CHAPTER SEVEN

COVARIANCE STRUCTURE ANALYSIS

7.1 INTRODUCTION

In this chapter I model canonical variate analysis (CVA) with longitudinal data using covariance structure analysis (COSAN) (McDonald 1978, 1980). If one assumes common canonical variates, then multivariate data with group structure imply a certain covariance structure. COSAN models this implied structure. An advantage of modeling with COSAN is that SAS software exists for analyzing it, obviating the need to program new algorithms.

I begin Chapter Seven with a description of the COSAN model in Section **7.2**. In Section **7.3** I express CVA with longitudinal data as a covariance structure. In Section **7.4** I show how CVA over time is parameterized in the COSAN framework. Lastly, in Section **7.5** I show a limited example based on the Shenandoah study previously described in Section **5.4.**

7.2 COVARIANCE STRUCTURE ANALYSIS

Covariance Structure Analysis (McDonald 1978, 1980, SAS 1990) is a model for analyzing positive definite or semidefinite matrices. Most commonly known models for analyzing covariance structures can be presented as special cases of COSAN, such as principal components analysis, confirmatory and exploratory factor analysis, and LISREL (Linear Structural RELations, Jöreskog 1989).

The general form of the COSAN model is

$$\mathbf{C} = \mathbf{F}_1\mathbf{P}_1\mathbf{F}_1' + \cdots + \mathbf{F}_m\mathbf{P}_m\mathbf{F}_m' \qquad (7.1)$$

where $\mathbf{C}$ is a symmetric positive definite or semi-definite matrix; each $\mathbf{F}_k$, $k = 1,\ldots,m$, is the product of s_k matrices, $\mathbf{F}_k = \mathbf{F}_{k1}\ldots\mathbf{F}_{ks_k}$; and each matrix $\mathbf{P}_k$ is symmetric. The matrices $\mathbf{P}_k$ can be of the form of an inverse of a matrix $\mathbf{H}_k$, that is, $\mathbf{P}_k = \mathbf{H}_k^{-1}$. The matrices $\mathbf{F}_{ki}$ above can be of the form of an inverse of a matrix $\mathbf{H}_{ki}$, or of the inverse of an identity matrix minus $\mathbf{H}_{ki}$; that is, $\mathbf{F}_{ki}$ can be of the form $\mathbf{F}_{ki} = \mathbf{H}_{ki}^{-1}$ or $\mathbf{F}_{ki} = (\mathbf{I} - \mathbf{H}_{ki})^{-1}$. Furthermore, a matrix can contain both parameter terms and constant terms. A parameter can be constrained to be a function of other parameters. Hence COSAN is a flexible model for analyzing covariance structures.

The general idea behind COSAN is that covariance structures can often be modeled as the crossproduct of matrices whose columns consist of the weights of components or the loadings of factors. Consider for example the factor analysis model, $\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$. $\boldsymbol{\Sigma}$ is modeled as the crossproduct matrix of $\mathbf{L}$, the matrix of factor loadings, (plus a diagonal matrix of specific variances $\boldsymbol{\Psi}$). The factor analysis model is expressed in COSAN as $\mathbf{C} = \mathbf{F}_1\mathbf{P}_1\mathbf{F}_1' + \mathbf{F}_2\mathbf{P}_2\mathbf{F}_2'$, where $\mathbf{F}_1$ is the matrix of factor loadings, $\mathbf{P}_1$ and $\mathbf{F}_2$ are restricted to be identity matrices, and $\mathbf{P}_2$ is restricted to be a diagonal matrix of specific variances. One can incorporate more complexity to get LISREL models by modeling the $\mathbf{F}_k$ to be the product of matrices $\mathbf{F}_k = \mathbf{F}_{k1}\ldots\mathbf{F}_{ks_k}$, and where appropriate by constraining these $\mathbf{F}_{ki}$ to be either the inverse of a matrix of parameter and constants, or the identity matrix minus the inverse of a matrix of parameter and constants (see Jöreskog 1989).

In COSAN restricting a matrix to be orthogonal is done indirectly with the Cayley (McDonald 1978) decomposition. For example, if one wants to restrict a matrix of parameters and constants $\mathbf{L}$ to be orthogonal, then set $\mathbf{L} = (\mathbf{I} - \mathbf{H}')^{-1}(\mathbf{I} - \mathbf{H})$, where $\mathbf{H}$ is skew symmetric with zeros as its diagonal elements. (Skew symmetric means $\mathbf{H}$ is parameterized such that $\mathbf{H} = -\mathbf{H}'$).

