FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

January 11, 1974

CANONICAL CORRELATION, MULTIPLE REGRESSION, AND SIMULTANEOUS SYSTEMS: SOME EQUIVALENCES AND THEIR IMPLICATIONS

Johny K. Johansson and Jagdish N. Sheth

#150

CANONICAL CORRELATION, MULTIPLE REGRESSION, AND SIMULTANEOUS SYSTEMS: SOME EQUIVALENCES AND THEIR IMPLICATIONS

by

Johny K. Johansson and Jagdish N. Sheth

University of Illinois

Introduction

The use and interpretation of results from canonical correlation analysis have always been difficult. This paper attempts to resolve some of the problems involved by establishing the connection between canonical correlation and multiple regression on one hand, and canonical correlation and simultaneous equation systems on the other. The hope is that with the relationships between these techniques clearly articulated, the researcher as well as the user of research findings will be in a better position to judge the possibilities and limitations of canonical correlation analysis, and to see where alternative techniques might provide additional information.

In what follows the multiple regression equivalences will be established first, and their implications spelled out. Then the relationships to simultaneous systems will be developed, and the consequences discussed. Finally, two empirical examples each illustrating different aspects of the use of the equivalences will be presented.

Canonical Correlation and Multiple Regression

Assume we deal with two dependent or criterion variables, y_1 and y_2, and three independent or predictor variables, x_1, x_2 and x_3.[1] These variables are all standardized with mean 0 and variance 1.

[1]The discussion that follows is directly generalizable to the case of p predictors and c criteria, $p \geq c$.

As is well known, a regression of one standardized, dependent variable upon another, standardized, independent variable yields a regression coefficient which is equal to the simple correlation between the two variables. If more than one, standardized, independent variable in introduced, the respective regression coefficients are equal to the corresponding partial correlation coefficients. (See, for example, Johnston, 1963, p. 30).

In canonical correlation we correlate a linear compound of the criterion variables with a linear compound of the predictors (and, successively, we develop additional linear compounds, orthogonal to the preceding ones and compute corresponding canonical correlations). These canonical correlations can then be seen as regression coefficients provided the linear compounds are appropriately standardized. We know that the variables entering each compound are standardized. Accordingly, the means of the compounds will all be zero: Any linear combination of variables with mean zero will itself have a mean of zero. As for the variance, we would need a standardization of the weights that will ensure a variance of 1. This is exactly the usual standardization $a' R_{y'y} a = 1$ imposed in canonical correlation (see, for example, Van de Geer, p. 157).[2]

Accordingly, we can write the canonical correlation analysis as the following system of equations:

$$a_{11}y_1 + a_{21}y_2 = c_1 (b_{11}x_1 + b_{21}x_2 + b_{31}x_3) + u_1$$

(1)

$$a_{12}y_1 + a_{22}y_2 = c_2 (b_{12}x_1 + b_{22}x_2 + b_{32}x_3) + u_2 \quad ,$$

[2]The superscript in a' denotes the transpose of the vector a; R denotes the correlation matrix.

where the $a_{i,j}$ and $b_{k,j}$, i, j, = 1,2, and k = 1,2,3, stand for the weights or loadings of the variables on the respective compound, and where c_i are the canonical correlations, here equal to the regression coefficients (because of the standardization the intercepts are of course zero). The u_i are random disturbances with zero means and constant variances. Their covariance is zero. Define the following matrices and vectors as

$$(2)\quad A' = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix},\ C = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix},\ B' = \begin{bmatrix} b_{11} & b_{21} & b_{31} \\ b_{12} & b_{22} & b_{32} \end{bmatrix},\ y' = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix},\ x' = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

$$u' = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

(In estimation, the element x_1, for example, will be a column vector of observations on the first independent variable). We can now write (1) as

$$(3)\quad yA = xBC + u$$

so that y can be expressed as

$$(4)\quad y = xBCA^{-1} + uA^{-1}$$

provided A^{-1} exists. But the rows of A correspond to the eigenvectors relating to each successive eigenvalue or canonical correlation.[3] Thus, A will always be nonsingular and its inverse exists. If we write

$$(5)\quad A^{-1} = \begin{bmatrix} a^{11} & a^{21} \\ a^{12} & a^{22} \end{bmatrix}$$

[3]These eigenvectors are not normalized to unit length -- the standardization used is the one ensuring unit variance -- and thus the A matrix is not orthogonal. It should be emphasized that the standardization to unit variance of the linear combinations could be replaced by, say, a normalization of the vector lengths; in such a case, however, the canonical analysis cannot be written as in (1).

the equation system defined by (4) becomes

$$(6)\quad \begin{aligned} y_1 &= (a^{11}c_1b_{11} + a^{12}c_2b_{12})x_1 + (a^{11}c_1b_{21} + a^{12}c_2b_{22})x_2 + (a^{11}c_1b_{31} + a^{12}c_2b_{32})x_3 + e_1 \\ y_2 &= (a^{21}c_1b_{11} + a^{22}c_2b_{12})x_1 + (a^{21}c_1b_{21} + a^{22}c_2b_{22})x_2 + (a^{21}c_1b_{31} + a^{22}c_2b_{32})x_3 + e_2 \end{aligned}$$

where $e_1 = a^{11}u_1 + a^{12}u_2$, $e_2 = a^{21}u_1 + a^{22}u_2$.

We see that the coefficients of the predictor or independent variables are composed of summed products, each product corresponding to a separate canonical correlation. The products themselves consist of the respective variables' weights on the canonical variates and the corresponding canonical root. Remembering that the y and x variables are standardized, it is clear that the coefficients of the predictors are the standardized or "beta" coefficients obtained in multiple regression.

