Canonical Correlation Analysis

LEARNING OBJECTIVES

Upon completing this chapter, you should be able to do the following:

- State the similarities and differences between multiple regression, factor analysis, discriminant analysis, and canonical correlation.
- Summarize the conditions that must be met for application of canonical correlation analysis.
- State what the canonical root measures and point out its limitations.
- State how many independent canonical functions can be defined between the two sets of original variables.
- Compare the advantages and disadvantages of the three methods for interpreting the nature of canonical functions.
- Define redundancy and compare it with multiple regression's R^2.

CHAPTER PREVIEW

Until recent years, canonical correlation analysis was a relatively unknown statistical technique. As with almost all of the multivariate techniques, the availability of computer programs has facilitated its increased application to research problems. It is particularly useful in situations in which multiple output measures such as satisfaction, purchase, or sales volume are available. If the independent variables were only categorical, multivariate analysis of variance could be used. But what if the independent variables are metric? Canonical correlation is the answer, allowing for the assessment of the relationship between metric independent variables and multiple dependent measures. Canonical correlation is considered to be the general model on which many other multivariate techniques are based because it can use both metric and nonmetric data for either the dependent or independent variables. We express the general form of canonical analysis as

$$\underset{\text{(metric, nonmetric)}}{Y_1 + Y_2 + Y_3 + \ldots + Y_n} = \underset{\text{(metric, nonmetric)}}{X_1 + X_2 + X_3 + \ldots + X_n}$$

This chapter introduces the researcher to the multivariate statistical technique of canonical correlation analysis. Specifically, we (1) describe the nature of canonical correlation analysis, (2) illustrate its application, and (3) discuss its potential advantages and limitations.

KEY TERMS

Before starting the chapter, review the key terms to develop an understanding of the concepts and terminology used. Throughout the chapter the key terms appear in **boldface**. Other points of emphasis in the chapter are *italicized*. Also, cross-references within the Key Terms appear in *italics*.

Canonical correlation Measure of the strength of the overall relationships between the linear composites (*canonical variates*) for the independent and dependent variables. In effect, it represents the bivariate correlation between the two canonical variates.

Canonical cross-loadings Correlation of each observed independent or dependent variable with the opposite *canonical variate*. For example, the independent variables are correlated with the dependent canonical variate. They can be interpreted like *canonical loadings*, but with the opposite canonical variate.

Canonical function Relationship (correlational) between two linear composites (*canonical variates*). Each canonical function has two canonical variates, one for the set of dependent variables and one for the set of independent variables. The strength of the relationship is given by the *canonical correlation.*

Canonical loadings Measure of the simple linear correlation between the independent variables and their respective *canonical variates*. These can be interpreted like factor loadings, and are also known as canonical structure correlations.

Canonical roots Squared *canonical correlations*, which provide an estimate of the amount of shared variance between the respective optimally weighted *canonical variates* of dependent and independent variables. Also known as *eigenvalues.*

Canonical variates Linear combinations that represent the weighted sum of two or more variables and can be defined for either dependent or independent variables. Also referred to as *linear composites*, linear compounds, and linear combinations.

Eigenvalues See *canonical roots*.

Linear composites See *canonical variates*.

Orthogonal Mathematical constraint specifying that the *canonical functions* are independent of each other. In other words, the canonical functions are derived so that each is at a right angle to all other functions when plotted in multivariate space, thus ensuring statistical independence between the canonical functions.

Redundancy index Amount of variance in a *canonical variate* (dependent or independent) explained by the other canonical variate in the *canonical function.* It can be computed for both the dependent and the independent canonical variates in each canonical function. For example, a redundancy index of the dependent variate represents the amount of variance in the dependent variables explained by the independent canonical variate.

What Is Canonical Correlation?

Multiple regression analysis is a multivariate technique which can predict the value of a single (metric) dependent variable from a linear function of a set of independent variables. For some research problems, however, interest may not center on a single dependent variable; rather, the researcher may be interested in relationships between sets of multiple dependent and multiple independent variables. **Canonical correlation analysis** is a multivariate statistical model that facilitates the study of interrelationships among sets of multiple dependent variables and multiple independent variables [5, 6]. Whereas multiple regression predicts a single dependent variable from a set of multiple independent variables, canonical correlation simultaneously predicts multiple dependent variables from multiple independent variables.

Canonical correlation places the fewest restrictions on the types of data on which it operates. Because the other techniques impose more rigid restrictions, it is generally believed that the information obtained from them is of higher quality and may be presented in a more interpretable manner. For this reason, many researchers view canonical correlation as a last-ditch effort, to be used when all other higher-level techniques have been exhausted. But in situations with multiple dependent and independent variables, canonical correlation is the most appropriate and powerful multivariate technique. It has gained acceptance in many fields and represents a useful tool for multivariate analysis, particularly as interest has spread to considering multiple dependent variables.

Hypothetical Example of Canonical Correlation

To clarify further the nature of canonical correlation, let us consider an extension of a simple example of multiple regression analysis. Assume that a survey was conducted to understand the relationships between family size and income as predictors of the number of credit cards a family would hold. Such a problem would involve examining the relationship between two independent variables and a single dependent variable.

Suppose the researcher then became interested in the broader concept of credit usage. To measure credit usage, the researcher considered not only the number of credit cards held by the family but also the family's average monthly dollar charges on all credit cards. These two measures were felt to give a much better perspective on a family's credit card usage. Readers interested in the approach of using multiple indicators to represent a concept are referred to discussions of Factor Analysis and Structural Equation Modeling. The problem now involves predicting two dependent measures simultaneously (number of credit cards and average dollar charges).

Multiple regression is capable of handling only a single dependent variable. Multivariate analysis of variance could be used, but only if all of the independent variables were nonmetric, which is not the case in this problem. Canonical correlation represents the only technique available for examining the relationship with multiple dependent variables.

The problem of predicting credit usage is illustrated in Table 8.1. The two dependent variables used to measure credit usage—number of credit cards held by the family and average monthly dollar expenditures on all credit cards—are listed at the left. The two independent variables selected to predict credit usage—family size and family income—are shown on the right. By using canonical correlation analysis, the researcher creates a composite measure of credit usage that consists of both dependent variables, rather than having to compute a separate regression equation for each of the dependent variables. The result of applying canonical correlation is a measure of the strength of the relationship between two sets of multiple variables (canonical variates). The measure of the strength of the relationship between the two variates is expressed as a canonical correlation coefficient (R_c). The researcher now has two results of interest: the canonical variates representing the optimal linear combinations of dependent and independent variables; and the canonical correlation representing the relationship between them.