For an example of modeling orthogonal matrices consider the principal components model. This is parameterized as

$$\mathbf{C} = (\mathbf{I} - \mathbf{H}')^{-1}(\mathbf{I} - \mathbf{H})\mathbf{P}(\mathbf{I} - \mathbf{H})'(\mathbf{I} - \mathbf{H}')^{-1'}.$$

$\mathbf{P}$ is a $p \times p$ diagonal matrix whose elements are the squares of the eigenvalues associated with the principal components. The $p \times p$ orthogonal matrix of principal components $\mathbf{V}$ is found as $\mathbf{V} = (\mathbf{I} - \mathbf{H}')^{-1}(\mathbf{I} - \mathbf{H})$, where $\mathbf{H}$ is a $p \times p$ skew symmetric matrix. In terms of the model given in (7.1), $\mathbf{C} = \mathbf{F}_{11}\mathbf{F}_{12}\mathbf{P}\mathbf{F}_{12}'\mathbf{F}_{11}'$, where $\mathbf{F}_{11}$ is the inverse of the identity minus $\mathbf{H}'$, and $\mathbf{F}_{12}$ is the identity minus **H**.

The COSAN model can be estimated using several different fit functions. These include maximum likelihood if one assumes the data follow a multivariate normal distribution. Other fit functions include unweighted least squares and generalized least squares. Estimates and test statistics can be obtained using SAS's Proc Calis package, which offers a variety of fit functions and convergence algorithms. All of the fit functions mentioned previously can be fit with Proc Calis. Proc Calis also offers the user the choice of the several optimization techniques. These include conjugate-gradient techniques, the Marquardt technique, and Newton-Raphson techniques. Furthermore, parameter constraints can be specified using SAS programming statements.

7.3 MODELING CANONICAL VARIATE ANALYSIS OVER TIME AS A COVARIANCE STRUCTURE

The CVA over time model that I present in this section is more akin to an extension of RA than of CVA, as the canonical variates are orthogonal in their weights as opposed to being uncorrelated. Essentially, I will model the between-groups covariance matrix. Start with the standard (non-longitudinal) case by performing a spectral decomposition of the between-groups covariance matrix. Let the between-groups covariance matrix for *p* Y-variables be denoted as CovB(**Y**), and let **V** denote a $p \times r$ columnwise orthonormal matrix of variate weights. Then

$$\text{CovB}(\mathbf{Y}) = \mathbf{V}\mathbf{D}^2\mathbf{V}',$$

where **D** is an $r \times r$ diagonal matrix whose i^{th} diagonal element is the square root of the between-groups variation of the i^{th} column of **V**, $\mathbf{v}_{\text{i}}$.

Now consider the multiple occasions case. (As a reminder, **X** and **Y** are assumed to be centered). The products matrix for $\mathbf{Y}_{\text{i}}$ and $\mathbf{Y}_{\text{j}}$ regressed on **X** is $\mathbf{Y}_{\text{i}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_{\text{j}}$, where $\mathbf{Y}_{\text{i}}$ is the Y-data at the i^{th} occasion. But $(n-1)^{-1/2}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'\mathbf{Y}_{\text{j}} = \mathbf{W}^*\mathbf{D}_{\text{j}}\mathbf{V}'$, where $\mathbf{W}^*$, and **V** are the redundancy variates for the X-variables and Y-variables as defined in Section **2.2.4**, and $\mathbf{D}_{\text{j}}$ is a diagonal matrix whose i^{th} diagonal element is the square root of the variance explained by the i^{th} variate; see equation (2.2). Now if one assumes that one has common variates at each of *g* occasions, then

$$(n-1)^{-1}\mathbf{Y}_{\text{i}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_{\text{j}} = \mathbf{V}\mathbf{D}_{\text{i}}\mathbf{D}_{\text{j}}\mathbf{V}'. \quad (7.2)$$

