Some Precautions in Using Canonical Analysis

Author(s): Zarrel V. Lambert and Richard M. Durand

Source: *Journal of Marketing Research*, Nov., 1975, Vol. 12, No. 4 (Nov., 1975), pp. 468-475

Published by: Sage Publications, Inc. on behalf of American Marketing Association

Stable URL: https://www.jstor.org/stable/3151100

Some Precautions in Using Canonical Analysis

ZARREL V. LAMBERT and RICHARD M. DURAND*

INTRODUCTION

The use of canonical correlation analysis in marketing research has been expanding substantially in recent years and for good reason (applications include [1, 2, 4, 6, 7, 8, 10, 11, 13, 14, 16, 17, 18, 20, 27]). However, canonical correlation like other analytical methods is not without certain limitations. It is important for these limitations to be kept in mind when choosing a technique for data analysis and when interpreting canonical results. Otherwise opportunities to distill additional important information from the data may be sacrificed or even worse faulty interpretations may occur. The purpose of this article is to describe and illustrate some potential shortcomings of canonical correlation analysis when it is used in marketing research.

One reason applications of canonical analysis have grown is because the technique provides for multivariate analysis of whole batteries or sets of variables as they relate to each other. Following convention these sets are often designated as criterion and predictor variables. More traditional methods such as bivariate and multiple correlation restrict analysis to only one criterion variable at a time. Consequently, when these methods are utilized the criterion side of the relationship must be analyzed in a univariate fashion.

Univariate analysis of criterion phenomena leaves something to be desired when the phenomena cannot be adequately expressed or measured by a single variable, which is often the case in marketing research. Frequently there is a potent conceptual basis for expecting both the criterion and predictor variables to be gestalt-like composite sets with relationships existing between the sets. In which case, any single criterion variable taken in isolation is at best indicative of only a part of the overall relationship. Therefore, an analytical procedure that examines only one criterion variable at a time in an integral set may fail to detect fully and describe accurately prevailing relationships.

As an example, it has been suggested that failure in the past to uncover strong relationships linking consumer behavior to personality may be partially attributable to the use of bivariate and multiple correlation methods [18]. It is said that these methods are inadequate because they do not analyze structures of interacting variables in their entirety.

While two or more sets of variables can be examined in their entirety by canonical correlation analysis, the statistical results are subject to certain vagaries and, depending on the research purpose, may not be amenable to straightforward interpretation. Generally speaking, the limitations of canonical analysis should be given careful attention whenever (1) interpretational validity depends on the amount of covariation between predictor and criterion variables, and (2) interpretations relate variation in specific criterion variables to variation in predictor variables. A more detailed examination of the limitations of canonical analysis follows after a brief overview of the technique.

TECHNIQUE SYNOPSIS

In canonical analysis, separate vectors of weights, say a_j and b_j, are derived which yield linear compounds of the predictor and criterion variables. These linear compounds are maximally correlated with each other and uncorrelated with other linear compounds which have maximum correlation. More than a single pair of a_j and b_j weight vectors can be derived from the observed data. The maximum number that can be extracted equals the number of variables in the smallest data set, predictor or criterion. Each pair, called a linear combination, consists of unique a_j and b_j weights. These weights when multiplied by the respective predictor and criterion variables produce maximally correlated linear compounds of these variables.

The linear compounds are termed canonical scores, p_i and c_i, in this article in an effort to improve clarity;

*Zarrel V. Lambert is Associate Professor of Marketing, University of Florida; Richard M. Durand is Assistant Professor of Marketing, Southern Illinois University at Carbondale. The authors are grateful to Professor Albert R. Wildt for his cogent comments and suggestions on an earlier draft.

Journal of Marketing Research
Vol. XII (November 1975), 468-75

a variety of other names including canonical variates, canonical variables, linear composite scores, and canonical components have been used elsewhere. The symbols p_i and c_i are used here to designate scores for the predictor and criterion variables respectively.

