CHAPTER EIGHT

CANONICAL VARIATE ANALYSIS OVER TIME

8.1 INTRODUCTION

In this chapter I present a model for CVA with measurements made over multiple occasions. In contrasts to earlier chapters, this chapter emphasizes statistical inference based on maximum likelihood methods. I shall call the models developed in this chapter CVA/time, though I distinguish between models with orthogonal canonical variates, CVA/time (orthogonal), and those with uncorrelated canonical variates, CVA/time (uncorrelated). CVA/time is suggested by Campbell and Tomenson`s (1983) model for CVA with multiple datasets, which hypothesizes that the group means lie on planes defined by canonical variates common to all datasets (see Section **2.4**). Analogously, CVA/time hypothesizes that the group means lie on planes defined by canonical variates which are common to all occasions.

The goal of the CVA/time model is to answer the question of what is and what is not changing over time when one has multivariate data with group structure. In particular it attempts to determine if the canonical variates are stable over time, and if they are, if the positions of the group means on the canonical variates are changing over time. Thus CVA/time is the only model in this dissertation that will estimate the group positions and develop hypothesis tests to determine if they are equal over time.

Chapter **Eight** is organized as follows. In Section **8.2** I make two preliminary points. In Section **8.3** I detail a model for group means in the space of orthogonal canonical variates, CVA/time (orthogonal). I also derive estimating equations for this model, discuss their solution and describe how to make statistical inferences. In Section **8.4** I use simulated data to test the methodology of Section **8.3**. In Section **8.5** I derive a model for group means in the space of uncorrelated canonical variates, CVA/time (uncorrelated). Uncorrelated variates entail assuming a particular structure for the within-groups covariance matrix. The estimation of this structure is also discussed in this section. In Section **8.6** I illustrate the methodology of Section **8.5** by analyzing a real dataset. Lastly, in Section **8.7** I compare CVA/time with several alternative methods for this type of data, with particular attention given to doubly multivariate repeated measures.

8.2 PRELIMINARIES

8.2.1 Orthogonal Versus Uncorrelated Variates

This chapter presents two models with differing assumptions about the structure of the group means. The first model to be discussed CVA/time (orth.), hypothesizes that the canonical variates are orthogonal to each other in their weights. It models the positions of the group means in the space of the untransformed data. The second model to be discussed, CVA/time (unc.), hypothesizes that the canonical variates are uncorrelated, which is consistent with the standard definition of canonical variates. It models the positions of the group means in the space transformed by the Mahalanobis transformation.. I shall present the CVA/time (orth.) model first because it is simpler.

The main reason to model orthogonal variates is that, unlike uncorrelated variates, they do not require the assumption that one has the same within-groups covariance structure at each occasion, an assumption which may be unrealistic. Beyond the issue of whether the within-groups covariance matrices are stable over time, there are important differences between the approaches whose implications the researcher needs to consider. These differences are analogous to the differences between canonical variate analysis and redundancy analysis, a topic which is discussed in Section **2.2.3**. Uncorrelated variates have the important advantage that they are scale invariant. They also are more closely related to the goal of optimizing the discrimination among the groups. Though CVA/time (unc.) does not explicitly maximize group discrimination over time (see Section **5.2.2** for something along this line), it is a generalization of CVA, which does. On the other hand, the CVA/time (orth.) model is not a true generalization of CVA, but is more akin to a generalization of redundancy analysis for grouped data (see Section **2.2.3**).

The situations where one may prefer orthogonal variates to uncorrelated variates when uncorrelated variables are feasible are the same as those where one would prefer to perform a redundancy analysis over a canonical correlation analysis. CVA may find group differences which are large in terms of discrimination but small in terms of between-groups variation explained. Or, the total variation explained may be of direct interest. For example, if the measurements made are directly comparable, such as if one had a battery of exams with the same scales, one may prefer to maximize the total variance explained by the group structure.

8.2.2 The Structure of the Data

A clarification of how the data is organized illuminates the discussion of the previous section and other issues not yet touched upon. The same variables measured at different occasions will be treated as distinct variables. Hence *tp* variables are effectively modeled, where *t* is the number of occasions and *p* is the number of variables measured at one occasion. This contrasts with Campbell & Tomenson's method which models distinct datasets of the same *p* variables.

It will be necessary to partition $\mathbf{\Sigma}$, the $tp \times tp$ within-groups covariance matrix, into t^2 $p \times p$ matrices $\mathbf{\Sigma}_{\mathrm{qs}}$, where $\mathbf{\Sigma}_{\mathrm{qs}}$ is the matrix of covariances between the measurements of the q^{th} and s^{th} occasions, $\mathrm{q}, \mathrm{s} = 1, \ldots, t$. The partitioning of $\mathbf{\Sigma}$ is shown below:

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}'_{21} & \cdots & \mathbf{\Sigma}'_{\mathrm{t}1} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \cdots & \mathbf{\Sigma}'_{\mathrm{t}2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{\mathrm{t}1} & \mathbf{\Sigma}_{\mathrm{t}2} & \cdots & \mathbf{\Sigma}_{\mathrm{tt}} \end{bmatrix}. \tag{8.1}$$

The model developed in Section **8.3** assumes no specific structure for $\mathbf{\Sigma}$. A consequence of this flexibility in $\mathbf{\Sigma}$ is that there is no common $p \times p$ within-groups covariance matrix, $\mathbf{\Sigma}_{\mathrm{qq}}$, by which to transform the data. In Section **8.5.2** a model will be introduced which assumes a structure for $\mathbf{\Sigma}$ that specifies common $\mathbf{\Sigma}_{\mathrm{qq}}$ and thus allows for modeling uncorrelated variates.