But (7.2) implies

$$\text{CovB}(\mathbf{Y}_1 : \mathbf{Y}_2 : \cdots : \mathbf{Y}_g) = \begin{bmatrix} \mathbf{V}\mathbf{D}_1^2\mathbf{V}' & \mathbf{V}\mathbf{D}_1\mathbf{D}_2\mathbf{V}' & \cdots & \mathbf{V}\mathbf{D}_1\mathbf{D}_g\mathbf{V}' \\ \mathbf{V}\mathbf{D}_2\mathbf{D}_1\mathbf{V}' & \mathbf{V}\mathbf{D}_2^2\mathbf{V}' & \cdots & \mathbf{V}\mathbf{D}_2\mathbf{D}_g\mathbf{V}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{V}\mathbf{D}_g\mathbf{D}_1\mathbf{V}' & \mathbf{V}\mathbf{D}_g\mathbf{D}_2\mathbf{V}' & \cdots & \mathbf{V}\mathbf{D}_g^2\mathbf{V}' \end{bmatrix}, \quad (7.3)$$

where $\text{CovB}(\mathbf{Y}_1 : \mathbf{Y}_2 : \cdots : \mathbf{Y}_g)$ indicates the between-groups covariance matrix for the p variables over g occasions. This between-groups covariance matrix follows a non-central, deficient Wishart distribution. Thus maximum likelihood estimates are not readily obtained. However, (7.3) can be estimated by the method of least squares.

In order to estimate with the method of maximum likelihood one must include the within-groups covariance matrix in the model, as the between-groups covariance matrix plus the within-groups covariance matrix yield the overall covariance structure, which does follow a Wishart distribution. A plausible assumption is that the within-groups covariance matrices at all occasions are proportional to **E**, where **E** is a $p \times p$ positive definite matrix. Then $\text{CovW}(\mathbf{Y}) = \mathbf{A} \otimes \mathbf{E}$, where $\text{CovW}(\mathbf{Y})$ is the $pg \times pg$ within-groups covariance matrix, and **A** is a $g \times g$ positive semi-definite matrix scaled such that $\text{trace}(\mathbf{A}) = g$.

The approach outlined above does not model **W**, the matrix of coefficients for the X-variables (which are group indicators). If one desires an estimate of **W**, a reasonable approach is to find **W** which minimizes the sums of squares fit to $\mathbf{S}_{\text{XYk}} = \mathbf{W}\mathbf{D}_\text{k}\mathbf{V}'$, for $\text{k} = 1, \ldots, g$. This is equivalent to a step in the alternating least squares algorithm for the PARAFAC (orth.) model.

A more complicated way to obtain **W** is to model the $\mathbf{S}_{\text{XY}}$ and $\mathbf{S}_{\text{XX}}$ matrices along with the $\mathbf{S}_{\text{YY}}$ matrices. By definition $\mathbf{S}_{\text{XX}} = \mathbf{W}^{-1'}\mathbf{W}^{-1}$ since $\mathbf{W}'\mathbf{S}_{\text{XX}}\mathbf{W} = \mathbf{I}$ and $\mathbf{S}_{\text{XY1}} = \mathbf{W}^{-1}\mathbf{D}_1\mathbf{V}'$ since $\mathbf{W}'\mathbf{S}_{\text{XY1}}\mathbf{V} = \mathbf{D}_1$. Hence the resulting model is

