Mislevy & Bock, 1982). The two IRT calibration programs use different estimation procedures. LOGIST uses joint maximum likelihood estimation. The default method of scoring subjects was selected for all BILOG computer runs. The default method of scoring was expectation a posteriori using a normal N(0, 1) Bayesian prior. The default priors were also used in the item parameter calibration: a log-normal prior on the discrimination estimates and no prior on the difficulty estimates.

These data were then evaluated to determine the ef. ct of confounding difficulty with dimensionality for both the compensatory and noncompensatory item sets. In addition, the effects of correlation between ability dimensions was studied.

Results

To estimate the LOGIST and BILOG orientation in the two dimensional ability plane, the ability estimates from each calibration run were first rescaled to the compensatory ability estimates for the $\rho_{\theta_1\theta_2} = 0.0$ case. The $\hat{\theta}$ for each calibration run were rank ordered and divided into twenty quantiles. The mean of the θ_1 and θ_2 parameters for each quantile were then calculated and plotted. These CENTROID plots were then examined to see if there was any curvelinearity suggesting that the composite (θ_1, θ_2) combination was not uniform across the univariate scale as predicted by the Reckase et al (1986) study.

The centroid plots for the LOGIST calibration of the four compensatory and four noncompensatory data sets are shown in Figures 4a and b. The BILOG counterparts are presented in Figures 5a and b.

Insert Figures 4a and b, 5a and 6 about here

The LOGIST orientation appears to be similar for each level of correlation and for each type of MIRT model. The BILOG centroids are noticeably more variable. For the BILOG centroids, as $\rho_{\theta_1\theta_2}$ approaches zero, the plot of the centroids increase in curvature. Thus, BILOG appears to be more sensitive to the confounding of difficulty and dimensionality. When the ability correlation is .9, the centroids for both calibration programs are almost linear. This is somewhat predictable because if the abilities are highly correlated, their response data would be expected to be unidimensional.

The correlations between $\hat{\theta}$ (univariate estimate) and each of the two abilities (θ_1 and θ_2) and the mean absolute difference (MAD) between $\hat{\theta}$ and each of θ_1 and θ_2 are shown in Tabl 2. Compared to the centroid plots, the data are much more alike for compensatory and noncompensatory data sets and for LOGIST estimates compared to BILOG's estimates. It is interesting to note that the univariate ability estimates correlate about equally with θ_1 and θ_2 for all levels of ability correlation and for each model. The correlations between θ_1 and $\hat{\theta}$, and θ_2 and $\hat{\theta}$ range from .59 ($\rho_{\theta_1\theta_2}$ = 0) to .95 ($\rho_{\theta_1\theta_2}$ = .9). These results parallel the mean absolute differences: as the correlation between ability dimensions increase the MAD values d crease. Thus as the data become more unidimensional, the MAD and correlational values support that the programs both appear to align the univariate scale about equidistant from the ability axes.

Insert Table 2 about here

For the compensatory data sets, correlations and MAD values between $\hat{a}$ (univariate discrimination) and a_1, a_2, and $\hat{b}$ (univariate difficulty) and d are shown in Table 3. As the correlation between abilities increases, the correlation between $\hat{a}$ and a_1 and $\hat{a}$ and a_2 approach zero for both LOGIST and BILOG. MAD values between the discrimination estimates and parameters were slightly higher for BILOG in all correlational conditions. For both programs, the correlation between b and d was .99 for all data sets. This would suggest very strongly that the pattern of difficulty between the individual items is recoverable to a high degree.

Insert Table 3 about here

Correlations and average MAD values between the discrimination and difficulty parameters and their estimates for the noncompensatory data sets are displayed in Table 4. The pattern of correlations between discrimination parameters and estimates is similar to that of the compensatory data. The correlations between $\hat{b}$ and b_1, are all .99, while the correlations between $\hat{b}$ and b_2 range from r = .38 to r = .42 for both LOGIST and BILOG. This suggests that for the noncompensatory data there is a tendency to measure one dimension more strongly. This may also be due to the restricted range of b_2 values.

Insert Table 4 about here

In both the compensatory and noncompensatory data sets, the $\hat{a}$'s correlated positively with a_1 and negatively with a_2 except for the $\rho_{\theta_1\theta_2} = .9$ case. Noticeable differences exist between the MAD values for the noncompensatory discrimination parameters and estimates for BILOG and LOGIST. For LOGIST the average absolute differences of both $a_1 - \hat{a}$ and $a_2 - \hat{a}$ range from .80 to .86, while the range is .32 to .38 for BILOG. For both calibration programs the correlations between $\hat{a}$ and a_2 are negative except for the $\rho_{\theta_1\theta_2} = .9$ case in which the pattern reverses.

Conclusion

Differences between the item response surfaces for each model when the item parameters are matched appear to be minimal and exist in places of the θ_1, θ_2 plane where very few subjects would be expected to be found. Mean p-values for the eight sets were identical and the matches on biserial correlations were almost identical for the $\rho_{\theta_1\theta_2} = .9$ case. Thus the least squares matching procedures appears to be an excellent method of matching the two MIRT models.

The confounding of difficulty with dimensionality, which was reported in the results of the Reckase et al. (1986) study, was replicated, however, only for the BILOG calibration of response data in which $\rho_{\theta_1\theta_2}$ was closer to 0.0. The "wrap around" effect of the θ_1, θ_2 centroids did not occur for any of the LOGIST estimation runs. Although it should be noted that in the Reckase study the items only measured θ_1 or θ_2. Whereas in this study each item measured a combination of θ_1 and θ_2 to varying degrees. Thus the confounding was not as great as in the Reckase et al. study. Another possible explanation may be the method of estimation. Perhaps the marginal maximum likelihood procedure of BILOG is more sensitive to the confounding of difficulty and dimensionality.

The confounding of difficulty appeared to have the same affect on the ability parameter estimates for both the compensatory and noncompensatory datasets. Despite different test information patterns, as seen in the INFLINE plots, the orientation of the centroids appeared to be the same for each calibrations programs estimation of the two MIRT models.