It is sometimes useful to exhibit directly the dependence of each of the criterion variables upon the derived linear compounds of the predictor variables. Then (5) can be written

$$(7)\quad \begin{aligned} y_1 &= a^{11}c_1\,(b_{11}x_1 + b_{21}x_2 + b_{31}x_3) + a^{12}c_2\,(b_{12}x_1 + b_{22}x_2 + b_{32}x_3) + e_1 \\ y_2 &= a^{21}c_1\,(b_{11}x_1 + b_{21}x_2 + b_{31}x_3) + a^{22}c_2\,(b_{12}x_1 + b_{22}x_2 + b_{32}x_3) + e_2. \end{aligned}$$

Because the criterion variables and the linear compounds are standardized with mean zero and unit variance, the coefficients of the compounds represent the partial correlation coefficients between the criterion variable on the left and each compound on the right. Furthermore, as each linear compound is orthogonal to the preceding one, these partial correlations are identical to the simple (zero-order) correlations.

In this way we see clearly what happens when we decide to exclude one pair of variates from consideration. One consequence is that the fundamental explanation of the dependent variables y is lowered. The amount is easily computed for each y-element, as its maximum R^2 minus the explained variance due to the included variates. Seen in this light, we could let our cutoff point be partly determined by a special interest in one of the dependent variables, say y, rather than the whole set.

Implications: Computation of the Redundancy Statistics

It is interesting to compare this formulation of the canonical correlation analysis with the so-called redundancy testing technique suggested by Stewart and Love (1968), and Miller (1969). The redundancy technique aims at testing the "goodness of fit" of the canonical variates by assessing the proportion of variation in the original variables reproduced by the derived canonical variates (in contrast to a test of the canonical correlations themselves, which is a test of the degree of explanation of the linear combinations of the original criterion variables). Accordingly, the first step in redundancy analysis consists of establishing the total variation to be explained, the estimated "shared" or "common" variation between the two sets of variables.

One approach to this estimation starts with the correlations of the criterion variables with the canonical variates of the predictor set (see Cooley & Lohnes, 1971, p. 172). Denoting this correlation r and writing

$b_1' = (b_{11}\ b_{21}\ b_{31})$,

we have for the first criterion variable y_1 and the first predictor variate:

(8) $$r_{y_1 b_1' x} = \frac{1}{N} \sum_{i=1}^{N} y_{1i}\ b_1' x_i \quad .$$

for sample observations $i = 1,\ldots,N$. But from (7) we see that

(9) $$r_{y_1 b_1' x} = a^{11} c_1 \quad .$$

Similarly, for the second criterion variable y_2 we get the correlation

$$(10) \quad r_{y_2\, b_1'x} = a^{21} c_1 \ .$$

The redundancy measure is then derived as

$$(11) \quad R_{c1} = \frac{1}{m_c} \sum_{j=1}^{m_c} r^2_{y_j\, b_1'x}$$
$$= ((a^{11}c_1)^2 + (a^{21}c_1)^2)/2 \ ,$$

where m_c refers to the number of criterion variables (in our case two), and where the subscript c_1 indicates that the redundancy measure here corresponds to the first canonical correlation.[4] Thus, the redundancy measure simply consists of the average squared correlation between the criterion variables and the predictor variate.

In the same manner we can derive a redundancy measure corresponding to the second canonical variate. From (7) we see directly that this measure becomes

$$(12) \quad R_{c2} = ((a^{12}c_2)^2 + (a^{22}c_2)^2)/2.$$

Then an overall redundancy measure is defined as the sum of the separate measures:

$$(13) \quad R_{tot} = R_{c1} + R_{c2}$$

From (7) and (6) it is immediately clear that this overall redundancy measure is equivalent to the average R^2 derived for the separate regressions of each criterion variable upon all the predictors.

---

[4] There is a parallell measure of redundancy which relates each predictor variable to the first criterion variate, but which will not necessarily be of the same magnitude. It is of no direct concern here.

Implications: Uncovering Heterogeneity in the Data

To see what information is generated by the canonical analysis but not dealt with in a multiple regress on approach, we need to consider the regression coefficients in (6). Concentrating upon each regression coefficient as a sum of products, we can say first that each product comprises the amount of covariation between the two variables "channelled" through the respective variate. This is equivalent to an evaluation of the weights in the usual interpretation approach.

When these products are all of the same sign, the interpretation is clear. Where one product is considerably larger than the others, one variate channels most of the covariation; when all products are of similar magnitude, the covariation is reproduced through all variates jointly. Where the sum is "large" the presence of the two variables adds to the redundancy and also to the canonical correlation -- although not necessarily to the same degree as we have seen. Where the sum is "small", relatively little is added to the overall redundancy measure -- but what happens to the canonical correlation?

Clearly, if every product is small, their sum would be relatively small as well. In this case the canonical correlation would presumably be small as well (although of course other pairwise correlations might be high enough to make the correlation significant -- with little contribution from the two variables under consideration here). On the other hand, a small sum could appear also when some product is large and positive, whereas some other product is also large but negative. That such a situation can arise in canonical correlation analysis can be attested to by many empirical results. In this case the canonical correlations relating to the different products would not only test out significantly -- if they did not the results could always be seen as random -- but their interpretation would have to draw quite heavily upon the two "inconsistent" variables.

These two variables seem to be related both positively and negatively at the same time in this case. Such a result can only appear when the sample observations are clustered around two orthogonal axes in the multidimensional space investigated. That is, there is heterogeneity in the data -- for some sub-group of the sample the relationship between the two variables is positive, for another it is negative. The analysis should proceed only after this situation is corrected, for example, by subdividing the sample into the sub-groups indicated, or by a rotation of the significant variates.[5]

Canonical Correlation and Simultaneous Equation Systems

But canonical correlation can also be viewed as a case of simultaneous equation systems. This is clear if we rewrite (3) as

(14) $yA - xBC - u = 0$,

which is a standard version of the structural form of a simultaneous equation system (see, e.g., Goldberger, 1964, Ch. 7). Then (4) clearly is to be seen as the reduced form of the system:

(4) $y = xBCA^{-1} + uA^{-1}$.