8.3 THE CVA/TIME (ORTHOGONAL) MODEL

In this section I develop a model for analyzing group structure with longitudinal multivariate data which I shall call CVA/time (orthogonal). CVA/time (orth.) is not a generalization of CVA/time, but rather a generalization of redundancy analysis. Beyond interest in its own right, the discussion of this model introduces basic ideas and methods which will be used later for the CVA/time model with uncorrelated variates. In particular, I introduce the concepts of common and unique variates, group positions, the methods of obtaining estimates, and statistical inference.

The model I develop in this section encompasses several possible cases. One basic model hypothesizes a given number of variates common to all occasions. A simple alternative to this model is one that hypothesizes that there are an equal number of variates specific or unique at each occasion. Henceforth I shall refer to the former as common variates and to the latter as unique variates. Unique variates are the natural alternative to common variates because they hypothesize variates that change over time, and the interest is to determine what is and what is not changing over time.

The model which I will refer to as CVA/time (orth.) hypothesizes both types of variates. However, even more complex models are possible. For example, one could hypothesize variates which are common to only a subset of the groups. Estimating equations can be derived for all of the possible models using the methods of calculus, though I shall derive them only for the CVA/time (orth.) model.

It is useful to give a simple example of a common variate model. Assume the positions of two group means can be plotted on one canonical variate which is common over two occasions, and assume the positions of the group means change over time. **Figure 8.1** shows the positions at the first occasion, and **Figure** Error! Reference source not found. shows the positions at the second occasion.

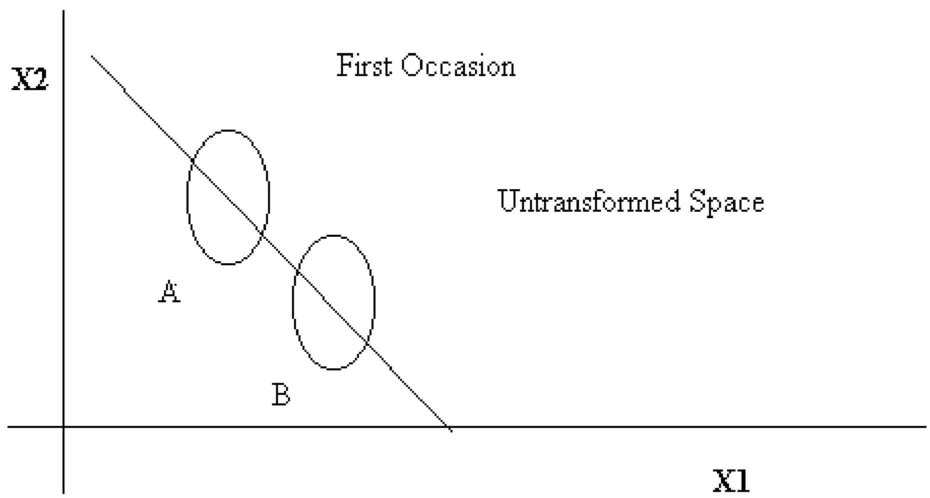

Figure 8.1

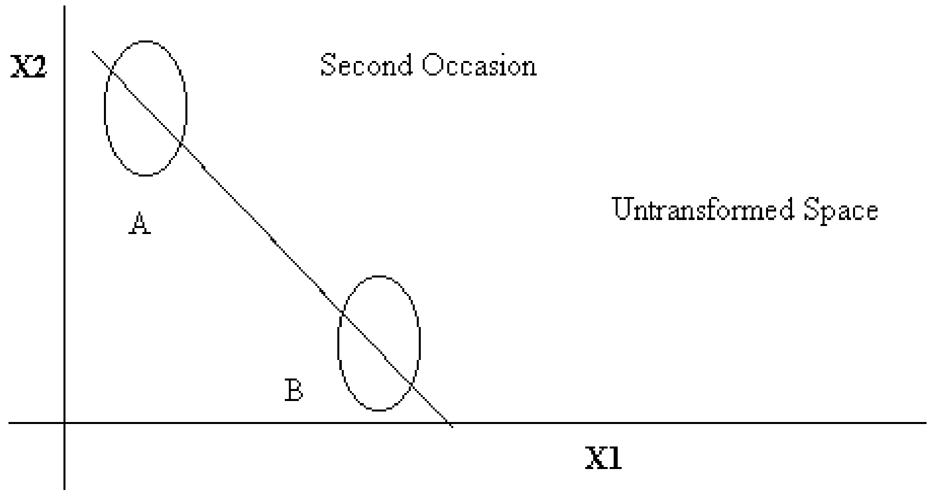

Figure 8.2

8.3.1 The CVA/Time Model with Orthogonal Variates

The CVA/time (orth.) model is specified as follows; assume the data follow the multivariate normal distribution; $\mathbf{x}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, where $\mathbf{x}_i$ is a tp vector of random variables, $\boldsymbol{\mu}_i$ is a tp vector of means, and $\boldsymbol{\Sigma}$ is a $tp \times tp$ covariance matrix. Further assume that $\boldsymbol{\mu}_i$ is completely determined by group membership, so that $\boldsymbol{\mu}_i \in \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_g\}$, depending on the group membership of the i^{th} observation.

