

---

## STATISTICAL DEVELOPMENTS AND APPLICATIONS

---

# Inferential Errors in Taxometric Analyses of Ordered Three-Class Constructs

ROBERT E. MCGRATH

*School of Psychology, Fairleigh Dickinson University*

Taxometric analysis (Waller & Meehl, 1998) has become a popular tool for investigating whether personality and psychopathology constructs are better modeled as dimensional or categorical. Monte Carlo evidence for the validity of taxometric methods, however, has been restricted to data that are derived either from latent dimensional or dichotomous models. There are various circumstances in which the correct model for a psychosocial construct could involve three or more ordered classes. It was demonstrated both mathematically and empirically that under these circumstances, the results of taxometric analyses can lead to incorrect conclusions about the population structure. Recommendations are provided for future studies intended to distinguish between categorical and dimensional structures.

Taxometric analysis refers to a set of statistical techniques introduced by Meehl and his associates (Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998) that has become a popular tool for addressing whether constructs in psychopathology and personality are inherently dichotomous or dimensional. Although there are other statistical methods that are intended to evaluate data structure, such as cluster analysis and finite mixture models, Meehl's discussions of taxometrics has spurred much of the recent interest in evaluating whether diagnostic and personality labels are in fact taxonic (e.g., Schmidt, Kotov, & Joiner, 2004). As a result, taxometric methods have become a particularly popular statistical approach for addressing such questions (see Haslam, 2003; Haslam & Kim, 2002; Ruscio, Haslam, & Ruscio, 2006; Schmidt et al., 2004). Since 1992, more than 100 articles have been published using taxometric analysis to evaluate whether latent structure is based on classes or factors, including several published in the *Journal of Personality Assessment* (Franklin, Strong, & Greene, 2002; Strong, Greene, Hoppe, Johnston, & Olesen, 1999; Strong, Greene, & Kordinak, 2002).

The statistical conclusion validity of taxometric analysis was established initially through a series of Monte Carlo studies (Meehl & Yonce, 1994, 1996). Taxometric methods were developed specifically to distinguish between latent dichotomous and dimensional structures, however, and simulation studies have focused exclusively on these two alternatives. In contrast, other statistical methods having to do with data structure, such as cluster analysis and finite mixture models, allow for the possibility that constructs can consist of more than two classes (polytomous structure). To date, no studies have been conducted evaluating the behavior of taxometric methods when a construct encompasses three or more classes.

Both Meehl (1999) and Ruscio and Ruscio (2002) discussed this issue, and concluded taxometric methods should, at least theoretically, be useful still. Both discussions assumed manifest indicators would be sensitive only to the discrimination between two of the classes, however, so that some indicators would distinguish only between classes A and B, while others would be sensitive to classes B versus C. If the classes demonstrate an ordinal structure in which the rank ordering of classes is consistent across indicators, however, it would seem more likely that valid indicators of a polytomous construct will be sensitive to multiple discriminations between adjacent classes.

For example, some measures of psychopathology and many measures of personality are constructed as bipolar scales. A higher than average score on a measure of introversion–extroversion may suggest an extroverted individual, while a lower than average score indicates introversion. If at the latent level both extroversion and introversion are categorically distinct from the normative level of social involvement, a well-constructed measure of the full bipolar spectrum should be sensitive to both discriminations.

Even in circumstances where the scale is unipolar, polytomous structure may be present if one or more classes in turn subsumes multiple classes. For example, although available research indicates the existence of a schizotypal taxon in the general population (e.g., Blanchard, Gangestad, Brown, & Horan, 2000; Erlenmeyer-Kimling, Golden, & Cornblatt, 1989), it is mute on whether individuals with schizophrenia represent an additional taxon within the schizotypal class. Going one step further, Blanchard, Horan, and Collins (2005) recently have found that a taxonic distinction exists among those meeting criteria for schizophrenia depending on the degree to which the disorder is characterized by negative symptoms. If a sample were gathered that included a reasonable sampling of normal, schizotypal, and schizophrenic individuals, and the indicators used are primarily sensitive to the severity of negative symptoms, it is possible that the results would reflect the presence of three and perhaps even four ordered classes.

---

Received July 27, 2005; Revised September 18, 2006.  
Address correspondence to Robert McGrath, School of Psychology, Fairleigh Dickinson University, 1000 River Road T-WH1-01, Teaneck, NJ 07666; Email: mcgrath@fd.edu

It is impossible to know how often the two conditions of (1) polytomous structure and (2) manifest indicators sensitive to the presence of more than two classes are met in practice. Given that there is reason to believe it probably does occur at times, it is worth considering how taxometric methods would behave under these circumstances. Unfortunately, a review of the published literature provides no indication that researchers have considered and ruled out the possibility of polytomous structure before electing to use taxometric methods. In the absence of such safeguards, one would hope the results from the taxometric analysis of polytomous data would produce results that are inconsistent with either a dimensional or dichotomous interpretation, but this is an empirical question.

The current study was conducted to investigate what is likely to happen when popular taxometric methods inadvertently are used in conjunction with polytomous ordinal data where the indicators are sensitive to multiple qualitative distinctions. This issue will be addressed both mathematically and empirically. To simplify matters, the analysis will be restricted to the three-class case, although the results provide insight into what would happen with more complex structures involving four or more classes.

The reference to inadvertent use is important for understanding the goals and methods of this study. The intention is not to develop or validate taxometric methods as a general approach to the detection of three or more ordered classes, although the potential for doing so will be discussed briefly later. The goal instead is to demonstrate for current users of taxometric methods the importance of considering the possibility of other latent structures besides dimensions and dichotomies before using taxometric methods. As the results will demonstrate, the failure to consider other possibilities potentially results in inferential errors.

## METHOD

### *File Generation*

A series of simulated data sets was generated for purposes of demonstrating the impact of ordered three-class structure on the outcomes of taxometric analyses. These data sets were developed using a generalization of a method described by Meehl and Yonce (1994) for the creation of two-class data sets. Although Meehl and Yonce also provided a method to generate dimensional data sets, numerous studies already have found that taxometric methods accurately can identify dimensional structure (e.g., Meehl & Yonce, 1994, 1996; Waller & Meehl, 1998), so there was no need to study dimensional data further. More sophisticated approaches have since been developed for generating simulated categorical data (e.g., Waller, Underhill, & Kaiser, 1999). The Meehl and Yonce (1994) algorithm is unique however, in terms of the number of studies demonstrating that taxometric methods accurately can detect taxonic structure in the resulting data sets.

The Meehl and Yonce (1994) algorithm requires setting four parameters for categorical data sets, the sample size; the base rate for each class, the mean separation between classes, and the degree of nuisance covariation within classes. The mean separation refers to the difference between means in adjoining classes. Larger separations between classes will produce greater covariation between indicators and also should make it easier for taxometric and other methods to detect the categorical structure.

Nuisance covariation refers to the degree of covariation within classes. Greater nuisance covariation also increases the degree of covariation between indicators, but it serves as noise that should interfere with the detection of categorical structure. In the Meehl and Yonce (1994) algorithm, nuisance covariation is introduced by the use of latent factor loadings.

Given the goals of the study, parameters were chosen because they were within a range that should allow the detection of categorical structure. To minimize the possible effect of sampling error on the findings, a single data set of 100,000 observations was generated for each combination of parameters to be examined. The other three parameters were set within a range that prior simulation studies on identifying categorical structure indicated would allow the detection of dichotomous structure (Beach, Amir, & Bau, 2005; Cleland, Rothschild, & Haslam, 2000; Meehl & Yonce, 1994; Steinley, 2004). Based on this review, the lower boundary for base rates was set to .10. Various base rates drawn from previous studies were used to develop rectangular, roughly bell-shaped, and skewed distributions reflecting three classes.

Similarly, prior discussions suggest separations between class means of 1.25 standard deviations represent the lower bound likely to be detectable by taxometric methods, although separations of 2.0 are better (e.g., Ruscio & Ruscio, 2004b). Finally, a factor loading of .001 was used by Meehl and Yonce (1994) to represent the case of no nuisance covariation, but loadings as high as .70 still produced detectable taxonic structures.

Crossing five base-rate distributions with the two nuisance covariation values (.001 and .70) and the four possible combinations of the two mean separations (1.25 and 2.0) produced the 40 scenarios summarized in Table 1. For each scenario, four indicators were generated. To simplify matters, all four indicators were developed using the same parameters.

To create the data sets, five random normal deviates ( $M = 0$ ,  $SD = 1.0$ ) were generated for each observation using SPSS version 12.0. A Visual BASIC program was developed that used the five deviates to create one observation's scores on four indicators, according to the following steps.

The first random deviate was multiplied by the factor loading in Table 1. This served as the seed value for all four indicators. The seed value was then added separately to the remaining four deviates weighted by

$$\sqrt{1 - \text{loading}^2}$$

to generate four different scores. This step created four correlated indicators without categorical structure. Finally, the first mean separation in Table 1 was added to the four indicators for a proportion of observations equal to the second base rate, and the two separation values were both added to a proportion of observations equal to the third base rate. As a result, each indicator included observations representing three distinct classes that were rank-ordered consistently across the four indicators.

For example, data set B4S1L1 (see Table 1) consisted of 100,000 observations in which 60,000 had a mean score of 0 on each of the four indicators, 20,000 observations had a mean score of 2, and 20,000 had a mean score of  $2 + 2 = 4$ . Within-class covariation was minimal in this data set, since the factor loadings were only .001, so covariation among the four

TABLE 1.—Data set patterns.

