#165

***Notes for*:**

Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports*, *78*,1091-1227. (Monograph Supplement 1-V78)

***For easier handling of this large monograph, main text and appendices are in two files*:**

165_MAXCOV_text_only.pdf (pp. 1091-1139 of the publication, including References)

165_MAXCOV_Appendices.pdf (pp. 1140-1227, Appendices A–I) .

TEXT (pp. 1091-1139)**: *Information subsequent to the initial publication*:**

p. 1112 & 1139, "Waller, Putnam, & Carlson, in press"— published:

Waller, N. G., Putnam, F. W., & Carlson, E. B. (1996) Types of dissociation and dissociative types: a taxometric analysis of dissociative experiences. *Psychological Methods*, *1*, 300-321.

p. 1129 & 1139 , "Waller & Meehl, in preparation"— published:

Waller, N. G., & Meehl, P. E. (1998) *Multivariate taxometric procedures: Distinguishing types from continua*. Newbury Park, CA: Sage.

p. 1130 & 1138, "(Meehl, in press)"— published:

Meehl, P. E. (1997) The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow, S. A. Mulaik, & J.H. Steiger (Eds.), *What if there were no significance tests?* (pp. 393-425). Mahwah, NJ: Erlbaum.

APPENDICES (pp. 1140-1227)**:**

LJY, January 2020

Psychological Reports, 1996, 78, 1091-1227. © Psychological Reports 1996
Monograph Supplement 1-V78

TAXOMETRIC ANALYSIS: II. DETECTING TAXONICITY USING COVARIANCE OF TWO QUANTITATIVE INDICATORS IN SUCCESSIVE INTERVALS OF A THIRD INDICATOR (MAXCOV PROCEDURE)[1,2]

PAUL E. MEEHL AND LESLIE J. YONCE

University of Minnesota *Minneapolis, Minnesota*

Summary.—Given three quantitative indicators of a conjectured latent taxon, a statistical function defined as the covariance of two indicators (designated for the procedure as the "output" indicators) computed within successive intervals along the third (designated as "input") indicator reveals whether the latent structure of the data is taxonic or not. If it is taxonic, latent parameters (base rate, hit rates, complement and taxon means) can be estimated, the latent distributions drawn, and subjects assigned to the taxon or the complement group. Several consistency tests are described. MAXCOV (MAXimum COVariance) is one of a related family of taxometric procedures in Meehl's Coherent Cut Kinetics Method.

CONTENTS

[1]We are grateful to Robert R. Golden, William M. Grove, Michael L. Raulin, and anonymous referees who made comments and suggestions to enhance and improve this monograph.
[2]Address enquiries to Paul E. Meehl, Psychology Department, University of Minnesota, N218 Elliott Hall, 75 East River Road, Minneapolis, MN 55414-0344.

This is the second in a series of articles about *coherent cut kinetics*, a system of procedures to determine whether the latent structure of a data set is taxonic and, if it is, to estimate the latent parameters associated with the taxon and complement (nontaxon) groups and to assign individual elements to membership in one or the other group.[3] The phrase 'coherent cut kinetics' refers to the epistemology and mathematics of the approach: we move cuts on a designated input variable and study the statistical behavior of other (output) variables on cases in the region of the cut and in regions demarcated by the cut. Inferring latent parameters (base rates, means, valid and false positive rates), we test the model and the numerical values for consistency over (a) different variables and (b) different procedures. We say 'kinetics' because the cuts move, 'coherent' because the inferences should be consistent.

MAXCOV (MAXimum COVariance) is one of the procedures used in the coherent cut kinetics method. It was first derived and conceived as a consistency test (Meehl, 1965), later as a taxon detector and parameter estimator (Meehl, 1968). The basic procedure has been described previously by Meehl (1973; Meehl & Golden, 1982), and MAXCOV or variants of it have been used by researchers to detect taxa in various areas (Golden & Meehl, 1973a; Lowrie & Raulin, 1990; and see discussion of dichotomous output indicators below). This article is an expansion on previous presentations, adding exemplary computer code and demonstrating the procedure with different Monte Carlo configurations.

Although MAXCOV will be presented here more or less as a standalone procedure, it is ideally used as one of the battery of taxometric proce-

[3]For discussion of the meaning, existence, and detection of taxa (real, nonarbitrary categories, types, entities) in personology and psychopathology, see Meehl (1992, 1995a), Meehl and Golden (1982), and methodological references cited in those papers. The first article in this series is about the MAMBAC procedure (Meehl & Yonce, 1994).

dures and consistency tests in Meehl's coherent cut kinetics method, each of which will contribute additional indications of taxonicity (or lack of it) and generate parameter estimates. The other procedures should confirm MAXCOV results or give improved estimates in some cases for which MAXCOV may not be the optimal procedure; likewise, MAXCOV will serve this function for other procedures in the method.

Selection of Indicators

MAXCOV requires three indicators, at least one of which is continuously distributed (in the usual social science sense of quantitative as contrasted with qualitative or dichotomous); if all the indicators are continuous, each may be used in turn as the "input" indicator, and the analysis is then done three times on the data set, yielding three MAXCOV output curves. It is desirable to have several continuous indicators so multiple curves can be generated and so all of the other procedures can be run to provide consistency checks and confirmation of results obtained from MAXCOV.

Selection or construction of indicators should preferably be theoretically motivated and guided by previous empirical research, taxometric or not. Conventional methods of inferring construct validity for a conjectured taxon, e.g., significant mean differences between diagnosed groups, are properly viewed as discovery aids, not as *assumptions* as that term is commonly used (Meehl, 1992; Meehl & Golden, 1982). Atheoretical taxometrics performed on a miscellany of "available" measures is not the optimal approach, but it is do-able and will be treated in the next article of this series. Although all correlates of taxonic indicators contribute to their construct-valid *interpretation*, the whole point of bootstrap taxometrics is that the statistical procedure largely "speaks for itself," given the mathematics and the numerical coherence (Meehl, 1995a). In this respect, it is similar to more familiar methods such as factor analysis and multidimensional scaling.

Each indicator should be selected to give good separation between the complement and taxon groups, i.e., to have good validity, and also to be uncorrelated with the other indicators within either the taxon or complement (nontaxon) group, i.e., to have no or little nuisance covariance. For most of the results reported here, we have used variables with 2 *SD* (latent expected) separation between the means of the complement and taxon groups and with negligible nuisance covariance. We have explored smaller separations and the addition of nuisance covariance to a limited extent, although much still needs to be done to assess the effects of these influences. A more detailed discussion of the selection of indicators and rationale behind our choices in these Monte Carlo tests may be found in Meehl and Yonce (1994). Although MAXCOV seems to be fairly robust, this should not be taken as a license for casualness in the initial selection of indicators.

Some psychologists may opine that separations of 1.50 *SD* or even 1.25

SD are unduly optimistic (or "perfectionistic"). We disagree. There are numerous and diverse domains of personality research in which means of taxa differ from those of complementary "controls" by 1.50 to 3.00 *SD* when carefully built measures, reliable criteria, and adequate sample sizes are employed. For example, such separations may be found in MMPI scores for nosological categories in psychopathology; malingering on the MMPI; psychometric deficits in diffuse brain damage; malingering in neuropsychological tests; IQs of business executives, brigadier generals, and college professors; self-confidence scores of general officers; "scientific" Strong Vocational Interest Blank scores of APA presidents; socialization scores of honored citizens versus shoplifters; interest scores of professional men and of skilled blue-collar workers; trade test scores; achievement test scores for different college majors; measures of religious belief and of political ideology in extreme groups; "femininity" scores of male homosexuals and of male Broadway actors; and morale scores of dissatisfied employees. In laboratory medicine, some biochemical tests are out 3, 4, or even 5 *SD* for diagnosed cases. The closer a fallible indicator is in the causal chain to the etiological agent (DNA, germ, tissue pathology, mental state or entity), the larger the separation will tend to be since each additional link in the chain of causal influences, being stochastic rather than nomological, attenuates the correlation between the initial members and the phenotypic terminus As was argued by Meehl and Yonce (1994) and by Meehl (1995a), it seems reasonable to urge taxon researchers to employ indicators that are at least as good as the weakest MMPI scales (Hy and Pa) are against pre-DSM diagnostic criteria. This amounts to a *T* score of 65. To achieve a hit rate of 75% on Gaussian distributions with equal groups and variances the taxon mean standard score is at 1.33, an MMPI *T* score of only 63, an unduly pessimistic separation.

Our advocacy of good separations (hardly a novel idea in psychometrics) and our optimism as to attaining them should not be construed as a negative thesis, that separations less than 1.50 or 1.25 *SD* cannot "work" taxometrically. No one knows the lower limit here. Meehl and Golden (1982) found that a 1.00 *SD* separation seemed safe for the consistency tests to detect unacceptable deviations of estimates from the true values. Lenzenweger and colleagues have obtained good results using the Chapman scales (which have about a 1.00 *SD* separation) to detect the schizotypal taxon among college students. It is noteworthy that the taxonic cases so identified show a *T* score of 78 on the MMPI Sc scale (Lenzenweger, 1993).

Ideally, one would prefer a direct observational check on nuisance covariance, but that would require a gold standard criterion of taxon membership, and, if we had such a criterion, we would (usually) not be doing bootstrap taxometrics. However, nuisance covariance does not present as insurmountable a problem as some might fear. First, in personology, psychopa-

thology, and other areas of "soft" psychology, correlations tend to be low—psychologists usually work hard to get them up to $r = .50$ or better. The typical phi coefficient between two items on the same MMPI scale is only $\phi = .12$ (Dahlstrom, personal communication, 1995). Thus, without data, the prior probability is that nuisance covariance will be small. Second, one can select indicators at different behavioral *levels* and *domains*, e.g., an MMPI score plus SPEM anomaly plus adiadochokinesia in researching schizotaxia, which are likely to be both relatively independent and theoretically interesting. Third, pairs of psychometric indicators can be item analyzed to reduce initial correlations. Nonpsychometric indicators not purifiable by item analysis may sometimes be subjected to nonlinear transformations of their metric to reduce the nuisance covariance term.

After using such indicator selection and manipulation techniques, how successful one has been in eliminating or reducing nuisance covariance can be checked by taxometric analysis of quasi-"pure" groups chosen by external criteria. For example, we might have a neurological indicator and a psychometric indicator of schizophrenia, each previously shown to be valid by conventional nontaxometric research on diagnosed patients. We then find the two candidate indicators to be negligibly correlated in a sample of carefully diagnosed (SADS interview, MMPI) schizophrenic probands; we study their clinically normal MZ twins and find the two candidate indicators to be negligibly correlated in that group also. We obtain similarly low correlations for carefully screened "normals."[4] These three "externally defined" categories will be sufficiently uncontaminated so that taxon/complement mixture cannot explain a sizable correlation, should we find one, between the candidate indicators. If the correlations are low, we may proceed with confidence to study a new group we conjecture contains a latent taxon that may be detected using the indicators we have chosen. Another fairly direct test of negligible nuisance covariance would be to form subgroups of taxon or complement members identified by one subset of indicators (say, x and y) and com-

[4]One cannot employ nonschizophrenic psychiatric controls for this purpose because some clinicians (e.g., Bleuler, Rado, Meehl) conjecture that a sizable proportion of such patients are unrecognized schizotypes decompensated in varying degrees and directions. Since our long-term research aim is to find indicators that detect such cases in the absence of florid schizophrenic signs, in the discovery phase (searching for valid indicators), it would be inconsistent to reject candidate indicators that, if valid for underlying schizotypy or schizotaxia, *should* properly correlate (due to taxon/complement mixture) among psychiatric patients not meeting, say, DSM criteria for schizophrenia (Meehl, 1990c). There is nothing "viciously circular" about this procedure. If the theory is correct, our theory-motivated indicator screening will enable us to discover its correctness taxometrically. If the theory is incorrect, nothing in the indicator selection process will produce an artefactual taxometric corroboration. That indicators x, y, z have been shown to be negligibly correlated in the three populations but not in a patient miscellany cannot somehow "force" them to display clear and consistent MAXCOV taxonic patterns when no taxon exists. That x, y, z were retained despite being correlated in a mixed nonschizophrenic psychiatric group has no mathematical tendency to make their relationships, even within that special population, be of the taxonic sort.

pute observable nuisance correlations of the other indicators (say, r_{zv}) within those homogenized groups. A small contamination of a complement subset by a few taxon cases, or vice versa, does not generate a large enough observed correlation in quasi-"pure" subjects to invalidate this check (see Appendix B, p. 1146).

Finally, we can rely on the statistics of the procedure and consistency tests. MAXCOV appears to be highly robust with respect to the independence conjecture, and consistency tests will tell us whether the model is too badly violated. Nuisance correlations up to .25 or .30 can be considered to be negligible in effect. When they are higher than that, researchers may elect to use the generalized procedure (Meehl, 1995b) or to rely on robustness if slight bias in base rate and other inferred latent values are not considered harmful. Such judgment calls, of course, depend on the aim of the study and the current state of knowledge. For example, merely showing that a conjectured taxon exists and that it has a base rate in a certain range, e.g., $.10 < P < .30$, might be valuable in appraising the DSM categories; whereas in testing a specific genetic model of schizotaxia, one would want $\hat{P}$ to be inferable with considerably higher precision than that. We repeat, however, it is preferable to select and construct indicators whose nuisance correlation is "safely" low, pretesting this auxiliary conjecture directly as described above.

Rationale for MAXCOV

The MAXimum COVariance procedure looks at the covariance[5] of two output indicators within successive intervals on a third input indicator. If the latent structure is taxonic (and there is sufficiently little or no nuisance covariance), that covariance graph increases in the vicinity of greater taxon-complement mixture; if the latent structure is nontaxonic, the covariance graph is relatively flat.

The core idea motivating the procedure is that, if two observable variables ("indicators") tend to discriminate, i.e., are valid for, a latent category ("taxon") and they do not covary otherwise (no "nuisance correlation" within the latent classes), then any observed correlation is due solely to category mixture. Hence, subpopulations that are "pure" (unmixed, consisting only of taxon or of complement class members) will exhibit no manifest indicator correlation, and "mixed" subpopulations will show correlations whose size varies with the mixture. The most "mixed" subpopulation being an even

[5]That most recent statistics texts used by psychologists do not index the term 'covariance' as such, without reference to analysis of covariance (ANCOVA), is sloppy writing or indexing, inexcusable pedagogy, perhaps attributable to the prevailing over-emphasis on Fisherian experimental designs. Cramér (1946), Kenney (1939), and Li (1975) index 'covariance,' as does Fisher himself in his *Statistical Methods for Research Workers* (1970). The shortest definition of the Pearson correlation coefficient is "$r =_{\text{def}}$ ratio of the covariance of two variables to the geometric mean of their variances." The dimensions of a covariance are those of the variables, e.g., dollars × IQ points, so, unlike r, the raw covariance is not a pure number.

split $p = q = \frac{1}{2}$ of taxon and complement cases, this latent subset yields the largest observed indicator correlation. If subsets of cases are ordered along a third ("input") indicator, the observed covariance of the first two ("output") variables in successive intervals along the input indicator will increase from zero to a maximum and then decrease to zero. The maximum is located at the even split, i.e., where there is the greatest "mixture," and, since this is where a diagnostic cut minimizes misclassifications (maximizing "hits") based on the input indicator, it is called the hitmax cut on that indicator. The observed covariance of the output indicators among cases in the region of that cut (hitmax interval) is used to solve for a latent parameter (K), which in turn is used to infer the latent taxon/complement proportions in each interval along the input variable; and these estimated proportions allow us to compute estimates of the taxon base rate P and the latent taxon and complement means. To accomplish this, we rely on the General Covariance Mixture Theorem

$$cov_{yz}(x) = pcov_{yzt} + qcov_{yzc} + pq(\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c) \qquad [1]$$

which says that, for any group of cases, the yz covariance is composed of the taxon proportion p multiplied by the covariance of yz in the taxon cases, plus the complement proportion q multiplied by the yz covariance in the complement cases, plus the product of the taxon proportion, the complement proportion, and the separations between the taxon and complement means on indicators y and z. The Covariance Mixture Theorem is general because it holds for situations when there is nuisance covariance, i.e., correlation within the taxon cases or within the complement cases, and it holds for any subset or region of the data. A proof of the Covariance Mixture Theorem underlying the MAXCOV procedure is given in Appendix A (pp. 1140-1145).