The canonical scores for a given linear combination, called a canonical root by some, can be expressed in the following way:

$$\begin{array}{ll} p_1 = a_1 x_{11} + \ldots + a_m x_{1m} & c_1 = b_1 y_{11} + \ldots + b_q y_{1q} \\ \vdots & \vdots \\ p_i = a_1 x_{i1} + \ldots + a_m x_{im} & c_i = b_1 y_{i1} + \ldots + b_q y_{iq} \end{array}$$

where:

x_m = the predictor variables;
y_q = the criterion variables;
i = the number of cases or observations in the sample;
m = the number of variables in the predictor set; and
q = the number of criterion variables.

As these equations show, each pair of a_j and b_j weight vectors produces a different set of p_i and c_i scores. The p_i scores are uncorrelated pairwise among linear combinations as are the c_i scores.

In the case of corresponding vectors, the canonical R_c value indicates the correlation between the p_i and c_i scores for that specific linear combination, i.e., pair of weights. Hence, the number of canonical correlations for a particular data sample is the same as the number of linear combinations extracted. Each of the R_c values can be tested for statistical significance.

Large numbers of predictor and criterion variables can be analyzed jointly resulting in a series of canonical R_c values. These R_c values differ in some important ways with familiar multiple correlation coefficients, R's, and should not be viewed in exactly the same manner, although conventional multiple correlation can be regarded as a special case of canonical analysis. These differences tend to make interpretation more difficult, and in that sense they constitute some of the limitations of canonical analysis.

LIMITATIONS

The subsequent sections focus on (1) the deficiency of the canonical R_c statistic as an indicator of the variance shared by the sets, (2) weight instability and correlation maximization, and (3) problems associated with attempts to partition the sets into correlated constructs. Illustrations of these limitations are drawn from a replication of an earlier study.

Shared Variance

The variance shared by two variables or sets is the amount of variance in one which is redundant with the other. It can be described as the intersection of the sets. In regression parlance, it is the proportion of variance in one which can be predicted by or attributed to variation in the other.

Unlike multiple correlation coefficients, canonical R_c values do not in themselves reveal the amount of criterion variance that is shared with the predictor variables. This characteristic sets the stage for potential misinterpretations. A danger lies in being accustomed to working with multiple correlation coefficients and being conditioned to viewing large coefficients as indicating that a substantial proportion of the criterion variance is accounted for or shared. In canonical analysis, very little of the criterion variance may be shared with the predictor variables although the canonical R_c values are high. This can occur because the canonical R_c's do not measure directly the correlation between the sets of variables [19, 25].

The R_c values measure the correlation between the p_i and c_i canonical scores. Thus, a canonical R_c for a particular linear combination reflects only the amount of variance shared by these canonical scores, not the variance shared by the original predictor and criterion variables. Consequently, the canonical correlations may be large in numerical terms and statistically significant although the original data sets have minimal shared variance.

Since canonical R_c values may be observed which are numerically much larger than previously reported bivariate and multiple correlation coefficients, there may be a temptation at times to presume that canonical analysis has uncovered substantial relationships of conceptual and practical significance. Before such conclusions are warranted, further analysis involving measures other than R_c's must be undertaken to determine the amount of redundancy between the observed variables. Otherwise, mistakes may result from interpreting what may be nothing more than the variance shared by the canonical scores.

For instance, a cursory glance at canonical results like those reported in the following statement, which exemplifies some found in the literature, may leave an impression which exaggerates the covariation in the observed variables.

> The analysis yielded two significant canonical correlation coefficients of .68 and .59 which accounted for 46 and 35% of the variance in the canonical variates.

The canonical results above pertain only to the covariation of canonical scores, not the observed data. It is conceivable in this instance that little of the variance in the criterion variables, say less than 10%, is shared with the predictors.

In published marketing studies employing canonical analysis, very few investigators have made a reported attempt to assess the amount of variance shared by the predictor and criterion variables. Typically, when statistically significant canonical R_c values have been observed, investigators have proceeded as if the sets of variables were themselves substantially correlated. Frequently their next step has been to search the variables within each set in an effort to say that particular criterion variables were related to certain variables in the predictor set. At times, subsets of variables have been grouped into so-called constructs like factors and then matched or linked across the predictor and criterion sets. Such interpretations may have little validity or explanatory value if, in actuality, the predictor variables account for little of the criterion variance.