The correla*ions among ability parameters and estimates suggest that as the relationship between the two ability dimensions become more linear, the data in a sense become unidimensional. As $\rho_{\theta_1\theta_2}$ approached .9, $\hat{\theta}$ correlated in the mid .90's with θ_1 and θ_2. This was confirmed by the plots of the θ_1, θ_2 centroids and the correlations of the discriminating parameters with their estimates. The correlations between $\hat{a}$ and a_1, and $\hat{a}$ and a_2 became closer as the correlation between θ_1 and θ_2 increased. Likewise as the $\rho_{\theta_1\theta_2}$ approached .9, the centroids appeared to align themselves along a 45° line. Both of these results suggest that θ_1 and θ_2 were being measured equally.

The plots of the θ_1, θ_2 centroids for the 20 $\hat{\theta}$ quantiles revealed differences between the two estimation programs. The centroid plot for LOGIST revealed only a slight confounding affect as $\rho_{\theta_1\theta_2}$ became closer to zero. However, θ_1, θ_2 centroids for BILOG's $\hat{\theta}$ display a much sharper wrapping around about the negative θ_2 axis and the positive θ_1 axis, especially when $\rho_{\theta_1\theta_2} = 0$. Thus it would appear that BILOG is more sensitive to the confounding of difficulty with dimensionality for both MIRT models.

Several directions for future research are suggested by this study. One area for future research would be to systematically vary test information with different two dimensional ability distributions to determine how the interaction of the two affects the orientation of the univariate ability scale in the two-dimensional plane. Also, the differences between maximum likelihood and marginal maximum likelihood estimation of multidimensional response data needs to be further explored.

References

Ansley, R. A., & Forsyth, T. N. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. Applied Psychological Measurement, 9, 37-48.

Hambleton, R. K., & Murray, L. N. Some goodness of fit investigations for item response models. In R. K. Hambleton (Ed.) Applications of item response theory, Vancover, B.C.: Educational Research Institute of British Columbia, 1982.

Mislevy, R. J. & Bock, R. D. (1982). BILOG, maximum likelihood item analysis and test scoring: Logistic model. Scientific Software, Inc.: Mooresville, IN.

Reckase, M. D. (1985, April). The difficulty of test items that measure more than one ability. Paper presented at the AERA Annual Meeting, Chicago.

Reckase, M. D., Carlson, J. E., Ackerman, T. A., & Spray, J. A. (1986, June). The interpretation of unidimensional IRT parameters when estimated from multidimensional data. Paper presented at the Psychometric Society Annual Meeting, Toronto.

Sympson, J. B. (1978). A model for testing with multidimensional items. In D. J. Weiss (Ed.) Proceedings of the 1977 Computerized Adaptive Testing Conference (pp. 82-98). Minneapolis; University of Minnesota, Department of Psychology, Psychometric Methods Program.

Traub, R. E. (1983). A priori consideration in choosing an item response model. In R. K. Hambleton (Ed.) Applications of item response theory (pp. 57-70). Vancover, B.C.: Educational Research Institute of British Columbia.

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). LOGIST User's guide. Princeton, NJ: Educational Testing Service.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, *8*, 125-145.

Table 1

Descriptive statistics of the multidimensional data sets (N = 1000, i = 40)

Data Type	$\rho(\theta_1\theta_2)$	Eigenvalues		KR-20	$\bar{p}$	Range of p		$\bar{r}$	Range of bis		Raw Score	
		λ_1	λ_2			Lo	Hi		Lo	Hi	$\bar{X}$	σ
COMP	.00	9.24	2.94	.91	.50	.16	.85	.64	.50	.71	20.15	8.64
	.3	10.84	2.59	.93	.50	.17	.84	.69	.57	.75	20.18	9.41
	.6	12.17	2.27	.94	.50	.18	.84	.73	.59	.79	20.15	10.04
	.9	13.38	2.00	.95	.50	.18	.83	.76	.61	.82	20.18	10.61
NCMP	.00	7.22	3.17	.88	.50	.16	.84	.56	.47	.64	20.03	7.64
	.3	9.52	2.69	.92	.50	.17	.83	.65	.57	.72	20.08	8.84
	.6	11.64	2.25	.94	.50	.17	.84	.71	.65	.76	20.13	9.82
	.9	13.53	1.98	.95	.50	.18	.83	.77	.69	.80	20.00	10.67

Note: Eigenvalues are those of the first and second principal components of the inter-item tetrachoric correlation.

Table 2

Correlations and mean absolute differences among $\hat{\theta}$, θ_1 and θ_2 by levels of correlation for compensatory and noncompensatory data sets

Calib	Data Type	$\rho(\theta_1, \theta_2)$	$r(\hat{\theta}, \theta_1)$	$r(\hat{\theta}, \theta_2)$	$\frac{\sum\lvert\theta_1 - \hat{\theta}\rvert}{k}$	$\frac{\sum\lvert\theta_2 - \hat{\theta}\rvert}{k}$
LOGIST	COMP	.00	.67	.64	.65	.67
		.3	.76	.76	.53	.53
		.6	.85	.85	.42	.42
		.9	.94	.94	.26	.27
BILOG		.00	.68	.64	.63	.65
		.3	.78	.76	.53	.54
		.6	.87	.86	.43	.44
		.9	.95	.95	.28	.28
LOGIST	NCMP	.00	.65	.60	.66	.70
		.3	.76	.72	.54	.58
		.6	.85	.84	.42	.43
		.9	.94	.94	.27	.28
BILOG		.0	.67	.59	.62	.67
		.3	.77	.73	.53	.56
		.6	.86	.85	.42	.44
		.9	.94	.94	.28	.29

ERIC Full Text Provided by ERIC

Table 3

Correlations and mean absolute differences between LOGIST and BILOG estimates and parameters under the compensatory model