Analyzing Relationships with Canonical Correlation

Canonical correlation analysis is the most generalized member of the family of multivariate statistical techniques. It is directly related to several dependence methods. Similar to regression, canonical correlation's goal is to quantify the strength of the relationship, in this case between the two sets of variables (independent and dependent). It corresponds to factor analysis in the creation of composites of variables. It also resembles discriminant analysis in its ability to determine independent dimensions (similar to discriminant functions) for each variable set, in this situation with the objective of producing the maximum correlation between the dimensions. Thus, canonical correlation identifies the optimum structure or dimensionality of each variable set that maximizes the relationship between independent and dependent variable sets.

Canonical correlation analysis deals with the association between composites of sets of multiple dependent and independent variables. In doing so, it develops a number of independent **canonical functions** that maximize the correlation between the **linear composites**, also known as **canonical variates**, which are sets of dependent and independent variables. Each canonical function is actually based on the correlation between two canonical variates, one variate for the dependent variables and one for the independent variables. Another unique feature of canonical correlation is that the variates are derived to maximize their correlation. Moreover, canonical correlation does not stop with the derivation of a single relationship between the sets of variables. Instead, a number of canonical functions (pairs of canonical variates) may be derived.

The following discussion of canonical correlation analysis is organized around a six-stage model-building process. The steps in this process include (1) specifying the objectives of canonical correlation, (2) developing the analysis plan, (3) assessing the assumptions underlying canonical correlation, (4) estimating the canonical model and assessing overall model fit, (5) interpreting the canonical variates, and (6) validating the model.

Stage 1: Objectives of Canonical Correlation Analysis

The appropriate data for canonical correlation analysis are two sets of variables. We assume that each set can be given some theoretical meaning, at least to the extent that one set could be defined as the independent variables and the other as the dependent variables. Once this distinction has been made, canonical correlation can address a wide range of objectives. These objects may be any or all of the following:

1. Determining whether two sets of variables (measurements made on the same objects) are independent of one another or, conversely, determining the magnitude of the relationships that may exist between the two sets.
2. Deriving a set of weights for each set of dependent and independent variables so that the linear combinations of each set are maximally correlated. Additional linear functions that maximize the remaining correlation are independent of the preceding set(s) of linear combinations.
3. Explaining the nature of whatever relationships exist between the sets of dependent and independent variables, generally by measuring the relative contribution of each variable to the canonical functions (relationships) that are extracted.

The inherent flexibility of canonical correlation in terms of the number and types of variables handled, both dependent and independent, makes it a logical candidate for many of the more complex problems addressed with multivariate techniques.

Stage 2: Designing a Canonical Correlation Analysis

As the most general form of multivariate analysis, canonical correlation analysis shares basic implementation issues common to all multivariate techniques. Discussions on the impact of measurement error, the types of variables, and their transformations that can be included are relevant to canonical correlation analysis as well.

The issues of the impact of sample size (both small and large) and the necessity for a sufficient number of observations per variable are frequently encountered with canonical correlation. Researchers are tempted to include many variables in both the independent and dependent variable set, not realizing the implications for sample size. Sample sizes that are very small will not represent the correlations well, thus obscuring any meaningful relationships. Very large samples will have a tendency to indicate statistical significance in all instances, even where practical significance is not indicated. The researcher is also encouraged to maintain at least 10 observations per variable to avoid "overfitting" the data.

The classification of variables as dependent or independent is of little importance for the statistical estimation of the canonical functions, because canonical correlation analysis weights both variates to maximize the correlation and places no particular emphasis on either variate. Yet because the technique produces variates to maximize the correlation between them, a variable in either set relates to all other variables in both sets. This allows the addition or deletion of a single variable to affect the entire solution, particularly the other variate. The composition of each variate, either independent or dependent, becomes critical. A researcher must have conceptually linked sets of the variables before applying canonical correlation analysis. This makes the specification of dependent versus independent variates essential to establishing a strong conceptual foundation for the variables.

Stage 3: Assumptions in Canonical Correlation

The generality of canonical correlation analysis also extends to its underlying statistical assumptions. The assumption of linearity affects two aspects of canonical correlation results. First, the correlation coefficient between any two variables is based on a linear relationship. If the relationship is nonlinear, then one or both variables should be transformed, if possible. Second, the canonical correlation is the linear relationship between the variates. If the variates relate in a nonlinear manner, the relationship will not be captured by canonical correlation. Thus, while canonical correlation analysis is the most generalized multivariate method, it is still constrained to identifying linear relationships.

Canonical correlation analysis can accommodate any metric variable without the strict assumption of normality. Normality is desirable because it standardizes a distribution to allow for a higher correlation among the variables. But in the strictest sense, canonical correlation analysis can accommodate even nonnormal variables if the distributional form (e.g., highly skewed) does not decrease the correlation with other variables. This allows for transformed nonmetric data (in the form of dummy variables) to be used as well. However, multivariate normality is required for the statistical inference test of the significance of each canonical function. Because tests for multivariate normality are not readily available, the prevailing guideline is to ensure that each variable has univariate normality. Thus, although normality is not strictly required, it is highly recommended that all variables be evaluated for normality and transformed if necessary.

Homoscedasticity, to the extent that it decreases the correlation between variables, should also be remedied. Finally, multicollinearity among either variable set will confound the ability of the technique to isolate the impact of any single variable, making interpretation less reliable.

Stage 4: Deriving the Canonical Functions and Assessing Overall Fit

The first step of canonical correlation analysis is to derive one or more canonical functions. Each function consists of a pair of variates, one representing the independent variables and the other representing the dependent variables. The maximum number of canonical variates (functions) that can be extracted from the sets of variables equals the number of variables in the smallest data set, independent or dependent. For example, when the research problem involves five independent variables and three dependent variables, the maximum number of canonical functions that can be extracted is three.

Deriving Canonical Functions

The derivation of successive canonical variates is similar to the procedure used with unrotated factor analysis. The first factor extracted accounts for the maximum amount of variance in the set of variables, then the second factor is computed so that it accounts for as much as possible of the variance not accounted for by the first factor, and so forth, until all factors have been extracted. Therefore, successive factors are derived from residual or leftover variance from earlier factors. Canonical correlation analysis follows a similar procedure but focuses on accounting for the maximum amount of the relationship between the two sets of variables, rather than within a single set. The result is that the first pair of canonical variates is derived so as to have the highest intercorrelation possible between the two sets of variables. The second pair of canonical variates is then derived so that it exhibits the maximum relationship between the two sets of variables (variates) not accounted for by the first pair of variates. In short, successive pairs of canonical variates are based on residual variance, and their respective canonical correlations (which reflect the interrelationships between the variates) become smaller as each additional function is extracted. That is, the first pair of canonical variates exhibits the highest intercorrelation, the next pair the second-highest correlation, and so forth.

One additional point about the derivation of canonical variates: as noted, successive pairs of canonical variates are based on residual variance. Therefore, each of the pairs of variates is **orthogonal** and independent of all other variates derived from the same set of data.

