

*Chapter 4*

---

Extension of the MAXCOV-  
HITMAX Taxometric  
Procedure to Situations of  
Sizable Nuisance Covariance

**Paul E. Meehl**  
University of Minnesota

THE PROBLEM OF  
MAXCOV-HITMAX IDEALIZATION

The MAXCOV-HITMAX taxometric procedure (Meehl, 1973; Meehl & Golden, 1982; Meehl & Yonce, 1996; various applications cited in Korfine & Lenzenweger, 1995; Lenzenweger & Korfine, 1992, 1995; Meehl, 1992, p. 135; Nicholson & Neufeld, 1994; Waller, Putnam, & Carlson, 1996; and nonpsychopathology applications by Gangestad & Snyder, 1985; Strube, 1989) utilizes the manifest statistics among three fallible quantitative indicators to (a) infer latent taxonicity, (b) estimate latent parameters (e.g., taxon base rate, specificity and sensitivity at a cut), and (c) make Bayes-Theorem classifications of individuals, with corresponding diagnostic confidence statements. The procedure relies on a theorem that, absent nuisance correlation within the latent classes, the covariance  $C_{yz}(x_i)$  of indicators  $y, z$  (*output pair*) computed on cases having a given score  $x_i$  on indicator  $x$  (*input indicator*) is a maximum when the  $x_i$ -interval is composed equally of taxon and complement (nontaxon) cases. This interval contains the  $x$ -value corresponding to the intersection of the two overlapping frequency functions where ordinates  $f_t(x)$  and  $f_c(x)$  are equal; thus, it is the  $x$ -cut that minimizes misclassifications (= hitmax cut  $x_i$  locating the *hitmax interval*  $\Delta x_i$ ), hence the acronym MAXCOV-HITMAX. That is, we locate the hitmax cut on  $x$  by studying the

behavior of the (yz)-covariance. While Monte Carlo runs and several real data studies show the procedure to be fairly robust under small or moderate departures from the strong auxiliary conjecture of zero nuisance covariance, this idealization is irksome and in some research situations cannot be well enough approximated to rely on robustness. The present chapter derives a more general algorithm free of this idealizing conjecture, called *generalized* MAXCOV (no longer HITMAX, because the covariance-maximizing interval located is not necessarily that containing the hitmax cut).

## GENERALIZED MAXCOV

The basic equation for deriving the MAXCOV algorithm (HITMAX or generalized) is a purely algebraic distribution-free decomposition identity, in fact a set-theoretical truism that holds even if the conjectured latent taxon has no real existence. (In that situation the search algorithm will reveal no clear maximum, and if a noise-determined pseudo-maximum—a chancy or artifactual “highest local jog” in the covariance graph—were foolishly treated seriously, the several consistency tests we rely on will be violated, protecting against erroneous taxonomic inference. For a general theoretical and methodological discussion of the taxonicity conjecture and its testing, see Meehl, 1992.) This basic equation is, ignoring sampling error,

$$C_{yz}(x_i) = pC_{yzt} + qC_{yzc} + p q (\bar{y}_t - \bar{y}_c) (\bar{z}_t - \bar{z}_c) \quad (1)$$

where

$C_{yz}(x_i)$  = manifest (yz)-covariance of cases having a score  $x_i$ ,  
hence, an empirical function of  $x$ ;

$C_{yzt}$  = (yz)-covariance within the taxon, a constant;

$C_{yzc}$  = (yz)-covariance within the complement, a constant;

$\bar{y}_t, \bar{y}_c, \bar{z}_t, \bar{z}_c$  = means of the taxon and complement classes on  
indicators  $y$  and  $z$ , respectively;

$p, q$  = proportions of the subset belonging to the taxon and  
complement classes, respectively.