Program	$\rho(\theta_1, \theta_2)$	$r_{\hat{a}, a_1}$	$r_{\hat{a}, a_2}$	$r_{\hat{b}, d}$	$\frac{\sum\lvert a_1-\hat{a}\rvert}{k}$	$\frac{\sum\lvert a_2-\hat{a}\rvert}{k}$	$\frac{\sum\lvert d-\hat{b}\rvert}{k}$
LOGIST	0.0	.30	-.30	-.99	.40	.45	2.09
	0.3	.26	-.26	-.99	.41	.45	2.09
	0.6	.17	-.17	-.99	.41	.44	2.09
	0.9	-.07	.07	-.99	.42	.43	2.09
BILOG	0.0	.26	-.26	-.99	.48	.52	2.09
	0.3	.18	-.18	-.99	.49	.50	2.10
	0.6	.19	-.19	-.99	.48	.50	2.10
	0.9	-.04	.04	-.99	.48	.48	2.10

ERIC Full Text Provided by ERIC

Table 4

Correlations and mean absolute differences between LOGIST and BILOG estimates and parameters of the noncompensatory model

Program	$\rho(\theta_1\theta_2)$	$r_{\hat{a}a_1}$	$r_{\hat{a}\ a_2}$	$r_{\hat{b}b_1}$	$r_{\hat{b}b_2}$	$\frac{\sum\|a_1-\hat{a}\|}{k}$	$\frac{\sum\|a_2-\hat{a}\|}{k}$	$\frac{\sum\|b_1-\hat{b}\|}{k}$	$\frac{\sum\|b_2-\hat{b}\|}{k}$
LOGIST	0.0	.31	-.23	.99	.42	.86	.82	.67	.87
	0.3	.27	-.22	.99	.41	.85	.81	.67	.87
	0.6	.19	-.14	.99	.40	.84	.80	.67	.86
	0.9	-.07	.06	.99	.38	.84	.80	.67	.85
BILOG	0.0	.28	-.21	.99	.42	.35	.38	.67	.86
	0.3	.19	-.17	.99	.41	.33	.37	.67	.86
	0.6	.19	-.20	.99	.40	.32	.35	.67	.86
	0.9	-.40	-.01	.99	.39	.32	.35	.67	.85

Figure Captions

Figure 1. Vectors representing the distance and direction from the origin to the point of maximum discrimination for the 40 generated compensatory items.

Figure 2. Diffic.lty (2a) and Discrimination (2b) parameter values for the 40 matched compensatory and noncompensatory items ($\rho_{\theta_1\theta_2} = 0.0$).

Figure 3. The item response surface and contour plot of matched compensatory and noncompensatory items: 1 (3a), 20 (3b) and 40 (3c).

Figure 4. Test information vectors at selected points in the ability plane for the compensatory (4a) and noncompensatory (4b) data sets.

Figure 5. A plot of the centroids for the LOGIST calibrated compensatory (5a) and noncompensatory (5b) response sets for each level of correlation among the two-dimensional abilities.

Figure 6. A plot of the centroids for the BILOG calibrated compensatory (6a) and noncompensatory (6b) response sets for each level of correlation among the two-dimensional abilities.