The strength of the relationship between the pairs of variates is reflected by the canonical correlation. When squared, the canonical correlation represents the amount of variance in one canonical variate accounted for by the other canonical variate. This also may be called the amount of shared variance between the two canonical variates. Squared canonical correlations are called **canonical roots** or **eigenvalues**.

Which Canonical Functions Should Be Interpreted?

As with research using other statistical techniques, the most common practice is to analyze functions whose canonical correlation coefficients are statistically significant beyond some level, typically .05 or above. If other independent functions are deemed insignificant, these relationships among the variables are not interpreted. Interpretation of the canonical variates in a significant function is based on the premise that variables in each set that contribute heavily to shared variances for these functions are considered to be related to each other.

The authors believe that the use of a single criterion such as the level of significance is too superficial. Instead, they recommend that three criteria be used in conjunction with one another to decide which canonical functions should be interpreted. The three criteria are (1) level of statistical significance of the function, (2) magnitude of the canonical correlation, and (3) redundancy measure for the percentage of variance accounted for from the two data sets.

Level of Significance

The level of significance of a canonical correlation generally considered to be the minimum acceptable for interpretation is the .05 level, which (along with the .01 level) has become the generally accepted level for considering a correlation coefficient statistically significant. This consensus has developed largely because of the availability of tables for these levels. These levels are not necessarily required in all situations, however, and researchers from various disciplines frequently must rely on results based on lower levels of significance. The most widely used test, and the one normally provided by computer packages, is the F statistic, based on Rao's approximation [3].

In addition to separate tests of each canonical function, a multivariate test of all canonical roots can also be used for evaluating the significance of canonical roots. Many of the measures for assessing the significance of discriminant functions, including Wilks' lambda, Hotelling's trace, Pillai's trace, and Roy's gcr, are also provided.

Magnitude of the Canonical Relationships

The practical significance of the canonical functions, represented by the size of the canonical correlations, also should be considered when deciding which functions to interpret. No generally accepted guidelines have been established regarding suitable sizes for canonical correlations. Rather, the decision is usually based on the contribution of the findings to better understanding of the research problem being studied. It seems logical that the guidelines suggested for significant factor loadings in factor analysis might be useful with canonical correlations, particularly when one considers that canonical correlations refer to the variance explained in the canonical variates (linear composites), not the original variables.

Redundancy Measure of Shared Variance

Recall that squared canonical correlations (roots) provide an estimate of the shared variance between the canonical variates. Although this is a simple and appealing measure of the shared variance, it may lead to some misinterpretation because the squared canonical correlations represent the variance shared by the linear composites of the sets of dependent and independent variables, and not the variance extracted from the sets of variables [1]. Thus, a relatively strong canonical correlation may be obtained between two linear composites (canonical variates), even though these linear composites may not extract significant portions of variance from their respective sets of variables.

Because canonical correlations may be obtained that are considerably larger than previously reported bivariate and multiple correlation coefficients, there may be a temptation to assume that canonical analysis has uncovered substantial relationships of conceptual and practical significance. Before such conclusions are warranted, however, further analysis involving measures other than canonical correlations must be undertaken to determine the amount of the dependent variable variance accounted for or shared with the independent variables [7].

To overcome the inherent bias and uncertainty in using canonical roots (squared canonical correlations) as a measure of shared variance, a **redundancy index** has been proposed [8]. It is the equivalent of computing the squared multiple correlation coefficient between the total independent variable set and each variable in the dependent variable set, and then averaging these squared coefficients to arrive at an average R^2. This index provides a summary measure of the ability of a set of independent variables (taken as a set) to explain variation in the dependent variables (taken one at a time). As such, the redundancy measure is perfectly analogous to multiple regression's R^2 statistic, and its value as an index is similar.

The Stewart-Love index of redundancy calculates the amount of variance in one set of variables that can be explained by the variance in the other set. This index serves as a measure of accounted-for variance, similar to the R^2 calculation used in multiple regression. The R^2 represents the amount of variance in the dependent variable explained by the regression function of the independent variables. In regression, the total variance in the dependent variable is equal to 1, or 100 percent. Remember that canonical correlation is different from multiple regression in that it does not deal with a single dependent variable but has a composite of the dependent variables, and this composite has only a portion of each dependent variable's total variance. For this reason, we cannot assume that 100 percent of the variance in the dependent variable set is available to be explained by the independent variable set. The set of independent variables can be expected to account only for the shared variance in the dependent canonical variate. For this reason, the calculation of the redundancy index is a three-step process. The first step involves calculating the amount of shared variance from the set of dependent variables included in the dependent canonical variate. The second step involves calculating the amount of variance in the dependent canonical variate that can be explained by the independent canonical variate. The final step is to calculate the redundancy index, found by multiplying these two components.

Step 1: The Amount of Shared Variance. To calculate the amount of shared variance in the dependent variable set included in the dependent canonical variate, let us first consider how the regression R^2 statistic is calculated. R^2 is simply the square of the correlation coefficient R, which represents the correlation between the actual dependent variable and the predicted value. In the canonical case, we are concerned with correlation between the dependent canonical variate and each of the dependent variables. Such information can be obtained from the canonical loadings (L_1), which represent the correlation between each input variable and its own canonical variate (discussed in more detail in the following section). By squaring each of the dependent variable loadings (L_i^2), one may obtain a measure of the amount of variation in each of the dependent variables explained by the dependent canonical variate. To calculate the amount of shared variance explained by the canonical variate, a simple average of the squared loadings is used.

Step 2: The Amount of Explained Variance. The second step of the redundancy process involves the percentage of variance in the dependent canonical variate that can be explained by the independent canonical variate. This is simply the squared correlation between the independent canonical variate and the dependent canonical variate, which is otherwise known as the canonical correlation. The squared canonical correlation is commonly called the canonical R^2.

Step 3: The Redundancy Index. The redundancy index of a variate is then derived by multiplying the two components (shared variance of the variate multiplied by the squared canonical correlation) to find the amount of shared variance that can be explained by each canonical function. To have a high redundancy index, one must have a high canonical correlation and a high degree of shared variance explained by the dependent variate. A high canonical correlation alone does not ensure a valuable canonical function. Redundancy indices are calculated for both the dependent and the independent variates, although in most instances the researcher is concerned only with the variance extracted from the dependent variable set, which provides a much more realistic measure of the predictive ability of canonical relationships. The researcher should note that while the canonical correlation is the same for both variates in the canonical function, the redundancy index will most likely vary between the two variates, as each will have a differing amount of shared variance.

What is the minimum acceptable redundancy index needed to justify the interpretation of canonical functions? Just as with canonical correlations, no generally accepted guidelines have been established. The researcher must judge each canonical function in light of its theoretical and practical significance to the research problem being investigated to determine whether the redundancy index is sufficient to justify interpretation. A test for the significance of the redundancy index has been developed [2], although it has not been widely utilized.