From this viewpoint one interesting question to be raised is to what extent canonical correlation estimates of the parameters in A, B, and C

[5]For a rotational solution to this problem, see Johansson and Sheth (1973). In the second empirical illustration presented later we give an example of a sample split solution.

are comparable to those derived using any of the available simultaneous techniques. To answer this, we will first have to deal with the identification problem: Can distinct estimates be derived for the parameters in (14)? If we remember that the system in (14) embodies the equations in (1) it is clear at one glance that this is not possible. The second equation incorporates exactly the same variables as the first one, so that a host of linear combinations (with at least one non-zero element) of the first equation will be observationally indistinguishable from the second equation.

An illustration of the problem can be given with reference to the reduced form of the system. Clearly, ordinary least squares can be applied to the reduced form as written in (8). In this way six parameter estimates can be generated (one for each of the x-variables in each of the two equations). As can be seen from (1), however, the number of parameters amounts to four a's, and two c's in addition to the six b's. Accordingly, from the reduced form estimates we cannot derive the structural form parameter estimates -- they are underidentified.

We need clearly six additional independent relationships between the parameters in order to derive the desired estimates. These identifying restrictions are to be found in a few arbitrary constraints routinely imposed when computing canonical correlations. First, as was mentioned earlier, one usually imposes a "normalizing" constraint. If a_1 and a_2 denote the first and second column, respectively, of the A matrix, and b_1, and b_2 are correspondingly defined for the B matrix, this constraint generates the following four equalities (R denotes the correlation matrix):[6]

(15) $a_1'R_{y'y}a_1 = 1,\ a_2'R_{y'y}a_2 = 1,\ b_1'R_{x'x}b_1 = 1,\ b_2'R_{x'x}b_2 = 1.$

[6]In multiple regression, the custom of writing the dependent variable with a

The two last relationships needed for identification are to be found in the requirement that the second pair of variates be orthogonal to the first. This leads to

(16) $R_{a_1'y'ya_2} = 0, \quad R_{b_1'x'xb_2} = 0$.

Since the constraints (15) as well as (16) are quadratic in the parameters we still have an indeterminacy.[7] This is reflected in the symmetry of a canonical solution, making the first and third quadrants equivalent, as are the second and fourth. Once the choice of direction of axes -- and thus of sign of the weights (choice of roots) -- has been made, the twelve parameters are identified.

Implications: Structural Considerations

This correspondence between canonical correlation and simultaneous systems immediately suggests some new ways of thinking about canonical correlations. For example, the arbitrary assumptions made necessary for identification could conceivably be replaced by some more structural relationships. In economics, the identifying restrictions derive most often from the theoretical exclusion of some variables from a given equation -- it might be that before a canonical analysis is applied some more thought as to such a priori restrictions could be made even in the cases where very little theory is developed. Then the identifying assumptions would be brought out more clearly, perhaps also resulting in easier interpretable results. As one straightforward example, it could be mentioned that if one were willing to assume $a_{21} = 0$ in (1), together with the constraints

---

[7] Since the original variables y and x, as well as the derived canonical variates are standardized, the correlations between them degenerates into simple sums of cross-products divided by the number of observations.

in (15), a recursive model would result with estimates of the a's and the b's easily available through ordinary least squares.[8] Such an approach becomes also feasible if an initial canonical correlation indicates that a_{21} is close to zero.

One resulting advantage of such an a priori structural analysis would be that alternative estimation methods -- two-stage least squares, limited information maximum likelihood methods, for example -- might be applicable. The increased range of possibilities in itself would be advantageous. Furthermore, although the sampling theory for these techniques is by no means completely worked out for small samples, the asymptotic characteristics of the estimates are well known in contrast to canonical correlation, where the significance of the weights remains an unsolved problem.

Implications: Use of Canonical Correlation Results

Approaching canonical analysis as a special case of simultaneous systems also yields new insights into the use of the results from the analysis. Up to now the analysis has mainly focused upon the structural form (3) which is clearly tantamount to the economist's focus upon the interrelationships of the phenomenon modelled. Here the usefulness of canonical results has been impaired by the difficulty of interpretation. As the present analysis shows, one reason for this difficulty is the reluctance to impose structural considerations early in the analysis. As a result, as hoc rationalizations after the results appear -- with often less theoretical rationale than would have been possible a priori -- are frequently needed to make the results useful for policy decisions.

---

[8] For these estimates to be unbiased, we also require $E(u_1 u_2) = 0$. This is a statistical assumption, however, not an identification constraint.

But as we have seen, the canonical analysis also embodies a reduced form of the system. Since this fact has not been clearly exhibited previously, some very useful information has been ignored. Basically, the reduced form is useful when the program is one of predicting the (future) values of the endogeneous variables. In such a case the version employed is that of (5). One might ask why such an analysis could not be carried out using two multiple regressions, one for each of the y's, and then simply use the estimated regression coefficients. The answer is that this is possible if one is willing to give up the extra information given by the structural form. In this case the information is incorporated in the structural form coefficients which determine the reduced form coefficients as depicted in (6). The knowledge of this relationship enables the decision maker to directly compute the effects of a change in one structural form parameter that might be within his control. One example would be where one structural form parameter measures the effect of an advertisement, the actual parameter value being dependent upon whether black-and-white or color ads were being used. Another advantage of knowing the structural form parameters and not only the reduced form is the added understanding of the underlying processes it yields. Thus, the parameters will not only have statistical properties but also yields substantive insights into the phenomenon studied. This second benefit of the structural form is of course small when there is very little theory to build into the structure.