In the original MAXCOV-HITMAX procedure on the idealizing conjecture of no nuisance covariance ( $C_{yzt} = C_{yzc} = 0$ ), the first two terms on the right vanish, so the left-hand manifest covariance equals the third term on the right. I call this the *validity mixture* term because its size depends on the two *crude validities* (mean taxon/complement differences, taken as constant absent sampling error or nontaxonic

moderator effects) and the taxonic mixture  $p:q$ . If either the  $y$  or  $z$  indicator had zero taxonic validity, or if there were no taxon/complement mix (i.e., if  $p = 0$  or  $q = 0$ ), this validity mixture term would also vanish in any such population or sample. Since the validity mixture term is maximized when  $p = q$ , plotting the graph of the observable  $C_{yz}(x_i)$  over  $x_i$ -intervals locates the hitmax interval by finding this graph's maximum. In that interval  $pq = 1/4$ , and knowing this allows us to solve for the mean difference product. Using that value we can write quadratics in  $p_i$  for all other intervals, solve for  $p_i$ , and proceed as explained in Meehl (1973) to estimate all of the latent values.

If sizable nuisance covariance exists, the first two terms on the right do not vanish and the hitmax interval cannot be located by finding the empirical maximum of  $C_{yz}(x)$  unless we know that the nuisance covariances are equal, whereby the sum  $p_i C_{yzt} + q_i C_{yzc}$  is invariant over  $x$ -intervals. Even then, however, we cannot solve for the validity product because we cannot parse the right side into its nuisance covariance component and its validity mixture component. We can, however, proceed as follows: Choosing a cut at sufficiently high values of  $x$ , we will have passed all or nearly all of the complement cases, so that cases lying above that cut will be nearly "pure taxon." Hence, in this region the validity mixture term vanishes, as does the complement nuisance component, and only the component  $p_i C_{yzt}$  (where  $p_i \simeq 1$ ) remains, giving us an estimate of the taxon nuisance covariance. Similarly, we estimate the complement nuisance covariance from cases lying in the extreme low region of the  $x$ -distribution. Returning to Equation 1, designating the validity product as an unknown parameter  $\theta$

$$\theta = (\bar{y}_t - \bar{y}_c) (\bar{z}_t - \bar{z}_c), \text{ a constant} = K. \tag{2}$$

Simplifying subscripts as  $C_{yzt} = C_t$ ,  $C_{yzc} = C_c$

and expanding Equation 1 in terms of  $p_i (= 1 - q_i)$ , we have (dropping the subscript  $i$ ),

$$C_{yz}(x) = p(C_t - C_c) + (p - p^2)\theta + C_c \tag{3}$$

differentiating with respect to  $x$  and setting = 0 for a maximum,

$$\frac{dC(x)}{dx} = \frac{dp}{dx}(C_t - C_c) + \frac{dp}{dx}\theta - 2p\frac{dp}{dx}\theta = 0 \quad [\text{equation corrected by author}] \tag{4}$$

dividing by  $\frac{dp}{dx}$  [ $\neq 0$  in region of interest]

$$(C_t - C_c) + \theta - 2p\theta = 0 \quad \text{at max} \tag{5}$$

so

$$p = \frac{\theta + (C_t - C_c)}{2\theta} \quad \text{at max.} \tag{6}$$

At this maximum,  $C_{yz}(x)$  has an observed numerical value,  $C_{\max}$ . So at that value,

$$p(C_t - C_c) + p\theta - p^2\theta = C_{\max} - C_c. \quad (7)$$

Plugging Equation 6 into Equation 7, we obtain

$$\left( \frac{\theta + (C_t - C_c)}{2\theta} (C_t - C_c) \right) + \left( \frac{\theta + (C_t - C_c)}{2\theta} \theta \right) - \left( \left[ \frac{\theta + (C_t - C_c)}{2\theta} \right]^2 \theta \right) + (C_c - C_{\max}) = 0, \quad (8)$$

which, with some straightforward tedious algebra, yields

$$\theta^2 + (2C_t + 2C_c - 4C_{\max})\theta + (C_t - C_c)^2 = 0 \quad (9)$$

a quadratic in  $\theta$ . Its roots are

$$\theta = (2C_{\max} - C_t - C_c) \pm 2[(C_t - C_{\max})(C_c - C_{\max})]^{1/2}. \quad (10)$$