Stage 5: Interpreting the Canonical Variate

If the canonical relationship is statistically significant and the magnitudes of the canonical root and the redundancy index are acceptable, the researcher still needs to make substantive interpretations of the results. Making these interpretations involves examining the canonical functions to determine the relative importance of each of the original variables in the canonical relationships. Three methods have been proposed: (1) canonical weights (standardized coefficients), (2) canonical loadings (structure correlations), and (3) canonical cross-loadings.

Canonical Weights

The traditional approach to interpreting canonical functions involves examining the sign and the magnitude of the canonical weight assigned to each variable in its canonical variate. Variables with relatively larger weights contribute more to the variates, and vice versa. Similarly, variables whose weights have opposite signs exhibit an inverse relationship with each other, and variables with weights of the same sign exhibit a direct relationship. However, interpreting the relative importance or contribution of a variable by its canonical weight is subject to the same criticisms associated with the interpretation of beta weights in regression techniques. For example, a small weight may mean either that its corresponding variable is irrelevant in determining a relationship or that it has been partialed out of the relationship because of a high degree of multicollinearity. Another problem with the use of canonical weights is that these weights are subject to considerable instability (variability) from one sample to another. This instability occurs because the computational procedure for canonical analysis yields weights that maximize the canonical correlations for a particular sample of observed dependent and independent variable sets [7]. These problems suggest considerable caution in using canonical weights to interpret the results of a canonical analysis.

Canonical Loadings

Canonical loadings have been increasingly used as a basis for interpretation because of the deficiencies inherent in canonical weights. **Canonical loadings**, also called canonical structure correlations, measure the simple linear correlation between an original observed variable in the dependent or independent set and the set's canonical variate. The canonical loading reflects the variance that the observed variable shares with the canonical variate and can be interpreted like a factor loading in assessing the relative contribution of each variable to each canonical function. The methodology considers each independent canonical function separately and computes the within-set variable-to-variate correlation. The larger the coefficient, the more important it is in deriving the canonical variate. Also, the criteria for determining the significance of canonical structure correlations are the same as with factor loadings in factor analysis.

Canonical loadings, like weights, may be subject to considerable variability from one sample to another. This variability suggests that loadings, and hence the relationships ascribed to them, may be sample-specific, resulting from chance or extraneous factors [7]. Although canonical loadings are considered relatively more valid than weights as a means of interpreting the nature of canonical relationships, the researcher still must be cautious when using loadings for interpreting canonical relationships, particularly with regard to the external validity of the findings.

Canonical Cross-Loadings

The computation of **canonical cross-loadings** has been suggested as an alternative to canonical loadings [4]. This procedure involves correlating each of the original observed dependent variables directly with the independent canonical variate, and vice versa. Recall that conventional loadings correlate the original observed variables with their respective variates after the two canonical variates (dependent and independent) are maximally correlated with each other. This may also seem similar to multiple regression, but it differs in that each independent variable, for example, is correlated with the dependent variate instead of a single dependent variable. Thus cross-loadings provide a more direct measure of the dependent–independent variable relationships by eliminating an intermediate step involved in conventional loadings. Some canonical analyses do not compute correlations between the variables and the variates. In such cases the canonical weights are considered comparable but not equivalent for purposes of our discussion.

Which Interpretation Approach to Use

Several different methods for interpreting the nature of canonical relationships have been discussed. The question remains, however: Which method should the researcher use? Because most canonical problems require a computer, the researcher frequently must use whichever method is available in the standard statistical packages. The cross-loadings approach is preferred, and it is provided by many computer programs, but if the cross-loadings are not available, the researcher is forced either to compute the cross-loadings by hand or to rely on the other methods of interpretation. The canonical loadings approach is somewhat more representative than the use of weights, just as was seen with factor analysis and discriminant analysis. Therefore, whenever possible the loadings approach is recommended as the best alternative to the canonical cross-loadings method.

Stage 6: Validation and Diagnosis

As with any other multivariate technique, canonical correlation analysis should be subjected to validation methods to ensure that the results are not specific only to the sample data and can be generalized to the population. The most direct procedure is to create two subsamples of the data (if sample size allows) and perform the analysis on each subsample separately. Then the results can be compared for similarity of canonical functions, variate loadings, and the like. If marked differences are found, the researcher should consider additional investigation to ensure that the final results are representative of the population values, not solely those of a single sample.

Another approach is to assess the sensitivity of the results to the removal of a dependent and/or independent variable. Because the canonical correlation procedure maximizes the correlation and does not optimize the interpretability, the canonical weights and loadings may vary substantially if one variable is removed from either variate. To ensure the stability of the canonical weights and loading, the researcher should estimate multiple canonical correlations, each time removing a different independent or dependent variable.

Although there are few diagnostic procedures developed specifically for canonical correlation analysis, the researcher should view the results within the limitations of the technique. Among the limitations that can have the greatest impact on the results and their interpretation are the following:

1. The canonical correlation reflects the variance shared by the linear composites of the sets of variables, not the variance extracted from the variables.
2. Canonical weights derived in computing canonical functions are subject to a great deal of instability.
3. Canonical weights are derived to maximize the correlation between linear composites, not the variance extracted.
4. The interpretation of the canonical variates may be difficult because they are calculated to maximize the relationship, and there are no aids for interpretation, such as rotation of variates, as seen in factor analysis.
5. It is difficult to identify meaningful relationships between the subsets of independent and dependent variables because precise statistics have not yet been developed to interpret canonical analysis, and we must rely on inadequate measures such as loadings or cross-loadings [7].

These limitations are not meant to discourage the use of canonical correlation. Rather, they are pointed out to enhance the effectiveness of canonical correlation as a research tool.

An Illustrative Example

To illustrate the application of canonical correlation, we use variables drawn from the a short survey of customers of the firm known as HATCO. The data consist of a series of measures obtained on a sample of 100 HATCO customers. The variables include ratings of HATCO on seven attributes (X_1 to X_7) and two measures reflecting the effects of HATCO's efforts (X_9, usage of HATCO products, and X_{10}, customer satisfaction with HATCO). A complete description of the HATCO survey is provided in Exhibit 1.

The discussion of this application of canonical correlation analysis follows the six-stage process discussed earlier in the chapter. At each stage the results illustrating the decisions in that stage are examined.

Stage 1: Objectives of Canonical Correlation Analysis

In demonstrating the application of canonical correlation, we use nine variables as input data. The HATCO ratings (X_1 through X_7) are designated as the set of independent variables. The measures of usage level and satisfaction level (variables X_9 and X_{10}) are specified as the set of dependent variables. The statistical problem involves identifying any latent relationships (relationships between composites of variables rather the individual variables themselves) between a customer's perceptions about HATCO and the customer's level of usage and satisfaction.