Data Set Label	BR	Sep	Load	$E(M)$	$E(Var)$	$E(r)$	Observed Skew ( $M$ )
B1S1L1	.33-.34-.33	2.00-2.00	0.001	2.00	3.64	0.73	0.00
B1S1L2	.33-.34-.33	2.00-2.00	0.700	2.00	3.64	0.86	0.00
B1S2L1	.33-.34-.33	1.25-2.00	0.001	1.50	2.77	0.64	0.15
B1S2L2	.33-.34-.33	1.25-2.00	0.700	1.50	2.77	0.82	0.15
B1S3L1	.33-.34-.33	2.00-1.25	0.001	1.75	2.77	0.64	-0.15
B1S3L2	.33-.34-.33	2.00-1.25	0.700	1.75	2.77	0.82	-0.14
B1S4L1	.33-.34-.33	1.25-1.25	0.001	1.25	2.03	0.51	0.00
B1S4L2	.33-.34-.33	1.25-1.25	0.700	1.25	2.03	0.75	0.01
B2S1L1	.20-.60-.20	2.00-2.00	0.001	2.00	2.60	0.62	-0.01
B2S1L2	.20-.60-.20	2.00-2.00	0.700	2.00	2.60	0.80	0.00
B2S2L1	.20-.60-.20	1.25-2.00	0.001	1.40	2.09	0.52	0.24
B2S2L2	.20-.60-.20	1.25-2.00	0.700	1.40	2.09	0.76	0.23
B2S3L1	.20-.60-.20	2.00-1.25	0.001	1.85	2.09	0.52	-0.23
B2S3L2	.20-.60-.20	2.00-1.25	0.700	1.85	2.09	0.76	-0.24
B2S4L1	.20-.60-.20	1.25-1.25	0.001	1.25	1.63	0.38	0.00
B2S4L2	.20-.60-.20	1.25-1.25	0.700	1.25	1.63	0.69	-0.01
B3S1L1	.10-.80-.10	2.00-2.00	0.001	2.00	1.80	0.44	-0.01
B3S1L2	.10-.80-.10	2.00-2.00	0.700	2.00	1.80	0.72	0.00
B3S2L1	.10-.80-.10	1.25-2.00	0.001	1.33	1.55	0.36	0.25
B3S2L2	.10-.80-.10	1.25-2.00	0.700	1.33	1.55	0.67	0.25
B3S3L1	.10-.80-.10	2.00-1.25	0.001	1.93	1.55	0.36	-0.25
B3S3L2	.10-.80-.10	2.00-1.25	0.700	1.93	1.55	0.67	-0.25
B3S4L1	.10-.80-.10	1.25-1.25	0.001	1.25	1.31	0.24	0.00
B3S4L2	.10-.80-.10	1.25-1.25	0.700	1.25	1.31	0.61	0.00
B4S1L1	.60-.20-.20	2.00-2.00	0.001	1.20	3.56	0.72	0.52
B4S1L2	.60-.20-.20	2.00-2.00	0.700	1.20	3.56	0.86	0.51
B4S2L1	.60-.20-.20	1.25-2.00	0.001	0.90	2.62	0.62	0.51
B4S2L2	.60-.20-.20	1.25-2.00	0.700	0.90	2.62	0.80	0.51
B4S3L1	.60-.20-.20	2.00-1.25	0.001	1.05	2.81	0.64	0.34
B4S3L2	.60-.20-.20	2.00-1.25	0.700	1.05	2.81	0.82	0.35
B4S4L1	.60-.20-.20	1.25-1.25	0.001	0.75	2.00	0.50	0.30
B4S4L2	.60-.20-.20	1.25-1.25	0.700	0.75	2.00	0.75	0.30
B5S1L1	.80-.10-.10	2.00-2.00	0.001	0.60	2.64	0.62	0.94
B5S1L2	.80-.10-.10	2.00-2.00	0.700	0.60	2.64	0.81	0.94
B5S2L1	.80-.10-.10	1.25-2.00	0.001	0.45	2.01	0.50	0.76
B5S2L2	.80-.10-.10	1.25-2.00	0.700	0.45	2.01	0.75	0.76
B5S3L1	.80-.10-.10	2.00-1.25	0.001	0.53	2.18	0.54	0.69
B5S3L2	.80-.10-.10	2.00-1.25	0.700	0.53	2.18	0.77	0.69
B5S4L1	.80-.10-.10	1.25-1.25	0.001	0.38	1.64	0.39	0.47
B5S4L2	.80-.10-.10	1.25-1.25	0.700	0.38	1.64	0.69	0.47

Note. Parameters were fixed across indicators in each data set. BR = base-rate pattern for the classes with the lowest, intermediate, and highest means, respectively; Sep = separation between the classes with the lowest and intermediate means, and intermediate and highest means, respectively; Load = factor loading used to create nuisance (within-class) correlation;  $E(M)$  = expected mean for each indicator;  $E(Var)$  = expected variance for each indicator;  $E(r)$  = expected correlation between each pair of indicators. The last column is the mean skew for the four indicators in the data set.

indicators was due almost exclusively to parallel differences in the means for the three classes.

Table 1 also provides the expected means, variances, and correlations for each data set. The latter two quantities were computed using a formula provided by Haertel (1990, Equation 12):

$$\text{cov}(yz) = \sum_i^k p_i \text{cov}(yz)_i + \sum_j^k \sum_{i < j}^k p_i p_j (\bar{y}_i - \bar{y}_j)(\bar{z}_i - \bar{z}_j). \quad (1)$$

Readers familiar with taxometric methods will recognize this as a generalization of the formula Meehl and Yonce (1996) called the General Covariance Mixture Theorem, which is the fundamental formula underlying several taxometric methods. A derivation of Haertel’s formula seems to be unavailable in the published literature. Since Equation 1 determines the effect of polytomous data on the outcomes from certain taxometric methods, a derivation is provided in Appendix A.

None of the actual data set means, variances, or correlations differed from the expectations provided in Table 1 by as much as .03, which is consistent with the accuracy of the algorithm used and the proposition that sampling error was minimized through large sample sizes. Mean skew of the indicators also was computed for each data set. These values varied between .26 and .95, and varied as a function of the base-rate distribution and the mean separations.

*Taxometric Analyses*

Each data set was examined using the four most popular taxometric methods. Mean Above Minus Below a Cut (MAMBAC; Meehl & Yonce, 1994) requires two indicators of the construct. Cut scores are set at successive points on one indicator (the input indicator). At each cut score, the mean score for a second indicator (the output indicator) is computed separately for those above and below the cut, and the difference between two means is computed. For dichotomous constructs without serious skew, a graph with input cut scores on the abscissa (i.e., Y- or

vertical-axis) and output mean differences on the ordinate (i.e., X- or horizontal-axis) should be hill-shaped, with the peak occurring where the sum of the hit rates for the two classes is at a maximum. When the latent construct is dimensional, the same graph should tend toward a U shape. Each pair of dimensional indicators can be evaluated twice, with one indicator used as the input indicator for one analysis and the output indicator for the other. With  $m = 4$  indicators it is therefore possible to generate

$$\frac{m!}{(m-2)!} = \frac{4!}{2} = 12$$

MAMBAC graphs.

Maximum Covariance (MAXCOV; Meehl & Yonce, 1996) requires one input and two output indicators. Observations are ordered and divided into overlapping subgroups or windows along the input indicator, and the covariance between the output indicators is computed within each window. For dichotomous constructs, the General Covariance Mixture Theorem suggests a graph with input windows on the abscissa and output covariances on the ordinate should resemble a hill that peaks in the window where the frequencies of the taxon and complement classes are approximately equal. If the construct is dimensional, the graph should be relatively flat or saw-toothed. Using four indicators it is possible to produce

$$\frac{m!}{2!(m-3)!} = \frac{4!}{2(1)} = 12$$

MAXCOV graphs.

Maximum Eigenvalue (MAXEIG; Waller & Meehl, 1998) is an extension of MAXCOV for circumstances where more than two indicators are available to serve as output. After dividing observations into overlapping windows on the input indicator, the eigenvalue for the first principal component based on the output indicators is computed within each window. The plot of eigenvalues as a function of window should follow the pattern described for MAXCOV. MAXEIG usually is conducted one time, with each indicator as the input indicator, resulting in  $m = 4$  graphs for each data set.

MAXCOV and MAXEIG also allow estimation of the Bayesian posterior probability of membership in the taxon class for each observation in the data set, assuming a taxon is present. The histogram of these probabilities provides another test of the taxonic hypothesis. A histogram in which the probabilities divide into two sets that cluster near 0 and 1 is considered supportive of the presence of two classes, though Ruscio and Ruscio (2004b) have argued that this pattern can occur even if the data are dimensional. If instead the histogram reveals a single cluster of probabilities, or if probabilities are distributed across the entire range from 0 to 1, the results are considered more consistent with a dimensional structure.

Latent Mode (L-Mode; Waller & Meehl, 1998) differs from the preceding methods in that all available indicators are treated as a single set. The indicators are factor analyzed, and the factor scores are generated for the first factor. Because of the reduction of measurement error, the density plot of the first factor scores should be bimodal if the data are taxonic and unimodal if the data are dimensional. One L-Mode factor analysis was conducted for each data set using the four indicators.

Taxometric methods also produce statistical results that are relevant to distinguishing between taxonic and dimensional data. An estimate of what the taxon base rate would be if the data are structurally dichotomous can be generated in association with each MAMBAC, MAXCOV, or MAXEIG graph. The density plot generated by the L-Mode procedure also can be used to produce three distinct estimates of the taxon base rate, two based on the distribution modes and one based on the attempted classification of observations as belonging to the taxon or complement class. If in fact a taxon is present in the data, all these estimates of taxon base rate should tend to converge on a single value. If, instead, the data are dimensional, there is no reason to expect convergence. Excessive variability in the taxon estimates is therefore indicative of dimensional structure, although no real standard has been suggested for what should be considered excessive.

Finally, Waller and Meehl (1998) also discussed the goodness of fit index (GFI) as a statistical indicator of structure. In the case of taxometric analysis, the GFI evaluates the degree of consistency between the observed variance-covariance matrix and the best estimate of that matrix assuming the data are in fact taxonic. They concluded that a GFI value  $\geq .90$  can be indicative of taxonic structure. Several studies have questioned this statistic's accuracy as an indicator of taxonic structure, however (Cleland et al., 2000; Ruscio & Ruscio, 2004b).

The proposition that data are structurally dichotomous is a hypothesis similar in form to those traditionally tested via null hypothesis significance tests (NHSTs). Taxometric methods are distinct from NHSTs, however, in that the conclusion one draws about the hypothesis is based not on a single test against a criterion for significance, but rather on whether there is a convergence about whether the hypothesis adequately accounts for the results across a series of discrete analyses. Meehl referred to this convergence criterion as *coherent cut kinetics* in the context of taxometrics (Meehl & Yonce, 1994), and as *consistency hurdles* in the context of a general alternative to NHSTs for hypothesis testing (Meehl, 1978). In taxometric analysis, the researcher is expected to examine the full array of findings—including graphical results, homogeneity of the taxon base rate estimates, and GFI values—to draw a conclusion about the hypothesis of taxonic structure based on consistency in the findings. Many researchers, however, interpret only the graphical results. In the present study, the graphical results will be examined first; then graphical and statistical results will be examined in combination. The use of coherent cut kinetics allowed for three possible outcomes for any data set in the present study:

1. The results converged to suggest a dichotomous structure, which would lead to an inaccurate interpretation in practice, that is, that the data are dichotomous.
2. The results converged to suggest a dimensional structure, which would lead to an incorrect interpretation, that is, that the data are dimensional.
3. The results did not converge on either interpretation. The manner in which such an outcome is likely to be treated in current practice will be discussed below.

Taxometric analyses were carried out using a software package developed by John Ruscio (Ruscio, Haslam, & Ruscio, 2006). This package is based on the R statistical programming

language, and was used in the present study in association with R version 2.2.1. In addition to the typical output associated with taxometric methods, the software allows computation of the GFI for each taxometric method.

### *Alternative Analyses*

As was noted earlier, taxometric methods are not the only analytic procedures that have been developed for the purpose of identifying categorical structure. Several other methods for estimating the correct number of classes reflected in data also were evaluated, with the purpose of comparing the outcomes with those resulting from taxometric analysis.

Cluster analysis is a heuristic statistical technique for categorizing observations, in that observations are clustered whether or not the data are truly categorical (just as factor analysis will produce factor scores regardless of whether the data are truly dimensional). A variety of procedures have been suggested, however, for evaluating the number of naturally occurring clusters in data analyzed using cluster analysis. Three such procedures found by Milligan and Cooper (1985) to be particularly effective in the presence of substantial error variability are available through SAS (SAS Institute, 2004).

All three are based on the use of hierarchical cluster analysis, which generates a series of models with the number of clusters first set to the number of observations (each observation treated as a cluster). In each subsequent step the two closest clusters are combined, to reduce the number of clusters by one, until the final step consists of a single cluster. The cubic clustering criterion (CCC) statistic is computed at each step. A  $CCC > 2$  is considered good evidence of the correct number of classes, while a value between 0 and 2 potentially is supportive. An elevated CCC value for the one-cluster solution would suggest a purely dimensional structure.