The model for the structure of the means given in equation (8.2) below specifies u variates for each occasion q; c of these variates, $\mathbf{v}_1, \ldots, \mathbf{v}_c$, are common to all occasions, where $\mathbf{v}_i$ indicates the i^{th} common variate. $u - c$ of these variates, $\mathbf{v}^q_{c+1}, \ldots, \mathbf{v}^q_u$, are unique to the q^{th} occasion, where $\mathbf{v}^q_i$ indicates the i^{th} variate of the set of variates for the q^{th} occasion. Thus the model for the g^{th} group mean, $\boldsymbol{\mu}_g$, $g = 1, \ldots, m$, is:

$$\boldsymbol{\mu}_g = \boldsymbol{\mu}_0 + \mathbf{v}_1 \otimes \mathbf{e}_{g,1} + \ldots + \mathbf{v}_c \otimes \mathbf{e}_{g,c} + \begin{bmatrix} e^1_{g,c+1}\mathbf{v}^1_{c+1} \\ \vdots \\ e^t_{g,c+1}\mathbf{v}^t_{c+1} \end{bmatrix} + \cdots + \begin{bmatrix} e^1_{g,u}\mathbf{v}^1_u \\ \vdots \\ e^t_{g,u}\mathbf{v}^t_u \end{bmatrix}, \qquad (8.2)$$

where $\boldsymbol{\mu}_g$ is a $pt \times 1$ vector of means for the g^{th} group, $\boldsymbol{\mu}_0$ is a $pt \times 1$ vector of overall means, $\mathbf{v}_i$ are c $p \times 1$ vectors of common variates, $\mathbf{v}^k_j$ are $(u - c)$ $t \times 1$ vectors of unique variates, $e^q_{g,i}$ is the score for the g^{th} group mean on the i^{th} canonical variate at the q^{th} occasion, and $\mathbf{e}_{g,i}$ is the $t \times 1$ vector whose elements are $e^q_{g,i}$.

Note the constraints on the parameters. First, the group positions for each occasion for each variate sum to zero, i.e., $\sum_{g=1}^{m} n_g e^q_{g,i} = 0$ for $q = 1, \ldots, t$, and $i = 1, \ldots, u$. This constraint is just a reflection of the fact that the model is centered by an overall mean, μ_0. Second, the common variates are mutually orthogonal:

$$\mathbf{V}_{\text{com}}'\mathbf{V}_{\text{com}} = \mathbf{I}_{c\times c},$$

where $\mathbf{V}_{\text{com}}$ is the matrix whose columns are the *c* common variates. Furthermore, within each set of unique variates for each occasion the variates are constrained to be mutually orthogonal, that is:

$$\mathbf{V}^{\text{q}\prime}\mathbf{V}^{\text{q}} = \mathbf{I}_{(u-c)\times(u-c)},$$

where $\mathbf{V}^{\text{q}}$ is the matrix whose columns are the $u - c$ unique variates for the q^{th} occasion, $\text{q} = 1,\ldots,t$. Finally, each variate in each set of unique variates is orthogonal with each common variate. Thus:

$$\mathbf{V}_{\text{com}}'\mathbf{V}^{\text{q}} = [0]_{c\times(u\text{-}c)}.$$

Note there is a limit on $u - c$, the number of unique variates one can have at each occasion. $u - c$ cannot be greater than the modular of $\frac{p}{t}$. For example, if $p = t$, then a model can hypothesize at most one unique variate at each occasion. Further, such a model is equivalent to a model with *p* common variates.

8.3.2 Sufficient Statistics

Before proceeding to develop the estimating equations I will show a result for grouped multivariate data that will simplify the later derivations. I will show that $\bar{\mathbf{x}}_{\text{g}}$ and **S**, the sample means and within-groups covariance matrix, are sufficient statistics for μ_{g} and Σ, and consequently for the parameters with which I later model μ_{g} and Σ. The likelihood equations for multivariate grouped data are as follows below, where **X** indicates the data matrix, μ indicates the parameters determining the mean of the variables, Σ indicates the parameters determining the covariance of the variables, $L(\mathbf{X}|\mu,\Sigma)$ indicates the likelihood of the data **X** given parameters μ and Σ, and $\mathbf{x}_{\text{ig}}$ is the i^{th} observation in the g^{th} group:

$$L(\mathbf{X}|\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{npt}{2}}|\Sigma|^{\frac{n}{2}}}\exp\left\{\tfrac{-1}{2}\left(\sum_{\text{g}=1}^{m}\sum_{\text{i}=1}^{n_{\text{g}}}(\mathbf{x}_{\text{gi}} - \mu_{\text{g}})'\Sigma^{-1}(\mathbf{x}_{\text{gi}} - \mu_{\text{g}})\right)\right\}. \tag{8.3}$$