Stages 2 and 3: Designing a Canonical Correlation Analysis and Testing the Assumptions

The designation of the variables includes two metric-dependent and seven metric-independent variables. The conceptual basis of both sets is well established, so there is no need for alternative model formulations testing different sets of variables. The seven variables resulted in a 13-to-1 ratio of observations to variables, exceeding the guideline of 10 observations per variable. The sample size of 100 is not felt to affect the estimates of sampling error markedly and thus should have no impact on the statistical significance of the results. Finally, for purposes of this example, assume that both dependent and independent variables were assessed for meeting the basic distributional assumptions underlying multivariate analyses and passed all statistical tests.

Stage 4: Deriving the Canonical Functions and Assessing Overall Fit

The canonical correlation analysis was restricted to deriving two canonical functions because the dependent variable set contained only two variables. To determine the number of canonical functions to include in the interpretation stage, the analysis focused on the level of statistical significance, the practical significance of the canonical correlation, and the redundancy indices for each variate.

Statistical and Practical Significance

The first statistical significance test is for the canonical correlations of each of the two canonical functions. In this example, both canonical correlations are statistically significant (see Table 8.2). In addition to tests of each canonical function separately, multivariate tests of both functions simultaneously are also performed. The test statistics employed are Wilks' lambda, Pillai's criterion, Hotelling's trace, and Roy's gcr. Table 8.2 also details the multivariate test statistics, which all indicate that the canonical functions, taken collectively, are statistically significant at the .01 level.

In addition to statistical significance, the canonical correlations were both of sufficient size to be deemed practically significant. The final step was to perform redundancy analyses on both canonical functions.

Redundancy Analysis

A redundancy index is calculated for the independent and dependent variates of the first function in Table 8.3. As can be seen, the redundancy index for the dependent variate is substantial (.751). The independent variate, however, has a markedly lower redundancy index (.242), although in this case, because there is a clear delineation between dependent and independent variables, this lower value is not unexpected or problematic. The low redundancy of the independent variate results from the relatively low shared variance in the independent variate (.276), not the canonical R^2. From the redundancy analysis and the statistical significance tests, the first function should be accepted.

The redundancy analysis for the second function produces quite different results (see Table 8.4). First, the canonical R^2 is substantially lower (.260). Moreover, both variable sets have low shared variance in the second function (.145 for the dependent variate and .082 for the independent variate). Their combination with the canonical root in the redundancy index produces values of .038 for the dependent variate and .021 for the independent variate. Thus, although the second function is statistically significant, it has little practical significance. With such a small percentage, one must question the value of the function. This is an excellent example of a statistically significant canonical function that does not have practical significance because it does not explain a large proportion of the dependent variables' variance.

The interested researcher should also consider factor analysis with attention to the discussion of scale development. Canonical correlation is in some ways a form of scale development, as the dependent and independent variates represent dimensions of the variable sets similar to the scales developed with factor analysis. The primary difference is that these dimensions are developed to maximize the relationship between them, whereas factor analysis maximizes the explanation (shared variance) of the variable set.

Stage *5*: Interpreting the Canonical Variates

With the canonical relationship deemed statistically significant and the magnitude of the canonical root and the redundancy index acceptable, the researcher proceeds to making substantive interpretations of the results. Although the second function could be considered practically nonsignificant, owing to the low redundancy value, it is included in the interpretation phase for illustrative reasons. These interpretations involve examining the canonical functions to determine the relative importance of each of the original variables in deriving the canonical relationships. The three methods for interpretation are (1) canonical weights (standardized coefficients), (2) canonical loadings (structure correlations), and (3) canonical cross-loadings.

Canonical Weights

Table 8.5 contains the standardized canonical weights for each canonical variate for both dependent and independent variables. As discussed earlier, the magnitude of the weights represents their relative contribution to the variate. Based on the size of the weights, the order of contribution of independent variables to the first variate is X_3, X_5, X_4, X_1, X_2, X_6, and X_7, and the dependent variable order on the first variate is X_{10}, then X_9. Similar rankings can be found for the variates of the second canonical function. Because canonical weights are typically unstable, particularly in instances of multicollinearity, owing to their calculation solely to optimize the canonical correlation, the canonical loading and cross-loadings are considered more appropriate.

Canonical Loadings

Table 8.6 contains the canonical loadings for the dependent and independent variates for both canonical functions. The objective of maximizing the variates for the correlation between them results in variates "optimized" not for interpretation, but instead for prediction. This makes identification of relationships more difficult. In the first dependent variate, both variables have loadings exceeding .90, resulting in the high shared variance (.855). This indicates a high degree of intercorrelation among the two variables and suggests that both, or either, measures are representative of the effects of HATCO's efforts.

The first independent variate has a quite different pattern, with loadings ranging from .061 to .765, with one independent variable (X_7) even having a negative loading, although it is rather small and not of substantive interest. The three variables with the highest loadings on the independent variate are X_5 (overall service), X_1 (delivery speed), and X_3 (price flexibility). This variate does not correspond to the dimensions that would be extracted in factor analysis, but it would not be expected to because the variates in canonical correlation are extracted only to maximize predictive objectives. As such, it should correspond more to the results from other dependence techniques. There is a close correspondence to multiple regression results with X_9 as the dependent variable. Two of these variables (X_3 and X_5) were included in the stepwise regression analysis in which X_9 (one of the two variables in the dependent variate) was the dependent variable. Thus, the first canonical function closely corresponds to the multiple regression results, with the independent variate representing the set of variables best predicting the two dependent measures. The researcher should also perform a sensitivity analysis of the independent variate in this case to see whether the loadings change when an independent variable is deleted (see stage 6).

The second variate's poor redundancy values are exhibited in the substantially lower loadings for both variates on the second function. Thus, the poorer interpretability as reflected in the lower loadings, coupled with the low redundancy values, reinforce the low practical significance of the second function.

Canonical Cross-Loadings

Table 8.6 also includes the cross-loadings for the two canonical functions. In studying the first canonical function, we see that both independent variables (X_9 and X_{10}) exhibit high correlations with the independent canonical variate (function 1): .855 and .877, respectively. This reflects the high shared variance between these two variables. By squaring these terms, we find the percentage of the variance for each of the variables explained by function 1. The results show that 73 percent of the variance in X_9 and 77 percent of the variance in X_{10} is explained by function 1. Looking at the independent variables' cross-loadings, we see that variables X_1 and X_5 both have high correlations of roughly .72 with the dependent canonical variate. From this information, approximately 52 percent of the variance in each of these two variables is explained by the dependent variate (the 52 percent is obtained by squaring the correlation coefficient, .72). The correlation of X_3 (.584) may appear high, but after squaring this correlation, only 34 percent of the variation is included in the canonical variate.