The amount of criterion variance shared with the predictors can be measured by the redundancy index, $\overline{R_{c/p}{}^2}$ shown below [3, 19]:

$$\overline{R_{c/p}{}^2} = \sum_{i=1}^{M_c} \lambda_i \left[\sum_{j=i}^{M_c} \frac{L_{ji}{}^2}{M_c} \right]$$

where:

λ_i = the squared canonical R_c for the ith linear combination,

L_{ji} = the loading between the jth variable and the ith linear combination;

M_c = the number of variables in the criterion set.

The index is summed over all linear combinations to determine the total proportion of variance that is shared. The disparity that can occur between the canonical R_c numerical values and the index of shared variance is illustrated by comparing these statistics in a study conducted by the authors. The study replicates a previous one on product usage and personality.

The earlier research, reported by Sparks and Tucker, analyzed usage of 17 products in relation to personality traits as measured by the Gordon Personal Profile and Inventory [18]. Their results included canonical R_c's of .606, .548, and .413 for the first three linear combinations. Upon replicating the study (N = 183), the authors observed canonical correlations of .491, .444, and .405 which, although somewhat lower, are not greatly dissimilar from the values reported by Sparks and Tucker. Based on the same test as used in the initial study, the first three replicated canonical R_c values are statistically significant at the .0047, .0252, and .0693 levels respectively. These resemble the significance levels in the first study. Stated in terms of covariance, the findings suggested rejection of the null hypothesis that the combined predictor and criterion covariance matrix equals zero.

The canonical R_c values viewed without additional analysis appear impressive in comparison to the bivariate and multiple correlation coefficients which have been reported by numerous other investigators of personality and consumer behavior. Interpreting these numerically large R_c's as an indication that canonical analysis has revealed strong ties between personality and consumer behavior, which heretofore have gone undetected, may be much in error. As described earlier, the large R_c values may be an artifact of the analytical procedure and not an indication of a strong association between product usage and personality. In fact, it is precisely this situation that emerges when the redundancy index is computed.

According to the index, only 6.2% of total variance in usage rates is shared with variance in personality traits. Thus, the amount of shared variance revealed by canonical analysis is equivalent to finding a bivariate or multiple correlation coefficient of about .25, had these correlation procedures been utilized. Correlation coefficients not a great deal lower than this have been found in the past.

As this example demonstrates, the explanatory capability of canonical analysis in accounting for criterion variance would have been grossly exaggerated in this instance if the canonical R_c's had been taken at face value and viewed like multiple correlation coefficients. The redundancy would have been overrated by approximately 400%!

Weight Instability

The values of the weights a_j and b_j which are used in computing the canonical scores are subject to considerable instability, or what might be termed variability, from sample to sample. Instability can occur because the computational procedure yields weights that maximize canonical score correlations for that particular sample of observed predictor and criterion values. Similarities exist with what has been called the "bouncing beta" problem in regression analysis. If the criterion and predictor observations vary from sample to sample, the weights will differ so as to produce the highest possible canonical correlations for each sample. Multicollinearity, correlation among variables within a set, also may contribute to weight instability.

Unstable weights suggest caution in interpreting the results of canonical analysis. Instability raises a serious question about the real meaning of the canonical correlations because the weights reflect each variable's contribution to the correlated elements, the canonical scores. If these contributions differ widely from sample to sample, one might ask what can be stated about the nature of overall relationships between predictor and criterion variables?

It should be noted parenthetically that some authors refer to the weights as canonical coefficients (see [9] for example). Some widely used computer programs also label weights as canonical coefficients. To avoid confusion, the term was not used in this article because other authors apply it to a different canonical statistic which will be discussed in the next section.

Since the weights are effected by sample estimates for the predictor and criterion variables, the a_j and b_j's themselves may be thought of as sample estimates.

This brings to mind the procedure that is available in regression analysis for establishing confidence intervals for estimated regression coefficients, also called weights. To the authors' knowledge, comparable procedures for canonical weights are not readily accessible. One major difference is that regression weights are estimated in relation to the observed criterion variable, Y. In canonical analysis, the weights are estimated in relation to derived values, not the actual observed criterion variables.