Although our Monte Carlo data were generated by a Gaussian algorithm assigning equal variances $SD_t^2 = SD_c^2 = 1$ to taxon and complement classes, none of the core derivations underlying MAXCOV are thus restrictive. The conjectured structure (not an "assumption," see Meehl, 1992, pp. 135-136; Meehl & Golden, 1982) is highly general, that of two overlapping unimodal frequency distributions. The mathematics speaks for itself, and it was developed by Meehl with psychopathology in mind, where skewness and heterogeneity of variance are common. One relies on Monte Carlo samples to get an idea of random error in inferring the latent parameters. Unequal variances and unequal nuisance correlation have little effect on trustworthiness of estimations (Meehl & Golden, 1982, Table 5.2, p. 163). We anticipate a later article in this series devoted wholly to "taxometrics under adverse conditions" that, in addition to weak taxa and nonlinear moderator effects, will study various extremes of distribution properties.

MAXCOV-HITMAX versus Generalized MAXCOV

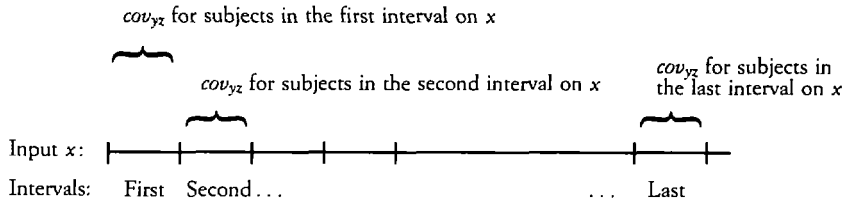
If there is no nuisance covariance, i.e., no correlation on the indicators *within* the taxon or complement groups, so that cov_{yzt} and cov_{yzc} in the covariance mixture formula are approximately zero, the interval of maximum covariance will be the place where half the cases are taxon members and half are complement members; the covariance is due solely to this mix (see Meehl, 1995a, p. 271). This is at the hitmax cut, and we then use the *MAXCOV-HITMAX* procedure to estimate various latent parameters. If there is nuisance covariance, the interval of maximum covariance will reflect both the taxon/complement mixture and also nuisance covariance; in that case we may use *generalized MAXCOV* procedures to estimate latent parameters (Meehl, 1995b). Although we will examine some configurations with nuisance covariance, we will focus on MAXCOV-HITMAX in this article.

The Monte Carlo Samples

Twenty-five Monte Carlo samples have been generated for each of various taxonic and nontaxonic configurations. For instance, there are taxonic configurations with different base rates and with different sample sizes and nontaxonic configurations with different factor loadings on the variables. Each sample has four continuously distributed variables (another procedure in the coherent cut kinetics method requires that many); although MAXCOV needs only three, more variables may be used when they are available. Each sample has a coded name indicating the particular configuration plus its unique sample number. The reader does not have to learn the coded names, but they are used in some places so that those who want to locate particular samples may do so. Monte Carlo results for only some samples are given in the text; results for all 25 of the samples for each configuration may be found in the appendices. A detailed description of the creation of these samples is given by Meehl and Yonce (1994, p. 1066-1068) when they were used in Monte Carlo tests of MAMBAC, a procedure which detects taxonicity and estimates parameters using two indicators. The reader who would like to see how a particular sample (or 25 samples for a particular configuration) performs on different taxometric procedures can compare them in the two articles.

Calculation of MAXCOV Values

The first step is to draw a MAXCOV graph. Using three continuously distributed indicators, designate one of them as the "input" indicator and demarcate successive intervals along that indicator; within each interval, calculate the covariance of the other two (designated as "output") indicators, and graph the resulting covariances across the intervals. The input indicator is used merely to locate the cases which are to be used in calculating the covariance between the output indicators.



Each continuous indicator is taken as the input indicator in turn; the other indicators are used in all possible pairs as output indicators. With the four continuous indicators (x, y, z, v) in our Monte Carlo samples, there are 12 possible input/output$_1$, output$_2$ combinations, hence 12 MAXCOV graphs for each sample. First, x is used as the input indicator and y and z are used as output indicators, denoted as $cov_{yz}(x)$; then x can be used as input with y and v as output, denoted as $cov_{yv}(x)$, and so on.

Cuts along the input variable can be made on the basis of either standard score units on the abscissa or number of subjects. Whichever of these ways we use, two considerations are relevant in choosing how coarse to make the intervals. Because the *shape of the graph* of covariances is of major interest, there should be enough intervals, i.e., enough points on the output graph, to make that clear. On that basis alone, the more points (intervals), the better. But, of course, we need to have enough subjects within each interval to calculate a fairly stable value. Unless otherwise noted, the Monte Carlo runs in this article used intervals on the abscissa defined in .25 *SD* units around the observed mean of the input distribution: all subjects falling below −2.00 *SD* were considered to be in the first interval, subjects falling in the interval −2.00 *SD* to −1.75 *SD* were considered to be in the second interval, and so on, to the top interval which was considered to be all subjects in the interval ≥ 2.00 *SD* above the mean of the input distribution. Once this was done, a second requirement was imposed to help ensure that there would be enough subjects for calculating covariances in the extreme intervals; the program checked both ends of the input distribution and accumulated subjects (moving forward from the low end, backward from the high end) in the predefined intervals until there were at least 15 cases. That located the beginning and ending intervals for a single output graph. Thus, for example, for a given input variable, the first interval might include all subjects below −1.75 *SD* from the observed mean of that variable. Fifteen was chosen (more or less arbitrarily, based on observations from initial Monte Carlo runs) as a minimal acceptable n for calculating a covariance. It was assumed that, since we were working in standard deviation units, once we had that many subjects, we would be "into the input distribution," so to speak, and could usually count on having sufficient subjects in successive middle

intervals to calculate a covariance for each interval. This resulted in about a dozen intervals along each input variable for samples with $N = 300$. Of course, some intervals within the range might have fewer than 15 subjects. For Monte Carlo tests of smaller sample sizes, the program was adjusted to handle a single interval with zero cases by averaging the covariances from the intervals on either side of it. This situation has occurred very rarely in our samples; we have encountered no cases where two successive "inside" intervals have zero subjects (which would cause our program to crash).

MAXCOV graphs are frequently "noisy" and benefit from smoothing. All of the MAXCOV graphs shown here are the raw data points with a smoothed curve drawn through them using the smooth function in S-Plus (Statistical Sciences, 1993) graphical software; this is Tukey's 4(3RSR)2H method (Tukey, 1977, Chapters 7 and 16). Weighted means smoothing ($^1/_2$ the value of a data point plus $^1/_4$ the value of the preceding and the succeeding points) also works well. A researcher might want to experiment with different smoothing techniques or with different parameter values for some techniques, e.g., with lowess the fraction of data used for smoothing each point is $^2/_3$ by default in S-Plus software, but MAXCOV curves show better definition with a smaller value.

The dimensions of $cov_{yz}(x)$ are set by the nature of the indicators—physical, psychological, economic, or whatever. A schizophrenia researcher might deal with covariates $y = T$ score on MMPI Scale 8 and $x =$ milliseconds lag in the smooth pursuit eye movement (SPEM) visual tracking task, yielding a MAXCOV graph whose hitmax interval peak could be $cov_{yz}(x_h) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)(20)(100) = 500$. The peak-valley difference here is 500 times that of our Monte Carlo graphs, where the hitmax interval value of $cov_{yz}(x_h)$ for a 2 *SD* separation is around 1.00. The threshold question "Taxonic or nontaxonic?" is presently answered inspectionally by investigators who have contemplated curves such as we present in this article. Meehl and Yonce (1994) demonstrated near perfect inspectional sorting of MAMBAC curves by both psychologists and nonpsychologists. The MAXCOV graphs are, if anything, even more clearly taxonic.[6] For all but the weakest situations, the difference between the taxonic and nontaxonic curves is so clear that any "reasonable"

[6]Three psychologists and two persons with no statistical training (BAs in Latin and in international relations) sorted 90 panels of MAXCOV curves from nontaxonic samples and taxonic samples of various base rates and separations—a total of 450 inspectional judgments—without error. Researchers who distrust inspection may wish to test the "flatness" of a nontaxonic-appearing curve by whatever statistic they prefer, e.g., *F* test or departure of a straight line from zero slope. We are working on an algorithm for the taxonic/nontaxonic decision for use in situations in which skilled user's inspection may err too often. The present plan is to fit a polynomial, probably a quartic, and examine the coefficients. "Reasonable" linear transformations of the metric seem to leave the taxonic/nontaxonic distinction fairly clear from the polynomi-

graphing procedure will yield patterns sufficiently similar to ours, keeping in mind that the curve shapes and hill-valley depths perceived by inspection depend on the graph's physical distances rather than the numerical units employed.

Although the raw metric of indicators can probably be safely employed, it may be desirable to *standardize* for reasons of long-run familiarity in the scientific community—so that many workers may develop a good inspectional flair. (Historical example: During the long period when the Stanford-Binet IQ was *defined* as the MA/CA ratio, its standard deviation varied somewhat over chronological age groups, but psychologists and teachers did develop a rough subjective-but-shared notion of its meaning after reading numerous research studies and working with many children.) The simplest suggestion, already adopted by some users of MAXCOV, is to standardize all indicators in (manifest) *SD* units. This will not relate the graph's hill-to-valley depths to the abscissa *SD* units in exactly the same ratio as ours because our standardization metric for these Monte Carlo runs is in latent *SD* units, a metric unknown to that investigator, but the difference involved is less than the difference in *our* Monte Carlo runs between large and small separations. (*Example*: Assume latent separations were to vary over studies from 1 to 2 *SD*s and base rates from $P = .10$ to $P = .50$. Then investigators would have to inflate the vertical dimension's numerical values by a factor of 1.18 to 2.00 to make the graphical distances comparable to ours, and they would have no way of accurately determining which dilation factor to employ.)

Pseudocode for calculating MAXCOV using four continuously distributed indicators and defining intervals on the basis of abscissa intervals and on the basis of deciles is given in Fig. 1. This was extracted from the Modula-2 (LOGITECH, Inc., Version 3.0) program used for the Monte Carlo graphs reported here. The code for calculating MAXCOV in successive deciles and plotting the graph with S-Plus, an interactive graphical analysis software program, is given in Fig. 2. Examples of the characteristic resulting taxonic and nontaxonic graphs are also shown there. It may be noted that using deciles to define the intervals is not a good idea when there is reason to expect a small base rate; this will be explained below.

Detection of Taxonicity With MAXCOV

If the underlying structure is taxonic, graphs of the conditional covariances generated by MAXCOV tend to be peaked, with the location of the peak depending on the latent base rate. A base rate of .50 (half taxon mem-

---

al coefficients, especially relying on the mathematical fact that the pairwise *ratios* of polynomial coefficients are invariant under linear transformation of the metric. Meanwhile, coherent results corroborate inspectional inference for doubtful graphs.

Pseudocode from MAXCOV program used in Monte Carlo runs

Input/output1,output2 combinations: xyz, xyv, xzv, yxz, yxv, yzv, zxy, zxv, zyv, vxy, vxz, vyz

FOR each input/output1,output2 combination DO
  Get covariance by .25 *SD* interval cuts on input
    Determine possible intervals around observed mean of input indicator:
      < -2.00; -2.00 to < -1.75; -1.75 to < -1.50; . . . 1.75 to < 2.00; ≥ 2.00
    Check the intervals at each end of the input distribution and move inward until at least 15 cases are accumulated at each end; this defines the FIRST and LAST interval for which a covariance will be calculated

    FOR each interval *i* from FIRST to LAST DO
      n := number of cases in interval *i*
      Σ_{out1} := sum of the output1 scores paired with cases in *i*
      Σ_{out2} := sum of the output2 scores paired with cases in *i*
      $\Sigma_{out1,out2}$:= sum of the product of output1 score and output2 score associated with each case in *i*

$$cov_{out1,out2} := \frac{(n)\,(\Sigma_{out1,out2}) - (\Sigma_{out1})\,(\Sigma_{out2})}{n^2}$$

    END (* for each interval *)

    Plot obtained covariances over intervals; smooth curve

  Get covariance by decile cuts on input
      Sort scores on input indicator, keeping output scores properly associated with each input score
      Determine the number of cases in each decile (depends on *N*)
      Proceed as above using successive deciles as the intervals

END (* for input/output combination *)

(* It will be helpful later to have saved the midpoint value, the covariance, and the observed number of cases in each interval *)

FIG. 1. Pseudocode for MAXCOV

bers and half complement members) gives a peak in the center of the observed input variable distribution. A nontaxonic latent structure results in graphs that are generally flat. Fig. 3 shows the MAXCOV curve shape for the error-free (Gaussian) situation when the taxon base rate is .50, the taxonic separation on each variable is 2σ, and there is no nuisance covariance.

MAXCOV curves from Monte Carlo samples are shown in Fig. 4. The panels in the top row are from 10 taxonic samples. Intervals on the input variable were defined by .25 *SD* units around the observed mean of the input variable. Each panel shows the 12 curves (offset for visual display) generated by four variables for a single sample. Superimposed on the points are smoothed curves using Tukey's 4(3RSR)2H twice method. MAXCOV curves from nontaxonic samples are shown in the lower row of panels. Because taxonic separation generates correlation between the variables (even though

S-Plus code for MAXCOV using decile intervals on the input indicator

```
# start with a data vector "xyz" composed of indicators x, y, and z arranged in three columns
# to use x as input, cov(yz) as output, sort on x, assumed to be in column 1 of "xyz"

    sortedx <- (xyz[sort.list(xyz[,1]),])

    N <- length(sortedx[,1])            # define N as the number of cases
                                        # (N=600 for the sample graphs below)
    MAXCOV <- vector ('numeric',N)      # define a vector to hold the MAXCOV values

# in successive intervals defined by deciles (N/10) on x, get the covariance of y, z

    for (i in seq(1,N-N/10,N/10) ) {
      MAXCOV[i] <- var(sortedx[(i : (i + N/10) ),2],sortedx[(i : (i + N/10) ),3]) }

# create a window, plot the MAXCOV values, and overlay the points with a smoothed curve

    win.graph ( )
    plot (MAXCOV[seq(1,N - N/10,N/10) ])
    lines (smooth (MAXCOV[seq(1,N-N/10,N/10) ]) )

# Examples of curves that might be obtained:
# (the nontaxonic example is plotted on a y-axis comparable to that of the taxonic case)
```

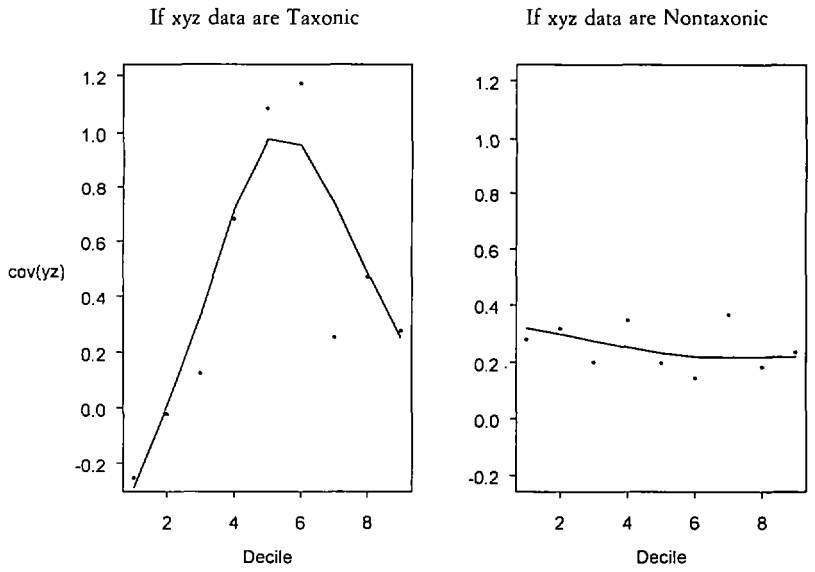


FIG. 2. S-Plus code and MAXCOV curves

they are uncorrelated within the taxon and within the complement groups), we imposed factor loadings on variables in the nontaxonic samples that would generate a comparable correlation. In the taxonic samples shown in

Fig. 4 the expected $r_{ij} = .50$ (for $P = .50$, 2 *SD* separation on each variable, and no nuisance covariance; see Appendix B, pp. 1146-1147). In the nontaxonic comparison samples, factor loadings of .707 on each variable generate the same expected $r_{ij} = .50$.