The pseudo- $F$  and pseudo- $t^2$  statistics also are computed at each step, and then they are inspected sequentially beginning with the most complex model and moving toward the one-cluster solution. A marked increase in the value of the pseudo- $F$  statistic is expected to correspond to the correct solution for the number of classes in the data. A marked increase in the pseudo- $t^2$  statistic is expected to correspond with the solution involving one fewer cluster than the correct solution.

There are several problems with this set of procedures. The pseudo- $F$  statistic cannot be computed for the one-cluster solution. The pseudo- $t^2$  statistic can, but since this statistic indicates the accuracy of the next largest solution, neither of these procedures can test directly for the presence of dimensional structure. None of the procedures are exclusive: several solutions can meet the criterion for a potentially correct solution. Finally, the three procedures often do not converge, leading Sarle and Kuo (1993) to recommend interpreting them on the basis of consensus if one exists.

More recently, Tonidandel and Overall (2004) developed a fourth procedure that in their preliminary analyses fared better than some methods for which Milligan and Cooper (1985) found positive results. Their method uses resampling to generate 100 data sets from the original data. Each sample then is divided into four independent subsamples, and hierarchical cluster analysis is conducted using each subsample. The cluster mean profiles from the four subsamples at each hierarchical level then are combined in a single higher-order cluster analysis. If the higher-

order clusters that result from this analysis each contain one mean profile from each subsample, then perfect replication has occurred. If several solutions result in perfect replication, the solution with the largest number of clusters is used as the best estimate of the number of clusters. This process is repeated for each of the resampled data sets, producing 100 estimates of the number of classes. The final step involves inspecting the best estimate of the number of clusters across the 100 replications. The value that defines the 80th percentile of these estimates is considered the best single estimate of the correct number of classes.

One useful feature of this Tonidandel and Overall (2004) procedure is that preliminary recommendations have been offered concerning the identification of purely dimensional structure. If the modal estimate of the number of clusters emerges in less than 65% of the resampled data sets, they concluded the results support the presence of dimensional structure.

Cluster analyses were computed using SAS v.9.1. By default, SAS prints CCC, pseudo- $F$ , and pseudo- $t^2$  statistics for solutions consisting of 30 clusters or fewer. A set of SAS macros are also available for generating the Tonidandel and Overall (2004) procedure. In all cases, data were cluster analyzed using Ward's minimum variance method. Resampled data sets each consisted of 10,000 observations.

Mixture models represent another alternative to taxometrics for the identification of categorical structure. These analyses were conducted using Fraley and Raftery's (2002, 2003) MCLUST package. Mixture models treat class membership as missing data and provide maximum-likelihood estimates of class membership using the expectation-minimization algorithm. MCLUST is distinctive in several ways. First, mixture models are sensitive to the choice of initial values for the maximum-likelihood estimation. MCLUST uses hierarchical agglomerative clustering to generate reasonable estimates for the initial values. Second, clustering procedures generally are sensitive to the shape of the clusters, for example, whether they are spherical or ellipsoidal. MCLUST examines models that differ both in the number of clusters and in the distributional characteristics of those clusters. MCLUST computes the Bayesian Information Criterion (BIC) as a goodness-of-fit statistic for each model. The BIC penalizes more complex models, so it treats parsimony as a characteristic of an optimal model. The model associated with the largest BIC value is presented as the best estimate of the number of classes underlying the data.

The version of MCLUST written for R was used in the present study. Because of computer limitations, a random subsample of 5,000 observations was drawn from each data set for the mixture-model analyses. By default, MCLUST examines models with the number of classes set from one to nine. The one-class model represents the purely dimensional case.

## RESULTS

Each data set served as the basis for 12 MAMBAC graphs, 12 MAXCOV graphs, 4 MAXEIG graphs, 1 histogram of MAXCOV and 1 histogram of MAXEIG Bayesian posterior probabilities, and 1 L-Mode graph. Figures 1–6 provide either the single graph for each data set, or a sample graph from the set. In all cases where a procedure generated multiple graphs, the results were very similar. This was to be expected given the

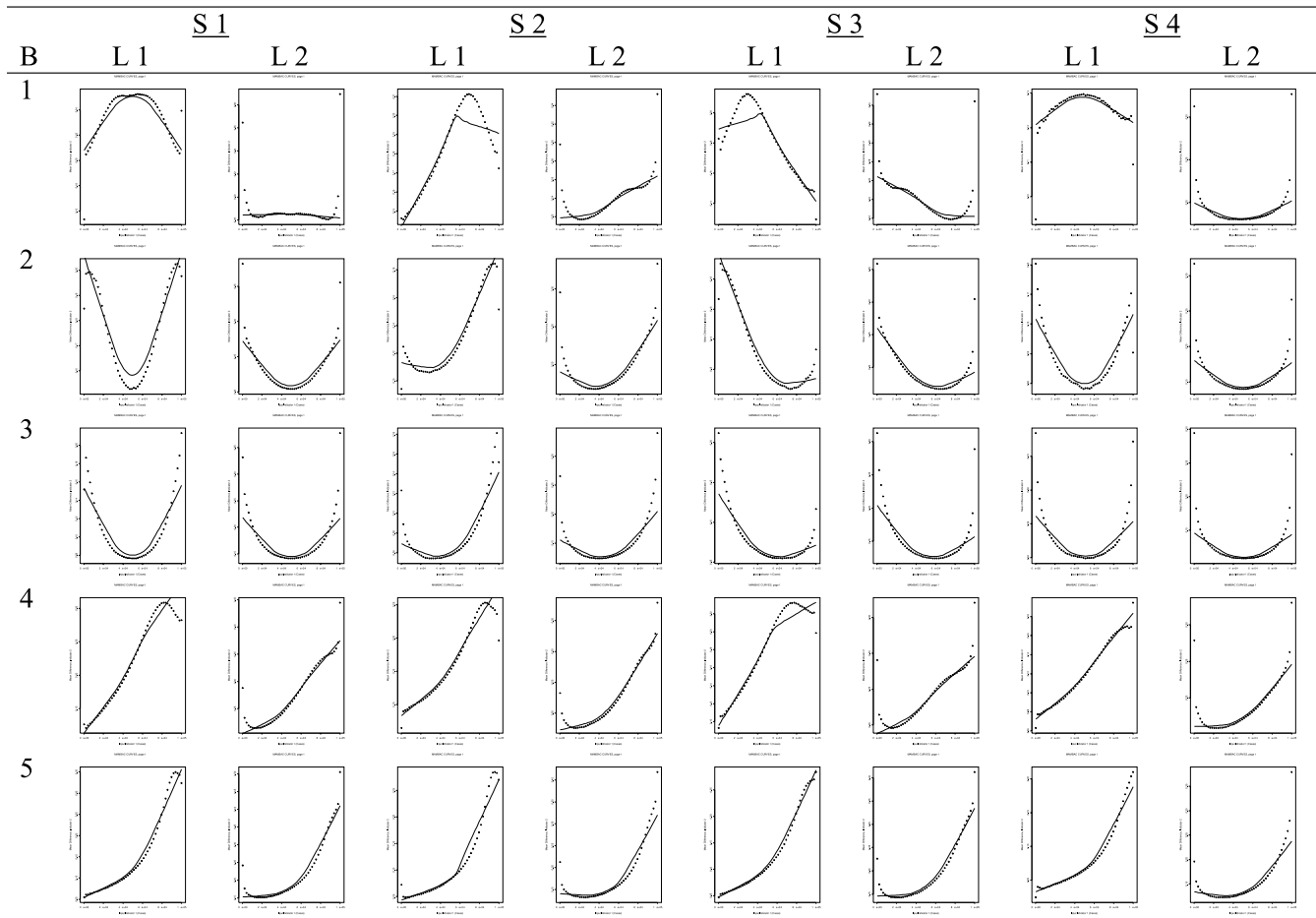


FIGURE 1.—Sample MAMBAC plot for each data set. Labels correspond to those in Table 1.

large sample size and the use of the same parameters for each indicator in a data set. The complete set of graphs generated for this study is available by request from the author.

MAXEIG, MAXCOV, and L-Mode graphs were smoothed using Tukey’s 3RS3R smoother, while lowest smoothing was used for MAMBAC. Dotted MAMBAC curves represent raw results, while smoothed curves are solid. Given the large sample size, the number of windows on the abscissa for MAXCOV and MAXEIG was increased to 5,000, and only the smoothed curve is presented to render the figures more legible.

For L-Mode graphs, taxometric software written using R finds the mode within each of two nonoverlapping ranges of scores within the distribution. These ranges usually are set from .001 to the lower limit of the factor scores, and from .001 to the upper limit (Waller & Meehl, 1998, Appendix D). In the bimodal and dimensional cases this is a reasonable approach. In the case where there is a third mode at zero, which should occur in symmetrical trichotomous distributions, the values .001 and .001 often were identified as the two modes, an outcome that caused an error in the taxometric program. Revising the standard L-Mode procedure so the medial boundaries for the lower and upper ranges were set to .5 and .5, respectively, eliminated this problem. Depending on the data set, this meant that the value identified for the lower mode was either .5 or  $Mo_3$ , while the upper mode was .5,  $Mo_1$ , or  $Mo_2$ .

Table 2 provides descriptive statistics associated with testing the hypothesis of taxonic structure for the 40 data sets. For MAMBAC, MAXCOV, and MAXEIG, the first column in the table provides the mean of the 4 (MAXEIG) or 12 (MAMBAC and MAXCOV) estimates of the taxon base rate, followed by the standard deviation of those estimates. For L-Mode, the first column instead is the mean of the two taxon base rate estimates derived from the modes of the curve, while the second column is the single base rate estimate derived from observation classification. For each taxometric method, the final column is the GFI.

The taxometric results will be discussed in two ways. Since the graphical output generally is treated as the central evidence concerning the structure of the data, the first section reviews the effect of trichotomous data on graphical output, so it is organized by the taxometric method. This section includes discussion of the effect of trichotomous data on the mathematical formulas that underlie the graphs. The second section addresses the issue of how a researcher might interpret the results if taxometric methods inadvertently are used with trichotomous data, and is organized by data set. This analysis is presented based on the graphical results alone, and on the integration of graphical and statistical outputs. The final section is a brief summary of outcomes from the four cluster-analytic procedures and from the finite mixture models.

TABLE 2.—Descriptive statistics for standard taxometric procedures.