This has the form

$$(a + b) \pm 2(ab)^{1/2} \quad (11)$$

since

$$a + b + 2(ab)^{1/2} = (a^{1/2} + b^{1/2})^2 \quad (12)$$

and

$$a + b - 2(ab)^{1/2} = (a^{1/2} - b^{1/2})^2 \quad (13)$$

the roots (Equation 11) are

$$\theta_1 = [(C_{\max} - C_t)^{1/2} + (C_{\max} - C_c)^{1/2}]^2 \quad (14)$$

$$\theta_2 = [(C_{\max} - C_t)^{1/2} - (C_{\max} - C_c)^{1/2}]^2 \quad (15)$$

Root  $\theta_1$  is selected by the physical situation that for the special case of zero nuisance covariance  $C_t = C_c = 0$  it yields the correct value  $\hat{K} = 4C_{\max}$  in the hitmax interval, whereas  $\theta_2$  gives an impermissible  $\hat{K} = 0$ .

Having found  $\theta$ , we proceed as in MAXCOV-HITMAX, using Equation 3 in each  $x$ -interval to get the interval's taxon rate  $p_i$ , the generalized quadratic algorithm for the taxon-proportion in an interval  $x_i$  being

$$p(x_i) = \frac{(K + C_t - C_c) \pm \sqrt{(K + C_t - C_c)^2 - 4K(C(x_i) - C_c)}}{2K}, \quad (16)$$

then  $N_i p_i = N_{ti}$  the interval's taxon frequency, then  $\sum N_{ti} = N_t$ , and, finally, base rate  $P = N_t/N$ . The latent frequencies having been computed for each  $x$ -interval, we can obtain directly latent means, standard deviations, skewness, and kurtosis if desired. For any triad  $(x, y, z)$  of indicator variables, three MAXCOV procedures exist (using either  $x, y$ , or  $z$  as input indicator) and the three inferred latent distributions are thus obtained, as in MAXCOV-HITMAX.

ROBUSTNESS OF  
THE ORIGINAL PROCEDURE

It is illuminating to ask why the original procedure is fairly robust under departures from the idealization of zero nuisance covariances within the categories (cf., even in my first technical report [Meehl, 1965, pp. 50-54], a reassuring numerical example for  $r_t = .40$ ,  $r_c = .20$ ; Meehl & Golden [1982, Table 5,2]; Meehl & Yonce [1996]). The General Covariance Mixture Formula for the observed ( $yz$ )-covariance in an  $x$ -interval includes the nuisance covariances

$$C_{yz}(x_i) = p_i C_t + q_i C_c + p_i q_i K$$

where  $K$  (treated as constant over intervals) is the product of the  $y$  and  $z$  separations,  $K = (\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c)$ , and  $p_i$ ,  $q_i$  are the taxon and complement proportions, respectively. Suppose the two nuisance covariances were equal,  $C_t = C_c = C$ , an auxiliary conjecture unlikely to be literally true in psychopathology but often an adequate approximation. Then the first two terms sum to  $C(p_i + q_i) = C$ , the neglected component being constant as we move through the intervals. The (erroneous, idealized) equation

$$C_{yz}(x_i) = K p_i q_i \tag{17}$$

employed instead of the correct relation

$$C_{yz}(x_i) = C + K p_i q_i \tag{18}$$

still locates the hitmax cut correctly by maximizing the variable term. However, solving for  $K$  via the hitmax interval relation  $p_i = q_i = 1/2$

$$C_{yz}(x_h) = 1/4 K \tag{19}$$

$$K = 4C_{yz}(x_h) \tag{20}$$

yields an inflated estimate  $\hat{K} > K$  [= true value of  $sep_y \cdot sep_z$ ]. Relying on this erroneous  $\hat{K}$  when we solve for the  $p_i$ s in the other  $x$ -intervals, the quadratic algorithm (treating  $C = 0$  again) is

$$p(x_i) = \frac{1}{2} \pm \left[ \frac{1}{4} - \frac{C_{yz}}{\hat{K}} \right]^{1/2} \tag{21}$$

where the observed interval covariance is, in *latent* terms,

$$C_{yz}(x_i) = C + p_i q_i K \quad (\text{true } K). \tag{22}$$

Then our approximation for each  $p_i$  is

$$p_i = \frac{1}{2} \pm \left[ \frac{1}{4} - \frac{C + p_i q_i \hat{K}}{C + \hat{K}} \right]^{1/2} \tag{23}$$