The final issue of interpretation is examining the signs of the cross-loadings. All independent variables except X_7 (product quality) have a positive, direct relationship. For the second function, two independent variables (X_4 and X_6), plus a dependent variable (X_{10}), are negative. The three highest cross-loadings of the first independent variate correspond to the variables with the highest canonical loadings as well. Thus all the relationships are direct except for one inverse relationship in the first function.

Stage 6: Validation and Diagnosis

The last stage should involve a validation of the canonical correlation analyses through one of several procedures. Among the available approaches would be (1) splitting the sample into estimation and validation samples, or (2) sensitivity analysis of the independent variable set. Table 8.7 contains the result of such a sensitivity analysis in which the canonical loadings are examined for stability when individual independent variables are deleted from the analysis. As seen, the canonical loadings in our example are remarkably stable and consistent in each of the three cases where an independent variable (X_1, X_2, or X_7) is deleted. The overall canonical correlations also remain stable. But the researcher examining the canonical weights (not presented in the table) would find widely varying results, depending on which variable was deleted. This reinforces the procedure of using the canonical loading and cross-loading for interpretation purposes.

A Managerial Overview

The canonical correlation analysis addresses two primary objectives: (1) the identification of dimensions among the dependent and independent variables that (2) maximize the relationship between the dimensions. From a managerial perspective, this provides the researcher with some insight into the structure of the different variable sets as they relate to a dependence relationship. First, the results indicate only a single relationship exists, supported by the low practical significance of the second canonical function. In examining this relationship, we first see that the two dependent variables are quite closely related and create a well-defined dimension for representing the outcomes of HATCO's efforts. Second, this outcome dimension is fairly well predicted by the set of independent variables when acting as a set. The redundancy value of .750 would be a quite acceptable R^2 for a comparable multiple regression. When interpreting the independent variate, we see that three variables, X_5 (overall service), X_1 (delivery speed), and X_3 (price flexibility) provide the substantive contributions and thus are the key predictors of the outcome dimension. These should be the focal points in the development of any strategy directed toward impacting the outcomes of HATCO.

Summary

Canonical correlation analysis is a useful and powerful technique for exploring the relationships among multiple dependent and independent variables. The technique is primarily descriptive, although it may be used for predictive purposes. Results obtained from a canonical analysis should suggest answers to questions concerning the number of ways in which the two sets of multiple variables are related, the strengths of the relationships, and the nature of the relationships defined.

Canonical analysis enables the researcher to combine into a composite measure what otherwise might be an unmanageably large number of bivariate correlations between sets of variables. It is useful for identifying overall relationships between multiple independent and dependent variables, particularly when the data researcher has little a priori knowledge about relationships among the sets of variables. Essentially, the researcher can apply canonical correlation analysis to a set of variables, select those variables (both independent and dependent) that appear to be significantly related, and run subsequent canonical correlations with the more significant variables remaining, or perform individual regressions with these variables.

Questions

1. Under what circumstances would you select canonical correlation analysis instead of multiple regression as the appropriate statistical technique?
2. What three criteria should you use in deciding which canonical functions should be interpreted? Explain the role of each.
3. How would you interpret a canonical correlation analysis?
4. What is the relationship among the canonical root, the redundancy index, and multiple regression's R^2?
5. What are the limitations associated with canonical correlation analysis?
6. Why has canonical correlation analysis been used much less frequently than the other multivariate techniques?

References

1. Alpert, Mark I., and Robert A. Peterson (1972), "On the Interpretation of Canonical Analysis." *Journal of Marketing Research* 9 (May): 187.
2. Alpert, Mark I., Robert A. Peterson, and Warren S. Marti (1975), "Testing the Significance of Canonical Correlations." *Proceedings, American Marketing Association* 37: 117–19.
3. Bartlett M. S. (1941), "The Statistical Significance of Canonical Correlations." *Biometrika* 32: 29.
4. Dillon, W. R., and M. Goldstein (1984), *Multivariate Analysis: Methods and Applications*. New York: Wiley.
5. Green, P. E. (1978), *Analyzing Multivariate Data*. Hinsdale, Ill.: Holt, Rinehart, & Winston.

6. Green, P. E., and J. Douglas Carroll (1978), *Mathematical Tools for Applied Multivariate Analysis*. New York: Academic Press.
7. Lambert, Z., and R. Durand (1975), “Some Precautions in Using Canonical Analysis.” *Journal of Marketing Research* 12 (November): 468–75.
8. Stewart, Douglas, and William Love (1968), “A General Canonical Correlation Index.” *Psychological Bulletin* 70: 160–63.

Annotated Articles

The following annotated articles are provided as illustrations of the application of canonical correlation analysis to substantive research questions of both a conceptual and managerial nature. The reader is encouraged to review the complete articles for greater detail on any of the specific issues regarding methodology or findings.

Schul, Patrick L., William M. Pride, and Taylor L. Little (1983), “The Impact of Channel Leadership Behavior on Intrachannel Conflict.” *Journal of Marketing* 47(3): 21–34.

This article uses canonical correlation analysis, a technique that allows for the investigation of multiple independent variable effects upon multiple dependent variables. In this article, the multivariate technique is applied to determine the effects of leadership style upon perceptions of intrachannel conflict. By examining the strength and direction of the relationship between leadership style and overall intrachannel conflict, canonical correlation analysis provides the researcher with information to improve distribution channel transactions. Findings of this nature also provide managers of channel member organizations with a means of structuring their conduct with other channel members. Three leadership styles—participative, directive, and supportive—which have been theorized to exhibit an inverse relationship with two forms of intrachannel conflict—administrative and product-service—are defined and measured. It is through this technique that researchers are able to explore the relationship between such multifaceted constructs as leadership and conflict.

Data from a sample of 349 franchised real estate brokers were analyzed to test for an association between channel leadership (as the predictor, or independent variables) and intrachannel conflict (as the criterion, or dependent variables). The authors are able to confirm that there exists a strong relationship between the type of channel leadership and intrachannel conflict by examining the redundancy index (indicates the amount of variance explained in a canonical variate by the other canonical variate) and the correlation between the two variates. Through individual analysis, all three leadership styles are shown to reduce conflict. Participative and supportive techniques facilitate understanding and acceptance of policies and procedures, whereas a directive style reduces role ambiguity. From the large cross-loading values, the results indicate that supportive leadership has a stronger relationship with reducing channel conflict than participative or directive leadership. For validation purposes, the authors run a canonical correlation analysis on an analysis and hold-out sample. Canonical weights are compared across the two samples, thereby providing an indication of the stability of this measure for the combined sample. These results indicate that a channel leader should implement a style that best fits the needs of the franchisee in order to minimize channel conflict.

Luthans, Fred, Dianne H. B. Welsh, and Lewis A. Taylor III (1988), “A Descriptive Model of Managerial Effectiveness.” *Group and Organization Studies* 13(2): 148–62.