Consider the part of the likelihood equation which is a function of $\mathbf{x}_{\text{gi}}$ and μ_{g} and call it K. Then K is:

$$\begin{aligned}\text{K} &= \sum_{\text{g}=1}^{m}\sum_{\text{i}=1}^{n_{\text{g}}}(\mathbf{x}_{\text{gi}} - \mu_{\text{g}})'\Sigma^{-1}(\mathbf{x}_{\text{gi}} - \mu_{\text{g}}) \\ &= \sum_{\text{g}=1}^{m}\sum_{\text{i}=1}^{n_{\text{g}}}\text{tr}(\Sigma^{-1}(\mathbf{x}_{\text{gi}} - \mu_{\text{g}})(\mathbf{x}_{\text{gi}} - \mu_{\text{g}})')\end{aligned}$$

$$= \sum_{g=1}^{m}\sum_{i=1}^{n_g} \mathrm{tr}(\mathbf{\Sigma}^{-1}((\mathbf{x}_{gi} - \bar{\mathbf{x}}_g) + (\bar{\mathbf{x}}_g - \boldsymbol{\mu}_g))((\mathbf{x}_{gi} - \bar{\mathbf{x}}_g) + (\bar{\mathbf{x}}_g - \boldsymbol{\mu}_g))')$$

Since $\sum_{i=1}^{n_g} \mathrm{tr}(\mathbf{\Sigma}^{-1}(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)(\bar{\mathbf{x}}_g - \boldsymbol{\mu}_g)') = 0$, for $g = 1, \ldots, m$,

$$\mathrm{K} = \sum_{g=1}^{m}\sum_{i=1}^{n_g} \mathrm{tr}(\mathbf{\Sigma}^{-1}(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)' + \mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}_g - \boldsymbol{\mu}_g)(\bar{\mathbf{x}}_g - \boldsymbol{\mu}_g)').$$

Recognizing that $\mathbf{S} = \frac{1}{n}\sum_{g=1}^{m}\sum_{i=1}^{n_g}(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)'$ and further rearrangement gives:

$$\mathrm{K} = \tfrac{1}{n}\mathbf{\Sigma}^{-1}\mathbf{S} + \sum_{g=1}^{m} n_g(\bar{\mathbf{x}}_g - \boldsymbol{\mu}_g)'\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}_g - \boldsymbol{\mu}_g).$$

Replacing K back into the likelihood equation one has the desired result:

$$L(\mathbf{X}|\boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{npt}{2}}|\Sigma|^{\frac{n}{2}}} \exp\left\{-\tfrac{1}{2}\left(\tfrac{1}{n}\mathbf{\Sigma}^{-1}\mathbf{S} + \sum_{g=1}^{m} n_g(\bar{\mathbf{x}}_g - \boldsymbol{\mu}_g)'\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}_g - \boldsymbol{\mu}_g)\right)\right\}. \quad (8.4)$$

From this form of the likelihood function it is clear that $\bar{\mathbf{x}}_g$ and **S** are sufficient statistics for $\boldsymbol{\mu}_g$ and $\mathbf{\Sigma}$ because the likelihood function is factored into a part which is a function of the sufficient statistics $\bar{\mathbf{x}}_g$ and **S**, and the parameters $\boldsymbol{\mu}_g$ and $\mathbf{\Sigma}$, and a part which is not a function of $\boldsymbol{\mu}_g$ and $\mathbf{\Sigma}$.

8.3.3 Estimating Equations

In this section I develop estimating equations for the CVA/time (orth.) model. Henceforth I will work with the log-likelihood equation instead of the likelihood equation. Let $l(\mathbf{X}|, \boldsymbol{\mu}, \mathbf{\Sigma})$ stand for the natural logarithm of the likelihood of the data **X** given parameters $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$. Then $l(\mathbf{X}|, \boldsymbol{\mu}, \mathbf{\Sigma})$ is:

$$l(\mathbf{X}|\boldsymbol{\mu}, \Sigma) = \frac{-npt}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma|$$
$$-\tfrac{1}{2}\left(\sum_{g=1}^{m}\sum_{i=1}^{n_g}(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)'\mathbf{\Sigma}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}_g) + \sum_{g=1}^{m} n_g(\bar{\mathbf{x}}_g - \boldsymbol{\mu}_g)'\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}_g - \boldsymbol{\mu}_g)\right). \quad (8.5)$$