FIGURE 1

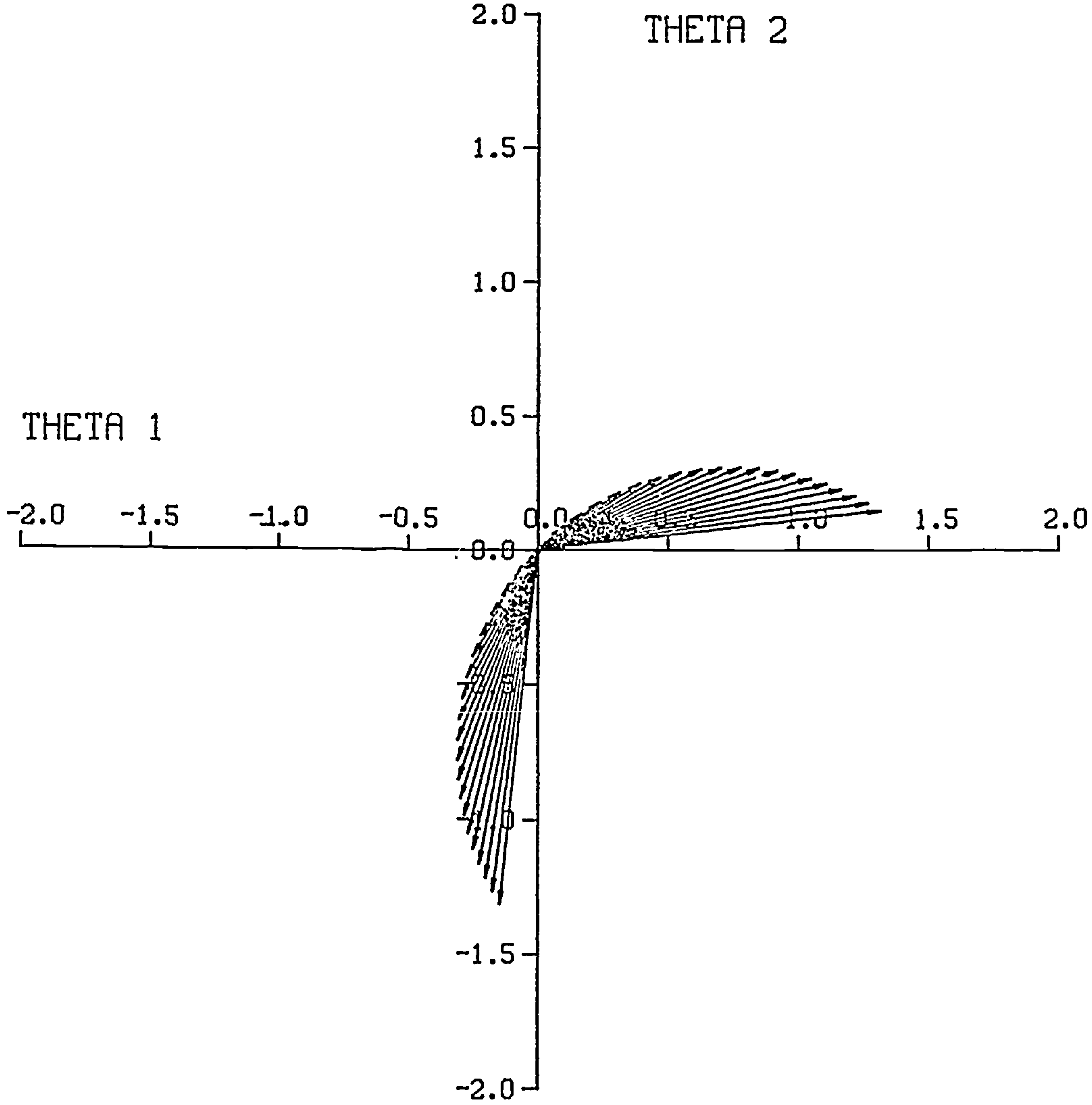


FIGURE 2

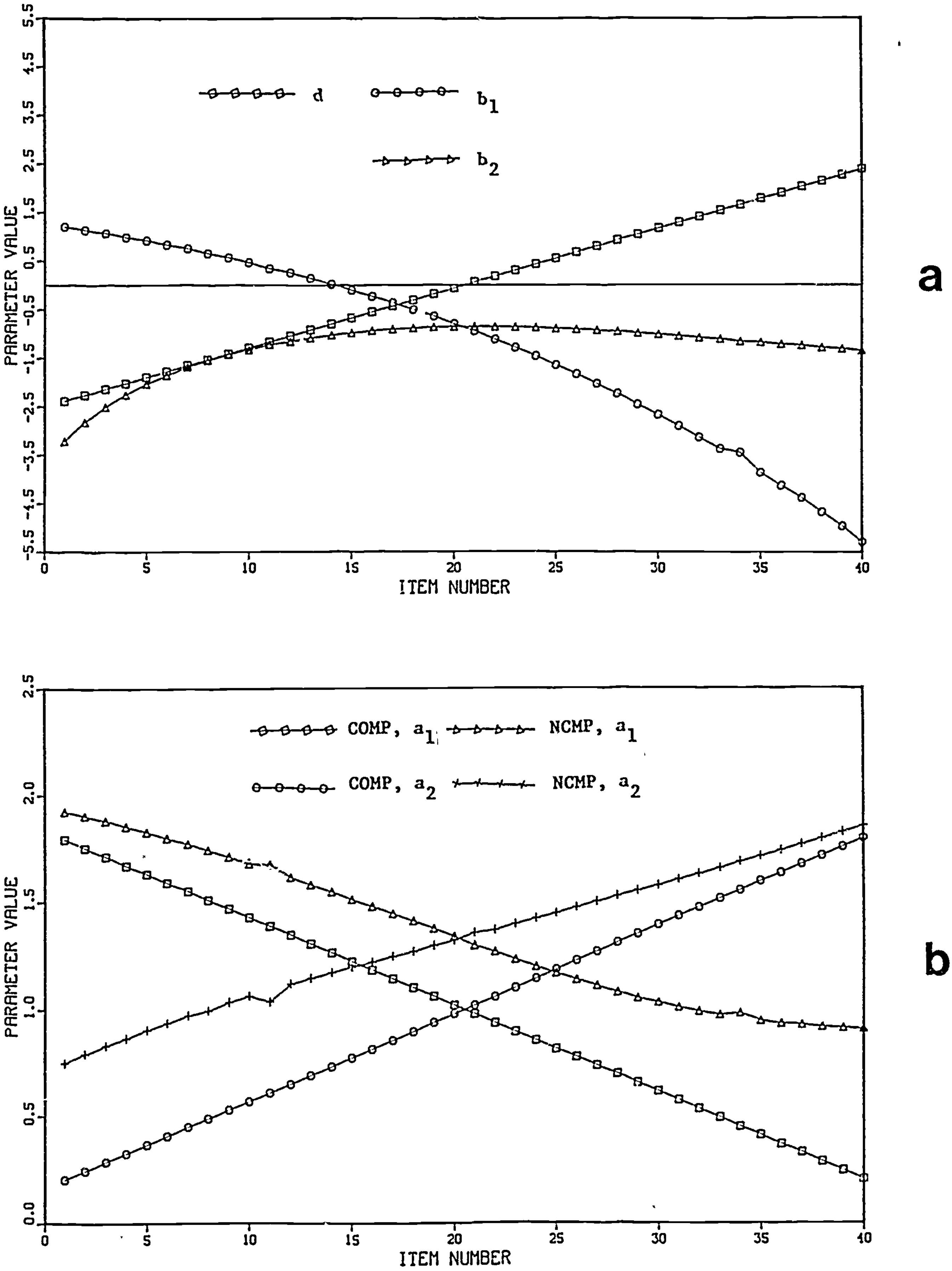

d
b_1
b_2
PARAMETER VALUE
5.5
4.5
3.5
2.5
1.5
0.5
-0.5
-1.5
-2.5
-3.5
-4.5
-5.5
0
5
10
15
20
25
30
35
40
ITEM NUMBER
a
COMP, a_1
NCMP, a_1
COMP, a_2
NCMP, a_2
PARAMETER VALUE
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
30
35
40
ITEM NUMBER
b