By conducting a canonical correlation analysis, this study seeks to determine which specific managerial activities relate to organizational effectiveness. The authors identify nine managerial activities and eight items on organizational subunit effectiveness. The activities of 78 managers are observed and recorded in order to measure engagement in the identified behaviors. To eliminate same-source bias that may be introduced if the managers rate subunit effectiveness, their subordinates (278 in all) rate subunit effectiveness. Whereas other studies examine the activities of successful managers (i.e., those on a fast promotion track), this study seeks to identify those behaviors that contribute to organizational effectiveness. The canonical correlation analysis between the frequency of the managerial activities and subordinate-reported subunit effectiveness is used in order to reveal the presence and strength of the relationship between the two sets of variables.

Results indicate a significant canonical variate (canonical correlation = .44); however, the strength of the relationship is not assessed (i.e., the redundancy index, which is a better measure of the ability of the predictor variables to explain variation in the criterion variables, is not reported). Interpretation of the results suggests a continuum of management orientation from quantity-oriented human resources to quality-oriented traditional. Quantity-oriented human resource managers are understood to focus on staffing and motivating or reinforcing activities and are perceived as having quantity performance in their units. These managers, however, have limited outside interaction, engagement in controlling and planning activities, or the perception of quality performance in their units. On the other hand, quality-oriented traditional managers are understood as having quality performance in their units, interacting with outsiders, and engaging in controlling and planning activities. Although the findings are not validated, the authors maintain that the results should assist organizational planners in identifying the necessary managerial skills for the desired organizational outcome (i.e., human resource activities may aid in the attainment of more output whereas traditional management activities may improve quality).

Van Auken, Howard E., B. Michael Doran, and Kil-Jhin Yoon (1993), "A Financial Comparison between Korean and U.S. Firms: A Cross–Balance Sheet Canonical Correlation Analysis." ***Journal of Small Business Management*** **31(3): 73–83.**

In this article, the authors seek to examine cross–balance sheet relationships and general financing strategies of small- to medium-sized Korean firms through the use of canonical correlation analysis. From a random sample of 45 Korean firms, financial position statements from 1988 were obtained for various asset, liability, and equity accounts. The relationships between the assets (cash, accounts receivable, inventories, and long-term assets) and liabilities and equities (accounts payable, other current liabilities, long-term debt, and equity) were explored using canonical correlation analysis. The study's design allowed the authors to compare the results of previously published works on Mexican and U.S. small- and medium-sized firms. This and previous studies have demonstrated that the financial strategies of small firms are influenced by economic conditions and cultural elements. The results of the study should aid in the decision making of small business owners who are developing financing strategies in similar economies.

The analysis resulted in all four canonical functions being significant. As a further assessment of the canonical functions, the authors calculate a redundancy index. The proportion of the asset variance accounted for by the liability variance is .60. The liability variance shared with the asset variance is .24. Because the canonical relationships were acceptable, the authors proceed to make the following interpretations about small- to medium-sized Korean firms: (1) they experience hedging, (2) they use collateral for loans, (3) their inventories are associated with accounts payable, and (4) they manage risk with the simultaneous use of lower leverage and greater liquidity balances. Compared to U.S. firms, Korean firms rely heavily on the use of current debt. The findings extend earlier studies, which suggest that the financial strategies of small firms in other countries are dependent on the marketing constraints of the country in which the firm operates.

Mahmood, Mo Adam, and Gary J. Mann (1993), "Measuring the Organizational Impact of Information Technology Investment: An Exploratory Study." ***Journal of Management Information Systems*** **10(1): 97–122.**

This article seeks to determine whether a relationship exists between information technology (IT) investment and the strategic and economic performance of the firm. From past research measuring the impact of IT on the organization, the authors determine which measures to include in the study. The predictor variables consist of five IT investment measures, and the criterion variables include six organizational strategic and economic performance measures. Canonical analysis is used for exploratory purposes with no specific hypotheses offered. The technique enables the researchers to determine the presence and magnitude of the association between multiple IT and organizational performance measures. From the results, the authors offer hypotheses and a model depicting the interrelationships between IT investment and performance.

The canonical correlation analysis is performed with a sample of 100 firms. The results indicate a significant relationship with 10.4 percent of the variation in the organizational performance measures explained by IT investment. Although only one of the five canonical functions is significant, the authors interpret the two functions that account for nearly 86 percent of the total explained variation (i.e., of the 10.4 percent). By examining the canonical loadings of the two functions, the authors are able to determine the relative importance of each variable. Altogether, the findings are interpreted as indicating that IT investment contributes to organizational performance when the firm invests in both equipment and employee IT training. These conclusions lead the authors to call for further research to test the interdependencies between IT investment and firm performance using different methods and samples.

Exhibit 1
Description of the HATCO Survey

The HATCO survey was a study conducted with 100 customers on 14 separate variables in a segmentation study for a business-to-business situation, specifically a survey of existing customers of HATCO. Three types of information were collected. The first type is the perception of HATCO on seven attributes identified in past studies as the most influential in the choice of suppliers. The respondents, purchasing managers of firms buying from HATCO, rated HATCO on each attribute. The second type of information relates to actual purchase outcomes, either the evaluations of each respondent's satisfaction with HATCO or the percentage of that respondent's product purchases from HATCO. The third type of information contains general characteristics of the purchasing companies (e.g., firm size, industry type). A brief description of the variables is provided below. A complete dataset of all 100 observations is also available at the *Multivariate Data Analysis* website (www.mvstats.com) under materials for the fifth edition.

Perceptions of HATCO

Each of the variables was measured on a graphic rating scale, where a 10-centimeter line was drawn between the endpoints, labeled "Poor" and "Excellent."

Respondents indicated their perceptions by making a mark anywhere on the line. The mark was then measured and the distance from 0 (in centimeters) was recorded. The result was a scale ranging from 0 to 10, rounded to a single decimal place. The seven HATCO attributes rated by each respondent are as follows:

- X_1 Delivery speed—amount of time it takes to deliver the product once an order has been confirmed
- X_2 Price level—perceived level of price charged by product suppliers
- X_3 Price flexibility—perceived willingness of HATCO representatives to negotiate price on all types of purchases
- X_4 Manufacturer's image—overall image of the manufacturer or supplier
- X_5 Overall service—overall level of service necessary for maintaining a satisfactory relationship between supplier and purchaser
- X_6 Salesforce image—overall image of the manufacturer's salesforce
- X_7 Product quality—perceived level of quality of a particular product (e.g., performance or yield)

Purchase Outcomes

Two specific measures were obtained that reflected the outcomes of the respondent's purchase relationships with HATCO. These measures include:

- X_9 Usage level—how much of the firm's total product is purchased from HATCO, measured on a 100-point percentage scale, ranging from 0 to 100 percent
- X_{10} Satisfaction level—how satisfied the purchaser is with past purchases from HATCO, measured on the same graphic rating scale as perceptions X_1 to X_7