$$\begin{bmatrix} \mathbf{S}_{\text{XX}} & \mathbf{S}_{\text{XY1}} & \mathbf{S}_{\text{XY2}} & \cdots & \mathbf{S}_{\text{XYg}} \\ \mathbf{S}_{\text{Y1X}} & \mathbf{S}_{\text{Y1Y1}} & \mathbf{S}_{\text{Y1Y2}} & \cdots & \mathbf{S}_{\text{Y1Yg}} \\ \mathbf{S}_{\text{Y2X}} & \mathbf{S}_{\text{Y2Y1}} & \mathbf{S}_{\text{Y2Y2}} & \cdots & \mathbf{S}_{\text{Y2Yg}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{\text{YgX}} & \mathbf{S}_{\text{YgY1}} & \mathbf{S}_{\text{YgY2}} & \cdots & \mathbf{S}_{\text{YgYg}} \end{bmatrix} = \begin{bmatrix} \mathbf{W}^{-1'}\mathbf{W}^{-1} & \mathbf{W}^{-1'}\mathbf{D}_1\mathbf{V}' & \mathbf{W}^{-1'}\mathbf{D}_2\mathbf{V}' & \cdots & \mathbf{W}^{-1'}\mathbf{D}_\text{k}\mathbf{V}' \\ \mathbf{W}^{-1'}\mathbf{D}_1\mathbf{V}' & \mathbf{V}\mathbf{D}_1^2\mathbf{V}' & \mathbf{V}\mathbf{D}_1\mathbf{D}_2\mathbf{V}' & \cdots & \mathbf{V}\mathbf{D}_1\mathbf{D}_\text{k}\mathbf{V}' \\ \mathbf{W}^{-1'}\mathbf{D}_2\mathbf{V}' & \mathbf{V}\mathbf{D}_2\mathbf{D}_1\mathbf{V}' & \mathbf{V}\mathbf{D}_2^2\mathbf{V}' & \cdots & \mathbf{V}\mathbf{D}_2\mathbf{D}_\text{k}\mathbf{V}' \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}^{-1'}\mathbf{D}_\text{k}\mathbf{V}' & \mathbf{V}\mathbf{D}_\text{k}\mathbf{D}_1\mathbf{V}' & \mathbf{V}\mathbf{D}_\text{k}\mathbf{D}_2\mathbf{V}' & \cdots & \mathbf{V}\mathbf{D}_\text{k}^2\mathbf{V}' \end{bmatrix}.$$

7.4 PUTTING CVA OVER TIME IN THE COSAN FRAMEWORK

Next I show how the CVA over time model is expressed in terms of the $\mathbf{F}_\text{i}$ and $\mathbf{P}_\text{i}$ matrices of (7.1). The model for the between-groups covariance matrix in (7.3) can be decomposed as in (7.4):

$$\begin{bmatrix} \mathbf{V} & & & \\ & \mathbf{V} & & \\ & & \ddots & \\ & & & \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 & & & \\ & \mathbf{D}_2 & & \\ & & \ddots & \\ & & & \mathbf{D}_\text{k} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 & & & \\ & \mathbf{D}_2 & & \\ & & \ddots & \\ & & & \mathbf{D}_\text{k} \end{bmatrix} \begin{bmatrix} \mathbf{V} & & & \\ & \mathbf{V} & & \\ & & \ddots & \\ & & & \mathbf{V} \end{bmatrix}'. \qquad (7.4)$$

Note that the rank of the center matrix and thus the rank of the product of the matrices is r, where r is the number of common variates, $r \leq \min(p, g - 1)$. Now, the second and fourth matrices in

(7.4), that is, the ones with block diagonals of $\mathbf{D}_1$ to $\mathbf{D}_g$, and the center matrix can be expressed directly in the COSAN model. For example, one could label the matrix of block diagonals of $\mathbf{D}_1$ to $\mathbf{D}_g$ as $\mathbf{F}_{11}$, and the center matrix as $\mathbf{P}_1$. However, **V** must be modeled indirectly, using the Cayley decomposition to constrain it to orthogonality. Hence let the matrix $\mathbf{I}_{(k)} \otimes \mathbf{V}$ be $\mathbf{I}_{(k)} \otimes (\mathbf{I}_{(p)} + \mathbf{H})^{-1} \mathbf{I}_{(k)} \otimes (\mathbf{I}_{(p)} - \mathbf{H})$. Let $\mathbf{I}_{(k)} \otimes (\mathbf{I}_{(p)} + \mathbf{H})^{-1}$ be labeled $\mathbf{F}_{12}$ and let $\mathbf{I}_{(k)} \otimes (\mathbf{I}_{(p)} - \mathbf{H})$ be labeled $\mathbf{F}_{13}$. Thus the between groups covariance is modeled as

$$\mathbf{F}_{12}\mathbf{F}_{13}\mathbf{F}_{11}\mathbf{P}_1\mathbf{F}_{11}'\mathbf{F}_{13}'\mathbf{F}_{12}'.$$