Data set label	MAMBAC			MAXCOV			MAXEIG			L-Mode		
	BR <i>M</i>	BR <i>SD</i>	GFI	BR <i>M</i>	BR <i>SD</i>	GFI	BR <i>M</i>	BR <i>SD</i>	GFI	BRM	BRC	GFI
B1S1L1	0.501	0.002	0.855	0.450	0.076	0.815	0.483	0.086	0.815	0.505	0.503	0.855
B1S1L2	0.500	0.001	0.419	0.498	0.082	0.382	0.491	0.084	0.371	0.511	0.507	0.419
B1S2L1	0.421	0.008	0.962	0.382	0.009	0.953	0.383	0.008	0.953	0.449	0.400	0.966
B1S2L2	0.454	0.001	0.606	0.382	0.020	0.572	0.391	0.013	0.569	0.336	0.382	0.610
B1S3L1	0.580	0.006	0.962	0.621	0.011	0.956	0.612	0.006	0.953	0.565	0.609	0.968
B1S3L2	0.544	0.001	0.603	0.591	0.018	0.583	0.589	0.005	0.584	0.658	0.614	0.607
B1S4L1	0.501	0.002	0.992	0.478	0.042	0.981	0.493	0.038	0.980	0.495	0.497	0.992
B1S4L2	0.497	0.001	0.746	0.515	0.050	0.715	0.525	0.047	0.716	0.503	0.501	0.746
B2S1L1	0.500	0.003	0.856	0.574	0.238	0.797	0.499	0.268	0.794	0.498	0.498	0.856
B2S1L2	0.500	0.001	0.509	0.400	0.186	0.362	0.395	0.204	0.377	0.501	0.499	0.509
B2S2L1	0.368	0.004	0.961	0.230	0.012	0.958	0.232	0.013	0.957	0.498	0.427	0.946
B2S2L2	0.425	0.001	0.673	0.276	0.014	0.538	0.281	0.012	0.517	0.496	0.458	0.656
B2S3L1	0.630	0.004	0.959	0.765	0.011	0.957	0.768	0.009	0.953	0.501	0.570	0.945
B2S3L2	0.575	0.002	0.670	0.714	0.025	0.564	0.708	0.031	0.568	0.497	0.537	0.651
B2S4L1	0.499	0.005	0.994	0.383	0.195	0.969	0.295	0.083	0.960	0.497	0.496	0.994
B2S4L2	0.503	0.002	0.808	0.474	0.168	0.705	0.532	0.135	0.757	0.504	0.504	0.808
B3S1L1	0.499	0.006	0.938	0.374	0.349	0.839	0.318	0.358	0.843	0.500	0.501	0.938
B3S1L2	0.500	0.001	0.686	0.651	0.279	0.227	0.511	0.348	0.352	0.502	0.502	0.686
B3S2L1	0.339	0.008	0.988	0.119	0.013	0.970	0.125	0.005	0.976	0.498	0.451	0.983
B3S2L2	0.430	0.001	0.794	0.194	0.010	0.677	0.198	0.008	0.665	0.498	0.472	0.786
B3S3L1	0.659	0.004	0.989	0.878	0.011	0.973	0.877	0.005	0.969	0.502	0.547	0.984
B3S3L2	0.566	0.001	0.791	0.796	0.020	0.630	0.795	0.008	0.647	0.500	0.525	0.784
B3S4L1	0.505	0.008	1.000	0.311	0.331	0.888	0.316	0.353	0.907	0.501	0.501	1.000
B3S4L2	0.502	0.002	0.887	0.618	0.262	0.617	0.465	0.297	0.739	0.502	0.502	0.887
B4S1L1	0.252	0.004	0.896	0.293	0.015	0.906	0.298	0.017	0.904	0.305	0.302	0.923
B4S1L2	0.353	0.001	0.486	0.340	0.033	0.447	0.328	0.030	0.408	0.543	0.427	0.310
B4S2L1	0.229	0.003	0.977	0.241	0.009	0.970	0.241	0.004	0.970	0.271	0.243	0.978
B4S2L2	0.348	0.002	0.613	0.273	0.012	0.528	0.280	0.015	0.541	0.499	0.409	0.500
B4S3L1	0.298	0.004	0.961	0.347	0.012	0.980	0.350	0.006	0.980	0.517	0.408	0.973
B4S3L2	0.387	0.002	0.638	0.377	0.020	0.619	0.386	0.020	0.616	0.377	0.378	0.639
B4S4L1	0.281	0.004	0.994	0.280	0.025	0.986	0.281	0.004	0.986	0.550	0.431	0.982
B4S4L2	0.396	0.001	0.752	0.318	0.018	0.688	0.326	0.043	0.686	0.505	0.449	0.719
B5S1L1	0.133	0.002	0.966	0.150	0.006	0.967	0.153	0.005	0.967	0.176	0.158	0.976
B5S1L2	0.273	0.001	0.545	0.195	0.008	0.397	0.198	0.007	0.395	0.502	0.376	0.291
B5S2L1	0.132	0.006	0.993	0.120	0.002	0.988	0.122	0.002	0.988	0.501	0.341	0.873
B5S2L2	0.303	0.002	0.672	0.178	0.007	0.551	0.178	0.007	0.522	0.501	0.411	0.555
B5S3L1	0.145	0.004	0.984	0.169	0.006	0.990	0.170	0.005	0.989	0.188	0.183	0.996
B5S3L2	0.306	0.001	0.668	0.211	0.004	0.542	0.214	0.006	0.550	0.503	0.400	0.520
B5S4L1	0.156	0.004	0.999	0.135	0.011	0.991	0.138	0.010	0.989	0.498	0.375	0.978
B5S4L2	0.360	0.001	0.794	0.210	0.020	0.621	0.203	0.003	0.611	0.495	0.438	0.756

Note. Data Set Labels correspond to those in Table 1. BR = estimates of taxon base rate; GFI = goodness of fit index; BRM = mean of the two base rate estimates derived from the L-Mode modes; BRC = base rate estimate derived from L-Mode classification of observations.

*The Effect of Trichotomous Data on Taxometric Graphs*

**MAMBAC.** Appendix B demonstrates that in the case of three ordered classes, the values on the ordinate of the MAMBAC curve are determined by

$$(p_{1a} - p_{1b})(\bar{y}_1 - \bar{y}_3) + (p_{2a} - p_{2b})(\bar{y}_2 - \bar{y}_3), \tag{2}$$

where  $p_{1a}$  is the proportion of observations above the cut belonging to the class with the highest mean,  $p_{2b}$  is the proportion of observations below the cut belonging to the intermediate class,  $\bar{y}_3$  is the mean score for observations in the class with the lowest mean, and so forth.

This formula creates a more complicated situation than is true in the two-class case. Low base rates for the extreme classes tend to produce elevations in the graph near the endpoints. This pattern is demonstrated in Figure 1. The rectangular distribution (B1: see Table 1 for data pattern labels) produced results that are likely to be interpreted as evidence of a dichotomy,

though this pattern was washed out when loadings were large (L2). As the base rates of the extreme classes declined, however, the corresponding portion of the curve became more elevated relative to the rest of the distribution. B4, which was mildly skewed, produced results that at times could be interpreted as evidence for a small taxon. The bell-shaped distributions, B2 and B3, produced graphs that were elevated at both endpoints, a result that would be interpreted as strongly suggestive of dimensional structure according to standard MAMBAC practice. B5, the most skewed distribution, produced results that would be interpreted as dimensional.

**MAXCOV and MAXEIG.** Since they are closely related statistical procedures, it is not surprising to find MAXCOV and MAXEIG produced parallel results, and so will be discussed in combination. According to Equation 1, if nuisance covariance is negligible, variations in the covariance between  $y$  and  $z$  across

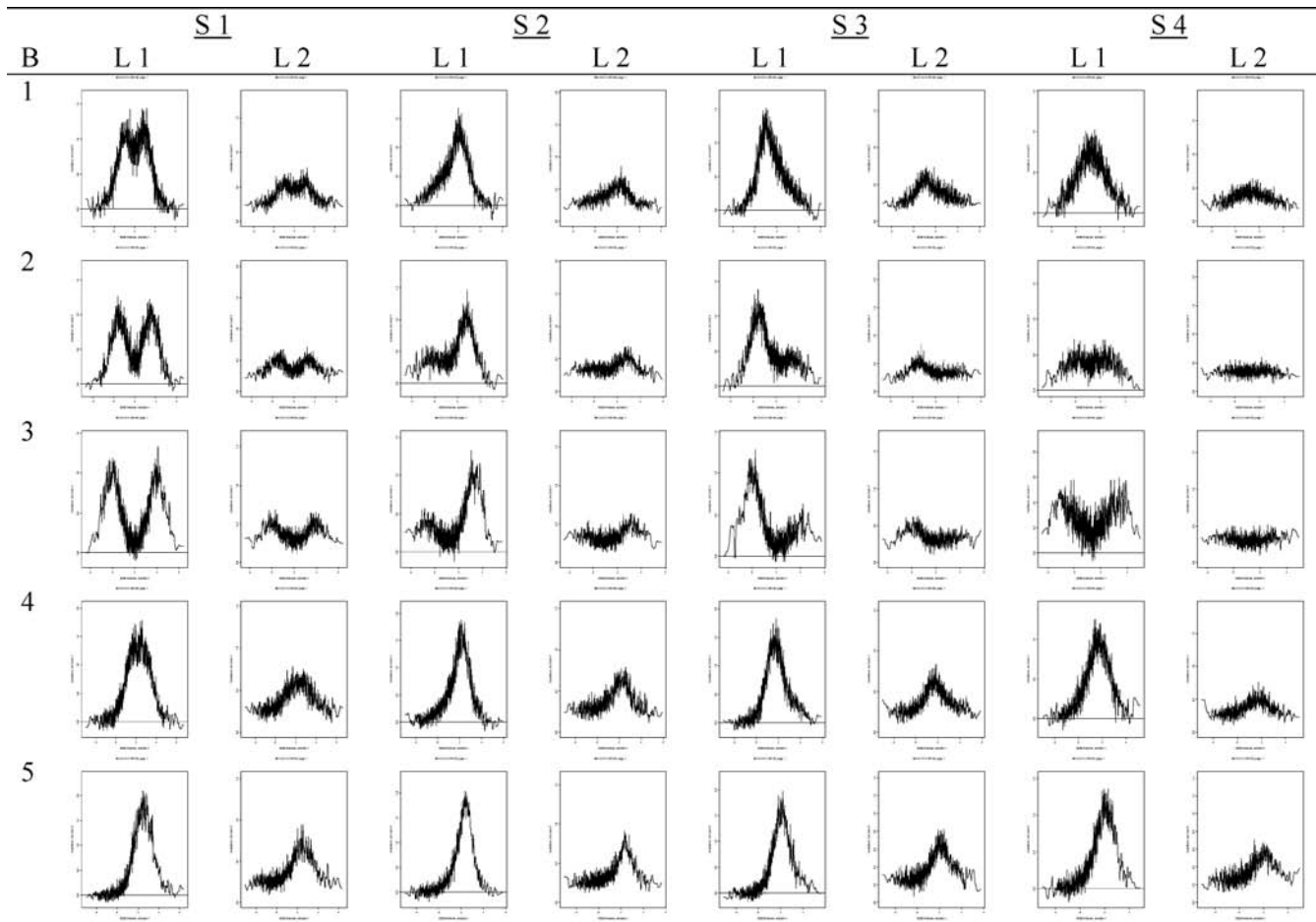


FIGURE 2.—Sample MAXCOV plot for each data set. Labels correspond to those in Table 1.

window depend on

$$p_1 p_2 (\bar{y}_1 - \bar{y}_2)(\bar{z}_1 - \bar{z}_2) + p_1 p_3 (\bar{y}_1 - \bar{y}_3)(\bar{z}_1 - \bar{z}_3) + p_2 p_3 (\bar{y}_2 - \bar{y}_3)(\bar{z}_2 - \bar{z}_3),$$

when there are three classes. Figures 2 and 3 demonstrate what happens in those circumstances. When the distribution of the classes is symmetrical and the separations between classes are reasonably large, the curve will tend toward a bimodal M shape (two convex curves and one concave curve), with the modes indicating the windows in which  $p_1$  and  $p_3$  are at their relative maxima. This pattern is unique to the trichotomous case. With smaller separations, higher levels of nuisance covariation, or more skewed distributions, however, the results instead tend to be consistent with what one would expect for a dichotomous construct.