First I derive the maximum likelihood estimator for $\boldsymbol{\mu}_o$. For convenience, let C denote terms not involving $\boldsymbol{\mu}_g$, and $\mathrm{F}(\mathbf{v}_i, \mathbf{v}_i^q, \mathbf{e}_{g,i})$ denote the terms in the model for $\boldsymbol{\mu}_g$ (8.2) that do not involve $\boldsymbol{\mu}_0$; hence $\boldsymbol{\mu}_g = \boldsymbol{\mu}_0 + \mathrm{F}(\mathbf{v}_i, \mathbf{v}_i^q, \mathbf{e}_{g,i})$. Then the log-likelihood is:

$$l(\mathbf{X}|\boldsymbol{\mu}, \Sigma) = \mathrm{C} - \tfrac{1}{2}\sum_{g=1}^{m} n_g\left(\bar{\mathbf{x}}_g - \boldsymbol{\mu}_0 - \mathrm{F}(\mathbf{v}_i, \mathbf{v}_i^q, \mathbf{e}_{g,i})\right)' \mathbf{\Sigma}^{-1}\left(\bar{\mathbf{x}}_g - \boldsymbol{\mu}_0 - \mathrm{F}(\mathbf{v}_i, \mathbf{v}_i^q, \mathbf{e}_{g,i})\right).$$

Taking the derivatives of the log-likelihood with respect to $\boldsymbol{\mu}_o$ yields the following:

$$\frac{\delta l(\mathbf{X}|\mu,\Sigma)}{\delta\mu_{o}} = -\tfrac{1}{2}\sum_{g=1}^{m}\left(n_{g}2\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{o} - 2n_{g}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}_{g} + 2n_{g}\boldsymbol{\Sigma}^{-1}\mathrm{F}(\mathbf{v}_{i},\mathbf{v}_{i}^{r},\mathbf{e}_{g,i})\right). \tag{8.6}$$

One sets these derivatives equal to zero to obtain the estimating equations for $\boldsymbol{\mu}_{o}$. The last term in (8.6) drops out as $\sum_{g=1}^{m} n_{g}\mathrm{F}(\mathbf{v}_{i},\mathbf{v}_{i}^{q},\mathbf{e}_{g,i}) = \mathbf{0}$, where $\mathbf{0}$ is a $pt \times 1$ vector of zeros, because $\sum_{g=1}^{m} e_{g,i}^{q} = 0$ for all q,i. Hence:

$$\sum_{g=1}^{m} n_{g}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{o} = \sum_{g=1}^{m} n_{g}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}_{g} .$$

Multiplying through by $\boldsymbol{\Sigma}$ gives the maximum likelihood estimate for $\boldsymbol{\mu}_{o}$, which I denote as $\hat{\boldsymbol{\mu}}_{o}$,

$$\hat{\boldsymbol{\mu}}_{o} = \left(\sum_{g=1}^{m} n_{g}\right)^{-1}\sum_{g=1}^{m} n_{g}\overline{\mathbf{x}}_{g} = n^{-1}\sum_{g=1}^{m} n_{g}\overline{\mathbf{x}}_{g} . \tag{8.7}$$

$\hat{\boldsymbol{\mu}}_{o}$ is just the average over all observations.

Next, I derive estimating equations for $\boldsymbol{\Sigma}$. Denote by C terms that are not a function of $\boldsymbol{\Sigma}$. Then $l(\mathbf{X}|\mu,\Sigma)$ is:

$$l(\mathbf{X}|\mu,\Sigma) = \mathrm{C} - \frac{n}{2}\log|\Sigma| - \tfrac{1}{2}\left(\sum_{g=1}^{m}\sum_{i=1}^{n_{g}}(\mathbf{x}_{gi}-\overline{\mathbf{x}}_{g})\Sigma^{-1}(\mathbf{x}_{gi}-\overline{\mathbf{x}}_{g})' + \sum_{g=1}^{m} n_{g}(\overline{\mathbf{x}}_{g}-\boldsymbol{\mu}_{g})\Sigma^{-1}(\overline{\mathbf{x}}_{g}-\boldsymbol{\mu}_{g})'\right).$$

Taking the derivative of $l(\mathbf{X}|\mu,\Sigma)$ with respect to $\boldsymbol{\Sigma}$ yields:

$$\begin{aligned}\frac{\delta l(\mathbf{X}|\mu,\Sigma)}{\delta\Sigma} &= -n\Sigma^{-1} + \tfrac{n}{2}\,\mathrm{diag}(\boldsymbol{\Sigma}^{-1})\\ &\quad + \sum_{g=1}^{m}\sum_{i=1}^{n_{g}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_{gi}-\overline{\mathbf{x}}_{g})(\mathbf{x}_{gi}-\overline{\mathbf{x}}_{g})'\boldsymbol{\Sigma}^{-1} - \tfrac{1}{2}\,\mathrm{diag}\left(\sum_{g=1}^{m}\sum_{i=1}^{n_{g}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_{gi}-\overline{\mathbf{x}}_{g})(\mathbf{x}_{gi}-\overline{\mathbf{x}}_{g})'\boldsymbol{\Sigma}^{-1}\right)\\ &\quad + \sum_{g=1}^{m} n_{g}\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}_{g}-\boldsymbol{\mu}_{g})(\overline{\mathbf{x}}_{g}-\boldsymbol{\mu}_{g})'\boldsymbol{\Sigma}^{-1} - \tfrac{1}{2}\,\mathrm{diag}\left(\sum_{g=1}^{m} n_{g}\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}_{g}-\boldsymbol{\mu}_{g})(\overline{\mathbf{x}}_{g}-\boldsymbol{\mu}_{g})'\boldsymbol{\Sigma}^{-1}\right).\end{aligned}$$