How then can the researcher measure the amount of instability that exists in the weights? The canonical R_c's are insufficient by themselves because similar R_c values conceivably can result from markedly different weights. Two analytical procedures may be employed for testing weight stability. Both involve either randomly dividing the overall sample of predictor and criterion observations into two subsamples or drawing separate samples.

To illustrate these two procedures which are described below, a split sample was utilized in the replicated study of product usage and personality. The full sample was randomly divided into analysis and validation subsamples having an N of 123 and 60 respectively.

One procedure for evaluating weight instability is to apply canonical analysis to each sample or subsample separately, thereby deriving separate weights for each. The weights for the two samples, then, are compared by simple correlation analysis to reveal the extent of agreement. Weights for the first linear combination in sample 1 are correlated with those for the first linear combination in sample 2. Then weights on the second linear combination are analyzed, and so on until all significant linear combinations are examined.

Application of this procedure in the replicated study resulted in product moment correlations of $-.053$, .198, and $-.307$ for the 17 criterion weights on the first three linear combinations. Corresponding correlations of .309, .483, and $-.579$ were found for the 8 predictor weights. These r coefficients revealed a large amount of weight instability. None of the coefficients were statistically significant at the .05 level. While large significant coefficients would have suggested the weights were relatively stable, they would not have precluded possible interpretational problems due to multicollinearity. More will be said about multicollinearity in later paragraphs.

A second procedure for assessing weight instability involves testing the weights derived for one sample by applying them to another sample to calculate canonical scores. The correlations between the p_i and c_i scores for the second sample are then determined and compared with the canonical R_c values for the first sample [26]. This kind of an approach has been utilized in behavioral science studies [21, 22]. Following this procedure in the replicated study, the weights computed for the analysis subsample were applied to the predictor and criterion data sets in the validation subsample.

A subsequent analysis of the resulting canonical scores indicated correlations of .068, .057, and .334 for the three linear combinations. The correlation coefficient for the third linear combination was the only one significant at .05 level using the conventional test. It has been suggested that interpretations are meaningless if the validation correlations are not statistically significant [21]. This criterion would seem to be the minimum for interpretation, and there may be instances when significant correlations are simply too small to justify an interpretation. In the current instance, little confidence can be placed in an interpretation of the third linear combination since the correlations were low and insignificant for the first two combinations. Furthermore, the $p_i - c_i$ correlations for the validation subsample were approximately 22 to 88% lower than the canonical R_c's of .561, .481, and .427 for the analysis subsample which produced the weights. Deflation of this magnitude suggests extreme care in interpreting results of canonical analysis. Interpretations would be especially hazardous if they inferred that particular criterion behavior was linked to certain predictor variables.

The above procedure like the first one described a test for weight stability, not for multicollinearity within the data sets. Conceivably weights may be stable despite the presence of multicollinearity. Therefore, upon observing stable weights, an additional test for multicollinearity may be conducted if correlations among the variables within a set would alter the interpretations.

It might be noted that multicollinearity in and of itself does not rule out meaningful interpretations provided the weights are not the basis for the interpretations. Other statistics not subject to the effects of multicollinearity may be utilized. These are described in the subsequent section.

Multicollinearity may manifest itself in weight instability since it can lead to weight estimates that are very sensitive to particular sample observations. However, the weights can be stable with the multicollinearity resulting in some variables being partialled out of the linear combination. When this happens, some variables are assigned small weights although they are substantially related to variables in the other set.

The magnitude of multicollinearity, regardless of whether the weights are stable or unstable, may be examined by computing the coefficient of multiple determination, R^2, between each variable and the remaining ones within the set. These coefficients will reveal which variables are most affected by multicollinearity [12]. The investigator can then decide if interpretations are warranted in view of the multicollinearity that is present.

Returning to the main question of weight instability, it is interesting to note that internal computation of weights for the validation subsample leads to canonical R_c values of .789, .729, and .660. These are 28 to 50% higher than their counterparts in the analysis subsample. They also have higher levels of significance.