A First Empirical Example: The Effects of Advertising

To illustrate the theoretical developments the preceding pages, a common empirical problem of advertising effect determination will be presented. We have data on two "effect" variables, brand awareness and attitude, and also observations on different media combinations for the brand. The problem is the one of relating the advertising outlays to the two effects in order to ascertain the strength of the relationship.[9]

A natural approach might seem to be the use of two separate regressions, one for each of the two effects. The results of such an analysis is depicted in Table 2. The main result seems to be a fairly strong and significant impact of TV and Newspaper advertising on the two effect measures. In allocating advertising funds between Network and Spot TV (and Newspaper) advertising, however, the firm faces a problem of choice of objectives: Should the emphasis be placed on awareness or on attitude? In case both effect measures are given some positive weight, the simplest solution would be to allocate the expenditures so as to maximize a linear combination of the two. If the weights are explicitly determined a priori, the procedure to follow would be one of regressing the obtained linear compound of the two effects upon the media variables and then use that relationship for allocation. When no explicit weights are assigned, one might allow a statistical technique to derive the linear combination and then carry out the regression. This is of course exactly what the canonical correlation technique will do.

The results from such a canonical analysis are depicted in Table 3. Only one canonical root is significant (at the .05 level), and the criterion weights clearly indicate that this first pair of variates reflect the awareness effect. Partly because of this asymmetry, and partly because concentration upon one pair of variates leaves out a certain amount of explicable

[9]The data used here comes from Johansson (1973), where the type and quality of the data are extensively discussed. The brand used is a national brand in a frequently purchased product class, and we have 26 monthly observations. The variables were measured in deviations from their respective means, since for the short time periods involved only fluctuations around "normal" levels were of interest. All models run were linear.

variance, however insignificant, management might feel a bit uneasy about using these canonical results without modification. One might, for example, argue that the two effect measures are neither independent (as implied in the separate regressions approach), nor completely equivalent (as a canonical analysis basically assumes). Rather, using the hierarchy of effects paradigm, one could argue for an effect from awareness to attitude. Then, advertising would affect attitude not only directly, but also indirectly by affecting awareness.

Such reasoning leads to a recursive model, where awareness is regressed upon advertising in one equation, and attitude is regressed upon advertising and awareness in a second equation. As we saw in the earlier theoretical development, such an approach can be visualized as constraining one of the criterion weights (a_{21}) to be equal to zero. From a management point of view, the approach has the advantage of clearly specifying the relationship between the two effects. It could be the case, for example, that the best way to increase attitude in a favorable direction would be to advertise so as to increase awareness first.

The results of a recursive analysis are presented in Table 4. As can be seen, the introduction of awareness as one explanatory variable in the attitude equation completely wipes out any effect of the media variables. The simple correlation between awareness and liking as depicted in Table 1 clearly gives the reason for this result. After awareness has been allowed for, very little attitude variation is left to explain. Clearly, one explanation for this would be that advertising only affects awareness, later indirect effects being channeled through the hierarchy. If this view is adopted, the correct model would consist of the awareness relation plus a simple regression relation of attitude upon awareness. Another view would be that awareness and liking are determined simultaneously as a result of advertising and perhaps other factors. Or, more precisely, for the monthly data available here, awareness and liking might for all practical purposes be considered simultaneously determined, as causal loops with feedbacks would be completed within the month.

As we saw in the theoretical development, this interpretation will lead to either a canonical correlation approach or a simultaneous equations approach depending upon the constraints one is willing to impose upon the model. The canonical analysis, as we know, imposes these constraints regardless of prior information. Thus, that technique seems preferable in the cases where such information is at a minimum (or, as we saw above, where the explicit aim of the analysis is the derivation of one, and only one, functional relationship between the explanatory variables and the effects).

When there is some prior information about the structure of the model, it will generally be desirable to incorporate this information into the model specification. In this particular case, we might, for example, augment the independent variables by introducing a brand purchasing variable in the awareness relation, on the assumption that the more purchases that have been generated for the brand in past months, the higher the awareness regardless of advertising. Similarly, a repeat purchase variable could be introduced in the attitude relation, to account for the favorable feedback from regular brand purchasers. By assumption, the coefficients of these two additional exogeneous variables would be zero in the relations where they do not appear, thus identifying the parameters of those relations. As was stated earlier, these types of constraints (zero-restrictions) are the more usual ones in econometrics.

The results of a two stage least squares estimation of the two-equation simultaneous specification appear in Table 5. Although the presented standard errors have only an asymptotic validity in this case so that few firm conclusions can be drawn for a sample of size 26, it is clear that the results do not differ much from the recursive runs. Thus, the major portion of explanation comes from the interdependence between the two endogeneous variables awareness and attitude. Since we know from the reduced form estimation (Table 2) that advertising can predict awareness and attitude rather well, it is clear that the effect of advertising is indirect, going primarily through awareness to attitude, with some feedback from attitude to awareness (presumably by attitude affecting a memory component relating to awareness). This feedback clearly occurs within the month.