Re-writing the erroneous values in terms of two error-multipliers  $m(x_i)$  and  $M$  on  $K$ ,

$$p_i = \frac{1}{2} \pm \left[ \frac{1}{4} - \frac{p_i q_i m(x_i) K}{MK} \right]^{1/2} \quad (24)$$

where  $M$  is constant over  $x$ -intervals but  $m(x_i)$  varies (since the fixed  $C$  added to the variable term  $Kp_i q_i$  results in a varying *proportional* error), we have then for our taxon proportion estimate in each interval  $x_i$

$$p(x_i) = \frac{1}{2} \pm \left[ \frac{1}{4} - \frac{p_i q_i m(x_i)}{M} \right]^{1/2}. \quad (25)$$

The variable term is correct only at  $x_h$ , since in all other intervals  $m(x_i) > M$ . We benefit from countervailing inflations in numerator and denominator, which partially explains the robustness but  $m(x_i)$  changes over intervals. In intervals other than the hitmax, the variable subtracted term is inflated; hence, the radicand is deflated. To the left of  $x_h$ , where one *adds* the radical to  $1/2$ , one overestimates  $p_i$ . To the right of  $x_h$ , where the radical is *subtracted* from  $1/2$ ,  $p_i$  is underestimated. Hence the values of  $\hat{N}_{ui} = n_i p(x_i)$  are inflated in the region  $x_i < x_h$  and deflated in the region  $x_i > x_h$ . For situations not too asymmetrical about hitmax (as when  $P \simeq .50$ ), the sums of these opposite errors tend to cancel out in estimating  $\hat{N}_t = \Sigma \hat{N}_{ui}$  from sums above and below  $x_h$ , and, hence,  $\hat{P} = \hat{N}_t / N$  is not badly estimated. We note that when  $P < .50$  (as usual in psychopathology research), the asymmetrical countervailing errors lead to overestimation of the base rate  $P$ .

The somewhat surprising robustness of the original procedure under departure from the zero nuisance covariance idealization is illustrated in Table 1, based on error-free data (Gaussian table) for  $\sigma_t = \sigma_c = 1$ , separations  $(\bar{y}_t - \bar{y}_c) = (\bar{z}_t - \bar{z}_c) = 2$ , and five nuisance covariance configurations. As the preceding equations and text require, the peak covariance shifts toward the right whenever nuisance covariance is greater in the taxon than in the complement group, and more so for smaller base rates. That  $C(yz)$  MAXCOV interval is, of course, not the hitmax interval unless the nuisance covariances are equal in taxon and complement. Even when the base rate is quite large ( $= .50$ ), estimated  $\hat{P}$  is unbiased when  $C_t = C_c$ . Large nuisance covariances and low base rate generates an upward bias in  $\hat{P}$  that is unacceptably large ( $\Delta \hat{P} = .07$ ), suggesting a need for the generalized procedure derived herein. The unavoidable tradeoff between bias (using the idealized procedure) and random sampling error (using the generalized procedure but often relying on somewhat unstable estimates of  $C_t$  and  $C_c$ ) remains to be examined over the parametric configuration space. It is conceivable that a simple standard correction (downward) for situations where the MAXCOV graph strongly suggests a low  $P$  would do as well as the generalized, more elegant approach developed above.