FIGURE 3a

Item 1

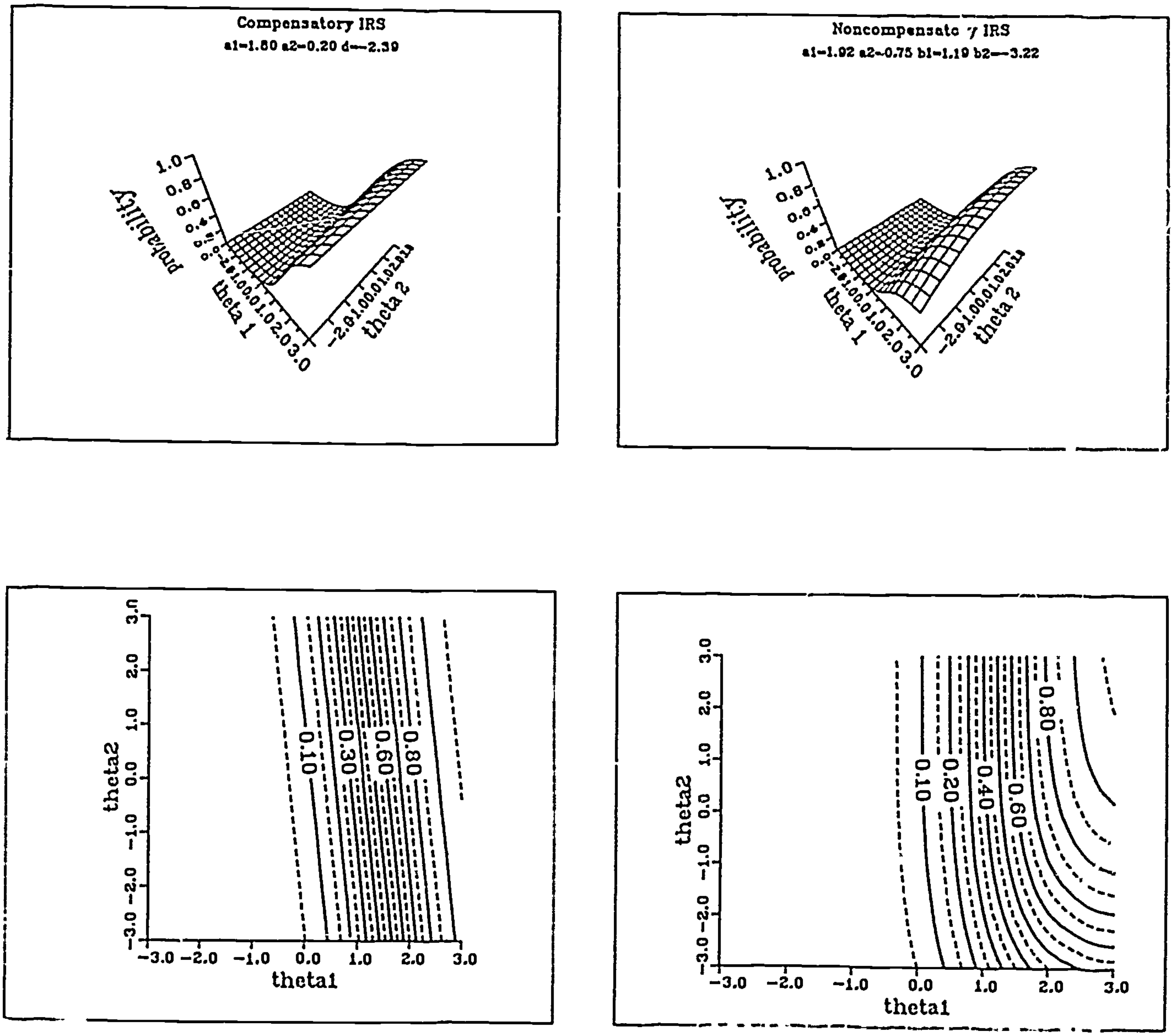


FIGURE 3b

Item 20

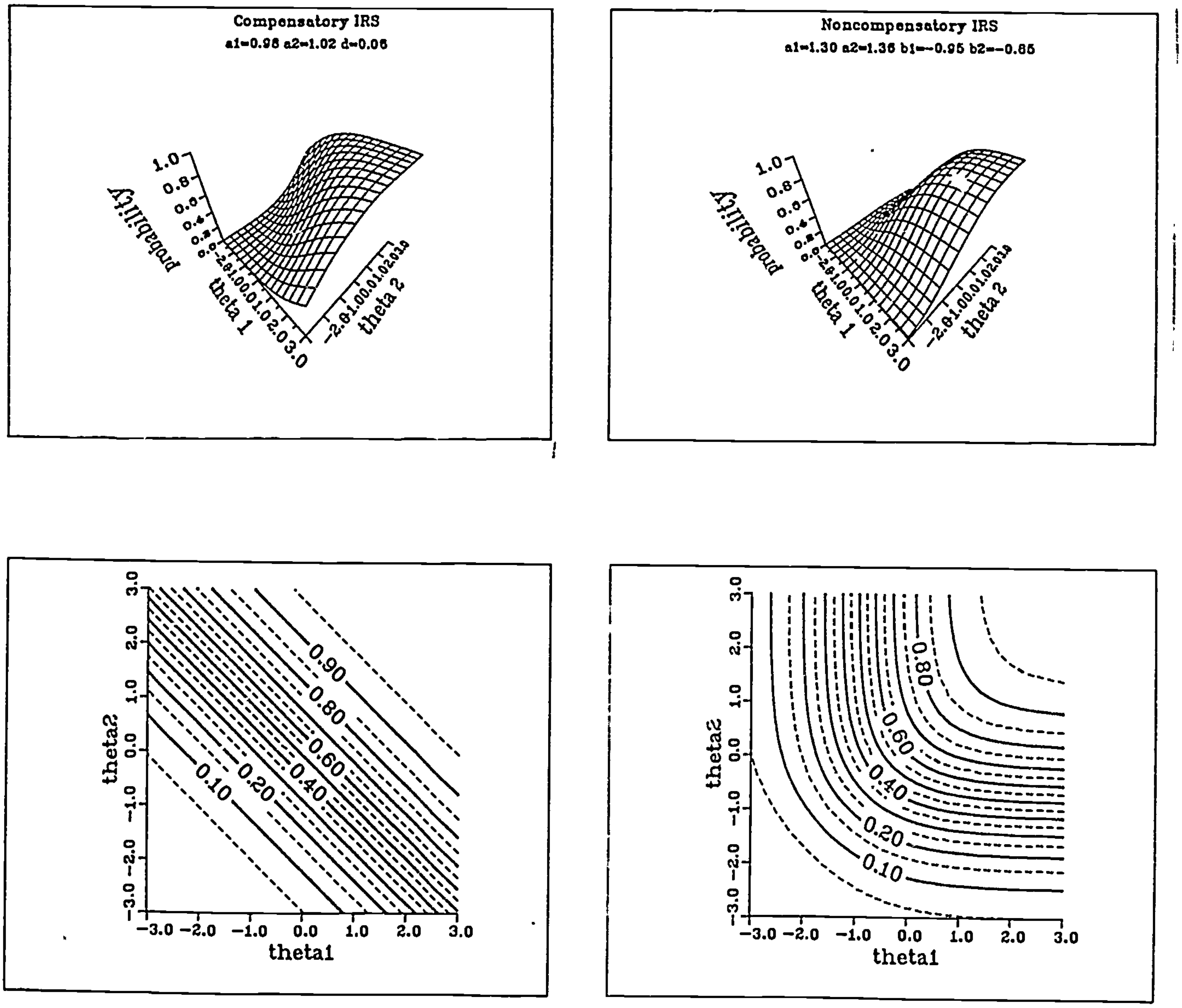


FIGURE 3c

Item 40

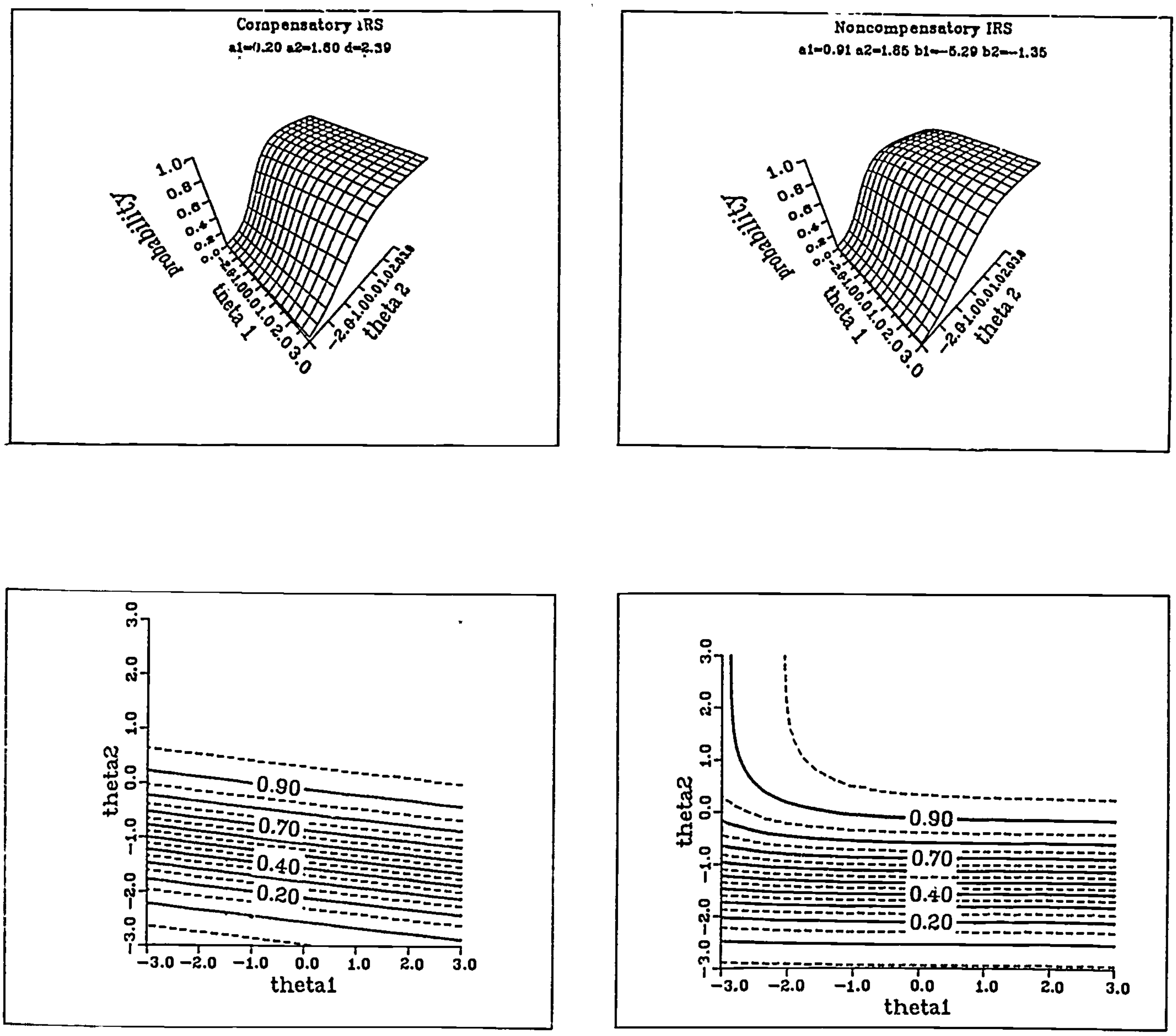