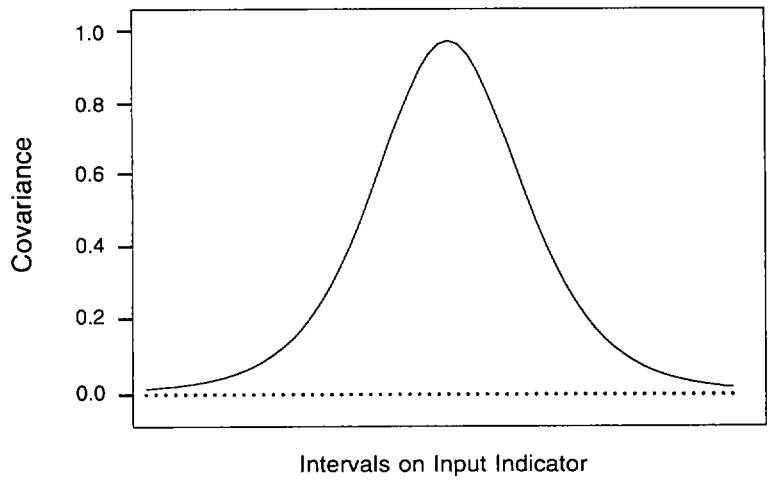


FIG. 3. MAXCOV error-free curve shape (solid line) for $P = .50$, 2σ separation on each indicator, and no nuisance covariance. The dotted line is the nontaxonic situation when $r_{ij} = .50$.

We have shown curves from only 10 samples of each configuration (taxonic and nontaxonic) in Fig. 4. They are unselected, i.e., they are merely the first 10 samples from each of these two Monte Carlo configurations. To see the difference between curves generated by a taxonic versus a nontaxonic configuration, any panel in the top row could be compared with any panel in the bottom row. Curves from all 25 samples that have been generated for these configurations may be found in Appendix C, pp. 1148-1177 (the first 10 panels there will be identical to those in Fig. 4).

Clearly MAXCOV detects taxonicity for the samples shown here. Curves from taxonic samples are peaked in the middle. The nontaxonic curves are relatively flat.

Effect of Different Methods of Cutting on the Input Distribution

Another way to define intervals along the input variable is by number of cases (histogram areas). Fig. 5 shows the same samples as in Fig. 4 but using deciles to determine the intervals along the input variable. This gives us 10 points for each MAXCOV graph and a predetermined number of cases within each interval. With larger samples, one might use finer cuts,

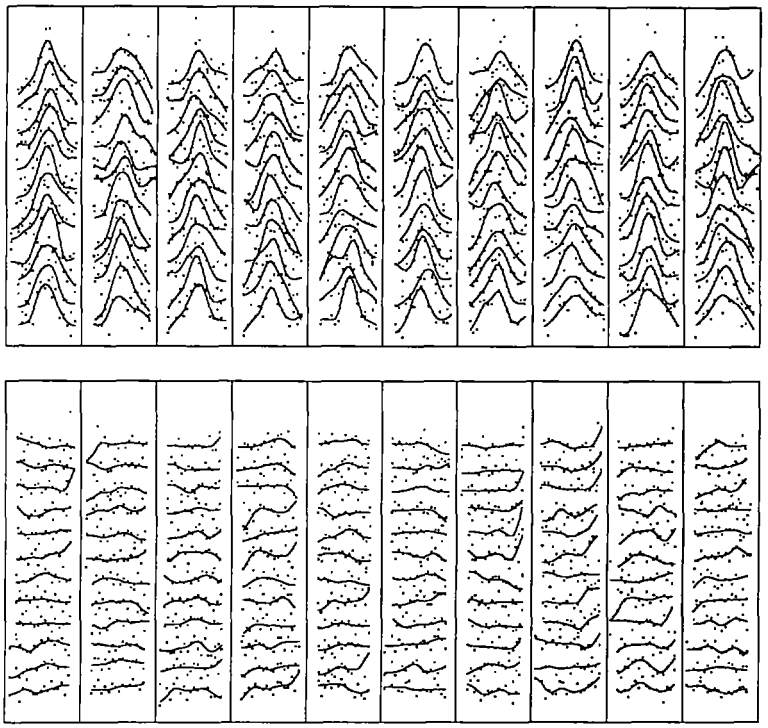


FIG. 4. Detection of taxonicity with MAXCOV using .25 *SD* intervals on the input indicator. Top panels are from taxonic samples: $N = 600$, $P = .50$, 2 *SD* separation on each indicator, no nuisance covariance, $r_{ij} = .50$ (because of complement-taxon mixture). Bottom panels are from nontaxonic samples: $N = 600$, $r_{ij} = .50$ (from factor loadings of .707 on each variable). These samples are also used in Fig. 5.

e.g., vigintiles, and increase the number of points on the output graph. When the base rate is near .50, there is no particular reason to prefer one way of defining the intervals. However, when the base rate becomes smaller, decile intervals can mask the peak interval. For instance, if the base rate were $P = .10$ and decile intervals were used, the most extreme situation would be if all the taxon members were in the last interval (an unlikely event, but one that could happen because of sampling error). In that case, there would be no interval with a mixture of taxon and complement members, hence no high point on the MAXCOV graph. In our experience with Monte Carlo samples, a small base rate gives a right-end cusp when decile intervals are used. Using abscissa intervals gives a better definition of the hitmax interval, and that is what we recommend when base rates are suspected to be small. Only MAXCOV graphs based on abscissa intervals are shown in the rest of this article.

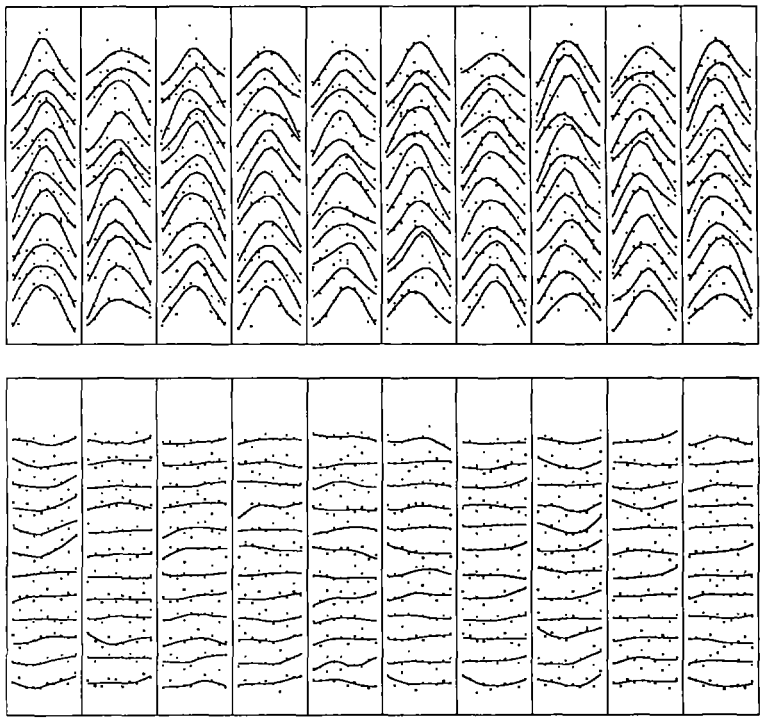


FIG. 5. Detection of taxonicity with MAXCOV using decile intervals on the input indicator. Top panels are from taxonic samples: $N = 600$, $P = .50$, 2 *SD* separation on each indicator, no nuisance covariance, $r_{ij} = .50$ (because of complement-taxon mixture). Bottom panels are from nontaxonic samples: $N = 600$, $r_{ij} = .50$ (from factor loadings of .707 on each variable). Curves from these samples are also used in Fig. 4.

Effect of Sample Size

Taxonicity may be detected with samples as small as $N = 100$ when conditions are otherwise favorable ($P = .50$, 2 *SD* separation on the variables, and no nuisance covariance), but larger samples are strongly recommended. Monte Carlo results in Fig. 6 show how a larger N increases the stability of the curves and results in more intervals, hence more points on the MAXCOV graph, when abscissa intervals are used. Larger samples also give clearer indications of latent nontaxonicity, as may be seen in Fig. 7. As will be seen later, estimates of the latent parameters are more accurate with larger samples.

Effect of Base Rate

As the taxon rate gets smaller, the MAXCOV peak shifts to the right (and if the taxon base rate were larger than .50, the peak would be shifted

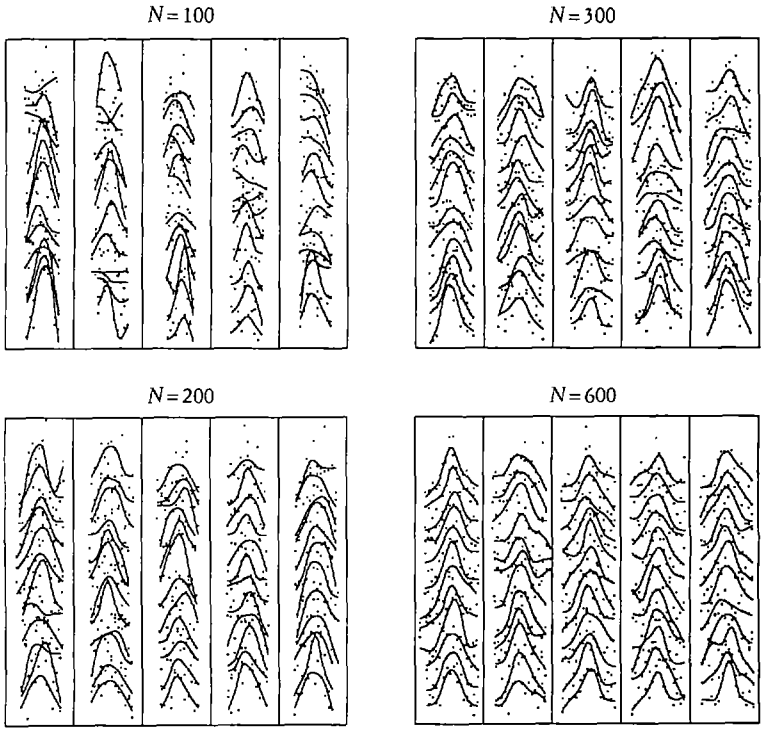


FIG. 6. Effect of sample size $N = 100$, 200, 300, 600. All samples shown here are taxonic: $P = .50$, 2 *SD* separation on each indicator, no nuisance covariance, $r_{ij} = .50$ (because of complement-taxon mixture). Curves from all 25 Monte Carlo samples for each sample size may be found in Appendix C (pp. 1148-1177).

to the left of center). The error-free MAXCOV curves in Fig. 8 show this progression; in each case there is 2σ separation on each variable and no nuisance covariance.

Fig. 9 shows Monte Carlo results for base-rate configurations: $P = .50$, $P = .25$, and $P = .10$. The peak of the MAXCOV curve should be at the location of the hitmax cut, and this is determined by the base rate. Larger samples show the effect more clearly, but curves from samples of $N = 300$ are usually clear (see Appendix C, pp. 1148-1177). The method of determining intervals becomes important with lower base rates. For instance, if $P = .10$ and deciles are used to define intervals, the MAXCOV curve will probably show only a cusp at the right end. Although a cusp often happens when abscissa intervals are used, particularly with smaller samples (see samples with $P = .10$ and $N = 300$ in Appendix C, p. 1162), there is a better chance that a peak will be discernible.

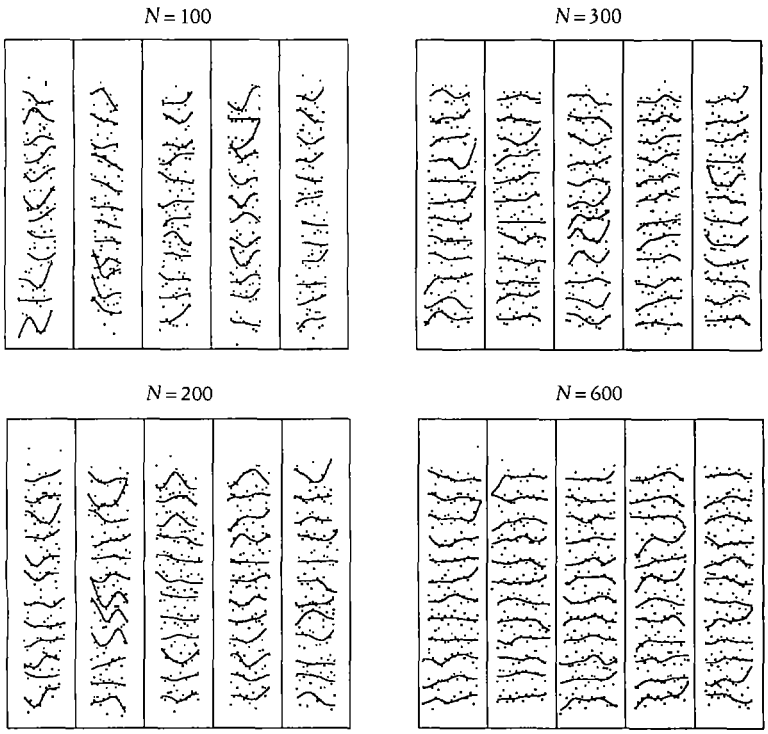


FIG. 7. Effect of sample size $N = 100, 200, 300, 600$. Samples shown here are nontaxonic: expected $r_{ij} = .50$ (factor loadings of .707 on each indicator).

Notice that the nontaxonic curves in the right-hand panels of Fig. 9 do not become taxonic-appearing with higher factor loadings. It has been alleged that this could be a source of pseudotaxonicity; in fact, higher loadings tend to be a source of error in the opposite direction. In taxonic situations, large intracategory nuisance covariances tend to attenuate MAXCOV peaks so that they look less clearly taxonic, leading possibly to false negative inferences. In nontaxonic situations, the higher the pairwise correlations, the lower will be the standard error of an output covariance, hence false positives arising from sampling error will tend to be fewer. Pending adequate investigation of the danger zone, researchers need not fear a protaxonic bias in the method arising from this source.[7]

[7]The same reasoning holds for the MAMBAC procedure (Meehl & Yonce, 1994). Large intracategory nuisance covariances tend to flatten the MAMBAC graphs for taxonic situations; higher factor loadings in nontaxonic situations lower the random fluctuation in means defined by MAMBAC cuts.

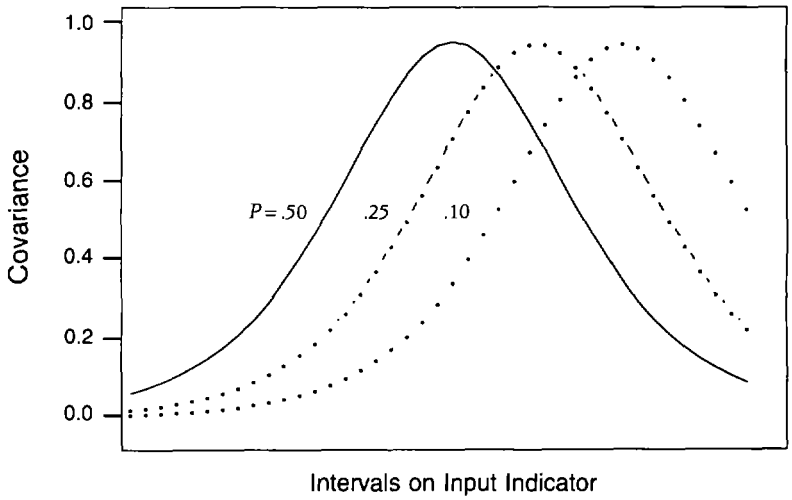


FIG. 8. MAXCOV error-free curve shapes for different base rates (no nuisance covariance and 2σ separation on each indicator)

Effect of Taxon Validity (Separation)

Reduced separation between the variables leads to a lower MAXCOV peak. Error-free curves for different separations are shown in Fig. 10.

The top panels of Fig. 11 show MAXCOV curves for the first ten Monte Carlo taxonic samples with a separation of only 1.5 *SD* on each of four variables. Curves in the lower panels are from nontaxonic samples with comparable expected correlation between variables. [See also Fig. 14 (p. 1115) which shows the effects of lower validities in addition to nuisance covariance.] Although the taxonic curves are less sharply peaked than they were with greater separation (compared with, e.g., taxonic curves in Fig. 4), they are still clearly distinguishable from nontaxonic graphs.

Effect of Nuisance Covariance

Error-free MAXCOV curves with different amounts of nuisance covariance are shown in Fig. 12. Curves are elevated as a whole, with the ends tapering off at a level on the *y*-axis that reflects the amount of nuisance covariance. Numbers to the left in Fig. 12 indicate the amount of nuisance covariance in the complement group for each curve; those on the right show nuisance covariance in the taxon. When nuisance covariance is the same in both groups (solid lines in Fig. 12), the curves are symmetrical. The MAXCOV peak shifts slightly to the right when there is greater covariance in the taxon group than in the complement (dotted curves).