In this case, if one were to give an overall judgment as to which model is most appropriate, it can be argued that the obvious simultaneity between awareness and attitude is best depicted through the canonical solution. With only one pair of variates being significant -- the second one exhibiting a very low chi-square value -- and with the weights easy to interpret, for most purposes the canonical analysis seems preferable here. Furthermore, as indicated earlier, should the criterion weights seem inappropriate, other weights and a consequent rotation can easily be accomodated.[10]

A Second Empirical Example: The Determinants of College Education

In a second empirical example some other features of the theoretical development will be illustrated. The data are taken from the Cooley & Lohnes

---

[10] Another factor which argues against choosing the simultaneous equation solution is the fact that the purchase and repeat variables fail to enter their respective equations significantly. Thus, since the introduction of these variables serve to identify the system one can argue that, strictly speaking, the two equations fail to be identified.

book (1971, Appendix B) and thus easily available for crosschecking purposes. They relate to high school students' college plans, subject interest, various test scores, and socio-demographic background.[11] The aim of the analysis here is simply to ascertain what factors determine the student's college plans and curriculum interests.

Since very little a priori structure to impose was available, the decision was to run a canonical correlation analysis. The criterion set consisted of two interest scores (Physical Science and Office Work), a Sociability Index, and a College Plans variable. The predictor set included different test scores, a socio-economic index, and general background variables (for the exact specification of the variables, see Tables 6 and 7). The results from the first canonical correlation run are depicted in Table 7.

At the .05 level, three canonical roots are significant. From the weights one can infer that the first pair of variates relate the two interest variables (Physical Science and Office Work) to Sex, Information test I, and perhaps the Mathematics test. The second pair of variates relate College Plans and again Office Work Interest to Sex, Mathematics, Socioeconomics and perhaps Information test I again. The third pair of variates, finally, relate Sociability and the two Interest variables to Sex, Socioeconomics, Information tests I and II, Mechanical Ability, and perhaps Reading Ability.

When the signs of the weights are accounted for, it is clear that several inconsistent relationships obtain in these results. From the

---

[11] For more information on the data, the so-called Talent data set, the reader is referred to the Cooley & Lohnes book. The variables used in the present analysis are listed in Table 6.

first pair of variates Physical Science Interest and Sex are seen as inversely related (females are scored high, males low on the sex variable), whereas they are positively related according to the third canonical function. Similarly, Office Work Interest and Information test I are negatively related in the first function, positively related in the second and third. A third example of inconsistent results is the relationship between Office Work Interest and Socioeconomic Status, which is positive according to the second function, but negative judging from the third function.

To resolve the inconsistencies the split sample approach advocated earlier clearly has its problems here when more than one inconsistent relationship is uncovered.[12] It is clear, however, that in the present case the variable sex becomes a natural choice for a subdivision of the total sample. For one, sex is dichotomous, resulting in an easy splitting procedure. Furthermore, and more important, sex turns out to be an important variable in all the three significant canonical functions. Finally, it would seem natural to hypothesize that the relationships between the reamining variables might be different depending upon the respondent's sex. This last argument constitutes a type of a priori reasoning that one might or might not want to use in the initial specification of the model analyzed.[13]

---

[12] This difficulty does not arise when the rotation approach suggested earlier is adopted.

[13] It should be made clear that the fact that sex enters strongly in all three significant functions does not imply that the relationships between the other variables are affected by sex. There is no necessary relation between our last two reasons for using sex as the split variable.

The results from the separate male and female runs are presented in Table 8. As can be seen, the results are quite similar for the two sexes contrary to expectations. Two pairs of variates are significant (at the .05 level) in both cases and the defining variables of the functions are very much the same. The first function in both cases relates College Plans and Physical Science Interest to Information test I and Mathematics. The second function relates College Plans and Physical Science Interest to Creativity, Mechanical Ability, Abstract Thinking, and Socioeconomic Status for the female group, and to Information test I, Reading Ability, Creativity, Abstract Thinking, and Socioeconomic Status for the male group. The main difference between the two groups seems to be that the different test weights for the second predictor set have changed signs in a few cases. As a result of the split, the Office Work Interest criterion variable fails to enter strongly in any function, an indication that it is fairly constant within the sex group (although different between the sexes as indicated in the earlier results).

Although for the female data the inconsistencies have been eliminated, the male data show an inconsistent relationship between College Plans and Information test I. In the first function their relationship is negative, in the second positive. Instead of another sample split, another approach was adopted in dealing with this inconsistency. On the basis of the results so far a two-equation simultaneous system was developed for each of the female and male data. The endogeneous variables were in both cases College Plans and Physical Science Interest, dropping the Office Work and the Sociability variables. The exogeneous variables were chosen on the basis of the weights on the predictor sets and consisted of the variables presented in Table 9. The two equations were normalized on the respective criterion variable with the greatest weight on the function corresponding to the equation. Although the choice of equation normalization is thus somewhat arbitrary -- it is a bit misleading to talk about a "College Plan equation" and a "Physical Science Interest equation" -- the procedure is only a convenience and imposes no essential restrictions, since the normalization can be easily reversed.[14]

The idea behind this approach was not only that the inconsistency might be removed. In addition, and more important, the structure of the simultaneous system was suggested by the canonical correlation results, and comprised thus a natural model building step. In no way should the proposed simultaneous model be seen as tested by these data that helped build it. Rather it should be seen simply as one way of exhibiting the relationships in the data more clearly.

The results from the two stage least squares estimation of the parameters of the simultaneous specifications are presented in Table 9. Without going into very much detail about the interpretation of the results, a few remarks can still be made. The inconsistency has disappeared, with the Information test I variable having a positive impact upon both endogeneous variables. The inconsistency seems to be reflected, however, by the two endogeneous variables themselves. Their partial relationship according to the first equation is positive, and according to the second

[14]Note that under this approach a recursive system might result whenever a significant function exhibits only one strongly weighted criterion variable. Similarly, where this happens for all significant functions, a set of separate regressions is suggested, an intuitively appealing result.

equation negative. The first equation coefficient for the endogeneous College Plans variable has a relatively large standard error for the males, however, making it insignificantly different from zero.[15]

The other results are as one could expect from the initial canonical correlations. Most of the variables enter strongly into the relationship, and when couched in terms of simultaneous equations rather than canonical analysis, the interpretation of the relations is considerably more straightforward, as one is able to develop a much firmer understanding of when the mean of an estimated weight or coefficient is high relative to its variance.