FIGURE 4

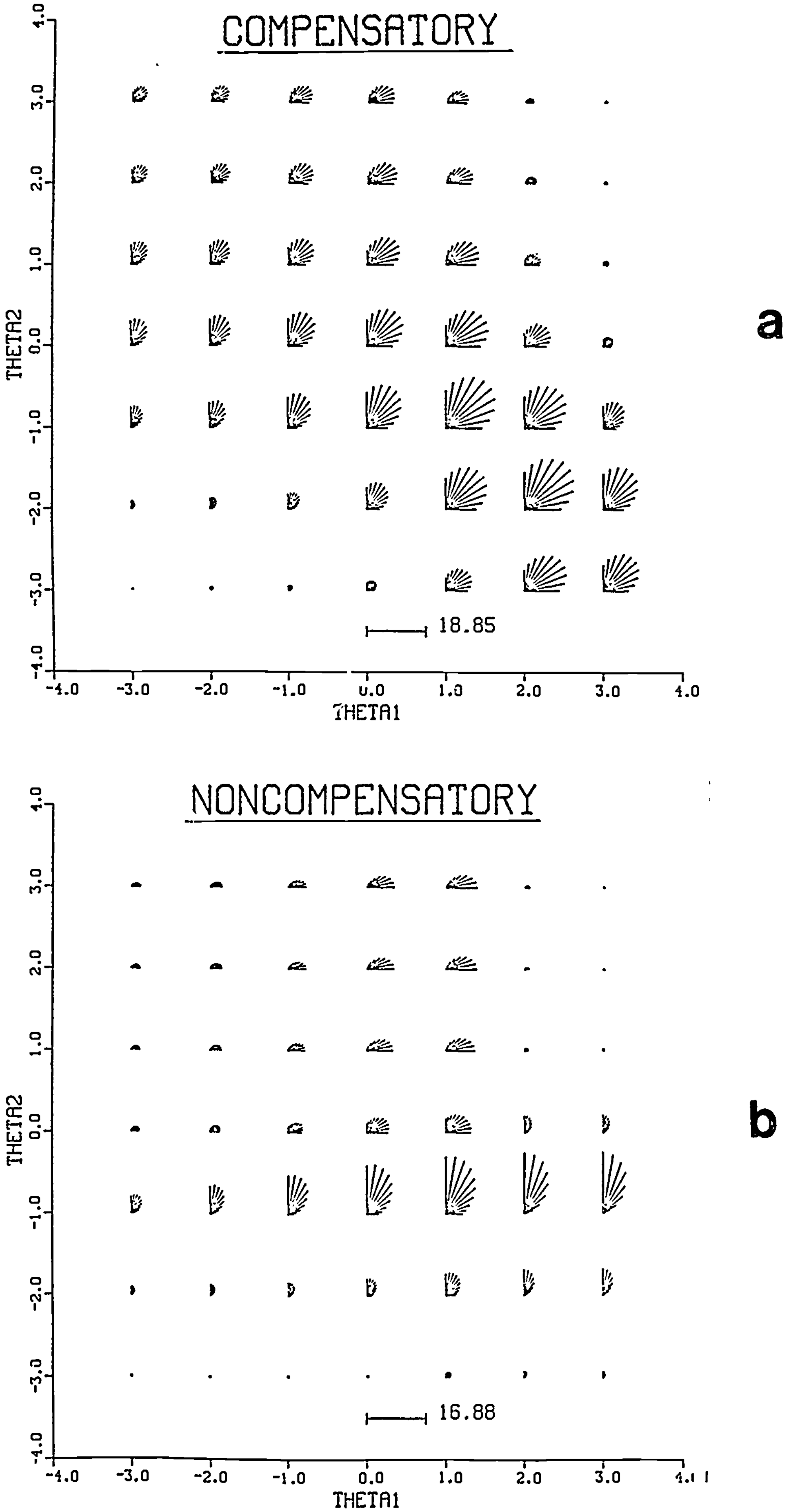

COMPENSATORY
THETA2
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
18.85
THETA1
-4.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0
4.0
a
NONCOMPENSATORY
THETA2
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
16.88
THETA1
-4.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0
4.0
b