Fig. 13 shows MAXCOV curves for Monte Carlo samples with a sepa-

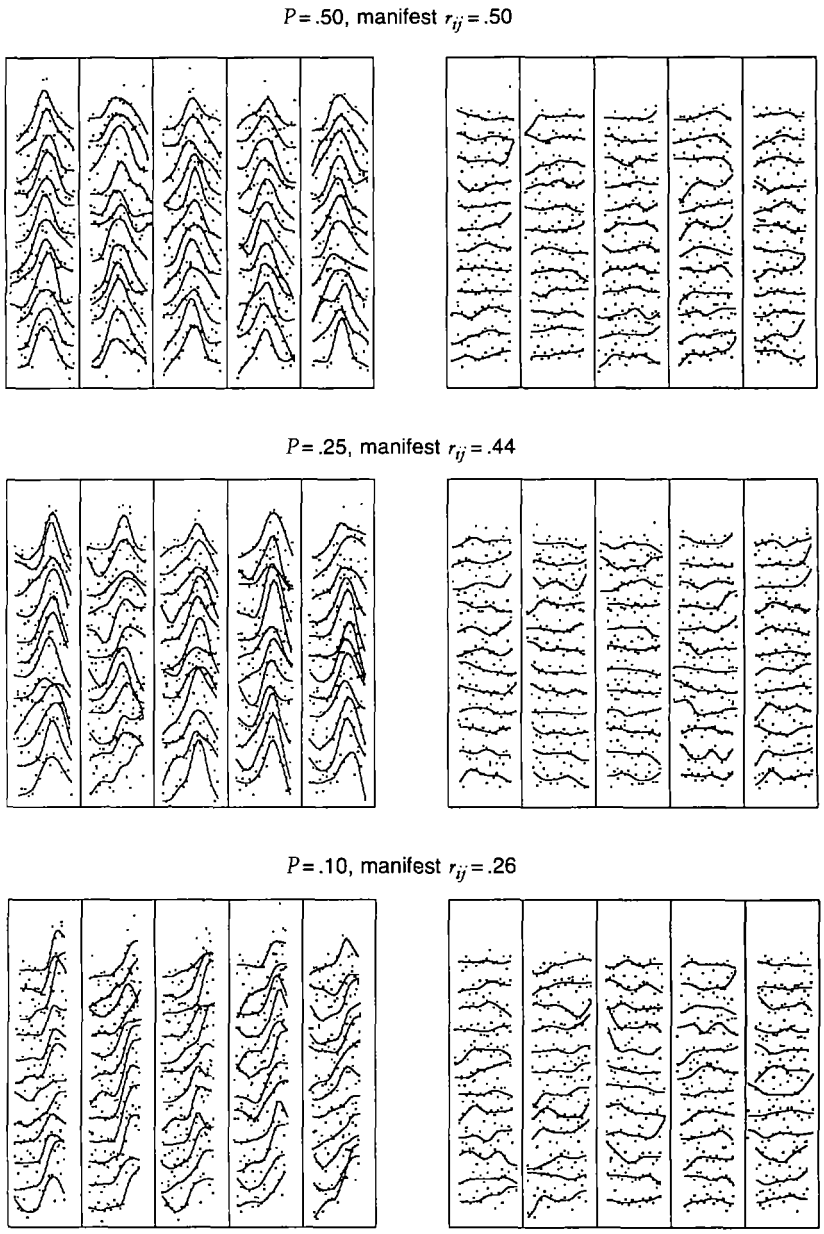


FIG. 9. Effect of base rate. MAXCOV graphs based on abscissa intervals. Samples on the left are taxonic: $N = 600$, 2 *SD* separation on each indicator, no nuisance covariance. Samples on the right are nontaxonic with expected r_{ij} matching that of the taxonic samples in each row.

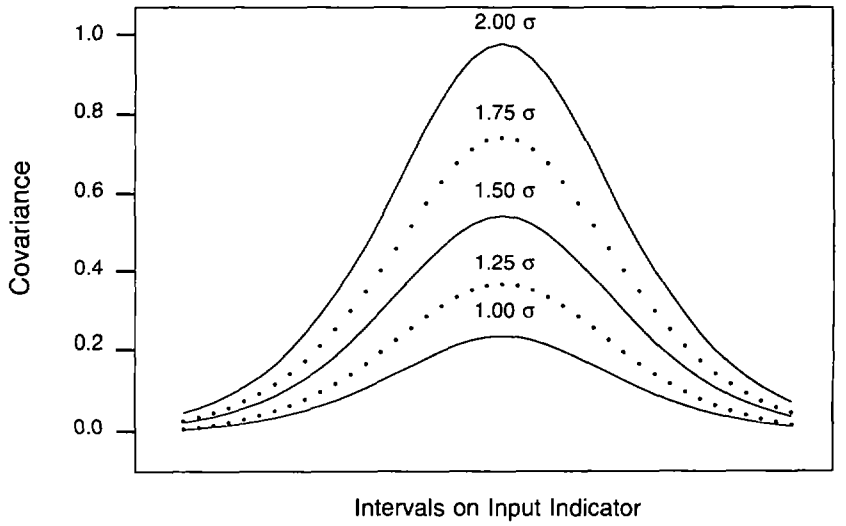


FIG. 10. MAXCOV error-free curves for different amounts of separation ($P = .50$, no nuisance covariance)

ration of 2 *SD* on each of the four indicators and nuisance covariance added within the taxon and complement groups ($N = 600$, $P = 50$). (See also Fig. 14 which illustrates a combination of nuisance covariance and various validities for four indicators). It is clear that MAXCOV is fairly robust with respect to nuisance covariance, confirming Monte Carlo results reported by Meehl and Golden (1982, Table 5.2, p. 163).

Combined Effects of Reduced Validities and Nuisance Covariance

The taxonic Monte Carlo samples in Fig. 14 (p. 1115) demonstrate the effects of both nuisance covariance and different validities for the four variables; these samples were generated to provide a deliberately difficult test for detecting taxonicity. Again, the MAXCOV peaks are usually attenuated, but the graphs are still distinct from those from nontaxonic samples. A naive ("nonstatistical") sorter would make no errors in classifying panels by inspection, having been shown a few examples.

Dichotomous Output Indicators

The use of dichotomous output indicators in MAXCOV was suggested by Meehl (1965, pp. 12-15). It is formally identical with the (preferred) quantitative output case because the General Covariance Mixture Theorem is a distribution-free algebraic identity, holding for any pair of real number variables, including those which take on only two values $y = 0$, $y = 1$. The numerator of a ϕ-coefficient $(p_{ij} - p_i p_j)$ is, of course, a covariance, literally. The

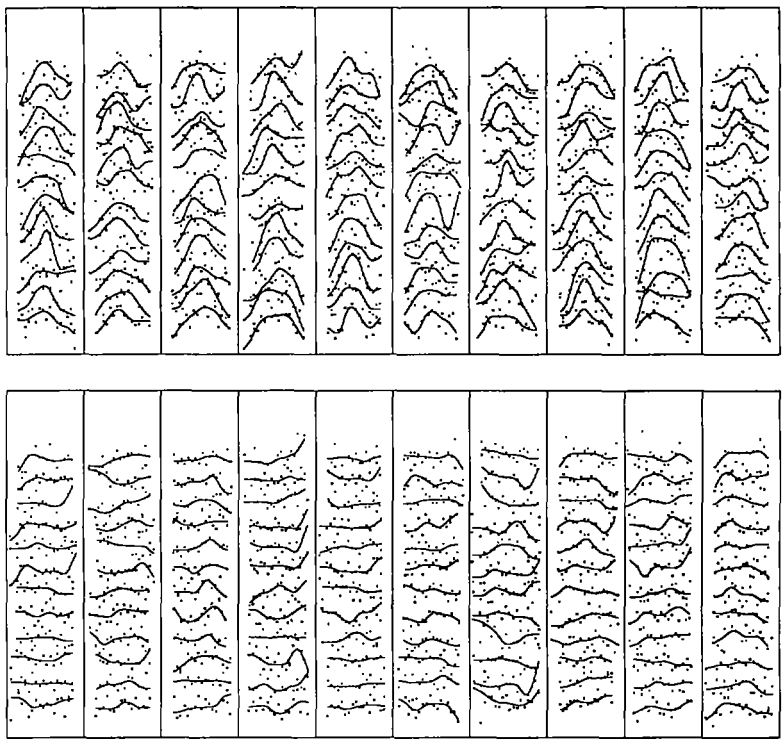


FIG. 11. Effect of reduced indicator validity on MAXCOV curves. Intervals are based on abscissa distances. Top panels are from taxonic samples: $N = 600$, $P = .50$, 1.5 *SD* separation on each indicator, no nuisance covariance, expected $r_{ij} = .36$ (because of complement-taxon mixture). Bottom panels are from nontaxonic samples: $N = 600$, expected $r_{ij} = .36$ (from factor loadings of .60 on each indicator).

proportions p_i and p_j are, algebraically, arithmetic means of variables that take on only values 0 and 1. Plotting $(p_{ij} - p_i p_j)$ is plotting a covariance $cov(yz) = (\frac{1}{n_t} \Sigma yz - \bar{y}\,\bar{z})$ when y and z happen to be two-valued.

Extensive Monte Carlo tests have not yet been done, but several studies have used MAXCOV with dichotomous output variables (e.g., Harris, Rice, & Quinsey, 1994; Haslam & Beck, 1994; Korfine & Lenzenweger, 1995; Lenzenweger & Korfine, 1992; Strube, 1989; Trull, Widiger, & Guthrie, 1990; Tyrka, Cannon, Haslam, Mednick, Schulsinger, Schulsinger, & Parnas, 1995; Waller, Putnam, & Carlson, in press). The first was by Gangestad and Snyder (1985), who used an 8-item scale of self-monitoring, repeatedly removing two items for use as dichotomous output indicators with the remaining six items serving as the quantitative input indicator. Because the sampling variance of a "mean" p_i is considerably larger (as a proportion of

its true value) than is the sampling variance of $\bar{y}$ as a proportion of its true value, Gangestad and Snyder averaged estimates of the latent frequencies over the item pair covariances to reduce the instability. Averaging fallible MAXCOV estimates of the same latent values obtained from different pairs is an obvious way to pool estimates that are not identical and to cut down on sampling error. We have not used averaging in this way in our Monte Carlo runs with quantitative output indicators.

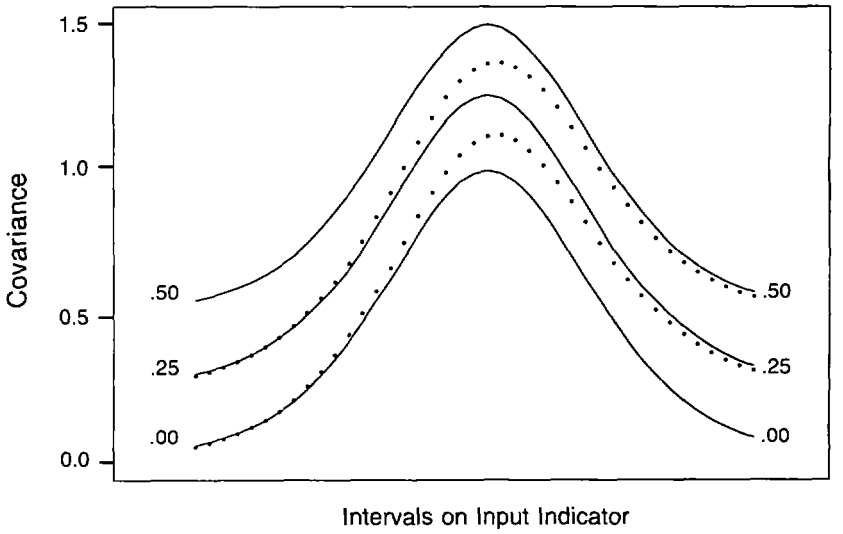


FIG. 12. Error-free MAXCOV curves for different amounts of nuisance covariance (P= .50, 2σ separation on indicators). Nuisance covariance in the complement group is indicated at the left of the curves; taxon nuisance covariance is at the right. Dotted curves have different amounts of covariance in the two groups.

Often researchers using dichotomous output indicators have included a nontaxonic control graph that is expected to be flat or additional graphs clearly expected to be taxonic and to have an expected base rate, hence to peak in a certain place. For instance, Trull, *et al.* (1990) generated a nicely flat (as expected) graph for dysthymia, a peaked (near the center, as expected for their sample) "control" graph using gender, along with their graph of the taxon of major interest, borderline personality disorder; the latter was cusped at the high end as expected for a small base rate taxon (though this curve was misinterpreted by those authors).

There has been some concern about the danger of spurious results when using dichotomous output pairs. What is the source of this concern? Why would $cov(yz)$ be peaked for $(p_{ij} - p_i p_j)$ but not if the variables are quantitative? If we are dealing with items which all have the same difficulty

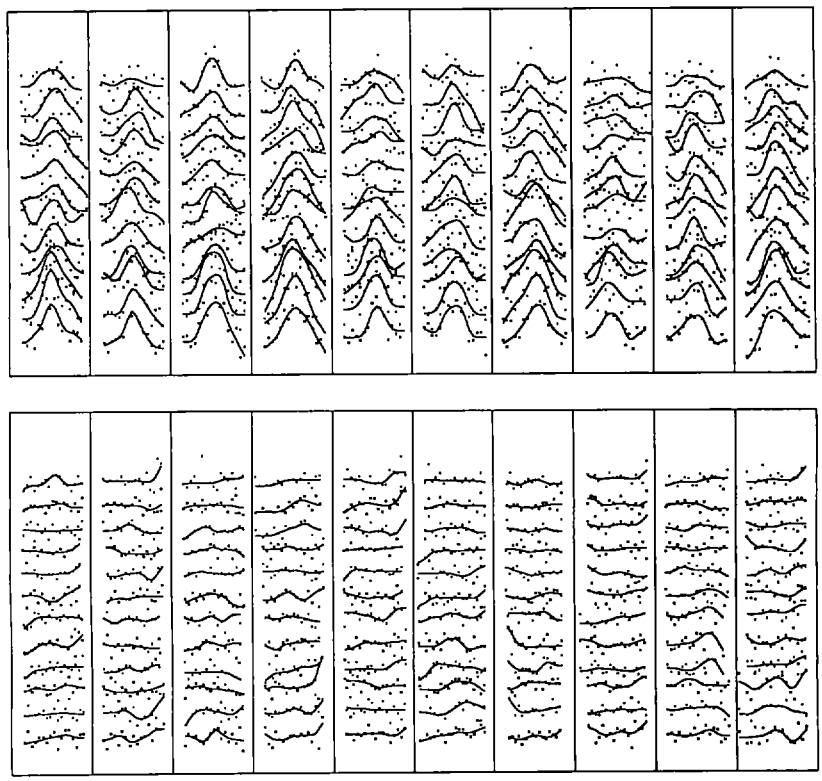


FIG. 13. Effect of nuisance covariance on MAXCOV curves. Intervals are based on abscissa distances. Top panels are from taxonic samples: $N = 600$, $P = .50$, 2.0 *SD* separation on each indicator, equal nuisance covariance in the complement and taxon groups generated by adding factor loadings $x = .70$, $y = .50$, $z = .40$, $v = .20$. Expected correlations (resulting from a combination of complement-taxon mixture and the factor loadings on the indicators): $r_{xy} = .68$, $r_{xz} = .64$, $r_{xv} = .57$, $r_{yz} = .60$, $r_{yv} = .55$, $r_{zv} = .54$. Bottom panels are from nontaxonic samples: $N = 600$, expected r_{ij} values equal to those in the taxonic samples by imposing factor loadings on: $x = .84$, $y = .79$, $z = .77$, $v = .69$.

level, very steep discrimination ogives are also needed to give spurious results. No real personality test items are like this, even if one tries to so construct them—they are never so "by accident." The easy check if this worry arises is, of course, to compute the item difficulties and plot the ogives.

Despite the impressive results that have been obtained by investigators using dichotomous outputs, we retain a strong preference for quantitative output indicators until more adequate Monte Carlo tests have been done. Since disparate marginal splits impose an upper limit on ϕ-coefficients, if for some reason such disparities were empirically correlated with item difficulty levels, an artefactual danger might arise when ϕ_{ij} between dichotomous out-

put indicators is the MAXCOV index. The danger of finding a "false taxonic" MAXCOV hump or cusp as a psychometric artefact due to an unfortunate pattern of dichotomous item parameters remains to be thoroughly explored.