Pre-multiply and post-multiply the above by $\boldsymbol{\Sigma}$, and set the equations equal to a matrix of zeros. Then the normal equations are solved when:

$$\hat{\boldsymbol{\Sigma}} = n^{-1}\sum_{g=1}^{m}\sum_{i=1}^{n_{g}}(\mathbf{x}_{gi}-\overline{\mathbf{x}}_{g})(\mathbf{x}_{gi}-\overline{\mathbf{x}}_{g})' + n^{-1}\sum_{g=1}^{m} n_{g}(\overline{\mathbf{x}}_{g}-\boldsymbol{\mu}_{g})(\overline{\mathbf{x}}_{g}-\boldsymbol{\mu}_{g})', \tag{8.8}$$

where $\hat{\boldsymbol{\Sigma}}$ denotes the estimate for $\boldsymbol{\Sigma}$.

One sees that $\hat{\boldsymbol{\Sigma}}$ is equal to **S**, the sample estimate of $\boldsymbol{\Sigma}$, plus an additional term which depends on the difference between the predicted and observed group means. Although the topic of obtaining estimates will be discussed later in Section **8.3.5**, it is worth mentioning now that this additional term will be small when the model is correctly specified, but inflated when the model is misspecified. **S**, on the other hand, is completely robust to model misspecification. Thus in practice using **S** may be preferable to $\hat{\boldsymbol{\Sigma}}$.

Next I derive the estimating equations for $\mathbf{v}_{\mathrm{i}}$, $\mathbf{v}_{\mathrm{i}}^{\mathrm{r}}$ and $\mathbf{e}_{\mathrm{g,i}}$. Substituting the means model for $\boldsymbol{\mu}_{\mathrm{g}}$ from equation (8.2) into equation (8.4) gives the likelihood equations for CVA/time (orth.). Denote the log-likelihood by $l(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})$, and the terms which include neither $\mathbf{v}_{\mathrm{i}}$, $\mathbf{v}_{\mathrm{i}}^{\mathrm{r}}$ nor $\mathbf{e}_{\mathrm{g,i}}$ by C. Then:

$$l(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \mathrm{C} - \tfrac{1}{2}\sum_{\mathrm{g}=1}^{m} n_{\mathrm{g}} \sum_{\mathrm{q}=1}^{t}\sum_{\mathrm{s}=1}^{t}\left(\sum_{\mathrm{a}=1}^{c}\sum_{\mathrm{b}=1}^{c} \mathrm{e}_{\mathrm{g,a}}^{\mathrm{q}}\mathbf{v}_{\mathrm{a}}'\boldsymbol{\Sigma}_{\mathrm{qs}}^{-1}\mathbf{v}_{\mathrm{b}}\mathrm{e}_{\mathrm{g,b}}^{\mathrm{s}} + 2\sum_{\mathrm{a}=1}^{c}\sum_{\mathrm{b}=c+1}^{u} \mathrm{e}_{\mathrm{g,a}}^{\mathrm{q}}\mathbf{v}_{\mathrm{a}}'\boldsymbol{\Sigma}_{\mathrm{qs}}^{-1}\mathbf{v}_{\mathrm{b}}^{\mathrm{s}}\mathrm{e}_{\mathrm{g,b}}^{\mathrm{s}} + \sum_{\mathrm{a}=c+1}^{u}\sum_{\mathrm{b}=c+1}^{u} \mathrm{e}_{\mathrm{g,a}}^{\mathrm{q}}\mathbf{v}_{\mathrm{a}}^{\mathrm{q}}\boldsymbol{\Sigma}_{\mathrm{qs}}^{-1}\mathbf{v}_{\mathrm{b}}^{\mathrm{s}}\mathrm{e}_{\mathrm{g,b}}^{\mathrm{s}}\right)$$
$$+ \tfrac{1}{2}\sum_{\mathrm{g}=1}^{m} n_{\mathrm{g}} \sum_{\mathrm{q}=1}^{t}\sum_{\mathrm{s}=1}^{t}\left(2\sum_{\mathrm{b}=1}^{c}(\overline{\mathbf{x}}_{\mathrm{g}}^{\mathrm{q}} - \boldsymbol{\mu}_{0}^{\mathrm{q}})'\boldsymbol{\Sigma}_{\mathrm{qs}}^{-1}\mathbf{v}_{\mathrm{b}}\mathrm{e}_{\mathrm{g,b}}^{\mathrm{s}} + 2\sum_{\mathrm{b}=c+1}^{\mathrm{k}}(\overline{\mathbf{x}}_{\mathrm{g}}^{\mathrm{q}} - \boldsymbol{\mu}_{0}^{\mathrm{q}})'\boldsymbol{\Sigma}_{\mathrm{qs}}^{-1}\mathbf{v}_{\mathrm{b}}^{\mathrm{s}}\mathrm{e}_{\mathrm{g,b}}^{\mathrm{s}}\right), \qquad (8.9)$$

where $\boldsymbol{\mu}_{0}^{\mathrm{q}}$ is the $p \times 1$ vector of means averaged over all groups for the q^{th} occasion, and $\overline{\mathbf{x}}_{\mathrm{g}}^{\mathrm{q}}$ is the $p \times 1$ vector of sample means for the g^{th} group.