FIGURE 5

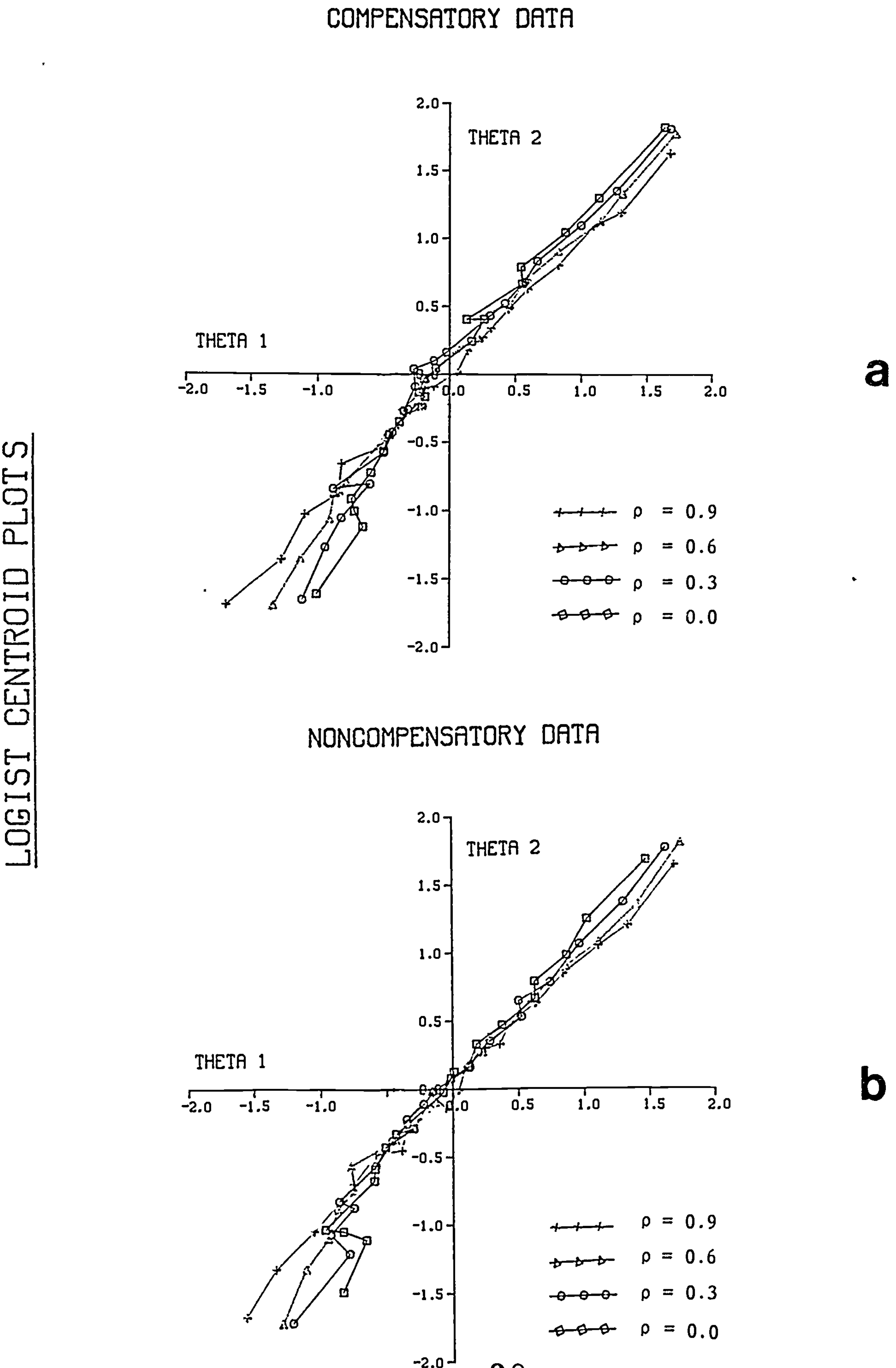
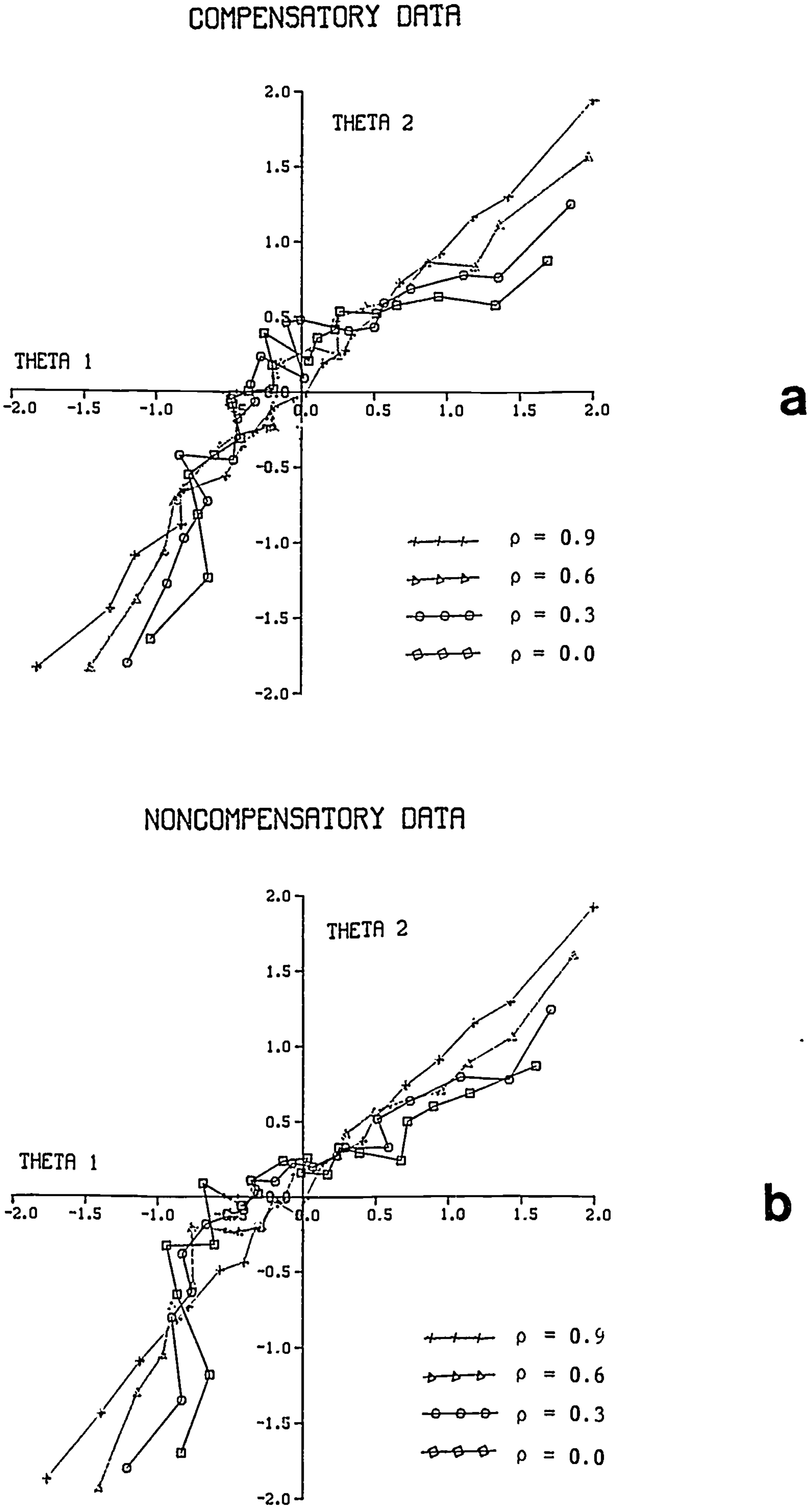

COMPENSATORY DATA
THETA 2
THETA 1
2.0
1.5
1.0
0.5
-0.5
-1.0
-1.5
-2.0
-2.0
-1.5
-1.0
0.0
0.5
1.0
1.5
2.0
ρ = 0.9
ρ = 0.6
ρ = 0.3
ρ = 0.0
a
NONCOMPENSATORY DATA
THETA 2
THETA 1
2.0
1.5
1.0
0.5
-0.5
-1.0
-1.5
-2.0
-2.0
-1.5
-1.0
0.0
0.5
1.0
1.5
2.0
ρ = 0.9
ρ = 0.6
ρ = 0.3
ρ = 0.0
b
LOGIST CENTROID PLOTS

FIGURE 6

BILOG CENTROID PLOTS



COMPENSATORY DATA

NONCOMPENSATORY DATA