The canonical R_c's for the analysis and validation subsamples tend to bracket those reported by Sparks and Tucker and run higher than those for the full sample. These four sets of R_c values which are listed in Table 1 further demonstrate the potential variability in canonical results.

Table 1
COMPARISON OF CANONICAL R_c VALUES

	Linear combinations		
Sample	*1*	*2*	*3*
Full ($N = 183$)	.491[a]	.444[a]	.405[a]
Analysis ($N = 123$)	.561[a]	.481	.427
Sparks & Tucker ($N = 173$)	.606[a]	.548[a]	.413[a]
Validation ($N = 60$)	.789[a]	.729[a]	.660

[a] $p < .10$.

Subset or Construct Interpretations

Interpretational hazards may be encountered with canonical analysis if, in order to accomplish the research goals, it is necessary to partition the full set of criterion variables into subsets and to say, then, how these subsets are related to the predictor variables or some subsets of them. Some researchers have attempted to connect subsets of criterion variables with subsets of predictors. At times, the individual subsets have been named or labeled like in factor analysis and then viewed as constructs. When subset interpretations are made, even though they may stop short of delineating constructs, it is implied, if not stated, that nontrivial correlations from a practical or theoretical standpoint exist between the subsets of criterion and predictor variables.

Problems in making flawless interpretations of this sort occur because commonly employed computational procedures fail to provide canonical counterparts to partial correlation coefficients. Consequently, subset interpretations typically have been based on the size of the canonical weights or the loadings (a statistic which is described below) of the variables on particular linear combinations. When used for this purpose, canonical weights and loadings may have serious limitations if they are taken at face value without additional validation.

Canonical weights, while indicating the contribution of each original variable to the canonical score p_i or c_i, do not necessarily show how each criterion variable is related to the predictors or vice versa. A comparatively large weight may enhance canonical score correlation even though the correlation of the variable to those in the other data set is quite low. Alternatively, a low weight may be derived for a variable because multicollinearity has resulted in the variable being partialled out of the equation, although it has a substantial correlation with one or more variables in the other set. Furthermore, negative weights may be attributed to some variables because these variables serve as suppressors by subtracting irrelevant components of other variables from the canonical score equation [15, 23]. Thus, a variable might receive a substantial negative weight although it has little or no negative correlation with the other set of variables. The interpretational ambiguity caused by the suppressing phenomenon is most commonly discussed in context of multiple regression analysis.

The deficiencies of utilizing weights for subset interpretations are magnified further if the weights are unstable. Weights that vary widely from sample to sample fail to provide the consistency necessary for meaningful generalizations about relationships among criterion and predictor variables. Generally speaking, weights are very inadequate and sometimes misleading indicators of relationships between predictor and criterion variables.

Another statistic that shall be called a canonical loading is frequently computed and used as a basis for subset interpretations. A loading measures the simple linear correlation of an observed variable in the criterion or predictor set with the set's canonical scores. Thus the loading reflects the variance which the observed variable, say criterion Y_q, shares with the canonical scores, say the c_i's. Since there is a vector of canonical scores for each linear combination, every observed variable has a separate loading for each linear combination.

Some authors apply the term canonical coefficient to this statistic. For the sake of clarity, it is called a loading in this article to differentiate it clearly from canonical weights since weights are also called canonical coefficients by some.

Loadings, when compared to weights, have an advantage of being largely free from the direct influence of multicollinearity. Unlike weights, they show the simple direct relationship with canonical scores while ignoring the correlation between other variables and the scores. Loading values, therefore, are unaffected by the partialling out of variables and the presence of suppressing phenomenon.

In making subset interpretations based on loadings, the point above which loading values are considered to be nontrivial is arbitrary. In past studies, it has tended to range from .28 to .45 with .30 being used most often. It should be noted that a loading of .30 means that 9% of the variance in the observed variable is shared with the canonical scores of that linear combination. One might ask if this is a cogent amount for purposes of interpretation since even less than

9% of the variance in the observed variable may be shared with the other data set. The coefficient of multiple determination, R^2, can be computed to measure the amount of variance in a particular variable which is shared with variables in the other set. However, it should be recognized that this type of analysis is a step back toward examining variables in one of the sets in a univariate manner.