Table 1 Error-Free MAXCOV Values With Various Combinations of Nuisance Covariance

Configuration		Max(cov)			Maximum Interval			Estimated <i>P</i>		
Cov <sub>c</sub>	Cov <sub>t</sub>	<i>P</i> = .50	.25	.10	<i>P</i> = .50	.25	.10	<i>P</i> = .50	.25	.10
.00	.00	1.00	1.00	1.00	1.0	1.5	2.1	.49	.25	.10
.00	.25	1.13	1.13	1.13	1.1	1.6	2.2	.46	.23	.09
.25	.25	1.25	1.25	1.25	1.0	1.5	2.1	.49	.27	.14
.25	.50	1.38	1.38	1.38	1.1	1.6	2.2	.46	.26	.13
.50	.50	1.50	1.50	1.50	1.0	1.5	2.1	.49	.29	.17

Unfortunately, that question involves unsettled epistemological—not purely mathematical—issues in statistical inference theory (e.g., Bayesian rectangularity of priors).

#### ESTIMATING THE NUISANCE COVARIANCES

The use of Equation 3 as described above requires that we have trustworthy estimates of the two nuisance covariances, as obtained from cases in the extreme (high and low) *x*-regions. Demarcating the “safe” *x*-regions is problematic, since we do not know how high a cut point must be to guarantee a negligible contamination of the taxon covariance estimate by the complement, and the same is true in the extreme low *x*-region. If we go out so far as to be very safe from taxon/complement mixture, the number of available cases in the demarcated region is dangerously small, except with very large initial samples. We have a tradeoff between random sampling error (demarcating a very extreme tail, and, hence, small *N*) and bias (systematic error due to appreciable complement class contribution to the manifest covariance). A rigorous analytic optimizing is presumably impossible absent knowledge of the taxonic separation, and I have not tried to derive one since latent Gaussian distributions cannot be safely assumed in psychopathology and, in general, will be false.

With large samples, the obvious demarcation criterion would be the high (and low) *x*-regions where the observed (*yz*)-covariance has clearly become flat with changes in *x*. Because sampling error becomes important with a smaller number of cases, the covariances within successive *x*-intervals begin to fluctuate (apart from the systematic effect of decreasing taxon mixture) as we move out into relatively “pure” extreme regions far from the *x*-mean. Our problem is to decide when, in aiming to reduce random error by raising the *N*, we have located the upper cut so far down that undesired contribution by the complement class has become appreciable. I cannot currently offer

a satisfactory answer to this. A crude nonparametric criterion would be the successive slopes of the interval covariance graph as we move downward. If the complement contribution is negligible in an  $x$ -region, increments and decrements of the  $(yz)$ -covariance will be random; hence, the slope *signs* (+) or (-) from interval to adjacent interval will be random. The probability of, say, 3 successive slopes being (- - -) is therefore  $p = .125$ , and, given a high prior that complement cases are being overtaken as we move down, this significance level might be a reasonable one at which to stop moving. Surely we should stop at (- - - -), where  $p < .07$ . Or, looking at it the other way around, attempting a “positive” argument for minimal contamination (via a “positive” argument for flatness), one might set up a crude criterion of “near-zero” slope of adjacent line segments, say  $\Delta C_x = C(x_i) - C(x_{i+1}) \leq k$ , then applying a sign test of “near-flatness” to the terminal segments. Another approach would be to define the  $x$ -intervals by blocks of equal frequency (deciles, or if  $N$  permits, vigintiles) and examine the Pearson  $r$ s in sets of adjacent blocks for homogeneity, testing their  $z_r$ -variance against the theoretical variance  $\sigma^2 = (N - 3)^{-1}$ . A third possibility asks, employing a suitable extrapolation algorithm, to what asymptote are the successive interval covariances converging as we move out? Perhaps a combination of procedures would be safest. Given thin data when far out, one must also hold down the effect of outliers (see, e.g., Cleveland, 1979, for an approach to this problem). It is reassuring that for sample size  $N = 300$ ,  $P = Q = \frac{1}{2}$ , and a 2-*SD* taxon separation, the two top deciles will be about 98% uncontaminated by the complement class, providing 60 cases for estimating the nuisance covariance.