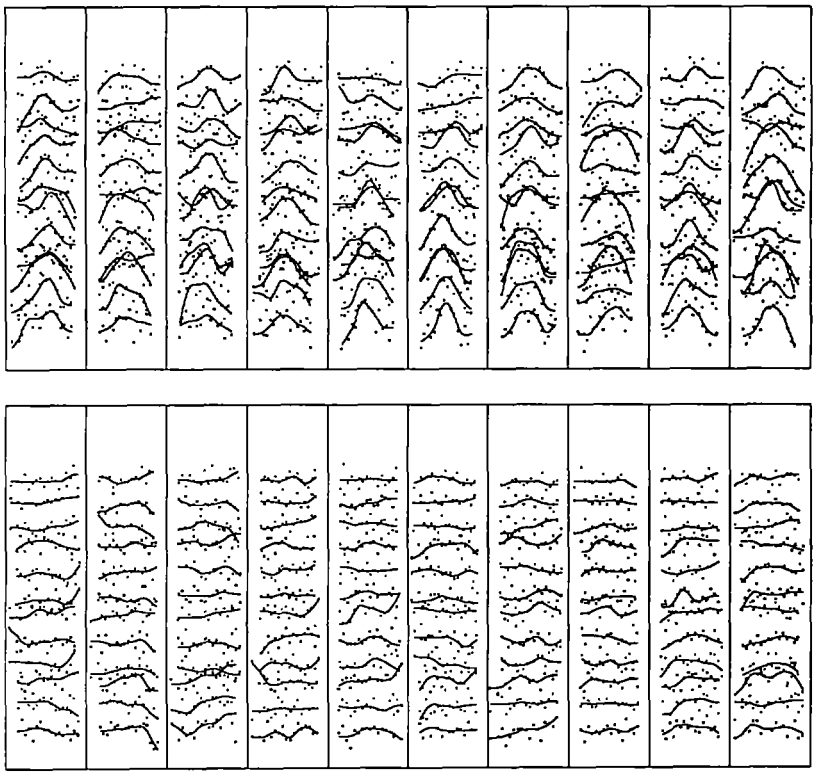


FIG. 14. Combined effects of nuisance covariance and various validities of 2.0 *SD* separation or less on MAXCOV curves. Intervals are based on abscissa distances. Top panels are from taxonic samples: $N=600$, $P=.50$. Factor loadings added to the taxonic samples: $x=.70$, $y=.50$, $z=.40$, $v=.20$; separations: $x=2.00$, $y=1.75$, $z=1.50$, $v=1.25$. Expected correlations (resulting from a combination of complement-taxon mixture and the factor loadings on the indicators): $r_{xy}=.65$, $r_{xz}=.58$, $r_{xv}=.46$, $r_{yz}=.52$, $r_{yv}=.41$, $r_{zv}=.37$. Bottom panels are from nontaxonic samples: $N=600$, expected r_{ij} values equal to those in the taxonic samples by imposing factor loadings on $x=.85$, $y=.76$, $z=.68$, $v=.54$.

THE HITMAX INTERVAL

Once we have determined from the MAXCOV curve shape that the latent structure of the data is taxonic, we can proceed to estimate the latent parameters. To do this we first estimate a latent (unobserved) *validity constant* value K, defined as the product of the differences between the latent

means of the output variables $(\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c)$. Of course, we do not know these latent means, hence we cannot know K directly; however, there is a way that we can estimate this value.

Theoretically, the maximum covariance occurs where there is the greatest mix of taxon and complement cases. When the base rate for the entire sample is .50, this maximum-mixture interval will be near the center of the MAXCOV curve; the maximum-mixture interval will be shifted to the right if the base rate is less than one-half (or to the left if the base rate is greater). If we were to call everyone falling above the maximum covariance interval a taxon member and everyone falling below it a member of the complement group, this would maximize our correct classifications, hence the term *hitmax* interval. Locating the hitmax interval is important because we can say certain things about it that make possible the parameter estimates to follow. Within this interval of maximum covariance is the point at which the latent distributions, one made up of complement cases and the other made up of the taxon cases, intersect. Thus, within the hitmax interval there are equal numbers of taxon and complement cases; 50% of the observed cases in that interval belong to each group. Fig. 15 shows a smoothed frequency distribution (what the researcher would observe) of a taxonic sample and the two latent curves that underlie it (not observed but only inferred by the researcher and to be estimated by the procedures outlined below). The hitmax interval is marked by vertical lines; the hitmax cut is that point on the x-axis where the two latent curves intersect within the hitmax interval.

We locate the hitmax interval by the peak of the MAXCOV curve, where the covariance is at a maximum. Within the hitmax interval, we have an observed (yz)-covariance (either smoothed or unsmoothed; unsmoothed values were used in the estimations reported here). From the covariance mixture formula, we know that within any interval

$$cov_{yz}(x) = pcov_{yzt} + qcov_{yzc} + pq(\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c) \qquad [2]$$

where p is the probability of taxon members within a single interval and q is the probability of complement membership. If we assume there is no nuisance covariance, i.e., that $cov_{yzt} \simeq 0$ and $cov_{yzc} \simeq 0$, then

$$\begin{aligned} cov_{yz}(x) &= p(0) + q(0) + pq(\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c) \\ &= pq(\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c) \ . \end{aligned} \qquad [3]$$

This is the *validity mixture* term, the one that is going to do the work for us. It is called that because, if there is no validity (separation on the indicators), the term will be zero; likewise, if there is no mixture of taxon and complement cases ($p = 0$ or $q = 0$), the term will be zero. Since the taxon ordinate equals the complement ordinate (because the distributions intersect) at the hitmax cut, the proportion of taxon cases in the hitmax interval equals the

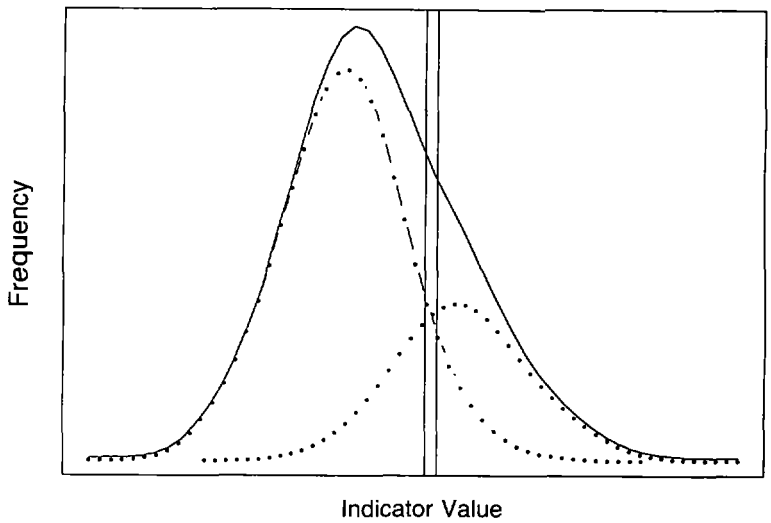


FIG. 15. A smoothed frequency distribution (solid line) and the frequency distributions for the complement (—) and taxon (· ·) distributions that make up the total sample. These curves were drawn from a taxonic Monte Carlo sample of $N = 1000$ with a base rate $P = .30$.

proportion of complement cases in that interval. So in that one interval we have

$$cov_{yzh} = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)(\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c) \ . \qquad [4]$$

If we define the validity constant K as the product of the separations, $(\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c)$, then in the hitmax interval

$$\begin{aligned} cov_{yzh} &= \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) K \\ &= \left(\frac{1}{4}\right) \cdot K \\ K &= 4cov_{yzh} \ . \end{aligned} \qquad [5]$$

Thus, the K value to use in making parameter estimations for a given triad of indicators is the covariance observed in the peak interval times 4. This calculation is included in the pseudocode given in Fig. 17 below.

Information about the accuracy of K estimates for our Monte Carlo samples is given in Appendix D (pp. 1178-1180). In general, estimates of K are more accurate with larger sample sizes, base rates closer to .50, and larger taxonic separations on the indicators. The effect of nuisance covariance is not clear from the limited Monte Carlo configurations we have tested. When it occurs in equal amounts within both groups and separations are good, it seems to have little effect on the K estimates, but when it occurs in varying amounts and in combination with reduced validities the effect is not clear.

When Latent Distributions Do Not Intersect

An unfavorable combination of low base rate and small separation may prevent the existence of a hitmax cut as defined because the latent frequency curves fail to intersect, the ordinate $f_c(x)$ exceeding $f_t(x)$ for all x, as illustrated in Fig. 16. There is then no cut x_c such that diagnoses of individuals scoring $x_t > x_c$ tend to be correct more often than not. The big complement distribution "swamps" the smaller taxon distribution throughout the empirical range, so that even the highest ordinate of $f_t(x)$ is exceeded by $f_c(x)$ at that point. If classification errors of both kinds are equally important (we do not here consider clinical disutilities but only theoretical research aims), it is never rational under this condition to diagnose an individual case as belonging to the taxon because $p_t(x_i) < \frac{1}{2}$ for all x_i (Meehl, 1956; Meehl & Rosen, 1955).

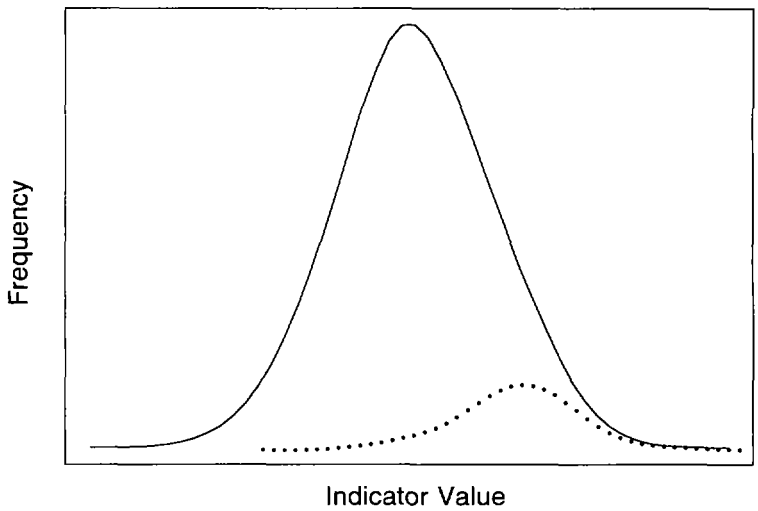


FIG. 16. A "swamped" taxon distribution (dotted line), falling wholly beneath the complement distribution (solid line)

But this undesirable state of affairs does not prevent a taxonic MAXCOV peak from appearing in the manifest graph of $cov_{yz}(x)$, since the validity mixture term $p_i q_i K$ will show a maximum when p_i is as close to $\frac{1}{2}$ as it can be. Despite $p < q$ therein, this MAXCOV interval is a "quasi-hitmax" interval, in the weak sense that were one to call all its cases taxon members—rational if it were a true hitmax interval, as it would be to call them all complement, or to flip a coin for each decision—the hit rate of

those calls would be as close to $\frac{1}{2}$ as could be achieved in any interval. Departing from calling "complement" for each case will yield fewer misses (although miss rate $> \frac{1}{2}$) in the MAXCOV interval than anywhere else.

The investigator, treating this apparent "hitmax" interval as usual, solves for the validity constant K_{yz} as if it were 4 cov_{max}; but since the pq relied on is actually $< \frac{1}{4}$, the $\hat{K}$ is too small. *Example*: Suppose the taxon ordinate at the taxon mode is only 4/5 as tall as the complement ordinate at that point, so the taxon probability in a small interval containing it is $4/9 = .44$. Then the pq wrongly assumed to be .25 is actually $(.44)(.56) = .2464$, a small discrepancy resulting in a propagated error of less than 2% in $\hat{K}$. Suppose the "swamping" is so bad that the complement ordinate is twice that of the taxon at the latter's peak; then $p_{t(max)} = .33$, $pq = .22$, yielding a propagated error of 11% in $\hat{K}$. This error in inferring K, while larger than one would readily accept, is nevertheless tolerable because it leads to partly countervailing errors in drawing the latent distributions. That there will be a sufficient number of intervals to the right of hitmax for some countervailing effect is plausible because a "weak" taxon[8] involves small separation, hence the taxon mean will usually be constrained toward the center—a MAXCOV hump rather than a top-interval cusp.

The swamping danger associated with small base rate will be warded off by sufficiently large separation, and conversely. Table 1 shows some representative combinations of base rates and separations that are just sufficiently adverse to begin swamping, i.e., yielding $P\phi_t(x_i) = f_t(x_i) = f_c(x_i) = Q\phi_c(x_i)$ where the ϕs are Gaussian densities at the taxon mode.

TABLE 1

SOME BASE RATE AND SEPARATION COMBINATIONS AT WHICH TAXON SWAMPING BEGINS

	SD Separation					
	1.00	1.25	1.50	1.75	2.00	2.25
P Value	.38	.31	.25	18	.12	.07

[8]We use the term 'weak' to describe a taxon whose taxometric detection is difficult due to low base rate, small indicator separation, or an unfavorable combination thereof. The extent to which such detection-weakness also tends to be associated with poorer accuracy in estimation of the latent parameters is unknown but we presume there is an appreciable relationship. Although this convention concerning 'weak' is conceptually epistemic—describing the state of evidence—that state of the evidence will, of course, occur under certain objective conditions, so it has an indirect (but not definitional) ontological overtone. For example, in a genetic taxon, such as schizotaxia, the base rate *P* reflects the gene frequency in a specified population; and the taxonic separation on a psychometric or neurological indicator depends on how many and how strong are the modifiers and potentiators that influence the causal chain running from DNA to the indicator.

It is conceivable that the largest $p_t(x_i)$ lies elsewhere than at the taxon distribution's mode, but it is very unlikely. For almost any weak taxon, the swamped taxon curve falls off to right of its peak faster than the complement, since even if we are in a region where the within-complement density $\phi_c(x)$ is declining faster than $\phi_t(x)$, e.g., nearer the former's flex point than the latter's, the absolute value of the derivative $|f_t'(x)| = |P\phi_t'(x)|$ will be smaller than $|f_c'(x)| = |Q\phi_c'(x)|$ given the "weak taxon" condition that $P \ll Q$.

This kind of situation needs thorough analytic and Monte Carlo investigation. Although we have not explored our Monte Carlo samples for possible swamping situations, neither has it emerged as a problem demanding attention. For now, we can only alert researchers to the possibility, suggesting that incoherency may detect it. For example, do the latent frequency distributions inferred via MAXCOV agree tolerably with those delivered by MAMBAC? A sizeable x-region of $cov(yz)$ near-flatness might suggest a swamping effect such that the latent distributions are declining at about the same rate, but we have no runs or analytic derivation of how special the conditions must be for this to occur. One would hope that, in a set of several indicators, some have separations sufficient to escape swamping even with a low base rate, permitting more trustworthy inferences when those indicators are used as input and appraising any questionable graphs in that light.

Estimating the Base Rate With MAXCOV

Once we have an estimate of K calculated from the hitmax interval, we can estimate the taxon probability p in any other interval. The covariance mixture formula holds in all intervals and for whatever values of p and q within a slice. Although we do not know the means of y and z for the taxon and complement groups, we do know that they will be some fixed values, hence the K which we have calculated via the covariance in the hitmax interval will be the same, neglecting sampling error, in any x interval we choose. We can write a quadratic equation in variable p_i (the probability of taxon members in an interval) for any interval x_i:

$$\begin{aligned} Kp_iq_i &= cov_{yz}(x_i) \\ Kp_i(1-p_i) - cov_{yz}(x_i) &= 0 \\ Kp_i - Kp_i^2 - cov_{yz}(x_i) &= 0 . \end{aligned} \qquad [6]$$

Reversing signs,

$$Kp_i^2 - Kp_i + cov_{yz}(x_i) = 0 \qquad [7]$$

in any interval x_i.

Given the observed covariance values that we have calculated and which we used to plot the MAXCOV curve, we solve the quadratic Equation [7] for each interval, using the quadratic algorithm

$$p_i = \frac{K \pm \sqrt{K^2 - 4K\, cov_{yzi}}}{2K}$$

$$= \frac{1}{2} \pm \sqrt{\frac{1}{4} - \frac{cov_{yzi}}{K}} \qquad [8]$$

and choosing the roots so that $p_i > \frac{1}{2}$ to the right of the hitmax interval, $p_i < \frac{1}{2}$ to the left. Once we have p_i for an interval, we multiply it by the observed interval frequency to get the estimated frequency of taxon cases in that interval,

$$n_{ti} = n_i p_i . \qquad [9]$$

Having the taxon frequency n_{ti} for each m interval, we sum them to get an estimated taxon total in the sample,

$$\hat{N}_t = \sum_{i=1}^{m} n_{ti} . \qquad [10]$$

The ratio of this value to the sample size is our estimate of the base rate

$$\hat{P} = \frac{\hat{N}_t}{N} . \qquad [11]$$

Pseudocode for MAXCOV base-rate estimates is shown in Fig. 17. We found it necessary to adopt two conventions when we calculated base-rate estimates for the Monte Carlo samples. First, small numbers and random error can result in negative covariances, especially at the ends of curves where the "true" covariance is approximately zero. For each interval, $cov = Kp_iq_i$; we know $K \neq 0$ (because the taxonic MAXCOV curve we have plotted tells us there is an underlying taxon with a mean greater than that of the complement group, assuming the indicators were scored initially in the customary positive direction); thus, if the covariance is zero, it must be the case that either $p_i = 0$ or $q_i = 0$. Our solution when negative covariances occurred was to infer "near purity" in such intervals and to assume that $p_i = 0$, i.e., no taxon members, all complement members, when a negative covariance occurred at the low end of the curve and $p_i = 1$, i.e., all taxon members, if it happened at the high end. Second, random error sometimes caused a covariance to be so large that cov/K was greater than .25, which would lead to taking the square root of a negative number. Our solution was to assume the term $\sqrt{.25 - cov/K}$ to be essentially zero (.001) when that occurred.