From the results of the simultaneous system one could perhaps go even further. In the male data, for example, eliminating the less significant variables from the relationships, one would end up with a recursive system. The first equation would relate Physical Science Interest to Information test I and Mathematics, and the second equation would remain as presently, except for the elimination of the Information test I variable. Even without that refinement, however, it is fairly clear that the runs depicted in Table 3 represent a reasonably accurate picture of the determinants of college education plans as recorded in the available data.

Conclusion

It is clear that there are other possible facets of the equivalence between canonical correlation, multiple regression, and simultaneous systems which will be uncovered as further research is done. Only some of the more obvious implications have been developed here, but they should be sufficient

[15] If the coefficient had been significant, the interpretation would have been that College Plans are positively affected by Physical Science Interests, but that those who plan to go to College have no particular interest in Physical Science. This is clearly the case for the female data.

to show the possibilities involved. Overall, the greatest value of the established equivalence lies perhaps in the bridging of the gap between multivariate statistics and analysis on the one hand, and econometric methods on the other. As such, this paper should serve so as to bring together two very useful sets of statistical techniques for the benefit of the applied researcher.

TABLE 1

The Advertising Effects Data[1]

Sample Size = 26

Correlation Matrix

AWARENESS OF BRAND 1	ATTITUDE TOWARDS BRAND 2	PURCHASE OF BRAND LAST TIME 3	REPEAT PURCHASE OF BRAND 4	MAGAZINE ADVERTISING 5	NETWORK TV ADVERTISING 6	SPOT TV ADVERTISING 7	NEWSPAPER ADVERTISING 8
1.000							
0.944	1.000						
0.556	0.626	1.000					
0.359	0.369	0.645	1.000				
-0.130	-0.136	0.154	0.310	1.000			
0.542	0.486	0.094	0.376	-0.263	1.000		
0.387	0.267	0.070	-0.014	-0.129	0.277	1.000	
0.293	0.348	0.293	0.291	-0.083	0.185	-0.290	1.000

Variables 1 through 4 are measured as sample proportions.
Variable 5 is measured in 10,000 dollars.
Variables 6 and 7 are measured in millions of dollars.
Variable 8 is measured in linage.

TABLE 2

The Separate Regression Runs[1]

DEP. VAR.	R	MAGAZINES	NETWORK TV	SPOT TV	NEWSPAPERS
AWARENESS	.67	.390	.017*	.040*	.052*
		(3.03)	(.008)	(.019)	(.032)
ATTITUDE	.61	.311	.012*	.023	.051*
		(2.30)	(.007)	(.015)	(.026)

[1]Unstandardized coefficients, with standard errors in parenthesis. Significance at the .05 level indicated by a star.

TABLE 3

The Canonical Correlation Run

FUNCTION	EIGENVALUE	CORRELATION	WILKS LAMBDA	CHI-SQUARE	DF
1	0.469	0.685	0.478	16.60	8
2	0.098	0.313	0.901	2.32	3

MATRIX OF CRITERION WEIGHTS - CRITERIA DOWN, CANONICAL FUNCTIONS ACROSS

		1	2
1	AWARENESS	-1.580	2.599
2	ATTITUDE	0.638	-2.974

MATRIX OF PREDICTOR WEIGHTS - PREDICTORS DOWN, CANONICAL FUNCTIONS ACROSS

		1	2
5	MAGAZINE	-0.092	0.196
6	NETWORK TV	-0.565	-0.078
7	SPOT TV	-0.632	0.523
8	NEWSPAPER	-0.443	-0.683

REDUNDANCY: .395 .015 Total Redundancy: .410

$$A^{-1} = \begin{bmatrix} -0.977 & -0.054 \\ -0.209 & -0.519 \end{bmatrix}$$

TABLE 4

The Recursive Model[1]

DEP. VAR.	R	MAGAZINES	NETWORK TV	SPOT TV	NEWSPAPERS	AWARENESS
AWARENESS	.67	.890	.017*	.040*	.062*	
		(3.03)	(.008)	(.019)	(.032)	
ATTITUDE	.95	-.367	-.001	-.008	.004	.761*
		(.993)	(.003)	(.007)	(.011)	(.071)

[1] Unstandardized coefficients, with standard errors in parenthesis. Significance at the .05 level indicated by a star.

TABLE 5

Two Stage Least Squares Estimates of the Simultaneous Model

EQUATION	MAGAZINES	NETWORK TV	SPOT TV	NEWSPAPERS	AWARENESS	ATTITUDE	PURCHASE	REPEAT
AWARENESS	.580	.004	.015	.015		1.10	-.027	
	(1.38)	(.006)	(.010)	(.017)		(.406)	(.600)	
ATTITUDE	-.536	-.004	-.014	-.005	.918			.022
	(1.29)	(.004)	(.009)	(.015)	(.139)			(.404)

TABLE 6

The TALENT Data[1]

Variable No.	Predictor Set
1	School Size (4 categories--based on number of seniors) 1. under 25 2. 25 - 99 3. 100 - 399 4. 400 or more
2	Age (nearest year)
3	Sex (1=male; 2=female)
4	Weight (lb) 01. 74 or less 02. 75 - 89 03. 90 - 104 04. 105 - 119 05. 120 - 134 06. 135 - 149 07. 150 - 164 08. 165 - 179 09. 180 - 194 10. 195 - 209 11. 210 - 224 12. 225 or more
6	Information Test, Part I (R-190)
7	Information Test, Part II (R-192)
8	English Test (R-230)
9	Reading Comprehension Test (R-250)
10	Creativity Test (R-260)
11	Mechanical Reasoning Test (R-270)
12	Abstract Reasoning Test (R-290)
13	Mathematics Test (R-340)
17	Socioeconomic Status Index (P-801)

[1]These data appear in Cooley & Lohnes (1971, Appendix B).