Similarly, when there are three classes, the Bayesian probabilities based on the assumption that two classes are present ideally would form three clusters, reflecting observations where there is consistent evidence across indicators of falling in the taxon class, observations where there is consistent evidence across indicators of falling in the complement class, and observations where results are mixed across indicators because they belong to the intermediate class. In the case of the sym-

metrical distributions, the results tended to be consistent with this expectation, though higher levels of nuisance covariation and lower separations tended to cloud the picture (Figures 4 and 5). Even in those instances where the histogram suggested three clusters, however, the results were problematic. First, the frequencies for the three clusters (the Y-axis variable) were consistently incorrect as estimates of the base rates for the three classes. For example, in Figure 4 the actual base rates for data set B2S1L1 were .20, .60, and .20. The probability of falling in the lowest cluster correctly was estimated to be approximately .20. The histogram instead suggested approximately 10% of the data set fell in the intermediate cluster, while almost 70% were included in the highest cluster. So the results cannot be used directly to estimate the true base rate of a three-class data set. Second, Waller and Meehl (1998, Figure 3.4) suggested a three-clustered histogram potentially is interpretable as evidence of nontaxonic data. This is the appropriate conclusion if the possibility of a three-class structure has been ruled out, as is true of their example, but it potentially is inaccurate if it has not.

*L-Mode.* In the case of three ordered classes, the smoothed density plot of the standardized factor scores should demonstrate three peaks. As demonstrated in Appendix C, those three modes



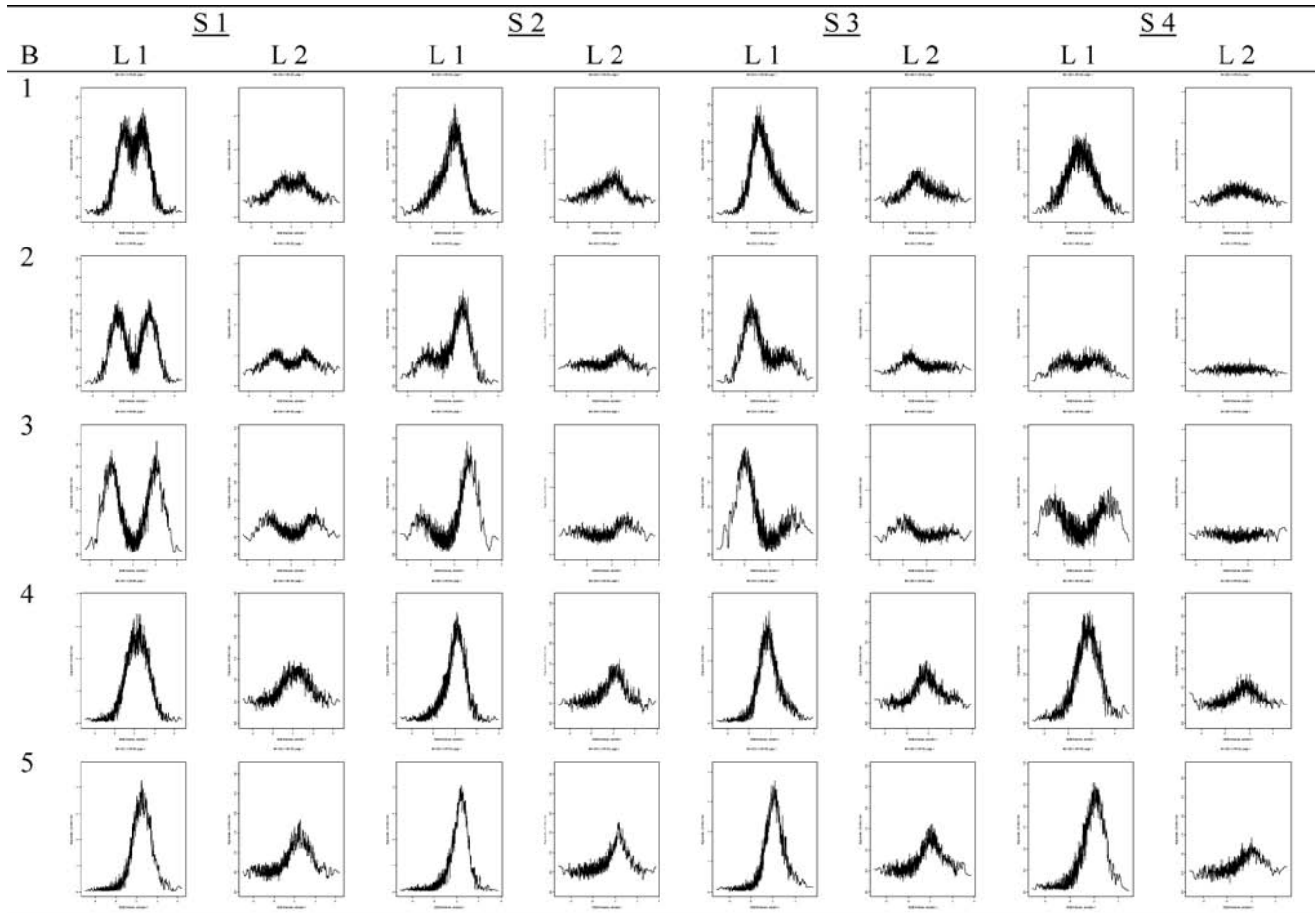


FIGURE 3.—Sample MAXEIG plot for each data set. Labels correspond to those in Table 1.

are expected to occur at the following locations:

$$E(Mo_1) = \frac{1 - (p_1 + .5p_2)}{\sqrt{p_1 + .25p_2 - (p_1 + .5p_2)^2}},$$

$$E(Mo_2) = \frac{.5 - (p_1 + .5p_2)}{\sqrt{p_1 + .25p_2 - (p_1 + .5p_2)^2}},$$

$$E(Mo_3) = \frac{-(p_1 + .5p_2)}{\sqrt{p_1 + .25p_2 - (p_1 + .5p_2)^2}},$$

although many factor score solutions will introduce bias into these estimates.

Figure 6 provides the L-Mode graph for each data set. In the best case, a pattern emerges revealing a “Halloween ghost” with arms raised, reflecting three prominent modes. Even when the results were not quite as clear, the pattern of three convex curves interspersed with two concave curves tended to emerge when the separations were high and the loadings were low. As the separations were reduced and the nuisance covariation was increased, the distributions became more bimodal, likely to be interpreted as a dichotomous structure, and ultimately unimodal, suggestive of dimensional structure.

*The Interpretation of Trichotomous Data via Taxometric Methods*

Table 3 provides summative results for the graphs associated with each data set. As described earlier, some graph shapes were idiosyncratic to trichotomous data, so these were reported separately. Histograms of Bayesian probabilities that produced three clusters were included in this category, although prior practice might have led researchers to interpret this outcome as consistent with dimensional structure. Since the multiple graphs generated by MAMBAC, MAXCOV, and MAXEIG largely were redundant, a single conclusion was drawn from each method to simplify the discussion.

The literature suggests researchers approach decision making in taxometric analysis several different ways. As the first taxometric methods to be discussed in full monographs (Meehl & Yonce, 1994, 1996), some studies have examined only MAMBAC and MAXCOV graphs (the Bayesian histogram was a later addition to the latter). For 7 of 40 data sets, the outcomes would have led the researcher to conclude that the data reflect a dichotomous structure, while 2 data sets supported a dimensional interpretation. In 11 cases the results for MAXCOV were inconsistent with either interpretation, while the two methods led to different conclusions in the remaining 20 data sets.

The results were more troubling when the full range of graphical outcomes is considered. If agreement among at least 5 of

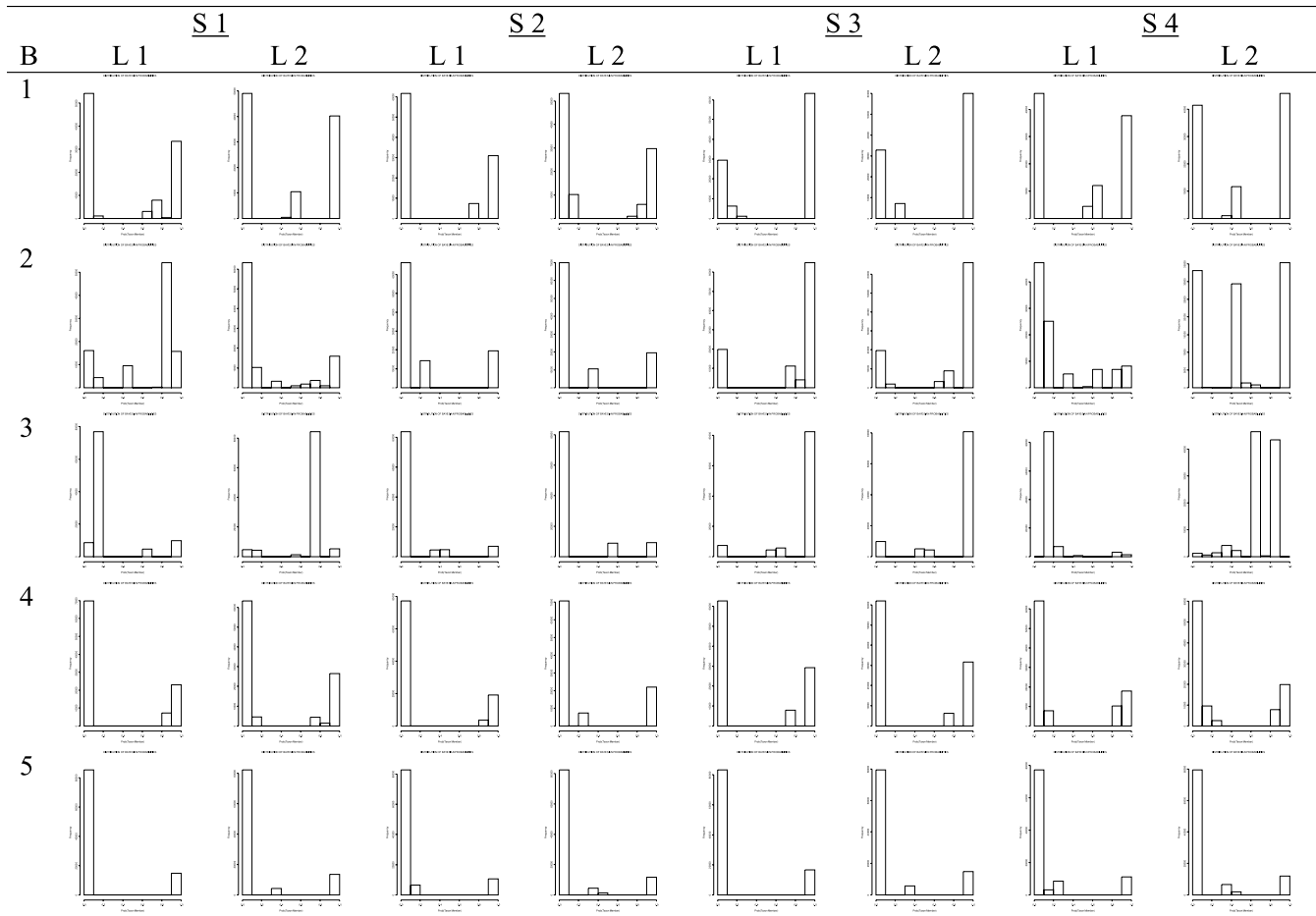


FIGURE 4.—Distribution of MAXCOV Bayesian posterior probabilities for each data set. Labels correspond to those in Table 1.

the 6 sets of outcomes is taken as a standard for a consistent outcome, 14 out of 40 data sets produced consistent support for a dichotomous interpretation, while 2 supported a dimensional interpretation. Some data sets consistently produced patterns that did not meet with expectations for either dimensional or dichotomous structure. Even if we set the standard for convergence at four outcomes (since there was no MAMBAC pattern idiosyncratic to trichotomous data), however, only 5 data sets produced consistent evidence of trichotomous structure.