Loadings like weights may be subject to considerable instability or variability from sample to sample. Instability suggests that the loadings and hence the relationships ascribed to them are sample specific, due to chance, or the result of extraneous factors. Interpretations which portray subset relationships based on unstable loadings obviously have doubtful external validity because they are based at least in part on sample-specific covariation.

An illustration of interpretational dangers posed by unstable loadings can be seen in the replicated study of product usage and personality. Variable loadings in the analysis and validation subsamples differ substantially from each other as do those for the full sample and the Sparks-Tucker study [18]. Some variables with high loadings in the analysis subsample have low loadings in the validation subsample and vice versa. For example, one criterion variable has a loading of .5702 on the first linear combination in the analysis group and a −.1669 in the validation subsample, a .7371 variation.

One measure of similarity in loadings between samples is their rank correlation on a given linear combination. That is, a variable with a relatively high loading on a particular linear combination, such as the first, in one sample might be expected to have a comparatively high loading on the same combination in a second sample if the loadings are stable.

The rank correlations reveal the instability in the replicated study. In the case of criterion variables, the correlation between the loadings in the analysis and validation subsamples is .2157 on the first linear combination, −.1127 on the second, and −.1863 on the third. None approach significance using the conventional test. Looking at the same variables, the rank correlations between the loadings in the full sample and those reported by Sparks and Tucker are −.2623, .3431, and .0931 for the first, second, and third linear combinations. In comparing the loadings, one should remember the canonical procedure employed in the replication was the same one utilized by Sparks and Tucker (for a description of the procedure, see [24]).

Aside from the lack of rank correlation among loadings, it might be asked if criterion and predictor variables which are heavily loaded, say above the often used .30 level, on a linear combination in one sample are also linked in a second sample. In other words, if interpretations matched certain criterion and predictor variables in one sample on the basis of loadings, would the interpretations be valid for the second sample?

The answer is clearly no for the replication and it further testifies to the deficiencies which may be present in interpretations based on loadings that vary substantially from one sample to another. As Table 2 shows, the criterion and predictor variables which are linked by virtue of loadings above .30 in the analysis subsample differ markedly from those linked in the validation subsample. A comparable amount of disparity exists between the full sample and results of the Sparks-Tucker research.

Loadings, being sample estimates, are open to potentially wide variation as witnessed above. As far as the authors know, there is no readily available procedure to form confidence intervals for loadings. Such a procedure would offer a basis for determining which loadings were significant for interpretational purposes and the amount of variation that might prevail in each. In the absence of such a procedure, it seems wise to evaluate loading stability across separately drawn samples or a split sample before placing confidence in loading interpretations.

TABLE 2
VARIABLES WITH LOADINGS OF .30 AND ABOVE

Sample	*Linear combinations*					
	1		*2*		*3*	
	Criterion	*Predictor*	*Criterion*	*Predictor*	*Criterion*	*Predictor*
Analysis (N = 123)	6,7,9	1,4,6	1,3,10,14, 17	2,3,4,8	1,2,3,5,17	1,4,7
Validation (N = 60)	1,3,11,17	1,5,6,7	12,13	2,4,7	2,4,5,9, 14,16	1,3,4,7
Full (N = 183)	2,3,4,7, 8,9,14,17	1,4	1,3,5,7	2,7	3,4,11,17	1,3
Sparks & Tucker (N = 175)	4,7,9,13	2,3,4	1,2,7,9,17	3,4,5	5,6,11	1,3,4,6
Cross-loadings (N = 183)	none	1,4	none	7	none	none

In terms of subset interpretations, another weakness of loadings, even if they are stable, is their failure to provide a direct measure of the covariation between a given variable, say Y_q, and those in the other data set, say the predictor variables. Interpretations that link certain criterion and predictor variables on the basis of their loadings infer these variables are meaningfully related because these particular criterion variables are correlated with their own canonical scores, which are correlated with canonical scores for the predictor set, which are correlated with certain predictor variables. In short, it is inferred that Y_q is meaningfully correlated with X_m since $Y_q \leftrightarrow c_i \leftrightarrow p_i \leftrightarrow X_m$. Consequently, there is a danger the actual relationships between the subsets of criterion and predictor variables may be much weaker than the interpretation might imply, considering that loadings as low as .30 are often used to infer these linkages.