The estimating equations for the common variates are considered next. The estimation of variates requires consideration of the constraints. The constraints for the orthogonality of the common variates are incorporated by the method of Lagrangian multipliers. The constraints with Lagrangian multipliers for the unit length of the common variates are as follows (note that here and in subsequent developments the constraints with Lagrangian multipliers are implicitly set to zero):

$$\sum_{\mathrm{a}=1}^{c}\frac{\gamma_{\mathrm{a}}}{2}(\mathbf{v}_{\mathrm{a}}'\mathbf{v}_{\mathrm{a}} - 1),$$

where γ_{a} are c Lagrangian multipliers. The constraints with Lagrangian multipliers for the orthogonality of the common variates are:

$$\sum_{\mathrm{a}=1}^{c}\sum_{\mathrm{b}=1}^{\mathrm{a}-1} \gamma_{\mathrm{ab}}\mathbf{v}_{\mathrm{a}}'\mathbf{v}_{\mathrm{b}},$$

where γ_{ab} are $c(c-1)/2$ Lagrangian multipliers. The constraints with Lagrangian multipliers for the orthogonality of each common variate with all of the unique variates are:

$$\sum_{\mathrm{q}=1}^{t}\sum_{\mathrm{a}=1}^{c}\sum_{\mathrm{b}=c+1}^{u}\gamma_{\mathrm{abq}}\mathbf{v}_{\mathrm{a}}'\mathbf{v}_{\mathrm{b}}^{\mathrm{q}}, \qquad (8.10)$$

where γ_{abq} are $tc(u-c)$ Lagrangian multipliers. Denote the log-likelihood modified to incorporate the constraints with Lagrangian multipliers by $l^{*}(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})$. Take the derivative of $l^{*}(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ with respect to $\mathbf{v}_{\mathrm{f}}$:

$$\frac{\delta l^{*}(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})}{\delta \mathbf{v}_{\mathrm{f}}} = \sum_{\mathrm{g}=1}^{m} n_{\mathrm{g}}\sum_{\mathrm{q}=1}^{t}\sum_{\mathrm{s}=1}^{t}\left(-\sum_{\mathrm{a}=1}^{c}\mathrm{e}_{\mathrm{g,a}}^{\mathrm{q}}\boldsymbol{\Sigma}_{\mathrm{qs}}^{-1}\mathbf{v}_{\mathrm{a}}\mathrm{e}_{\mathrm{g,f}}^{\mathrm{s}} - \sum_{\mathrm{b}=c+1}^{u}\mathrm{e}_{\mathrm{g,f}}^{\mathrm{q}}\boldsymbol{\Sigma}_{\mathrm{qs}}^{-1}\mathbf{v}_{\mathrm{b}}^{\mathrm{s}}\mathrm{e}_{\mathrm{g,b}}^{\mathrm{s}} + (\overline{\mathbf{x}}_{\mathrm{g}}^{\mathrm{q}} - \boldsymbol{\mu}_{0}^{\mathrm{q}})'\boldsymbol{\Sigma}_{\mathrm{qs}}^{-1}\mathrm{e}_{\mathrm{g,f}}^{\mathrm{s}}\right)$$
$$+\gamma_{\mathrm{f}}\mathbf{v}_{\mathrm{f}} + \sum_{\substack{\mathrm{a}=1 \\ \mathrm{a}\neq\mathrm{f}}}^{c}\gamma_{\mathrm{af}}\mathbf{v}_{\mathrm{a}} + \sum_{\mathrm{q}=1}^{t}\sum_{\mathrm{b}=c+1}^{u}\gamma_{\mathrm{fbq}}\mathbf{v}_{\mathrm{b}}^{\mathrm{q}}.$$

Setting these derivatives equal to a vector of zeros yields the estimating equations to solve for $\mathbf{v}_{\mathrm{f}}$.

Next I derive the estimating equations for the unique variates. The constraints with Lagrangian multipliers are as follows, starting with those which constrain the unique variates to unit length (note they are implicitly set to zero):

$$\sum_{a=c+1}^{u} \sum_{q=1}^{t} \frac{\gamma_{aq}}{2}\left(\mathbf{v}_{a}^{q\prime}\mathbf{v}_{a}^{q} - 1\right),$$

where γ_{aq} are $(u - c)t$ Lagrangian multipliers. The constraints for the mutual orthogonality of each unique variate with the other unique variates of the same occasion are:

$$\sum_{q=1}^{t} \sum_{a=c+1}^{u} \sum_{b=c+1}^{a-1} \gamma_{abq}\mathbf{v}_{a}^{q\prime}\mathbf{v}_{b}^{q},$$

where γ_{abq} are $t(u-c)(u-c-1)/2$ Lagrangian multipliers. The constraints for the orthogonality of each common variate with all of the unique variates are already given in equation (8.10).