To see how the error matrix, $\text{CovW}(\mathbf{Y}) = \mathbf{A} \otimes \mathbf{E}$, would be expressed in the terms of the $\mathbf{F}_i$ and $\mathbf{P}_i$ matrices of (7.1), notice that:

$$\mathbf{A}\otimes\mathbf{E} = \begin{bmatrix} \mathbf{L} & & & \\ & \mathbf{L} & & \\ & & \ddots & \\ & & & \mathbf{L} \end{bmatrix} \begin{bmatrix} a_{11}\mathbf{I}(p) & a_{12}\mathbf{I}(p) & \cdots & a_{1g}\mathbf{I}(p) \\ a_{21}\mathbf{I}(p) & a_{22}\mathbf{I}(p) & \cdots & a_{2g}\mathbf{I}(p) \\ \vdots & \vdots & \ddots & \vdots \\ a_{g1}\mathbf{I}(p) & a_{g2}\mathbf{I}(p) & \cdots & a_{gg}\mathbf{I}(p) \end{bmatrix} \begin{bmatrix} \mathbf{L}' & & & \\ & \mathbf{L}' & & \\ & & \ddots & \\ & & & \mathbf{L}' \end{bmatrix}$$

where **L** is a $p \times p$ matrix. Thus **E** is not directly modeled by COSAN, but is determined as $\mathbf{E} = \mathbf{L}\mathbf{L}'$. One can call the matrix of block diagonals of **L** matrices $\mathbf{F}_{21}$, and the center matrix $\mathbf{P}_2$. These terms are added to those that model the between groups variation. Thus the model for the total covariance is the between groups covariance plus the within groups covariance:

$$\mathbf{F}_{12}\mathbf{F}_{13}\mathbf{F}_{11}\mathbf{P}_1\mathbf{F}_{11}'\mathbf{F}_{13}'\mathbf{F}_{12}' + \mathbf{F}_{21}\mathbf{P}_2\mathbf{F}_{21}'.$$

It is necessary to mention a further detail. In the CVA over time model one will usually desire a solution with fewer variates than *p*, say *r*. Thus one would like to model orthonormal **V**, where **V** has $r < p$ columns. But the Cayley decomposition requires a square matrix. Hence one needs to model dummy canonical variates in the matrix **V**. One fixes these dummy variates by setting certain elements of **V** equal to zero. However, this must be done indirectly because in the COSAN model one directly estimates the matrix of parameters **H**. The details on how to do this are found in Appendix Five.

7.5 AN EXAMPLE

In this section I revisit the data from the Shenandoah study which was described in Section **5.4**. Because of unresolved difficulties in programming in Proc Calis and in executing programs that model large covariance matrices, the example will be limited to modeling one canonical variate over the first two occasions, August and September of 1981. Since the error variance was previously determined to be non-homogeneous (see **Table 5.1**), the between-groups covariance matrix is being modeled with the fit function unweighted least squares. The SAS program is given in Appendix Five.

The weights for the first canonical variate are presented in **Figure 7.1**, along with the first canonical variate of the PARAFAC (orth.) solution for comparison. These weights are similar, as are the estimated roots of the variances explained by each variate, which are 2.47 and 2.65 for

COSAN, in contrast to 2.25 and 2.52 for the PARAFAC (orth.). The similarity in the parameter estimates is evidence that both methods are correctly estimating the common variate, though it will be necessary to estimate a larger COSAN model with data from all the occasions to have convincing evidence of this.

Measurement	COSAN Estimates	PARAFAC (orth.) Estimates
discharge	-0.024	0.036
conductivity	**0.411**	**0.369**
pH	0.046	0.113
temperature	0.097	0.132
Ca^{++}	**0.415**	**0.383**
Mg^{++}	**0.434**	**0.365**
Na^{+}	**0.312**	**0.364**
K^{+}	-0.069	-0.158
alkalinity	**0.467**	**0.394**
$SO_4^{=}$	0.052	0.14
Cl^{-}	**0.220**	**0.265**
$SiO_4^{=}$	**0.293**	**0.363**
NO_3^{-}	0.023	0.133
NH_4^{+}	-0.001	0.035

Figure 7.1 Estimates for the COSAN and PARAFAC Models

7.6 CONCLUDING REMARKS

In summary, COSAN offers a flexible and powerful modeling tool for modeling CVA with longitudinal data. However, work needs to be done to overcome programming and estimation difficulties. One problem is that the SAS system uses up all the available memory when large covariance matrices are analyzed. Another is that when writing code each element of each matrix must be specified, making programming a laborious task for modeling large matrices. This is seen in the SAS code in Appendix Five. A way to solve this problem is to write a macro that sets up the program code.

If the difficulties mentioned in the previous paragraph are overcome, then the model can be further developed. For example, modeling uncorrelated canonical variates would be useful, as would a model that hypothesized that some variates be unique to each occasion.