In those cases where the graphical results might have been interpreted as evidence of dichotomous structure, the statistical outcomes in Table 2 also tended to be consistent with that interpretation. This is not surprising, since those estimates are derived from the graphs. For example, all taxometric methods except L-Mode produced graphs for data set B1S2L1 that were consistently supportive of dichotomous structure. Furthermore, the MAMBAC, MAXCOV, and MAXEIG graphs all suggested the base rate of the two classes would be approximately equal. According to Table 2, there was relatively little variability in taxon base rate estimates across or within taxometric methods, all converging on a value around .40. Furthermore, all four GFI values exceeded .90 by a substantial amount. In cases where the various graphical results converge on one interpretation, the statistical outcomes simply tend to reinforce that interpretation.

From both perspectives, the majority of data sets produced inconsistent results or results that were idiosyncratic to trichotomous data. It is important to consider, however, how such evidence likely is to be used in current practice. It was noted earlier that the Bayesian histograms identifying three clear clusters probably would be interpreted as evidence of dimensional structure. This example reflects a problem with current practice in taxometric analysis. Based on simulation studies restricted to dimensional and dichotomous data, taxometric researchers tend to treat these as the only two possible options. One consequence of this practice is the tendency to use the hypothesis of dimensional structure as if it were a null hypothesis that can be rejected only if there is consistent evidence of dichotomous structure (Ruscio & Ruscio, 2004a). As a result, outcomes that vary between supporting a dichotomous and dimensional interpretation are likely to be treated as evidence of dimensional structure. In the case where patterns consistently emerge that are idiosyncratic to trichotomous data, the failure to consider other structures might lead the researcher to decide not to pursue publication at all.

In summary, the application of current taxometric methods to trichotomous data resulted in either inconsistent outcomes or outcomes supportive of trichotomous structure in the majority of data sets. Currently, the former finding may well be interpreted as evidence of dimensional structure, while the latter simply may lead the researcher to give up the project completely. In

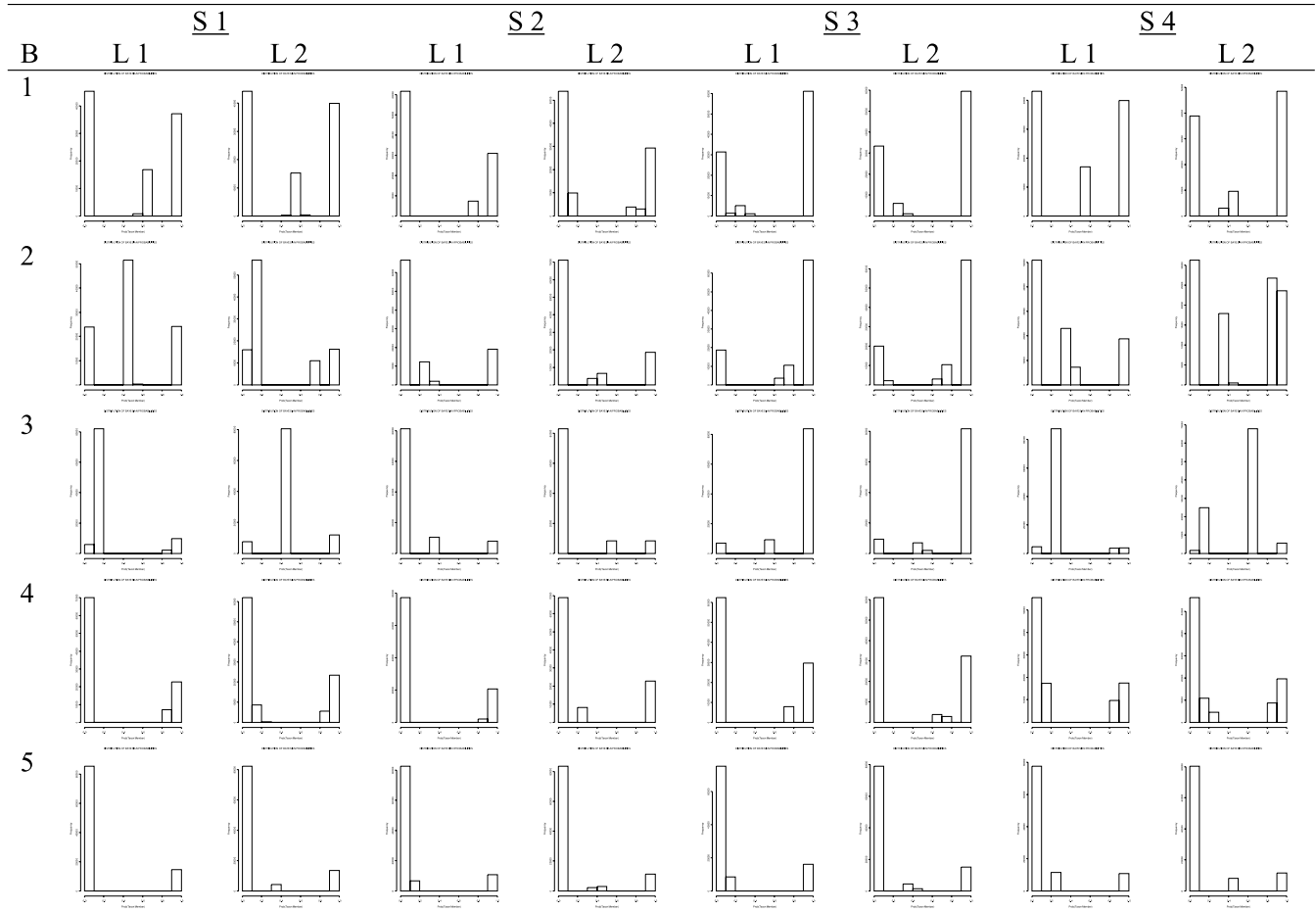


FIGURE 5.—Distribution of MAXEIG Bayesian posterior probabilities for each data set. Labels correspond to those in Table 1.

a number of data sets, the results consistently supported the conclusion that the data were dichotomous or dimensional. In all cases, current practice in taxometric research would lead the researcher to an incorrect or unfortunate conclusion.

*Alternative Analyses*

As noted earlier, the three methods for evaluating class structure that were available through SAS (CCC, pseudo- $F$ , and pseudo- $t^2$ ) do not always converge, so it has been recommended that they be used in conjunction with one another to draw a conclusion about structure. At least two procedures offered evidence of three true clusters in 19 of the data sets. Accuracy was greater in symmetrical distributions: 15 out of 24 distributions were correctly identified as having three clusters, while only 4 of 16 skewed distributions were. The degree of nuisance covariation seemed less of a factor: 11 of 20 data sets with small factor loadings and 8 out of 20 with large loadings were correctly identified.

The Tonidandel and Overall (2004) procedure, which is the most recently introduced alternative, performed poorly with the present simulations. Only the rectangular data sets with small loadings were identified correctly as including three data clusters. In all other cases the results suggested dimensional structure.

The most effective method proved to be MCLUST, which correctly identified 26 of the data sets as consisting of three classes. Nuisance covariation seemed more problematic for MCLUST than for the SAS cluster-analytic methods: 18 out of 20 data sets with factor loadings of .001 were correctly identified, but only 8 of 20 data sets with loadings of .70 were. The only two cases where a data set with small loadings was not correctly identified occurred in the most skewed sample when at least one separation was set to 1.25.

Although none of the three approaches emerged as a “gold standard” for the presence of trichotomous structure, a combination of the three SAS cluster-analytic options for drawing inferences about structure and the MCLUST package functioned best. If the data were assumed to be trichotomous on the basis of either MCLUST or the SAS procedures suggesting three-class structure, then 30 out of 40 data sets correctly were identified. Of course, the present study cannot address what the false positive rate for this strategy would be if the sample included other structures beside trichotomous data.

DISCUSSION

Taxometric analysis was developed specifically to differentiate between dimensional and dichotomous structure. When the data are limited to these two options, taxometric methods seem to perform quite well; in fact, evidence suggests it has several

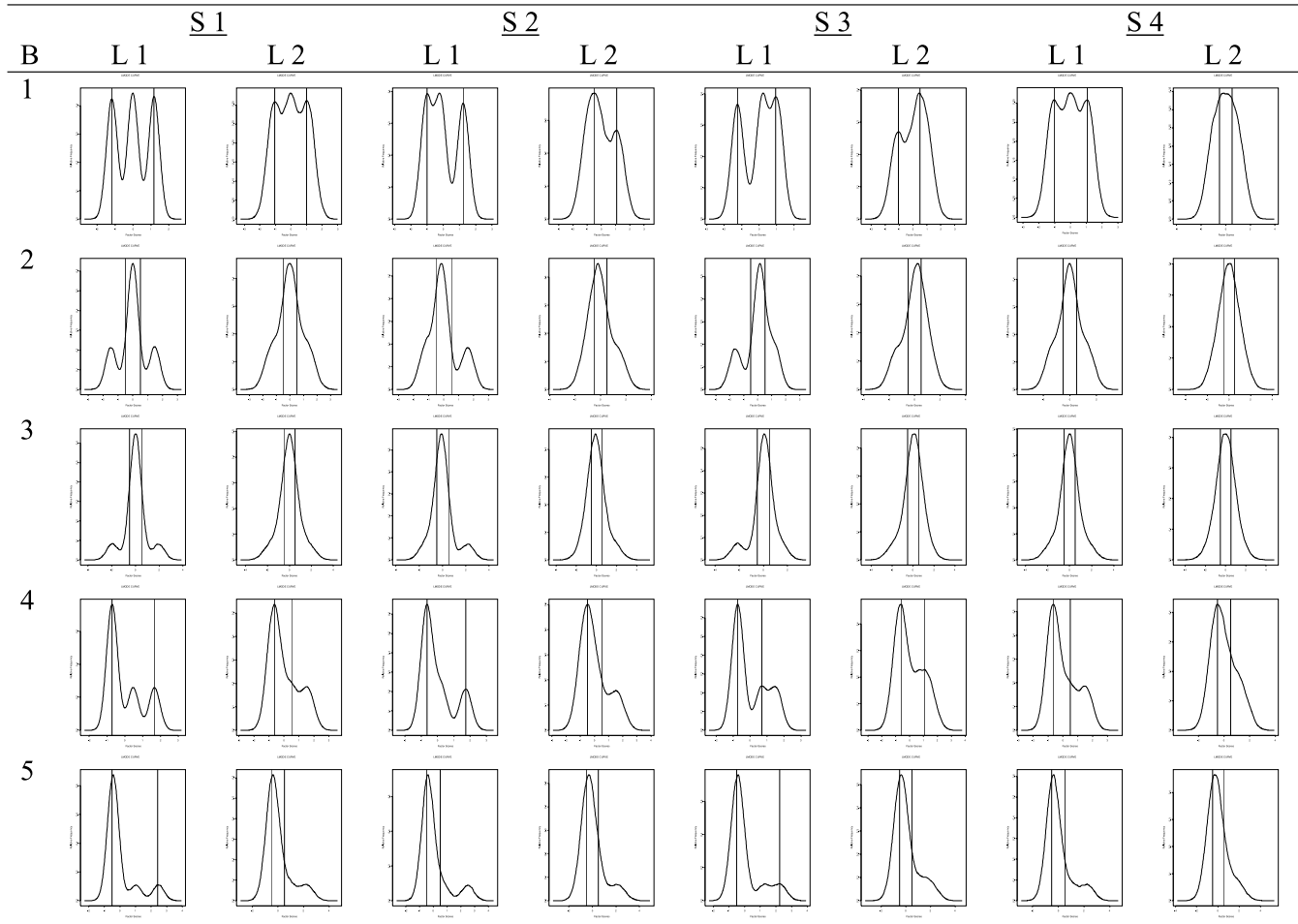


FIGURE 6.—L-Mode plot for each data set. Labels correspond to those in Table 1.

advantages over alternative methods such as cluster analysis and finite mixture models and often produces more accurate results (Ruscio et al., 2006).