Table 2 shows the average base-rate estimates for different taxonic configurations. The 12 individual estimates for each input/output combination for each taxonic Monte Carlo sample may be found in Appendix E (pp. 1181-1195).

```
PROCEDURE EstimateBaseRate
(* This code presupposes a file containing covariances calculated for each interval
   and the observed number of cases in each interval *)
 FOR each input/output1,output2 combination
   Read file containing calculated covariance for each interval and
       number of observed cases n within each interval
   NTaxon := 0 (* reset counter for taxon members *)
   hitmax := interval with the maximum covariance (* locate largest covariance *)
   K := covariance in the hitmax interval * 4

   FOR each interval
     IF (interval = hitmax) THEN nTaxon[i] := .5 * n[i]
         (* in the hitmax interval, half the cases are taxon members *)

     ELSE IF (interval < hitmax)   (* for all intervals below the hitmax interval *)
         IF (covariance < 0) THEN pTaxon[i] := 0
             (* a convention to handle negative covariances on the low end *)

         ELSE IF (.25 - covariance/K) < .001 THEN pTaxon[i] := .5 - .001
         (* if cov/K is greater than .25, call the term essentially zero to avoid taking the
             square root of a negative number *)

         ELSE pTaxon[i] := .5 - √.25 - covariance/K
             (* subtract from .5 if interval is below the hitmax interval *)

         nTaxon[i] := pTaxon[i] * n[i]
         (* number of taxon members in this interval is pTaxon for the interval
             times the number of cases observed in the interval *)

     ELSE IF (interval > hitmax)   (* for all intervals above the hitmax interval *)
         IF (covariance < 0) pTaxon[i] := 1
             (* a convention to handle negative covariances on the high end *)

         ELSE IF (.25 - covariance/K) < .001 THEN pTaxon[i] := .5 + .001
         (* if cov/K is greater than .25, call the term essentially zero to avoid taking the
             square root of a negative number *)

         ELSE pTaxon[i] := .5 + √.25 - covariance/K
             (* add to .5 if interval is above the hitmax interval *)
         nTaxon[i] := pTaxon[i] * n[i]

     NTaxon := NTaxon + nTaxon[i]  (* sum the taxon estimates for each interval *)

   IF (i is last interval) P_estimate := NTaxon/N

END EstimateBaseRate
```

FIG. 17. Pseudocode for MAXCOV base-rate estimates

Frequency distributions of base rate estimates of all the Monte Carlo samples are shown in Fig. 18. If the MAXCOV curve is not taxonic, it would serve no purpose to calculate "base rate" estimates; however, the algebra goes through even when the underlying situation is nontaxonic, and such estimates may be seen in the upper part of Fig. 18. (The average estimates calculated for nontaxonic samples are given in Appendix F, p. 1196).

TABLE 2
AVERAGE BASE-RATE ESTIMATES (AVERAGES OVER 25 SAMPLES PER MONTE CARLO CONFIGURATION)

Sample Configuration		*N*	True Base Rate	Estimated Base Rate	
				M	*SD*
A1-50-20	2 *SD* separation on each variable	100	.50	.49	.04
A2-50-20	and no nuisance covariance	200	.50	.49	.03
A3-50-20		300	.50	.49	.02
A6-50-20		600	.50	.49	.02
A3-50-15	1.5 *SD* separation, no nuisance	300	.50	.51	.06
A6-50-15	covariance	600	.50	.50	.04
N3-50-20	Nuisance covariance in both	300	.50	.49	.04
N6-50-20	groups, 2 *SD* separation	600	.50	.49	.04
D3-50-v1	Various separations and nuisance	300	.50	.50	.07
D6-50-v1	covariance	600	.50	.49	.07
A3-25-20	2 *SD* separation, no nuisance	300	.25	.24	.02
A6-25-20	covariance	600	.25	.24	.02
A3-10-20	2 *SD* separation, no nuisance	300	.10	.13	.03
A6-10-20	covariance	600	.10	.10	.01

ESTIMATING LATENT MEANS AND VALIDITY (SEPARATION)

In the process of estimating the base rate, we have obtained estimates of the number of taxon cases within the intervals. Subtracting these from the observed frequency in each interval gives us an estimate of the number of complement cases per interval. If we multiply the estimated taxon frequency by the interval midpoints and divide by the total estimated taxon frequency, we get an estimate of the mean of the latent taxon distribution. Similarly, we can use the estimated complement numbers in each interval to get an estimated complement mean. Subtracting these estimates of the latent means gives us an estimate of the taxonic separation. Pseudocode for these steps is shown in Fig. 19.

Because the taxon number in an interval is gotten by an estimated taxon probability for that interval, the result is not necessarily a whole number. We opted to use the fractional numbers when calculating the estimated means. Another way would be to round to whole numbers of taxon or complement cases before multiplying by the interval midpoints. Our impression is that it makes little difference, but we have not explored how much difference it does make with our samples.

Table 3 shows the estimated latent means for our Monte Carlo samples. With four indicators, each sample gives three estimates of a latent mean for each indicator; there is an estimate for x when y and z are the output variables, another when y and v are used, and a third when z and v are used. We took the average of these three as the estimate of the latent mean for a

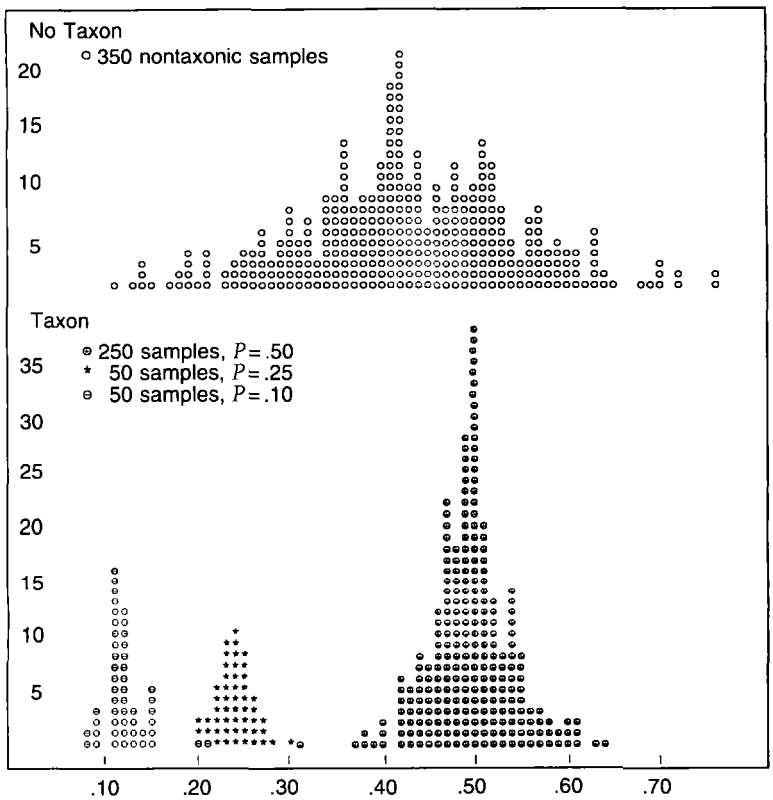


FIG. 18. Frequency distributions of MAXCOV base-rate estimates for all of the Monte Carlo samples

single sample. In Table 3 the values are further averaged across the 25 Monte Carlo samples for each configuration, and it is those means and standard deviations that are presented there. The estimates for individual samples may be found in Appendix G (p. 1197). All of the true complement means for the configurations listed in Table 3 are about zero, so those estimated means should be about the same. (Remember, we know the true complement and taxon means for these samples because we constructed them; the researcher would not have access to the true state of affairs.) The taxon means should all be about 2.00 except for samples with reduced validity, in which case the taxon means should be about 1.50, and samples with a combination of reduced validity and nuisance covariance; in the latter case, expected means for the indicators are given in the table. The estimates are generally better with larger sample sizes and no nuisance covariance. The

notable exception is samples with a low base rate ($P = .10$); the complement estimates in such samples are good, but the taxon estimates (which should be about 2.00 in our samples) are lowered.

```
PROCEDURE EstimateMeansAndSeparation
(* This code presupposes a file containing interval midpoints, total number of cases in
   each interval, and the estimated number of taxon cases in each interval *)

   FOR each input/output1,output2 combination
     Read file containing midpoints of intervals, total observed number of cases n[i] in
       each interval, and estimated number of taxon cases nTaxon[i] in each interval

   FOR each interval
     nComplement[i] := n[i] - nTaxon[i]      (* subtract taxon estimate from total n *)
   END (* for each interval *)               (* to get complement n in each interval *)

   NTaxon := Σ nTaxon[i]
   NComplement := Σ nComplement[i]

   meanTaxon := Σ (midpoint[i] * nTaxon[i]) / NTaxon
   meanComplement := Σ (midpoint[i] * nComplement[i]) / NComplement

   separation := meanTaxon - meanComplement

END EstimateMeansAndSeparation
```

FIG. 19. Pseudocode for estimates of means and separation

As noted before, the algebra works whether or not the MAXCOV curves look taxonic. If parameter estimates were attempted when the latent situation is nontaxonic, estimates of the "complement and taxon means" would be located more or less symmetrically around the observed distribution mean (see Appendix H, p. 1212); but without a taxonic MAXCOV curve these have no meaning.

To estimate the validity, the taxonic separation, we simply subtract the complement from the taxon mean. The results for the Monte Carlo samples, averaged over the 25 samples for each configuration, are given in Table 4 (results for individual samples in Appendix I, p. 1213). Expected separations should be about 2.00 *SD* in each case except for configurations with reduced validities ('A. . . –15' and 'D. . . ' sample codes). Obviously, taxonic separation tends to be underestimated, more so with smaller samples, small base rates, and nuisance covariance. Notice that we now have another way of estimating the validity constant K. This provides us with a consistency test for the K estimates and will be discussed below.

DRAWING THE LATENT CURVES

Having estimated the number of taxon members within each interval and, by subtraction from the observed frequencies, the complement mem-

TABLE 3
ESTIMATES OF LATENT MEANS

Monte Carlo Configuration		Complement								Taxon							
		x		*y*		*z*		*v*		*x*		*y*		*z*		*v*	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Sample Size																	
$N = 100$	A1-50-20	.19	.14	.19	.22	.14	.15	.23	.17	1.88	.21	1.82	.21	1.83	.14	1.91	.16
$N = 200$	A2-50-20	.11	.14	.04	.16	.05	.18	.07	.14	1.99	.14	1.95	.14	1.90	.09	1.92	.17
$N = 300$	A3-50-20	.09	.15	.05	.10	.02	.14	.07	.11	1.99	.13	1.99	.11	1.92	.14	1.97	.10
$N = 600$	A6-50-20	.06	.13	.05	.11	.05	.08	.04	.08	1.97	.10	1.98	.09	1.96	.09	2.02	.07
Different Base Rates																	
$P = .25$	A3-25-20	.06	.10	.03	.10	.08	.09	.05	.11	1.92	.18	1.91	.15	1.93	.13	1.88	.13
	A6-25-20	.03	.07	.05	.08	.06	.06	.06	.09	1.94	.11	1.98	.12	1.90	.10	1.93	.13
$P = .10$	A3-10-20	−.01	.10	.01	.06	.01	.08	−.01	.07	1.55	.23	1.73	.20	1.54	.20	1.62	.21
	A6-10-20	.00	.05	.04	.05	.02	.06	.02	.05	1.78	.15	1.79	.22	1.78	.14	1.82	.19
Reduced Validity (taxonic separation is 1.5 *SD*, expected taxon mean is 1.50)																	
	A3-50-15	−.02	.20	.03	.16	−.06	.13	.00	.17	1.49	.19	1.48	.14	1.48	.15	1.46	.18
	A6-50-15	.01	.15	.04	.12	.03	.14	.03	.15	1.49	.10	1.51	.14	1.41	.14	1.50	.12
Nuisance Covariance																	
	N3-50-20	.14	.17	.14	.13	.20	.18	.13	.15	1.88	.14	1.90	.17	1.94	.17	1.88	.15
	N6-50-20	.07	.15	.10	.14	.15	.13	.15	.19	1.93	.16	1.95	.12	1.89	.11	1.86	.11
Nuisance Covariance + Reduced Validities (expected taxon means: $x = 2.00$, $y = 1.75$, $z = 1.50$, $v = 1.25$)																	
	D3-50-v1	.12	.23	.09	.16	.05	.14	.08	.19	1.83	.22	1.62	.20	1.41	.17	1.26	.13
	D6-50-v1	.20	.24	.16	.16	.11	.12	.08	.16	1.77	.24	1.66	.16	1.47	.14	1.23	.09

TABLE 4
ESTIMATES OF TAXONIC SEPARATION

Monte Carlo Configuration		Separation							
		x		*y*		*z*		*v*	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Sample Size									
$N=100$	A1-50-20	1.69	.21	1.63	.21	1.69	.18	1.68	.17
$N=200$	A2-50-20	1.87	.12	1.90	.15	1.85	.18	1.84	.17
$N=300$	A3-50-20	1.90	.13	1.94	.11	1.90	.17	1.89	.12
$N=600$	A6-50-20	1.91	.18	1.93	.11	1.91	.09	1.98	.08
Different Base Rates									
$P=.25$	A3-25-20	1.86	.17	1.88	.16	1.86	.11	1.84	.11
	A6-25-20	1.91	.10	1.93	.14	1.84	.11	1.87	.17
$P=.10$	A3-10-20	1.56	.21	1.72	.21	1.54	.23	1.63	.19
	A6-10-20	1.78	.05	1.75	.23	1.75	.12	1.80	.18
Reduced Validity (expected separation is 1.5 *SD*)									
	A3-50-15	1.51	.17	1.46	.17	1.53	.15	1.46	.17
	A6-50-15	1.49	.15	1.48	.16	1.38	.20	1.48	.17
Nuisance Covariance									
	N3-50-20	1.74	.19	1.76	.17	1.74	.14	1.75	.15
	N6-50-20	1.86	.24	1.85	.14	1.74	.20	1.72	.25
Nuisance Covariance + Reduced Validities (expected separation: $x=2.00$, $y=1.75$, $z=1.50$, $v=1.25$)									
	D3-50-v1	1.71	.23	1.54	.22	1.36	.16	1.18	.16
	D6-50-v1	1.57	.31	1.50	.25	1.36	.19	1.15	.18

bers, one could draw the latent complement and taxon distributions. Fig. 20 illustrates latent curves drawn for one input indicator using the three possible output combinations for a Monte Carlo sample in which the base rate is .50, there is good separation on each indicator, and no nuisance covariance is present. This sample was unselected; it is merely the first one in that configuration. The taxon frequencies estimated to be in each interval (and used to draw the curves in Fig. 20) are shown in Table 5. As can be seen by the correlations of the estimated and true taxon frequencies, the MAXCOV procedure does a very good job here.