Purchaser Characteristics

The five characteristics of the responding firms used in the study, some metric and some nonmetric, are as follows:

- X_8 Size of firm—size of the firm relative to others in this market. This variable has two categories: 1=large, 0=small
- X_{11} Specification buying—extent to which a particular purchaser evaluates each purchase separately (total value analysis) versus the use of specification buying, which details precisely the product characteristics desired. This variable has two categories: 1=employs total value analysis approach, evaluating each purchase separately; 0=use of specification buying
- X_{12} Structure of procurement—method of procuring or purchasing products within a particular company. This variable has two categories: 1=centralized procurement, 0=decentralized procurement
- X_{13} Type of industry—industry classification in which a product purchaser belongs. This variable has two categories: 1=industry A, 0=other industries
- X_{14} Type of buying situation—type of situation facing the purchaser. This variable has three categories: 1=new task, 2=modified rebuy, 3=straight rebuy

TABLE 1
Canonical Correlation of Credit Usage (Number of Credit Cards and Usage Rate) with Customer Characteristics (Family Size and Family Income)

Survey Variables

Measures of Credit Usage	Measures of Customer Characteristics
• Number of credit cards held	• Family size
• Average monthly dollar expenditures on credit cards	• Family income

Canonical Analysis Elements

Composite of Dependent Variables	Canonical Correlation	Composite of Independent Variables
Dependent canonical variate	R_c	*Independent canonical variate*

Adapted from Chapter 8, *Multivariate Data Analysis, 5th edition* by Joseph F. Hair, Jr., Rolph E. Anderson, Ronald L. Tatham and William C. Black.

TABLE 2
Canonical Correlation Analysis Relating Levels of Usage and Satisfaction with HATCO to Perceptions of HATCO

Measures of Overall Model Fit for Canonical Correlation Analysis

Canonical Function	Canonical Correlation	Canonical R^2	F Statistic	Probability
1	.937	.878	30.235	.000
2	.510	.260	5.391	.000

Multivariate Tests of Significance

Statistic	Value	Approximate F Statistic	Probability
Wilks' lambda	.090	30.235	.000
Pillai's trace	1.138	17.348	.000
Hotelling's trace	7.535	48.441	.000
Roy's gcr	.878		

TABLE 3
Calculation of the Redundancy Indices for the First Canonical Function

Variate/Variables	Canonical Loading	Canonical Loading Squared	Average Loading Squared	Canonical R^2	Reundancy Index[a]
Dependent variables					
X_9 Usage level	.913	.834			
X_{10} Satisfaction level	.936	.876			
Dependent variate		1.710	.855	.878	.751
Independent variables					
X_1 Delivery speed	.764	.584			
X_2 Price level	.061	.004			
X_3 Price flexibility	.624	.389			
X_4 Manufacturer image	.414	.171			
X_5 Overall service	.765	.585			
X_6 Salesforce image	.348	.121			
X_7 Product quality	-.278	.077			
Independent variate		1.931	.276	.878	.242

[a] The redundancy index is calculated as the average loading squared times the canonical R^2.

TABLE 4
Redundancy Analysis of Dependent and Independent Variates for Both Canonical Functions

Standardized Variance of the Dependent Variables Explained by

	Their Own Canonical Variate (Shared Variance)		The Opposite Canonical Variate (Redundancy)		
Canonical Function	Percentage	Cumulative Percentage	Canonical R^2	Percentage	Cumulative Percentage
1	.855	.855	.878	.751	.751
2	.145	1.000	.260	.038	.789

Standardized Variance of the Independent Variables Explained by

	Their Own Canonical Value (Shared Variance)		The Opposite Canonical Variate (Redundancy)		
Canonical Function	Percentage	Cumulative Percentage	Canonical R^2	Percentage	Cumulative Percentage
1	.276	.276	.878	.242	.242
2	.082	.358	.260	.021	.263

TABLE 5
Canonical Weights for the Two Canonical Functions

	Canonical Weights	
	Function *1*	Function *2*
Standardized canonical coefficients for the independent variables		
X_1 Delivery speed	.225	2.965
X_2 Price level	.103	2.868
X_3 Price flexibility	.569	.160
X_4 Manufacturer image	.348	-1.456
X_5 Overall service	.445	1.530
X_6 Salesforce image	-.051	.736
X_7 Product quality	.001	.478
Standardized canonical coefficients for the dependent variables		
X_9 Usage level	.501	1.330
X_{10} Satisfaction level	.580	-1.298

TABLE 6
Canonical Structure for the Two Canonical Functions

	Canonical Loadings	
	Function *1*	Function *2*
Correlations between the independent variables and their canonical variates		
X_1 Delivery speed	.764	.109
X_2 Price level	.061	.141
X_3 Price flexibility	.624	.123
X_4 Manufacturer image	.414	-.626
X_5 Overall service	.765	.222
X_6 Salesforce image	.348	-.199
X_7 Product quality	-.278	.219
Correlations between the dependent variables and their canonical variates		
X_9 Usage level	.913	.408
X_{10} Satisfaction level	.936	-.352

	Canonical Cross-Loadings[a]	
	Function *1*	Function *2*
Correlations between the independent variables and dependent canonical variates		
X_1 Delivery speed	.716	.056
X_2 Price level	.058	.072
X_3 Price flexibility	.584	.063
X_4 Manufacturer image	.388	-.319
X_5 Overall service	.717	.113
X_6 Salesforce image	.326	-.102
X_7 Product quality	-.261	.112
Correlations between the dependent variables and independent canonical variates		
X_9 Usage level	.855	.208
X_{10} Satisfaction level	.877	-.180

[a]The canonical cross-loadings are provided by SAS because SPSS does not report the cross-loadings.

TABLE 7
Sensitivity Analysis of the Canonical Correlation Results to Removal of an Independent Variable

		Results after Deletion of		
	Complete Variate	X_1	X_2	X_7
Canonical correlation (R)	.937	.936	.937	.937
Canonical root (R^2)	.878	.876	.878	.878
Independent Variate				
Canonical loadings				
X_1 Delivery speed	.764	omitted	.765	.764
X_2 Price level	.061	.062	omitted	.061
X_3 Price flexibility	.624	.624	.624	.624
X_4 Manufacturer image	.414	.413	.414	.415
X_5 Overall service	.765	.766	.766	.765
X_6 Salesforce image	.348	.348	.348	.348
X_7 Product quality	-.278	-.278	-.278	omitted
Shared variance	.276	.225	.322	.309
Redundancy	.242	.197	.282	.271
Dependent Variate				
Canonical loadings				
X_9 Usage level	.913	.915	.914	.913
X_{10} Satisfaction level	.936	.934	.935	.936
Shared variance	.855	.855	.855	.855
Redundancy	.750	.749	.750	.750