These two options, however, do not exhaust the universe of possible data structures. The present study explored what happens when researchers inadvertently use taxometric methods when the construct they are investigating consists of three ordered classes and the indicators selected are sensitive to both categorical distinctions. The results suggest it would not be unusual for researchers using current standards for the interpretation of taxometric methods to conclude incorrectly that the data are dichotomous or dimensional. Alternatively, the researcher might conclude the results are uninterpretable, when in fact they suggest a more complex structure to the data.

It was stated earlier that there is no way to know how frequently psychologists interested in latent structure actually are dealing with ordinal polytomous data, and with manifest indicators that are sensitive to each qualitative distinction. It is therefore unclear to what extent this is an important threat to statistical conclusion validity for researchers interested in using taxometric methods to detect the true structure of psychosocial constructs. Even so, the present findings should raise concern about the use of taxometric methods when the possibility of polytomous structure exists.

Several recommendations emerge from these findings. Methodological commentaries on taxometric methods regularly recommend their use only when there is sufficient reason to consider the issue of dimensional versus categorical latent structure is an important one for the construct of interest (Lenzenweger, 2004; Ruscio et al., 2006). The results of the present study would suggest even further reflection is necessary about the potential for polytomous structure before a researcher uses taxometric methods. If there is a reasonable possibility of three or even more classes given the measures and population under investigation, it would be best to think twice before relying solely or primarily on taxometric methods.

Second, the risk of including more than two classes is likely to be smaller if the population is relatively homogeneous. Similarly, the risk of using measures that are sensitive to more than one qualitative difference is likely to be lower when measures reflect the full network of subdomains associated with the target construct. Earlier in the article it was suggested that research finds the schizotypal dimension may encompass three to four classes. This finding was only possible because the samples involved spanned the spectrum from normal to negative symptom schizophrenia. It also depended on the use of indicators specifically reflective of negative symptoms. As the sample becomes more homogeneous, or the indicators become more diverse,

TABLE 3.—Results from graphical analyses.

Label	MAMBAC	MAXCOV		MAXEIG		L-Mode	Totals		
		Graphs	Bayes	Graphs	Bayes		Dim	Dich	Other
B1S1L1	Dich	Other	Dich	Other	Other	Other	0	2	4
B1S1L2	Dim	Other	Other	Other	Other	Other	1	0	5
B1S2L1	Dich	Dich	Dich	Dich	Dich	Other	0	5	1
B1S2L2	Dim	Dich	Dich	Dich	Dich	Dich	1	5	0
B1S3L1	Dich	Dich	Dich	Dich	Dich	Other	0	5	1
B1S3L2	Dim	Dich	Dich	Dich	Dich	Dich	1	5	0
B1S4L1	Dich	Dich	Other	Dich	Other	Other	0	3	3
B1S4L2	Dim	Dich	Other	Dich	Other	Dim	2	2	2
B2S1L1	Dim	Other	Other	Other	Other	Other	1	0	5
B2S1L2	Dim	Other	Dim	Other	Dich	Dim	3	1	2
B2S2L1	Dim	Other	Dich	Other	Dich	Dich	1	3	2
B2S2L2	Dim	Dich	Other	Dich	Other	Dim	2	2	2
B2S3L1	Dim	Other	Dich	Other	Dich	Dich	1	3	2
B2S3L2	Dim	Dich	Dich	Dich	Dich	Dim	2	4	0
B2S4L1	Dim	Other	Dim	Other	Other	Dim	3	0	3
B2S4L2	Dim	Dim	Other	Dim	Other	Dim	4	0	2
B3S1L1	Dim	Other	Other	Other	Dich	Other	1	1	4
B3S1L2	Dim	Other	Dich	Other	Other	Dim	2	1	3
B3S2L1	Dim	Other	Other	Other	Other	Dich	1	1	4
B3S2L2	Dim	Dich	Other	Dich	Other	Dim	2	2	2
B3S3L1	Dim	Dich	Other	Other	Other	Dich	1	2	3
B3S3L2	Dim	Dich	Other	Dich	Other	Dim	2	2	2
B3S4L1	Dim	Other	Dich	Other	Dich	Dim	2	2	2
B3S4L2	Dim	Dim	Dim	Dim	Other	Dim	5	0	1
B4S1L1	Dich	Dich	Dich	Dich	Dich	Other	0	5	1
B4S1L2	Dim	Dich	Dich	Dich	Dich	Dich	1	5	0
B4S2L1	Dich	Dich	Dich	Dich	Dich	Dich	0	6	0
B4S2L2	Dim	Dich	Dich	Dich	Dich	Dich	1	5	0
B4S3L1	Dich	Dich	Dich	Dich	Dich	Other	0	5	1
B4S3L2	Dim	Dich	Dich	Dich	Dich	Dich	1	5	0
B4S4L1	Dich	Dich	Dich	Dich	Dich	Dich	0	6	0
B4S4L2	Dim	Dich	Dich	Dich	Dich	Dim	2	4	0
B5S1L1	Dim	Dich	Dich	Dich	Dich	Other	1	4	1
B5S1L2	Dim	Dich	Other	Dich	Other	Dich	1	3	2
B5S2L1	Dim	Dich	Dich	Dich	Dich	Dich	1	5	0
B5S2L2	Dim	Dich	Other	Dich	Other	Dich	1	3	2
B5S3L1	Dim	Dich	Dich	Dich	Dich	Dich	1	5	0
B5S3L2	Dim	Dich	Other	Dich	Other	Dim	2	2	2
B5S4L1	Dim	Dich	Dich	Dich	Dich	Dich	1	5	0
B5S4L2	Dim	Dich	Other	Dich	Other	Dim	2	2	2

Note. Data Set Labels correspond to those in Table 1. Dim = likely to be interpreted as supportive of dimensional structure, Dich = likely to be interpreted as supportive of taxonic structure, Other = the pattern is idiosyncratic to three-class data.

the two conditions underlying the simulations presented in this study (polytomous structure and indicators sensitive to all discriminations) become less likely.

Third, many researchers have relied on one to three sets of graphical taxometric outputs. The present results suggest the value of using as many taxometric methods as are available. Ideally, data consisting of  $k$  classes would produce an L-Mode graph with  $k$  convex curves, histograms of Bayesian posterior probabilities demonstrating  $k$  clusters, and MAXCOV and MAXEIG graphs demonstrating  $k - 1$  convex curves. In fact, the present results suggest that even one of these outcomes should lead the researcher to consider seriously the possibility of  $k$ -class structure.

Fourth, one might conclude from these results that researchers interested in detecting even a dichotomous structure would be better off avoiding the use of taxometric methods altogether, but this is an undesirable conclusion. Taxometric methods consistently have been found to be more accurate at distinguishing between dichotomous and dimensional data than alternatives such as cluster analysis and mixture models (e.g., Cle-

land et al., 2000; Waller & Meehl, 1998), and the latter tend to have more requirements in terms of the population distribution. Furthermore, the variety of consistency tests that have been developed for use in taxometric analysis is unparalleled in statistical practice. Finally, the present results suggest no method provides a gold standard for detecting polytomous structure.

A better alternative would suggest expanding the concept of consistency tests to encompass nontaxometric alternatives. Convergence across multiple statistical models clearly would present a stronger case for whatever conclusions are appropriate. That being said, the likelihood of convergence across statistical methods is not that great because they approach the detection of data structure from very different perspectives. While taxometric analysis requires that each indicator be independently sensitive to the categorical discrimination, cluster analysis allows for the possibility that some distinctions are based on combinations of indicators. Cluster analysis and mixture models treat the discrimination of dimensional from class models as secondary to the issue of determining the number of classes. Where

Tonidandel and Overall (2004) tend to select more over less complex models, the reverse is true for MCLUST. Given these differences, lack of convergence between alternative approaches to detecting structure may be the norm rather than the exception. The issue of how best to identify data structure remains an unresolved issue.

Finally, although the purpose of this study was to raise caution about the use of taxometric methods in circumstances where the construct is neither dichotomous nor dimensional, it is reasonable to ask whether it would be possible to develop generalized versions of the taxometric methods that are relevant in cases of more than two classes. This would be a potentially valuable addition to the current stable of statistical methods. As the preceding makes clear, however, doing so would require substantial revision of both the formalism and the methodology underlying taxometric methods. In the meantime, researchers should be more thoughtful about whether taxometric methods offer an appropriate statistical approach, and whether there are other conclusions possible besides the two typically considered in relation to taxometric analyses. Here, as in all circumstances, Meehl (1999, p. 172) was correct in stating that “no statistic is self-interpreting.”

#### ACKNOWLEDGMENT

I am grateful to John Ruscio for his permission to use the statistical package described in this article.