TABLE 6 (cont'd)

Variable No.	Criterion Set
5	Plan College Full-time? (STB 301) 1. Definitely will go 2. Almost sure to go 3. Likely to go 4. Not likely to go 5. Definitely will not go
14	Sociability Inventory (R-601)
15	Physical Science Interest Inventory (P-701)
16	Office Work Interest Inventory (P-713)

TABLE 6 (cont'd)

SAMPLE SIZE = 505

CORRELATION MATRIX

	1	2	3	4	5	6	7	8	9
1	1.000								
2	-0.233	1.000							
3	0.051	-0.027	1.000						
4	-0.060	0.020	-0.596	1.000					
5	-0.113	0.084	0.101	-0.084	1.000				
6	0.053	-0.117	-0.328	0.263	-0.480	1.000			
7	0.107	-0.156	-0.150	0.146	-0.381	0.822	1.000		
8	0.029	-0.185	0.255	-0.157	-0.270	0.464	0.489	1.000	
9	0.066	-0.149	0.014	-0.020	-0.379	0.685	0.719	0.553	1.000
10	0.012	-0.138	-0.138	0.131	-0.283	0.640	0.568	0.385	0.598
11	-0.033	-0.046	-0.536	0.361	-0.253	0.652	0.506	0.137	0.424
12	0.085	-0.144	-0.042	-0.046	-0.231	0.475	0.457	0.405	0.539
13	0.044	-0.141	-0.175	0.160	-0.484	0.755	0.618	0.567	0.614
14	0.009	-0.093	0.093	-0.085	-0.082	-0.011	0.090	0.085	0.014
15	0.068	-0.066	-0.480	0.326	-0.379	0.572	0.401	0.081	0.273
16	0.000	-0.017	0.587	-0.383	0.183	-0.348	-0.220	0.098	-0.083
17	0.179	-0.159	-0.000	0.009	-0.372	0.385	0.377	0.227	0.294

	10	11	12	13	14	15	16	17
10	1.000							
11	0.519	1.000						
12	0.422	0.462	1.000					
13	0.537	0.528	0.490	1.000				
14	-0.001	-0.075	0.011	0.031	1.000			
15	0.330	0.528	0.207	0.489	0.060	1.000		
16	-0.126	-0.352	-0.069	-0.192	0.069	-0.239	1.000	
17	0.222	0.181	0.185	0.349	0.137	0.208	-0.097	1.000

TABLE 7

The First Canonical Correlation Run

FUNCTION	EIGENVALUE	CORRELATION	WILKS LAMBDA	CHI-SQUARE	DF
1	0.646	0.804	0.238	712.09	52
2	0.251	0.501	0.673	196.17	36
3	0.068	0.261	0.899	52.62	22
4	0.034	0.185	0.965	17.38	10

MATRIX OF CRITERION WEIGHTS - CRITERIA DOWN, CANONICAL FUNCTIONS ACROSS

	1	2	3	4
5	-0.152	-0.924	0.124	0.537
14	-0.094	0.105	-0.860	0.510
15	0.659	-0.046	0.440	0.759
16	-0.516	0.563	0.539	0.460

MATRIX OF PREDICTOR WEIGHTS - PREDICTORS DOWN, CANONICAL FUNCTIONS ACROSS

	1	2	3	4
1	0.076	0.079	0.221	-0.078
2	0.003	-0.008	0.105	-0.310
3	-0.515	0.574	0.706	0.341
4	0.014	-0.057	0.083	-0.254
6	0.649	0.256	0.811	-1.001
7	-0.114	-0.161	-1.111	1.131
8	-0.147	-0.013	-0.190	-0.159
9	-0.022	0.175	0.224	-0.919
10	-0.077	-0.004	0.084	0.268
11	0.041	-0.138	0.673	0.772
12	-0.063	-0.045	-0.120	-0.205
13	0.220	0.530	0.118	0.605
17	0.048	0.333	-0.577	-0.231
Redundancy:	.233	.057	.014	.007

Total Redundancy: .311

$$A^{-1} = \begin{bmatrix} -0.485 & -0.089 & 0.094 & -0.711 \\ -0.812 & 0.228 & 0.170 & 0.414 \\ 0.130 & -0.796 & 0.213 & 0.379 \\ 0.296 & 0.553 & 0.479 & 0.422 \end{bmatrix}$$

TABLE 8

The Male Runs

FUNCTION	EIGENVALUE	CORRELATION	WILKS LAMBDA	CHI-SQUARE	DF
1	0.549	0.741	0.338	243.79	52
2	0.153	0.391	0.751	64.23	36
3	0.074	0.273	0.887	26.80	22
4	0.040	0.201	0.959	9.35	10

MATRIX OF CRITERION WEIGHTS - CRITERIA DOWN, CANONICAL FUNCTIONS ACROSS

	1	2	3	4
5	-0.450	-0.899	-0.508	-0.228
14	-0.183	0.376	-0.936	0.141
15	0.777	-0.792	-0.356	-0.234
16	-0.248	-0.154	0.116	0.999