As an alternative to conventional loadings, Cooley and Lohnes have mentioned the possibility of correlating each of the observed criterion variables with the predictor canonical scores, and each observed predictor variable with the criterion canonical scores [5]. These correlations can be computed for each significant linear combination, yielding for each observed variable what shall be called cross-loadings here. The cross-loadings bridge across one of the intermediate relationships inherent in the conventional loadings and thus provide a somewhat more direct measure of criterion-predictor subset relationships. However, the approach does not eliminate the basic problem which is one of identifying meaningful subset relationships while lacking precise canonical statistics for doing so.

It seems cross-loadings would tend to be more conservative than conventional loadings and perhaps offer more solidity on which to base interpretations. The conservative feature of cross-loadings can be clearly seen in the replication.

The cross-loading values were substantially lower than the conventional loadings. In the full sample, the reduction on the criterion variables ranged from 47 to 56% on the first linear combination, 49 to 70% on the second, and 34 to 76% on the third.

The impact of this conservatism on subset interpretations is illustrated dramatically by comparing the criterion and predictor variables linked by cross-loadings with those linked by conventional loadings. For purposes of comparison, these linkages were based on the often used postulate that loading values of .30 and above are sufficient for drawing inferences about criterion-predictor relationships.

As shown in Table 2, the subset interpretations based on the two types of loadings would differ drastically. Conventional loadings on the first linear combination would suggest eight criterion variables (2, 3, 4, 7, 8, 9, 14, and 17) were associated with two predictors (1 and 4). All eight of these criterion variables would drop from the inferred relationship of cross-loadings were used. Cross-loadings on the second combination would delete all four of the criterion variables included by conventional loadings and one of the two predictors. On the third linear combination, all criterion and predictor variables would be dropped. In short, cross-loadings would indicate that relationships could not be ascribed to subsets of criterion and predictor variables. Whereas conventional loadings would suggest several such relationships involving 16 of the 25 criterion and predictor variables which were examined in the study.

CONCLUDING COMMENTS

Canonical correlation analysis can be an invaluable tool when appropriately employed in marketing research, and the intent of this article is not to discourage its use or to disparage the meaningful work by Sparks and Tucker. It is hoped that by illuminating limitations of canonical analysis its effectiveness as a research tool will be enhanced by reducing the possibility of misinterpretations. An awareness of the limitations also lessens the chance canonical analysis will be utilized when some other technique would yield a better picture of the marketing phenomenon being investigated. In either event, the effect indirectly facilitates an expansion of marketing knowledge by contributing to the productiveness of research efforts.

Canonical analysis provides the researcher with a tool for consolidating into a composite measure what otherwise might be an imponderable number of bivariate correlations between sets of variables. It is particularly useful in revealing overall relationships prevailing among criterion and predictor variables. This can be especially valuable if there is little *a priori* knowledge about relations between predictor and criterion variables.

Additional analyses become necessary if the researcher wishes to do more than identify overall relationships, such as formulate conclusions that depend on the magnitude of covariation between predictor and criterion variables or infer linkages among subsets of the variables. Before making interpretations like these, it is wise to establish that a substantial amount of variance in the criterion variables is shared with the predictors, that supporting statistics like weights and loadings are relatively stable from sample to sample, and that sizable correlations in fact exist between predictor-criterion subsets.

REFERENCES

1. Alpert, Mark I. "Personality and the Determinants of Product Choice," *Journal of Marketing Research*, 9 (February 1972), 89-92.
2. ______. "A Canonical Analysis of Personality and the Determinants of Automobile Choice, "*Combined Proceedings*. Chicago: American Marketing Association, 1971, 312-6.
3. ______ and Robert A. Peterson. "On the Interpretation of Canonical Analysis," *Journal of Marketing Research*,

9 (May 1972), 187-92; Working Paper, No. 71-45, University of Texas at Austin, May 1971.