#### REFERENCES

- Beach, S. R. H., Amir, N., & Bau, J. J. (2005). Can sample-specific simulations help detect low base-rate taxonomicity? *Psychological Assessment, 17*, 446–461.
- Blanchard, J. J., Gangestad, S. W., Brown, S. A., & Horan, W. P. (2000). Hedonic capacity and schizotypy revisited: A taxometric analysis of social anhedonia. *Journal of Abnormal Psychology, 109*, 87–95.
- Blanchard, J. J., Horan, W. P., & Collins, L. M. (2005). Examining the latent structure of negative symptoms: Is there a distinct subtype of negative symptom schizophrenia? *Schizophrenia Research, 77*, 151–165.
- Cleland, C. M., Rothschild, L., & Haslam, N. (2000). Detecting latent taxa: Monte Carlo comparison of taxometric, mixture model, and clustering procedures. *Psychological Reports, 87*, 37–47.
- Erlenmeyer-Kimling, L., Golden, R. R., & Cornblatt, B. A. (1989). A taxometric analysis of cognitive and neuromotor variables in children at risk for schizophrenia. *Journal of Abnormal Psychology, 98*, 203–208.
- Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association, 97*, 611–631.
- Fraley, C., & Raftery, A. E. (2003). Enhanced software for model-based clustering, discriminant analysis, and density estimation: MCLUST. *Journal of Classification, 20*, 263–286.
- Franklin, C. L., Strong, D. R., & Greene, R. L. (2002). A taxometric analysis of the MMPI-2 Depression Scales. *Journal of Personality Assessment, 79*, 110–121.
- Haertel, E. H. (1990). Continuous and discrete latent structure models for item response data. *Psychometrika, 55*, 477–494.
- Haslam, N. (2003). Categorical versus dimensional models of mental disorder: The taxometric evidence. *Australian and New Zealand Journal of Psychiatry, 37*, 696–704.
- Haslam, N., & Kim, H. C. (2002). Categories and continua: A review of taxometric research. *Genetic, Social, and General Psychology Monographs, 128*, 271–320.
- Lenzenweger, M. F. (2004). Consideration of the challenges, complications, and pitfalls of taxometric analysis. *Journal of Abnormal Psychology, 113*, 10–23.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 806–834.
- Meehl, P. E. (1999). Clarifications about taxometric method. *Applied and Preventive Psychology, 9*, 165–174.
- Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonomicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports, 74*, 1059–1274.
- Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonomicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports, 78*, 1091–1227.
- Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50*, 159–179.
- Ruscio, J., Haslam, N., & Ruscio, A. M. (2006). *Introduction to the taxometric method: A practical guide*. Mahwah, NJ: Erlbaum.
- Ruscio, J., & Ruscio, A. M. (2002). A structure-based approach to psychological assessment: Matching measurement models to latent structure. *Assessment, 9*, 4–16.
- Ruscio, J., & Ruscio, A. M. (2004a). A conceptual and methodological checklist for conducting a taxometric investigation. *Behavior Therapy, 35*, 403–447.
- Ruscio, J., & Ruscio, A. M. (2004). A nontechnical introduction to the taxometric method. *Understanding Statistics, 3*, 151–194.
- Sarle, W. S., & Kuo, A.-H. (1993). *The MODECLUS procedure: SAS Technical Report P-256*. Cary, NC: SAS Institute.
- SAS Institute. (2004). *SAS Online Doc 9.1.3*. Cary, NC: Author.
- Schmidt, N. B., Kotov, R., & Joiner, T. E., Jr. (2004). *Taxometrics: Toward a new diagnostic scheme for psychopathology*. Washington DC: American Psychological Association.
- Steinley, D. (2004). Properties of the Hubert-Arable Adjusted Rand Index. *Psychological Methods, 9*, 386–396.
- Strong, D. R., Greene, R. L., Hoppe, C., Johnston, T., & Olesen, N. (1999). Taxometric analysis of impression management and self-deception on the MMPI-2 in child-custody litigants. *Journal of Personality Assessment, 73*, 1–18.
- Strong, D. R., Greene, R. L., & Kordinak, S. T. (2002). Taxometric analysis of impression management and self-deception in college student and personnel evaluation settings. *Journal of Personality Assessment, 78*, 161–175.
- Tonidandel, S., & Overall, J. E. (2004). Determining the number of clusters by sampling with replacement. *Psychological Methods, 9*, 238–249.
- Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Thousand Oaks, CA: Sage.
- Waller, N. G., Underhill, J. M., & Kaiser, H. A. (1999). A method for generating simulated plasmodes and artificial test clusters with user-defined shape, size, and orientation. *Multivariate Behavioral Research, 34*, 123–142.

#### APPENDIX A

##### Derivation of the Generalized General Covariance Mixture Theorem

The formula for the covariance of two variables equals:

$$\begin{aligned} \text{cov}(yz) &= \frac{\sum(y - \bar{y})(z - \bar{z})}{N} \\ &= (\bar{y}\bar{z}) - (\bar{y})(\bar{z}). \end{aligned}$$

That is, the covariance equals the difference between the mean cross product and the product of the means. In the  $k$ -class case, the terms of this second equation can be restated as follows:

$$\begin{aligned} (\bar{y}\bar{z}) &= p_1(\bar{y}\bar{z})_1 + p_2(\bar{y}\bar{z})_2 + \dots + p_k(\bar{y}\bar{z})_k, \\ \bar{y} &= p_1\bar{y}_1 + p_2\bar{y}_2 + \dots + p_k\bar{y}_k, \\ \bar{z} &= p_1\bar{z}_1 + p_2\bar{z}_2 + \dots + p_k\bar{z}_k. \end{aligned}$$

The generalization of the General Covariance Mixture Theorem to the polytomous case indicates that the covariance of two

indicators of a common taxonic construct is expected to equal the following:

$$\begin{aligned}
\text{cov}(yz) &= p_1(\bar{y}\bar{z})_1 + p_2(\bar{y}\bar{z})_2 + \dots p_k(\bar{y}\bar{z})_k - (p_1\bar{y}_1 + p_2\bar{y}_2 \\
&\quad + \dots p_k\bar{y}_k)(p_1\bar{z}_1 + p_2\bar{z}_2 + \dots p_k\bar{z}_k) \\
&= p_1(\bar{y}\bar{z})_1 + p_2(\bar{y}\bar{z})_2 + \dots p_k(\bar{y}\bar{z})_k \\
&\quad - \left[ \sum_i^k p_i^2 \bar{y}_i \bar{z}_i + \sum_i^k \sum_{j \neq i}^k p_i p_j \bar{y}_i \bar{z}_j \right] \\
&= p_1(\bar{y}\bar{z})_1 + p_2(\bar{y}\bar{z})_2 + \dots p_k(\bar{y}\bar{z})_k \\
&\quad + \sum_i^k (p_i - p_i - p_i^2) \bar{y}_i \bar{z}_i - \sum_i^k \sum_{j \neq i}^k p_i p_j \bar{y}_i \bar{z}_j \\
&= p_1(\bar{y}\bar{z})_1 - p_1(\bar{y}_1)(\bar{z}_1) + p_2(\bar{y}\bar{z})_2 - p_2(\bar{y}_2)(\bar{z}_2) \\
&\quad + \dots p_k(\bar{y}\bar{z})_k - p_k(\bar{y}_k)(\bar{z}_k) + \sum_i^k p_i(1 - p_i) \bar{y}_i \bar{z}_i \\
&\quad - \sum_i^k \sum_{j \neq i}^k p_i p_j \bar{y}_i \bar{z}_j \\
&= p_1 \text{cov}(yz)_1 + p_2 \text{cov}(yz)_2 + \dots p_k \text{cov}(yz)_k \\
&\quad + \sum_i^k \sum_{j \neq i}^k p_i p_j \bar{y}_i \bar{z}_i - \sum_i^k \sum_{j \neq i}^k p_i p_j \bar{y}_i \bar{z}_j \\
&= \sum_i^k p_i \text{cov}(yz)_i + \sum_i^k \sum_{j \neq i}^k (p_i p_j \bar{y}_i \bar{z}_i - p_i p_j \bar{y}_i \bar{z}_j \\
&\quad - p_j p_i \bar{y}_j \bar{z}_i + p_j p_i \bar{y}_j \bar{z}_j) \\
&= \sum_i^k p_i \text{cov}(yz)_i \\
&\quad + \sum_{j < i}^k \sum_{i < j}^k p_i p_j (\bar{y}_i - \bar{y}_j)(\bar{z}_i - \bar{z}_j).
\end{aligned}$$

Since the variance of a variable is its covariance with itself, the expected variance for variable  $y$  is

$$\text{var}(y) = 1 + \sum_{j < i}^k \sum_{i < j}^k p_i p_j (\bar{y}_i - \bar{y}_j)^2,$$

and the expected correlation between  $y$  and  $z$  is given by

$$\begin{aligned}
r_{yz} &= \frac{\text{cov}(yz)}{\sqrt{\text{var}(y)\text{var}(z)}} \\
&= \frac{\sum_i^k p_i \text{cov}(yz)_i + \sum_{j < i}^k \sum_{i < j}^k p_i p_j (\bar{y}_i - \bar{y}_j)(\bar{z}_i - \bar{z}_j)}{\sqrt{\left[ 1 + \sum_{j < i}^k \sum_{i < j}^k p_i p_j (\bar{y}_i - \bar{y}_j)^2 \right] \left[ 1 + \sum_{j < i}^k \sum_{i < j}^k p_i p_j (\bar{z}_i - \bar{z}_j)^2 \right]}}.
\end{aligned}$$

## APPENDIX B

*Determination of MAMBAC Mean Differences in the Trichotomous Case*

If output variable  $y$  reflects membership in three ordered classes, then the mean above ( $\bar{y}_a$ ) and below ( $\bar{y}_b$ ) the cut will equal

$$\bar{y}_a = p_{1_a} \bar{y}_1 + p_{2_a} \bar{y}_2 + p_{3_a} \bar{y}_3,$$

$$\bar{y}_b = p_{1_b} \bar{y}_1 + p_{2_b} \bar{y}_2 + p_{3_b} \bar{y}_3,$$

where, for example,  $\bar{y}_1$  is the mean of class 1 (the class with the highest mean), and  $p_{1_a}$  is the probability of membership in class 1 among those above the cut. For any cut on the input variable, the difference between the means above and below the cut, therefore, equals

$$\begin{aligned}
\bar{y}_a - \bar{y}_b &= p_{1_a} \bar{y}_1 + p_{2_a} \bar{y}_2 + p_{3_a} \bar{y}_3 \\
&\quad - (p_{1_b} \bar{y}_1 + p_{2_b} \bar{y}_2 + p_{3_b} \bar{y}_3) \\
&= p_{1_a} \bar{y}_1 + p_{2_a} \bar{y}_2 + (1 - p_{1_a} - p_{2_a}) \bar{y}_3 \\
&\quad - p_{1_b} \bar{y}_1 - p_{2_b} \bar{y}_2 - (1 - p_{1_b} - p_{2_b}) \bar{y}_3 \\
&= p_{1_a} \bar{y}_1 + p_{2_a} \bar{y}_2 + \bar{y}_3 - p_{1_a} \bar{y}_3 - p_{2_a} \bar{y}_3 \\
&\quad - p_{1_b} \bar{y}_1 - p_{2_b} \bar{y}_2 - \bar{y}_3 + p_{1_b} \bar{y}_3 + p_{2_b} \bar{y}_3 \\
&= (p_{1_a} - p_{1_b})(\bar{y}_1 - \bar{y}_3) + (p_{2_a} - p_{2_b})(\bar{y}_2 - \bar{y}_3).
\end{aligned}$$

## APPENDIX C

*Determination of Modal Standardized Factor Scores in the Trichotomous Case*

For the  $k$ -class case, the expected unstandardized factor score for the different classes will be distributed equally through the interval 0 to 1. As Waller and Meehl (1998) noted, when  $k = 2$ , this is equivalent to dummy coding the classes as 0 and 1. In the three-class case it results in the values 0, .5, and 1. The mean of the factor scores is derived from

$$\begin{aligned}
E(X) &= 1p_1 + .5p_2 + 0p_3 \\
&= p_1 + .5p_2,
\end{aligned}$$

and the variance is equal to

$$\begin{aligned}
E(X^2) - E(X)^2 &= 1^2 p_1 + .5^2 p_2 + 0^2 p_3 - (p_1 + .5p_2)^2 \\
&= p_1 + .25p_2 - (p_1 + .5p_2)^2.
\end{aligned}$$

Based on this information and the derivation provided in Waller and Meehl (pp. 51–52), the three modes for the standardized factor scores are expected to fall at the following locations:

$$\begin{aligned}
E(M_{O_1}) &= \frac{1 - (p_1 + .5p_2)}{\sqrt{p_1 + .25p_2 - (p_1 + .5p_2)^2}}, \\
E(M_{O_2}) &= \frac{.5 - (p_1 + .5p_2)}{\sqrt{p_1 + .25p_2 - (p_1 + .5p_2)^2}}, \\
E(M_{O_3}) &= \frac{-(p_1 + .5p_2)}{\sqrt{p_1 + .25p_2 - (p_1 + .5p_2)^2}}.
\end{aligned}$$