CLASSIFYING INDIVIDUALS

In the first publication on MAXCOV, Bayes' Rule was applied in classifying individuals, using the inferred valid and false (+) and (−) rates determined by the hitmax cut on each indicator (Meehl, 1973, p. 214). This use of coarsely defined cut-determined probabilities throws away the information provided by varying taxon probabilities per interval. Using the latter in Bayes' formula often leads to a different classification decision than the cruder approach. If an individual's two dichotomous signs x^+, y^+ result from

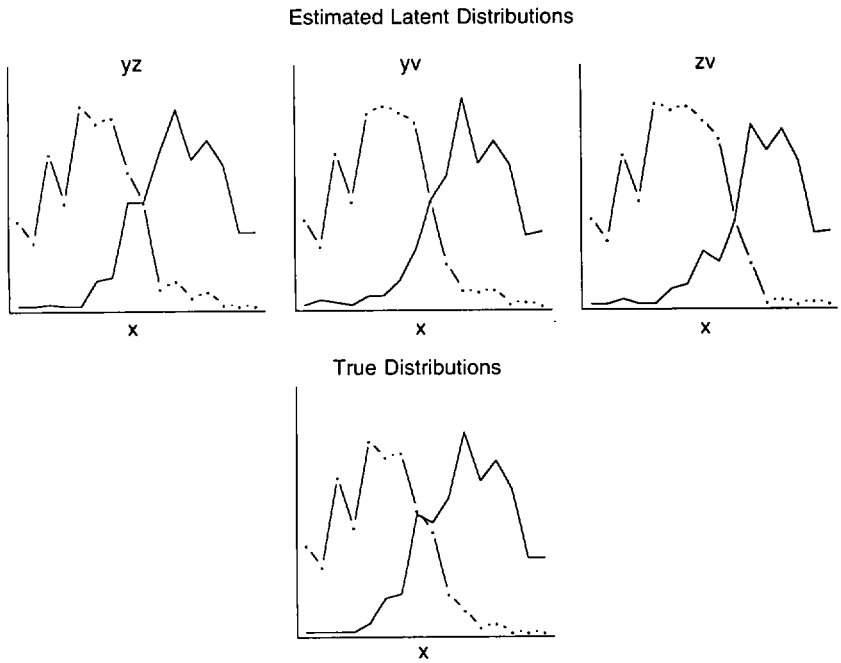


FIG. 20. Estimated latent complement and taxon distributions on x using different output indicators and the true complement and taxon distributions

scores $x_i > x_h$ and $y_i > y_h$ being in the first intervals above the x and y hitmax cuts, and the third sign z^- is associated with a z-indicator score lying in a very low z-interval, the coarse classification is "wrong," i.e., will err far more often than the fine one. Failure to mention this obvious point was due to two extenuating circumstances. At the time of that publication, Minnesota clinicians were preoccupied with optimizing cutting scores on MMPI keys and with configural rules that were usually inherently dichotomous, e.g., a specified Hathaway code, satisfying a Meehl-Dahlstrom rule, decision "profile invalid" by raw $F > 16$ or Gough's dissimulation scale *Di*, the Marks-Seeman (1963) profile types which classed cases as "in" or "out" of a category. The emphasis was on valid and false (+) and (−) rates—the epidemiologists' *sensitivity* and *specificity*—as yielded by a cutting score, profile pattern, or dichotomous sign or symptom found in the medical chart or interviewer checklists. Second, the author had little confidence in the statistical reliability of slice-probabilities (with the modest sample sizes then being used); adding taxon tallies over slices above and below x_h seemed safe, but multiplying three unstable proportions estimated intraslice seemed dangerous. It is now

clear that this seemingly "cautious" line of thought entails discarding a large amount of information. We have not yet assessed the accuracy of these different approaches to classifying individuals with our Monte Carlo samples. Just how trustworthy the inferred latent distributions (smoothed and raw) are is a complicated question we defer to subsequent publications.

TABLE 5
NUMBER OF TAXON CASES ESTIMATED TO BE IN EACH INTERVAL ALONG THE *x* INPUT VARIABLE USING DIFFERENT OUTPUT INDICATOR PAIRS (MONTE CARLO SAMPLE A6-50-20.1; 2 *SD* SEPARATION ON EACH INDICATOR, NO NUISANCE COVARIANCE)

Interval	Total *n*	True Taxon *n*	Estimated by Output Pair		
			yz	*yv*	*zv*
≤1	21	0	0	0	0
2	16	0	0	1	0
3	37	0	0	1	1
4	28	0	0	0	0
5	45	2	0	2	0
6	50	8	6	2	4
7	52	9	7	6	5
8	58	28	25	13	13
9	49	26	25	25	10
10	41	32	37	31	20
11	53	48	47	49	43
12	37	36	35	34	37
13	43	41	40	39	42
14	34	34	34	34	34
15	18	18	13	17	17
16	18	18	13	18	18
Total	600	300	282	272	244
Correlation With True *n*			.99	.97	.94

CONSISTENCY TESTS

A *consistency test* is any numerical procedure capable of falsifying a conjecture as to latent (compositional or causal) structure or, given a structure, capable of indicating that one's estimates of the latent parameters are untrustworthy. We rely on theorems asserting equalities or inequalities between numerical values summarizing observations or among latent values inferred from observational statistics. There are several kinds of consistency tests (Meehl, 1995a, p. 272; Waller & Meehl, in preparation) that differ with respect to the mix of manifest and latent values with distinct inferential paths. The essential feature of a consistency test is that the theorem into which observed or inferred numerical values are inserted (to see whether the equality or inequality is satisfied) is not a mathematical identity—necessarily satisfied by any numerical values we assign—but is derived from postulates that implicitly define the structural model we are examining.

For example, if x, y are observable variables ("indicators"), the relation $r^2 = 1 - \frac{\sigma^2_{y \cdot x}}{\sigma^2_y}$ holds for any least squares fitted straight line $y = a + bx$ (even if a parabola should have been fitted instead!), hence it cannot serve as a consistency test of any structural conjecture or parameter estimate. The General Covariance Mixture Theorem (Equation 1 *supra*) is in latent terms, but it is an algebraic identity, satisfied by any arbitrary partition of cases into two classes. Contrast these two identities with the grand covariance consistency test

$$\text{Cov}(yz) = PQ(\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c) \qquad [12]$$

which is not an algebraic identity *except within the postulated taxonic model.* If there is in fact no latent taxon, the taxometrically inferred values P, Q, $\bar{y}_t$, $\bar{y}_c$, $\bar{z}_t$, $\bar{z}_c$ do not *denote* (have no referent). In that case, there is no reason to expect that the left-hand term, when computed directly from the manifest variables, will "match" the right-hand numerical value when it is concocted from these six nondenoting numbers.

How can a consistency test be nonredundant, not a mere tautology, since it relies on a theorem in the formalism? The short answer is that the theorem is not a theorem of general mathematics but of the formal postulates that define the latent structure. The interpretive text (Meehl, 1990a, p. 109; 1990b, pp. 3-5) motivates the postulates and licenses some steps in the derivations; hence, satisfying multiple consistency tests can function to corroborate the substantive theory that the text asserts. Of course, the *strength* of this corroboration depends on how antecedently improbable (a Salmonean "strange coincidence," Meehl, 1990a) the numerical relation would be, absent any latent theory, given one's background knowledge, e.g., usual numerical range of the several observables (Meehl, in press). Consistency tests are based on *vertical derivability* (from structural postulates "downward") of theorems together with absence of *horizontal derivability* (numerical equalities required by general mathematical identities without the adjoined taxonomic postulates).[9]

The crucial role of consistency tests in Meehl's coherent cut kinetics method has been emphasized repeatedly (Golden & Meehl, 1973b; Meehl, 1965, 1992, 1995a; Meehl & Golden, 1982; Meehl & Yonce, 1994). Coher-

[9]A helpful clarifier here is Carnap's (1939) distinction between the *general calculus* (mathematics, logic) and the *special calculus* (additional formal postulates of a theory in empirical science). These adjoined postulates can be set out without their interpretive text so, although a reader would recognize their special character, one would not know which empirical science is being formalized. The postulates are formulable *in* the general calculus, uninterpreted, but they are not theorems *of* the general calculus. This question ties in with the metatheoretical concepts of construct validity, implicit definition, and bootstrapping, discussed by Meehl (1995a).

ent cut kinetics relies on the dependence of a statistic on the *taxon/complement mixture* in subsamples of cases defined by an "input" indicator. Sometimes the subsamples may be defined by successive intervals, sometimes by cases above and below successive cuts along the input variable. We will describe briefly several possible consistency tests that could be used. Some involve procedures previously published; for others we offer the core intuitive reasoning and illustrative Monte Carlo results, reserving further and more rigorous mathematical development for subsequent articles.

Agreement Between Parameters Obtained in Different Ways

One consistency test suggested by Meehl (1965, 1995a, p. 272; Meehl & Golden, 1982, p. 165, Eq. 24) compares the observed grand covariances with those constructed using the estimates of base rate and latent separations. If the parameter estimates are trustworthy and there is no nuisance covariance, we should be able to reconstitute the covariances. If there is no nuisance covariance, the grand covariance may be inferred by

$$\mathrm{Cov}_{xy} = PQ\bar{\Delta}x\,\bar{\Delta}y\,, \qquad [13]$$

and once we have estimates of P and the separations for x and y, we can calculate Cov_{xy} that way. How do these different estimates of Cov_{xy} compare? Table 6 shows summary data for the reconstituted minus observed covariances for different sample configurations. In our samples, the reconstituted covariance nearly always underestimates the observed covariance. With larger samples and base rates of .50 and .25 there is less discrepancy. As we would expect, there are larger discrepancies when there is nuisance covariance, because the model is then violated. This is a rich area for investigation. There are obviously relations between taxonicity, sample size, and amount of nuisance covariance. The correlations between reconstituted and observed covariances are higher in taxonic samples, but they are lowered due to range restrictions when samples are larger (and estimations are better). Perhaps it is possible to use the reconstituted/observed covariance discrepancy as another way of estimating nuisance covariance. The relationships also suggest further coherence checks: a researcher with adequate sample size can construct a new sample and see whether the discrepancy changes in a predicted manner.

MAMBAC

The MAXCOV-HITMAX procedure begins by locating the x-interval in which the observed (yz)-covariance is a maximum, inferring that in this interval the mix of taxon and complement cases is greatest ($p_i = q_i = \frac{1}{2}$). In the MAMBAC (for Mean Above Minus mean Below A Cut) procedure (Meehl & Yonce, 1994) we look at the other side of the coin, reasoning that

TABLE 6

RECONSTITUTED COVARIANCES MINUS OBSERVED COVARIANCES, MEAN DIFFERENCES FOR 25 SAMPLES IN EACH CONFIGURATION

Monte Carlo Configuration		Taxonic Samples												r
		xy		xz		xv		yz		yv		zv		
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
Sample Size														
$N=100$	(A1-50-20)	−.26	.12	−.27	.12	−.28	.11	−.28	.14	−.30	.15	−.29	.12	.63
$N=200$	(A2-50-20)	−.14	.09	−.15	.11	−.15	.08	−.12	.11	−.16	.10	−.13	.13	.59
$N=300$	(A3-50-20)	−.08	.07	−.09	.08	−.08	.07	−.11	.10	−.10	.06	−.10	.07	.61
$N=600$	(A6-50-20)	−.09	.10	−.09	.09	−.08	.10	−.06	.07	−.04	.08	−.06	.05	.35
Different Base Rates														
$P=.25$														
$N=300$	(A3-25-20)	−.10	.10	−.12	.09	−.13	.07	−.12	.07	−.13	.10	−.10	.07	.40
$N=600$	(A6-25-20)	−.08	.06	−.09	.08	−.09	.07	−.06	.08	−.09	.09	−.10	.11	.36
$P=.10$														
$N=300$	(A3-10-20)	−.09	.08	−.09	.08	−.05	.07	−.08	06	−.06	.09	−.06	.05	.28
$N=600$	(A6-10-20)	−.09	.06	−.09	.05	−.08	.06	−.07	08	−.08	.05	−.08	.05	.28
Reduced Validity (taxonic separation is 1.5 *SD*, expected taxon mean is 1.50)														
$N=300$	(A3-50-15)	−.03	.11	.01	.10	−.03	.12	−.01	.12	−.01	.09	−.03	.09	.22
$N=600$	(A6-50-15)	−.00	.08	−.06	.09	−.02	.09	−.04	.10	−.02	.10	−.06	.11	.21
Nuisance Covariance														
$N=300$	(N3-50-20)	−.58	.14	−.49	.13	−.39	.14	−.42	.11	−.35	.12	−.35	.11	.14
$N=600$	(N6-50-20)	−.52	.14	−.48	.15	−.36	.16	−.43	.13	−.33	.15	−.34	.15	.35
Nuisance Covariance + Reduced Validities (expected taxon means: $x=2.00$, $y=1.75$, $z=1.50$, $v=1.25$)														
$N=300$	(D3-50-v1)	−.57	.16	−.46	.14	−.28	.14	−.32	.10	−.21	.12	−.15	.10	.62
$N=600$	(D6-50-v1)	−.65	.17	−.52	.16	−.31	.13	−.36	.14	−.20	.11	−.17	.10	.48

(continued on next page)

TABLE 6 (CONT'D)

RECONSTITUTED COVARIANCES MINUS OBSERVED COVARIANCES, MEAN DIFFERENCES FOR 25 SAMPLES IN EACH CONFIGURATION

Monte Carlo Configuration		Nontaxonic Samples												
		xy		*xz*		*xv*		*yz*		*yv*		*zv*		*r*
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
r_{ij} = .50 (taxonic samples *P* = .50, 2 *SD* separation)														
N = 100	(C100)	−.26	.09	−.28	.09	−.24	.10	−.28	.09	−.25	.10	−.27	.09	−.16
N = 200	(C200)	−.27	.10	−.27	.08	−.25	.08	−.28	.10	−.26	.11	−.26	.09	−.08
N = 300	(C300)	−.35	.08	−.34	.07	−.34	.05	−.37	.07	−.36	.07	−.35	.06	.06
N = 600	(C600)	−.41	.06	−.41	.07	−.42	.06	−.40	.07	−.39	.07	−.40	.06	.15
r_{ij} = .43 (taxonic samples *P* = .25)														
N = 300	(F300)	−.26	.09	−.27	.09	−.24	.07	−.25	08	−.23	.10	−.25	.09	−.06
N = 600	(F600)	−.34	.07	−.34	.07	−.34	.06	−.32	08	−.32	.07	−.32	.07	.02
r_{ij} = .26 (taxonic samples *P* = .10)														
N = 300	(E300)	−.07	.09	−.06	.09	−.04	.10	−.06	.09	−.07	.08	−.04	.09	.02
N = 600	(E600)	−.11	.07	−.10	.08	−.08	.08	−.10	.08	−.11	.08	−.09	.08	.05
r_{ij} = .36 (taxonic samples separation 1.5 *SD*)														
N = 300	(B300)	−.21	.09	−.18	.08	−.18	.07	−.20	.10	−.19	.07	−.17	.08	−.07
N = 600	(B600)	−.27	.06	−.24	.06	−.26	.08	−.26	.07	−.26	.07	−.25	.06	.13
r_{ij} = varies (taxonic samples have nuisance covariance)														
N = 300	(N=300)	−.48	.08	−.47	.09	−.41	.09	−.44	.08	−.39	.09	−.37	.09	.11
N = 600	(N=600)	−.59	.07	−.56	.06	−.50	.06	−.54	.07	−.49	.06	−.45	.05	.05
r_{ij} = varies (taxonic samples have nuisance covariance + reduced validities)														
N = 300	(D300)	−.44	.08	−.37	.06	−.28	.08	−.34	.07	−.28	.08	−.22	.06	.26
N = 600	(D600)	−.52	.09	−.46	.08	−.34	.09	−.42	.07	−.33	.07	−.28	.07	.02

less mixture within cases above (or below) a sliding x-cut tends to increase the observed separation of y-means, the output statistic $\bar{d}_y(x)$ being graphed to discern that. The Monte Carlo samples used in this article are identical to those used in Meehl and Yonce (1994). The text and appendices of that monograph and the present one make it possible for any given sample to be examined for its performance in these different procedures.

MAXSLOPE

Another procedure, MAXSLOPE (Grove & Meehl, 1993), is conceptually similar to MAXCOV but requires only two variables and looks at the regression slope of one over successive intervals along a second (input) indicator. We have not run MAXSLOPE on all of the samples described in this article, but Fig. 21 illustrates the differences in the regression slopes for y on x for the first sample in a taxonic and a nontaxonic situation. MAXSLOPE also locates the hitmax cut and estimates the base rate and latent parameters.