When  $P \ll \frac{1}{2}$  (as in most psychopathology research) it is possible to demarcate a fairly large lower region of the  $x$ -distribution that will consist of nontaxon cases, with a subsample size permitting a trustworthy estimate of  $C_c$ . However, when the output covariance maximum (the taxonic “hump”) is markedly displaced from center (due to small  $P$ ), there may not be any flat  $x$ -region at the high side, and the graph may continue to fall or rise as we move farther out on  $x$ . In that situation it may be appropriate to employ an extrapolation procedure, inferring the asymptote of  $C_{yz}(x_i) = (p_i C_t + q_i C_c + p_i q_i K_{yz})$  as  $p_i \rightarrow 1$ ,  $q_i \rightarrow 0$ , and  $p_i q_i \rightarrow 0$ . This asymptote estimates the desired nuisance covariance  $C_t$  for the taxon.

When the base rate is quite small (e.g.,  $P \leq .10$ ) the MAXCOV graph may not have a local maximum, since, unless the sample size is *very* great, there is not sufficient room far enough out to get past the hitmax interval (where  $p_i \simeq q_i$ ) so that the validity mixture term  $K p_i q_i$  can begin to decline again. Error-free MAXCOV curves are shown in Figure 1. Nuisance covariance in the complement group is

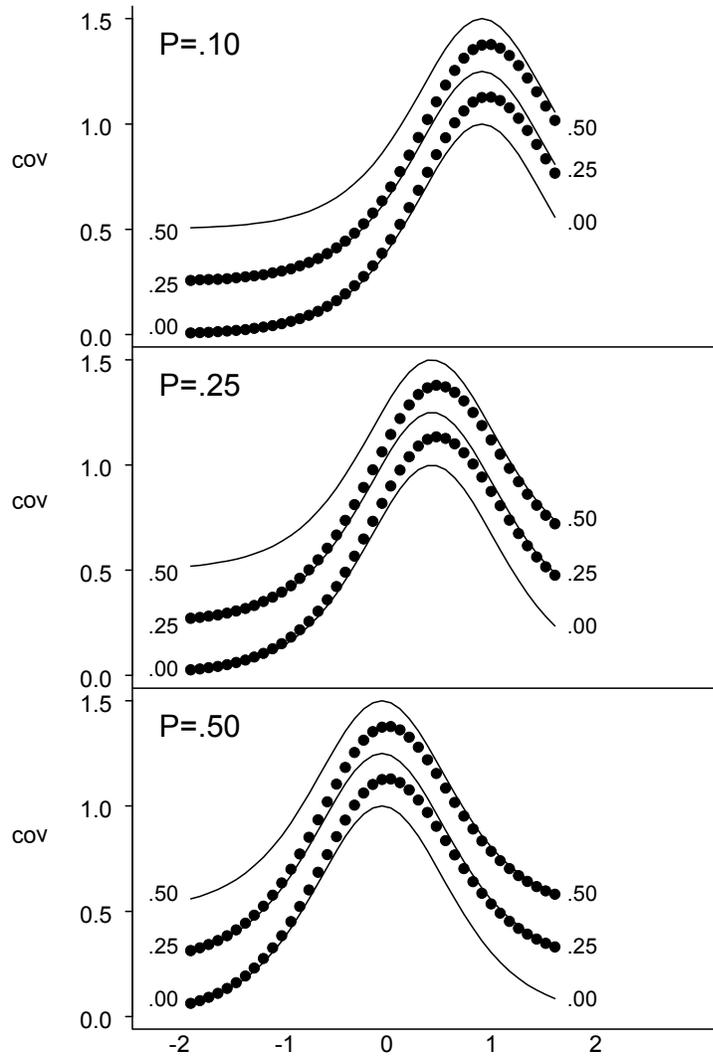


Figure 1 Error-free MAXCOV curves with different amounts of nuisance covariance

indicated at the left of the curves; taxon nuisance covariance is on the right. The MAXCOV peak shifts to the right with decreasing base rate (panels for  $P = .50, .25, .10$ ) and also whenever there is greater nuisance covariance in the taxon group than in the complement (the latter situation is indicated by the dotted curves in each panel). Monte Carlo runs on artificial data (incorporating random error) for  $P = .10$  and a 2-*SD* separation may display either a (Tukey-smoothed) hump or a cusp (Meehl & Yonce, 1996). One cannot even be sure that  $p_i \simeq 1/2$  in the “top” interval, since  $p_i$  may still be rising in the region; so  $p_i q_i < 1/4$ , rather than  $= 1/4$  (as it is in a true hitmax interval where the taxon/complement mix is even).