MATRIX OF PREDICTOR WEIGHTS - PREDICTORS DOWN, CANONICAL FUNCTIONS ACROSS

	1	2	3	4
1	0.079	0.078	0.415	-0.359
2	0.071	0.082	0.229	-0.209
4	0.020	0.177	0.235	-0.276
6	0.701	-0.849	0.914	-0.999
7	-0.201	0.284	-1.199	-0.002
8	-0.133	0.267	-0.623	-0.777
9	-0.083	0.644	0.767	0.758
10	0.116	-0.653	-0.428	0.137
11	-0.028	-0.075	0.108	0.256
12	-0.188	0.483	0.323	-0.278
13	0.626	0.159	0.092	0.768
17	0.136	0.624	-0.310	0.200
Redundancy:	.183	.027	.016	.011

Total Redundancy: .237

$$A^{-1} = \begin{bmatrix} -0.747 & 0.073 & 0.878 & 0.024 \\ -0.595 & 0.416 & -0.334 & -0.274 \\ -0.132 & -0.886 & -0.276 & 0.030 \\ -0.262 & 0.186 & 0.198 & 0.960 \end{bmatrix}$$

TABLE 8 (cont'd)

The Female Runs

FUNCTION	EIGENVALUE	CORRELATION	WILKS LAMBDA	CHI-SQUARE	DF
1	0.434	0.658	0.426	223.09	52
2	0.148	0.384	0.754	73.98	36
3	0.067	0.259	0.885	31.99	22
4	0.050	0.225	0.949	13.67	10

MATRIX OF CRITERION WEIGHTS - CRITERIA DOWN, CANONICAL FUNCTIONS ACROSS

	1	2	3	4
5	-0.490	-0.777	0.361	-0.469
14	0.025	0.232	0.911	0.342
15	0.577	-0.838	0.074	0.219
16	-0.330	-0.079	-0.367	0.933

MATRIX OF PREDICTOR WEIGHTS - PREDICTORS DOWN, CANONICAL FUNCTIONS ACROSS

	1	2	3	4
1	0.165	-0.174	0.177	0.417
2	-0.080	0.039	-0.334	-0.108
4	-0.001	-0.145	-0.092	-0.283
6	1.059	-0.119	-0.418	-0.935
7	-0.183	-0.021	1.276	0.284
8	-0.209	0.211	-0.618	0.652
9	0.042	0.143	-0.587	-0.520
10	-0.294	0.610	-0.178	0.098
11	0.059	-0.921	-0.168	0.282
12	-0.005	-0.257	0.091	0.290
13	0.314	0.179	0.225	0.365
17	0.091	0.538	0.157	0.155
Redundancy:	.150	.028	.016	.011

Total Redundancy: .215

$$A^{-1} = \begin{bmatrix} -0.760 & 0.011 & 0.761 & -0.566 \\ -0.559 & 0.186 & -0.602 & -0.208 \\ 0.246 & 0.907 & 0.043 & -0.219 \\ -0.219 & 0.377 & 0.235 & 0.766 \end{bmatrix}$$

TABLE 9

Two Stage Least Squares Estimation of a Simultaneous Model

MALES

Equation	Inform. Test I	Reading	Creativity	Abstract	Mathematics	Socioeconomic	College Plans	Physical Science Interest
Physical Science Interest	.091				.336		.982	
	(.026)				(.121)		(1.55)	
College Plans	.008	-.047	.105	-.093		-.032		-.131
	(.008)	(.016)	(.036)	(.046)		(.012)		(.040)

FEMALES

Equation	Inform. Test I	English	Creativity	Mechanical	Abstract	Mathematics	Socio-economics	College Plans	Physical Science Interest
Physical Science Interest	.198	-.151	-.505			.285		3.50	
	(.043)	(.063)	(.186)			(.103)		(1.56)	
College Plans		-.024	-.128	.187	.062		-.038		-.218
		(.012)	(.040)	(.059)	(.052)		(.011)		(.046)

REFERENCES

Alpert, M.I., and R.A. Peterson, "On the Interpretation of Canonical Analysis", Journal of Marketing Research, May 1972.

Anderson, T.W., An Introduction to Multivariate Statistical Analysis. New York: Wiley, 1958.

Cooley, W.W. and P.R. Lohnes, Multivariate Data Analysis, Wiley, 1971.

Dhrymes, P.J., Econometrics: Statistical Foundations and Applications. New York: Harper & Row, 1970.

Goldberger, A., Econometric Theory, Wiley, 1964.

Hannan, E.J., "Canonical Correlation and Multiple Equation Systems in Economics", Econometrica, Vol. 35, No. 1 (January, 1967).

Hooper, J.W., "Simultaneous Equations and Canonical Correlation Theory", Econometrica, Vol. 35, 1967, pp. 245-256.

Hotelling, H., "Relations between Two Sets of Variates", Biometrika, December 1935.

Johansson, J.K., "Returns to Scale in Advertising Media". Unpublished Doctoral Dissertation, University of California, Berkeley (1972).

Johansson, J.K., and J.N. Sheth, "Canonical Correlation Analysis of Competitive Market Structure", Proceedings, AIDS National Meeting, Boston, Nov. 1973.

Johnston, J. Econometric Methods, McGraw Hill, 1963.

Miller, J.K., "The Development and Application of Bi-Multivariate Correlation", Unpublished Doctoral Dissertation, SUNY at Buffalo, 1969.

Miller, J.K., and S.D. Farr, "Bimultivariate Redundancy: A Comprehensive Measure of Interbattery Relationship", Multivariate Behavioral Research July 1971.

Morrison, D.F., Multivariate Statistical Methods, McGraw Hill, 1967.

Morrison, D.F., "Significance Tests of Linear Combinations of P-Variates", Journal of the American Statistical Association, September 1971.

Stewart, D. and W. Love, "A General Canonical Correlation Index", Psychological Bulletin, September 1968.

Van de Geer, J.P., Multivariate Analysis for the Social Sciences, Freeman, 1971.

Vinod, H.D., "Econometrics of Joint Production", Econometrica, Vol. 36, No. 2 (April, 1968).