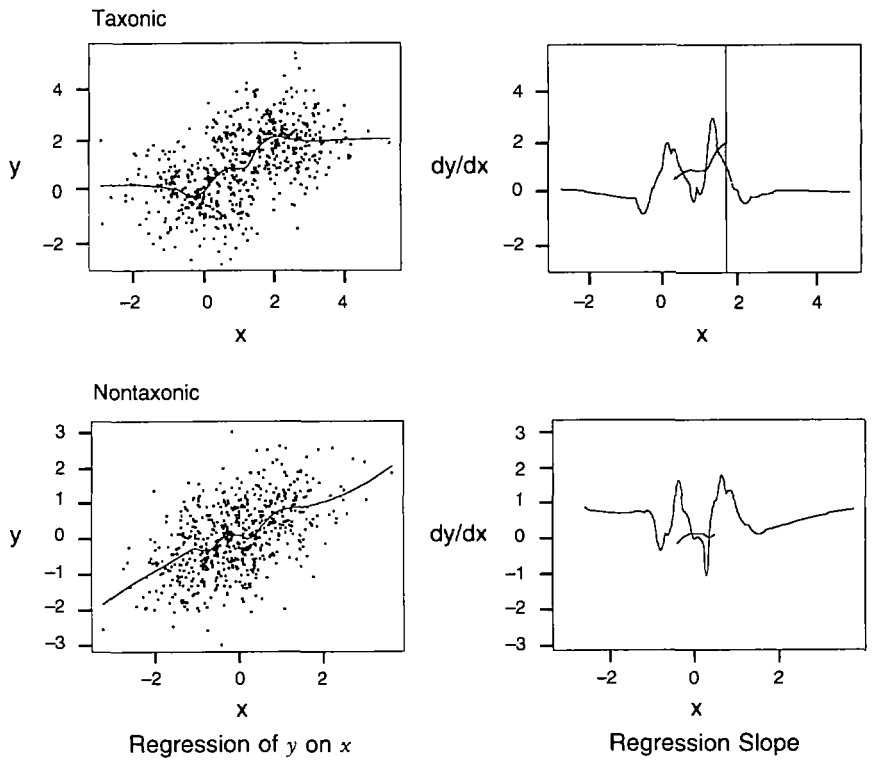


FIG. 21. MAXSLOPE graphs for taxonic ($N = 600$, $P = .50$, 2 *SD* separation) and nontaxonic ($N = 600$, $r_{ij} = .50$) Monte Carlo samples. The hitmax cut is shown for the taxonic sample.

Minimizing $SS_b + SS_a$

The dispersion of (output) indicator y is a composite of the latent class dispersions and the taxonic separation $(\bar{y}_t - \bar{y}_c)$. By analogy with MAMBAC and MAXCOV combined, if we locate an x-cut so as to minimize the "mix" within the two subgroups identified as above or below x_c, the resulting two y-dispersions should tend to be smaller than if the cut yielded less "purified" subsets above and below the cut. Monte Carlo runs and the quasi-proof by Meehl (1968, pp. 32-39) indicate the sums of squares are better than the variances. Rigorous proofs will be presented in a subsequent article. Fig. 22 shows results for taxonic samples with different base rates, together with a shallow dish for the nontaxonic case. Here the $P = .50$ taxonic graph ratio $(SS_{MAX} - SS_{MIN}) / SS_{MIN}$ is over twice that of the nontaxonic dish ("depth"), but whether a safe, robust decision rule for smaller separations is formulable awaits further investigation. Unpublished Monte Carlo work by Robert Golden encourages us to consider shifting $SS_b + SS_a$, when combined with study of the ordinary $y \cdot x$ regression line, from a consistency test to a main search procedure. Meanwhile, the statistic should be treated as a marginal consistency test.

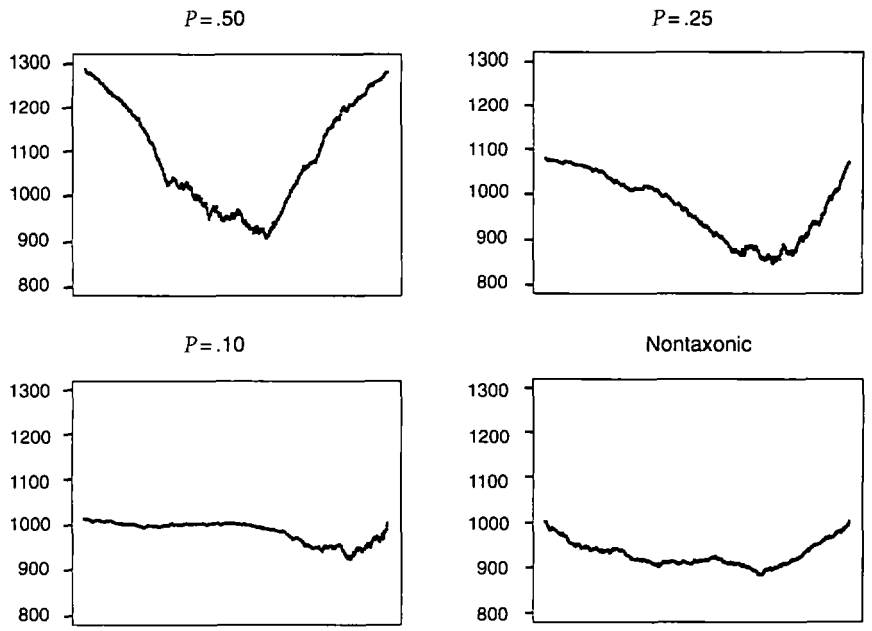


FIG. 22. Sums of squares below plus sums of squares above successive cuts on an "input" indicator for different base rates (all with $N = 600$ and 2 SD separations) and for a nontaxonic sample ($N = 600$, $r_{ij} = .50$)

Maximizing $\sigma^2_{y \cdot x}$ *in* x-*interval*

The structural relations motivating MAMBAC and MAXCOV suggest yet another procedure that employs an input indicator to maximize an output statistic. Analogous to the general covariance mixture theorem on which MAXCOV-HITMAX relies (a variance is a "self-covariance"), there is a relation involving only a single output indicator, thus:

$$\sigma^2_{y \cdot x} = p_i \sigma^2_{yt} + q_i \sigma^2_{yc} + p_i q_i (\bar{y}_t - \bar{y}_c)^2 . \qquad [14]$$

If the variances are equal, this quantity is maximized in the x-interval where $p_i = q_i = \frac{1}{2}$ (Robert R. Golden, personal communications, 2/15/75, 12/7/85). In the nontaxonic situation we expect no such clear "hump" in the output graph, and if the dependence of y on x is a linear homoscedastic regression, while $\bar{y}_i$ has nonzero slope, the graph of $\sigma^2_{y \cdot x}$ is flat. Fig. 23 displays these relations for three base rates and for the nontaxonic case plotted on comparable axes. Whether minimizing $SS_b + SS_a$ and maximizing $\sigma^2_{y \cdot x}$ in x-interval can also be employed as main procedures (yielding accurate estimates of P, hitmax cut, valid and false (+) rates, and separations) awaits investigation.

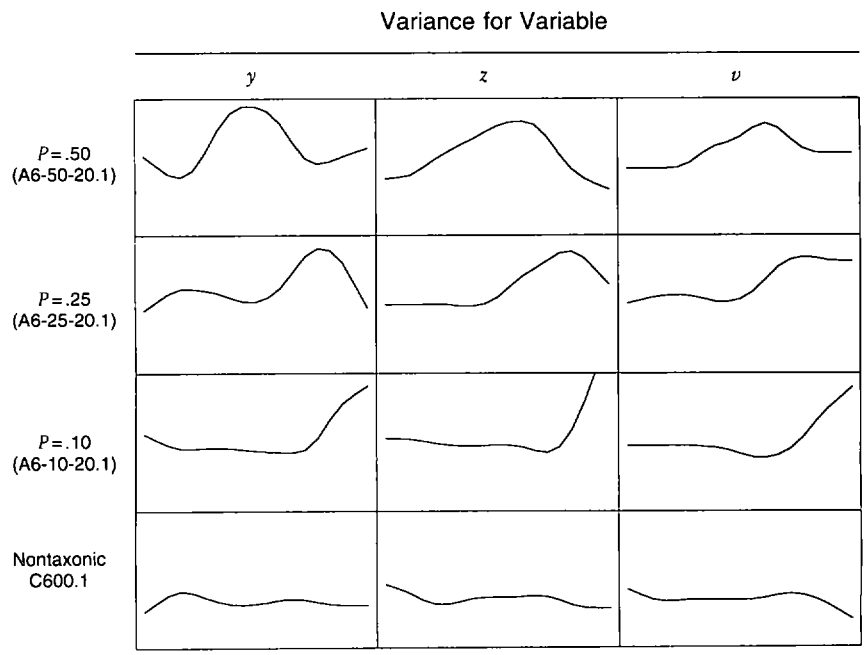


FIG. 23. Variance of several output variables in vigintiles along an input variable x for different Monte Carlo configurations

RESUMÉ

We have demonstrated the MAXCOV procedure with a range of Monte Carlo configurations. If one conjectures the existence of a latent taxon (type, categorical entity) of which three or more quantitative observable variables are fallible indicators, the MAXCOV-HITMAX procedure provides a test of the taxonic conjecture and estimates of the latent parameters. Relying on the General Covariance Mixture Theorem, the procedure locates the hitmax interval (optimal diagnostic cut) on an "input" indicator by graphing the covariance of two "output" variables, solves for a latent constant which is equal to the product of the separations of the two latent means, and uses that constant to solve quadratics to obtain the taxon frequency in each interval, thereby drawing the two latent distributions. No "criterion" is required, and the procedure is self-validating by the coherence of quantitative values inferred by different epistemic paths. Given trustworthy (coherent) inferred values, individuals are classifiable via Bayes' Rule. The underlying derivations are free of problematic conventional assumptions about distributions, e.g., normality, equal variance, although our Monte Carlo runs in this monograph were generated by a Gaussian algorithm. The procedure appears satisfactorily robust under departures from the idealization of within-category independence; a modified procedure is available (Meehl, 1995b) for situations of sizeable nuisance covariance. Unlike cluster methods, MAXCOV does not deliver a taxon when the latent generating structure is nontaxonic. As with all statistical procedures, it is neutral with respect to substantive issues concerning the causal origin of taxonicity.

REFERENCES

CARNAP, R. (1939) Foundations of logic and mathematics. *International Encyclopedia of Unified Science*, 1(3).

CRAMÉR, H. (1946) *Mathematical methods of statistics*. Princeton, NJ: Princeton Univer. Press.

FISHER, R. A. (1970) *Statistical methods for research workers*. (14th ed.) Edinburgh: Oliver & Boyd.

GANGESTAD, S., & SNYDER, M. (1985) "To carve nature at its joints": on the existence of discrete classes in personality. *Psychological Review*, 92, 317-349.

GOLDEN, R., & MEEHL, P. E. (1973a) *Detecting latent clinical taxa: IV. An empirical study of the maximum covariance method and the normal minimum chi-square method using three MMPI keys to identify the sexes*. (Report No. PR-73-2) Minneapolis, MN: Univer. of Minnesota, Research Laboratories of the Department of Psychiatry.

GOLDEN, R., & MEEHL, P. E. (1973b) *Detecting latent clinical taxa: V. A Monte Carlo study of the maximum covariance method and associated consistency tests*. (Report No. PR-73-3) Minneapolis, MN: Univer. of Minnesota, Research Laboratories of the Department of Psychiatry.

GROVE, W. M., & MEEHL, P. E. (1993) Simple regression-based procedures for taxometric investigations. *Psychological Reports*, 73, 707-737.

HARRIS, G. T., RICE, M. E., & QUINSEY, V. L. (1994) Psychopathy as a taxon: evidence that psychopaths are a discrete class. *Journal of Consulting and Clinical Psychology*, 62, 387-397.

HASLAM, N., & BECK, A. T. (1994) Subtyping major depression: a taxometric analysis. *Journal of Abnormal Psychology*, 103, 686-692.

KENNEY, J. F. (1939) *Mathematics of statistics, Part II*. New York: Van Nostrand.

KORFINE, L., & LENZENWEGER, M. F. (1995) The taxonicity of schizotypy: a replication. *Journal of Abnormal Psychology*, 104, 26-31.

LENZENWEGER, M. F. (1993) Explorations in schizotypy and the psychometric high-risk paradigm. In L. J. Chapman, J. P. Chapman, & D. C. Fowles (Eds.), *Progress in experimental personality and psychopathology research*. Vol. 16. New York: Springer. Pp. 66-116.

LENZENWEGER, M. F., & KORFINE, L. (1992) Confirming the latent structure and base rate of schizotypy: a taxometric analysis. *Journal of Abnormal Psychology*, 101, 567-571.

LI, C. C. (1975) *Path analysis: a primer*. Pacific Grove, CA: Boxwood Press.

LOWRIE, G. S., & RAULIN, M. L. (1990) Search for schizotypic and nonschizotypic taxonomies in a college population. Paper presented at the Sixty-first Annual Convention of the Eastern Psychological Association, Philadelphia, PA, March 29-April 1, 1990.

MEEHL, P. E. (1956) Wanted—a good cookbook. *American Psychologist*, 11, 263-272. [Reprinted in P. E. Meehl, *Psychodiagnosis: selected papers*. Minneapolis, MN: Univer. of Minnesota Press, 1973. Pp. 63-80.]

MEEHL, P. E. (1965) *Detecting latent clinical taxa by fallible quantitative indicators lacking an accepted criterion*. (Report No. PR-65-2) Minneapolis, MN: Univer. of Minnesota, Research Laboratories of the Department of Psychiatry.

MEEHL, P. E. (1968) *Detecting latent clinical taxa: II. A simplified procedure, some additional hitmax cut locators, a single-indicator method, and miscellaneous theorems*. (Report No. PR-68-2) Minneapolis, MN: Univer. of Minnesota, Research Laboratories of the Department of Psychiatry.

MEEHL, P. E. (1973) MAXCOV-HITMAX: a taxonomic search method for loose genetic syndromes. In P. E. Meehl, *Psychodiagnosis: selected papers*. Minneapolis, MN: Univer. of Minnesota Press. Pp. 200-224.

MEEHL, P. E. (1990a) Appraising and amending theories: the strategy of Lakatosian defense and two principles that warrant using it. *Psychological Inquiry*, 1, 108-141, 173-180.

MEEHL, P. E. (1990b) *Corroboration and verisimilitude: against Lakatos' "sheer leap of faith."* (Working Paper, MCPS-90-01) Minneapolis, MN: Univer. of Minnesota, Center for Philosophy of Science.

MEEHL, P. E. (1990c) Toward an integrated theory of schizotaxia, schizotypy, and schizophrenia. *Journal of Personality Disorders*, 4, 1-99.

MEEHL, P. E. (1992) Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality*, 60, 117-174.

MEEHL, P. E. (1995a) Bootstraps taxometrics: solving the classification problem in psychopathology. *American Psychologist*, 50, 266-275.

MEEHL, P. E. (1995b) Extension of the MAXCOV-HITMAX taxometric procedure to situations of sizeable nuisance covariance. In D. Lubinski & R. V. Dawis (Eds.), *Assessing individual differences in human behavior: new concepts, methods, and findings*. Palo Alto, CA: Davies-Black. Pp. 81-92. [This is the first publication describing the generalized MAXCOV procedure.]

MEEHL, P. E. (in press) The problem is epistemology, not statistics: replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow (Ed.), *What if there were no significance tests?* Hillsdale, NJ: Erlbaum.

MEEHL, P. E., & GOLDEN, R. (1982) Taxometric methods. In P. Kendall & J. Butcher (Eds.), *Handbook of research methods in clinical psychology*. New York: Wiley. Pp. 127-181.

MEEHL, P. E., & ROSEN, A. (1955) Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52, 194-216. [Reprinted in P. E. Meehl, *Psychodiagnosis: selected papers*. Minneapolis, MN: Univer. of Minnesota Press, 1973. Pp. 32-62.]

MEEHL, P. E., & YONCE, L. J. (1994) Taxometric Analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports*, 74, 1059-1274.

SNYDER, M., & GANGESTAD, S. (1986) On the nature of self-monitoring: matters of assessment, matters of validity. *Journal of Personality and Social Psychology*, 51, 125-137.

STATISTICAL SCIENCES, INC. (1993) *S-Plus for Windows (Version 3.1)*. Seattle, WA: Statistical Sciences, Inc.

STRUBE, M. J. (1989) Evidence for the *type* in Type A behavior: a taxometric analysis. *Journal of Personality and Social Psychology*, 56, 972-987.

TRULL, T. J., WIDIGER, T. A., & GUTHRIE, P. (1990) Categorical versus dimensional status of borderline personality disorder. *Journal of Abnormal Psychology*, 99, 40-48.

TYRKA, A. R., CANNON, T. D., HASLAM, N., MEDNICK, S. A., SCHULSINGER, F., SCHULSINGER, H., & PARNAS, J. (1995) The latent structure of schizotypy: premorbid indicators of a taxon of individuals at risk for schizophrenia-spectrum disorders *Journal of Abnormal Psychology*, 104, 173-183.

TUKEY, J. W. (1977) *Exploratory data analysis*. Reading, MA: Addison-Wesley.

WALLER, N. G., & MEEHL, P. E. (in preparation) *Multivariate taxometric procedures: distinguishing types from continua*. Newbury Park, CA: Sage.

WALLER, N. G., PUTNAM, F. W., & CARLSON, E. B. (in press) Types of dissociation and dissociative types: a taxometric analysis of dissociative experiences. *Psychological Methods*.

Accepted April 1, 1996.