My suggestion for this unfavorable case is an iterative bootstraps procedure, but I lack empirical evidence that it will work satisfactorily. The iterative sequence proceeds as follows. Writing the general covariance mixture equation for the observed  $x_i$ -interval covariance  $C_{yz}(x_i)$  in latent terms,

$$C_{yz}(x_i) = p_i C_t + q_i C_c + p_i q_i K_{yz}, \quad (26)$$

we take as first approximation that the nuisance covariances  $C_t, C_c$  are equal [=  $C_{yz}$ ] and that the taxon proportion  $p_i$  in the  $C_{\max}$  interval, whether a hump or a cusp, is  $p_i = 1/2$ . Then

$$C_{yz(\max)} = C_{yz} + 1/4 K. \quad (27)$$

To use this relation we need to estimate the common covariance  $C_{yz}$ , which we get by calculating directly the observed covariance of the bottom half or third of the  $x$ -distribution, safely assumed almost wholly uncontaminated by taxon cases when the base rate is small and the maximum covariance is a hump or cusp located far to the right. Thus, if  $N = 300$ , there are 150 cases below the  $x$ -median, yielding a trustworthy “direct” estimate of  $C_c (= C_{yz})$ .

Putting this  $\hat{C}$  in Equation 27 we solve for  $K$ ,

$$\hat{K}_{yz} = 4(C_{\max} - \hat{C}_{yz}). \quad (28)$$

Given these estimates  $\hat{C}_{yz}$  and  $\hat{K}$ , writing the equation for each  $x$ -interval in the unknown  $p_i$ ,

$$C_{yz}(x_i) = \hat{C}_{yz} + p_i q_i \hat{K}, \quad (29)$$

we solve for  $p_i, q_i$  per interval. In each interval, we compute the taxon *frequency*  $N_{ti}$ , given the interval’s total frequency  $N$  (observed),

$$\hat{N}_{ti} = \hat{p}_i N_i \quad (30)$$

and  $\hat{N}_{ci} = N_i - \hat{N}_{ti} \quad (31)$

From these latent interval frequencies we compute estimates of the latent means  $\bar{x}_t$  and  $\bar{x}_c$  and of the base rate  $\hat{P}_x = \frac{1}{N} \sum \hat{N}_{ti}$  over all  $n$ -intervals.

Conducting this sequence with  $y$  as input and  $C_{xz}(y_i)$  as output, we get another estimate  $\hat{P}_y$ , and of latent  $y$ -means  $\bar{y}_t, \bar{y}_c$ ; similarly, we use  $z$  as input to get another estimate  $\hat{P}_z$  and of the latent  $z$ -means  $\bar{z}_t, \bar{z}_c$ .

From the latent  $y$ - and  $z$ -means we now *reconstitute* the  $\hat{K}_{yz}$  (instead of inferring it from the observed  $(yz)$ -covariance with  $x$  as input) and, taking as a revised base rate estimate the average  $\hat{P}_{yz} = \frac{1}{2} (\hat{P}_y + \hat{P}_z)$ , write a new grand covariance mixture equation for the *whole group* of cases,

$$C_{yz}(x) = \hat{P}_{yz} C_t + \hat{Q}_{yz} \hat{C}_c + \hat{P}_{yz} \hat{Q}_{yz} \hat{K}_{yz} \quad (32)$$

now treating the taxon covariance as an unknown (hence no carat on it in Equation 32) and solving for it (instead of assuming it =  $C_c$ ).