4. Baumgarten, Steven A. and L. Winston Ring. "An Evaluation of Media Readership Constructs and Audience Profiles by Use of Canonical Correlation Analysis," *Combined Proceedings*. Chicago: American Marketing Association, 1971, 584-8.
5. Cooley, William W. and Paul R. Lohnes. *Multivariate Data Analysis*. New York: John Wiley & Sons, 1971.
6. Darden, William R. and Fred D. Reynolds. "Shopping Orientations and Product Usage Rates," *Journal of Marketing Research*, 8 (November 1971), 505-8.
7. Farley, John U. and L. Winston Ring, "'Empirical' Specification of a Buyer Behavior Model," *Journal of Marketing Research*, 11 (February 1974), 89-96.
8. Frank, Ronald E. and Charles E. Strain. "A Segmentation Research Design Using Consumer Panel Data," *Journal of Marketing Research*, 9 (November 1972), 385-90.
9. Green, Paul E., Michael H. Halbert, and Patrick J. Robinson. "Canonical Analysis: An Exposition and Illustrative Application," *Journal of Marketing Research*, 3 (February 1966), 32-9.
10. Gensch, Dennis H. and Ronald Turner. "Expanding the Definition of Potential Customer Using Modified Canonical Correlation," *Combined Proceedings*. Chicago: American Marketing Association, 1972, 499-504.
11. Kernan, Jerome B. "Choice Criteria, Decision Behavior, and Personality," *Journal of Marketing Research*, 5 (May 1968), 155-64.
12. Johnston, J. *Econometric Methods*. New York: McGraw-Hill, 1972.
13. Lessig, V. Parker. "Relating Multivariate Measures of Store Loyalty and Store Image," *Combined Proceedings*. Chicago: American Marketing Association, 1972, 305-9.
14. ______ and John O. Tollefson. "Market Segmentation Through Numerical Taxonomy," *Journal of Marketing Research*, 8 (November 1971), 480-7.
15. NcNemar, Quinn. *Psychological Statistics*. New York: John Wiley & Sons, 1962.
16. Neilson, Richard P. and John Stanton. "Type of Information Sensitivity, Fatalism and Spending Behavior," *Combined Proceedings*. Chicago: American Marketing Association, 1972, 400-5.
17. Pruden, Henry O. and Robert A. Peterson. "Personality and Performance-Satisfaction of Industrial Salesman," *Journal of Marketing Research*, 8 (November 1971), 501-4.
18. Sparks, David L. and W. T. Tucker. "A Multivariate Analysis of Personality and Product Use," *Journal of Marketing Research*, 8 (February 1971), 67-70.
19. Stewart, Douglas and William Love. "A General Canonical Correlation Index," *Psychological Bulletin*, 70 (September 1968), 160-3.
20. Sweeney, Timothy W., H. Lee Mathews, and David T. Wilson. "An Analysis of Industrial Buyers' Risk Reducing Behavior: Some Personality Correlates," *Combined Proceedings*, American Marketing Association, 1972, 217-21.
21. Thorndike, Robert M., and David J. Weiss. "Stability of Canonical Components," *Proceedings*. 78th Annual Convention of the American Psychological Association, 1970, Vol. 5, Part 1, 107-8.
22. ______ and Rene V. Dawis. "Multivariate Relationships between a Measure of Vocational Interests and a Measure of Vocational Needs," *Journal of Applied Psychology*, 52 (December 1968), 491-6.
23. Van de Geer, John P. *Introduction to Multivariate Analysis for the Social Sciences*. San Francisco: W. H. Freeman, 1971, 156-70.
24. Veldman, Donald J. *Fortran Programming for the Behavioral Sciences*. New York: Holt, Rinehart and Winston, 1967.
25. Wood, Donald A. "Toward the Interpretation of Canonical Dimensions," *Multivariate Behavioral Research*, 7 (October 1972), 477-82.
26. ______ and James A. Erskine. "Analytical Strategies in Canonical Correlation: A Review and Application," Working paper, Graduate School of Business, Indiana University, undated.
27. Worthing, Parker M., M. Venkatesan, and Steve Smith. "A Modified Approach to the Exploration of Personality and Product Use," *Combined Proceedings*. Chicago: American Marketing Association, 1971, 363-7.