Setting up equations analogous to Equation 32 with  $y$  and  $z$  as inputs, we obtain new approximations for  $C_{txz}$  and  $C_{txy}$  in the same manner. These values can then be employed in the generalized procedure described at the beginning of this chapter. If there is a MAXCOV “hump,” that is clearly appropriate. If the graph is a cusp, we do not have an *observed* mathematical maximum, but relying on the error-free results, it may be safe to proceed as if the top interval is “like a maximum” for use of the generalized MAXCOV equation.

There is no vicious “circularity” here, because we obtain the  $\bar{y}_t, \bar{y}_c, \bar{z}_t, \bar{z}_c$  components for reconstituting  $K_{yz}$  when we employ  $y$  and  $x$  as *input* indicators, but the  $\hat{P}$  and  $\hat{C}_t(yz)$  are inferred, employing the general equation with  $x$  as input variable. Since the formalism does not force convergence of these different epistemic paths to latent parameters, iterative agreement indicates the final estimates to be accurate and further corroborates the structural conjecture upon which the equations are predicated.

Estimating the two nuisance covariances is obviously a complicated, difficult matter which colleagues and I are currently exploring analytically and with Monte Carlo runs on artificial data. In this chapter I have only sketched out what appear to be the main options. The availability and appropriateness of each undoubtedly depends on the parametric configuration, especially sample size, base rate, and taxonic separation—only the first being accurately known to the investigator. Assuming the threshold question of taxonicity has been answered, one asks whether the MAXCOV graphs (three or more in number, e.g., 12 graphs given four indicators) display a flat region at the high end of the input variable. This inspectional impression must be corroborated by appropriate statistical tests (e.g., sign test on slopes of line segments,

variance test on  $z_r$ -transformed ( $yz$ ) correlations in the flat-looking  $x$ -intervals). If these tests indicate no change in the region, one concludes that the cases lying therein are almost “pure taxon,” largely uncontaminated by complement class cases, so pooling these intervals one can compute  $C_t$  directly from the observations. In cases of low base rate, the iterative procedure may be an alternative. Monte Carlo investigation of estimating nuisance covariance with these procedures should provide more helpful guidelines for researchers.

*I am grateful to Niels G. Waller and Leslie J. Yonce for helpful comments on this chapter, and to the latter for work on the chapter table and figure.*

#### REFERENCES

- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*, 829–836.
- Gangestad, S., & Snyder, M. (1985). “To carve nature at its joints”: On the existence of discrete classes in personality. *Psychological Review*, *92*, 317–349.
- Korfine, L., & Lenzenweger, M. F. (1995). The taxonicity of schizotypy: A replication. *Journal of Abnormal Psychology*, *104*, 26–31.
- Lenzenweger, M. F., & Korfine, L. (1995). Tracking the taxon: On the latent structure and base rate of schizotypy. In A. Raine, T. Lencz, & S. A. Mednick (Eds.), *Schizotypal personality* (pp. 135–167). New York: Cambridge University Press.
- 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.
- Meehl, P. E. (1965). *Detecting latent clinical taxa by fallible quantitative indicators lacking an accepted criterion* (Rep. No. PR–65–2). Minneapolis: University of Minnesota, Research Laboratories, Department of Psychiatry.
- Meehl, P. E. (1973). MAXCOV-HITMAX: A taxonomic search procedure for loose genetic syndromes. In P. E. Meehl, *Psychodiagnosis: Selected papers* (pp. 200–224). Minneapolis: University of Minnesota Press.
- 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., & Golden, R. (1982). Taxometric methods. In P. Kendall & J. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 127–181). New York: Wiley.
- 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.
- Nicholson, I. R., & Neufeld, R. W. J. (1994). The problem of dissecting schizophrenia: Evidence for a dimension of disorder. Manuscript submitted for publication.
- Strube, M. J. (1989). Evidence for the Type in Type A behavior: A taxometric analysis. *Journal of Personality and Social Psychology*, *56*, 972–987.
- 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.