Now take the derivative of $l(\mathbf{X}|\mu,\Sigma)$ with respect to $\mathbf{v}_{f}^{r}$. These derivatives yield the estimating equations for solving for $\mathbf{v}_{f}^{r}$ when they are set to zero:

$$\frac{\delta l^{*}(\mathbf{X}|\mu,\Sigma)}{\delta \mathbf{v}_{f}^{r}} = \sum_{g=1}^{m} n_{g} \sum_{q=1}^{t}\left(-\sum_{a=1}^{c} e_{g,a}^{q}\boldsymbol{\Sigma}_{qr}^{-1}\mathbf{v}_{a}e_{g,w}^{r} - \sum_{b=c+1}^{u} e_{g,b}^{q}\boldsymbol{\Sigma}_{qr}^{-1}\mathbf{v}_{w}^{r}e_{g,w}^{r} + (\bar{\mathbf{x}}_{gq} - \boldsymbol{\mu}_{0}^{q})\boldsymbol{\Sigma}_{qr}^{-1}e_{g,f}^{r}\right)$$

$$+\gamma_{rf}\mathbf{v}_{f}^{r} + \sum_{s=1}^{c}\gamma_{afr}\mathbf{v}_{a} + \sum_{b=c+1}^{u}\gamma_{\phi\beta\rho}\mathbf{v}_{b}^{r}.$$

Lastly I derive estimating equations for the group positions, the $e_{g,b}^{s}$ terms, beginning with those corresponding to the common variates. The constraints for these terms are $\sum_{g=1}^{m} n_{g}e_{g,b}^{s} = 0$, for $s = 1,\ldots,t$ and $b = 1,\ldots,c$. They are handled in the estimation by letting $e_{m,b}^{s} = -\sum_{h=1}^{m-1} e_{h,b}^{s}$ for $s = 1,\ldots,t$ and $b = 1,\ldots,c$. One takes the derivative of $l(\mathbf{X}|\mu,\Sigma)$ with respect to $e_{w,f}^{r}$ for $w \neq m$, and sets these derivatives equal to zero to obtain the estimating equations for the $e_{g,b}^{s}$ terms:

$$\frac{\delta l(\mathbf{X}|\mu,\Sigma)}{\delta e_{w,f}^{r}} = \sum_{q=1}^{t}\left(-\sum_{a=1}^{c} e_{w,a}^{q}\mathbf{v}_{f}^{\prime}\boldsymbol{\Sigma}_{qr}^{-1}\mathbf{v}_{a} - \sum_{b=c+1}^{u} e_{w,b}^{q}\mathbf{v}_{f}^{\prime}\,\boldsymbol{\Sigma}_{qr}^{-1}\mathbf{v}_{b}^{s} + \left(\bar{\mathbf{x}}_{w}^{q} - \boldsymbol{\mu}_{0}^{q}\right)\boldsymbol{\Sigma}_{qr}^{-1}\mathbf{v}_{f}\right).$$

The estimating equations for the group positions corresponding to the unique variates are handled similarly to those corresponding to the common variates. The constraints are identical to those of the common variates. Take the derivative of $l(\mathbf{X}|\mu,\Sigma)$ with respect to $e_{w,f}^{r}$ for $h \neq m$:

$$\frac{\delta l(\mathbf{X}|\mu,\Sigma)}{\delta e_{w,f}^{r}} = \sum_{q=1}^{t}\left(2\sum_{a=1}^{c} e_{w,a}^{q}\mathbf{v}_{a}^{\prime}\boldsymbol{\Sigma}_{qr}^{-1}\mathbf{v}_{f}^{r} + 2\sum_{b=c+1}^{u}\mathbf{v}_{f}^{\prime}\boldsymbol{\Sigma}_{qr}^{-1}\mathbf{v}_{b}^{s}e_{w,b}^{s} + 2\bar{\mathbf{x}}_{w}^{q}\boldsymbol{\Sigma}_{qr}^{-1}\mathbf{v}_{f}^{r}\right).$$

Set these derivatives equal to zero, yielding the estimating equations.

8.3.4 Unchanging Group Positions

Another model of possible interest to the researcher is one that hypothesizes that the scores for the group means on the common variates, $e^{q}_{g,a}$, do not change; i.e., they are equal at different occasions, $e^{q}_{g,a} = e^{s}_{g,a} \;\; \forall\, q \neq s$. Let $e_{g,a}$ denote the unchanging score of the g^{th} group for the a^{th} common variate. The likelihood equation for this model is obtained by substituting $e_{g,a}$ for $e^{q}_{g,a}$ in the likelihood equation for CVA/time (orth.) (8.9). The constraints on the $e_{g,a}$ and the manner of handling them are the same as for the $e^{q}_{g,a}$ terms in Section **8.3.3**. The estimating equations for the $e_{g,a}$ are found by taking the derivative of $l(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ with respect to $e_{g,f}$ and setting it equal to zero:

$$\frac{\delta l(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})}{\delta e_{h,w}} = -n_h \sum_{q=1}^{t}\sum_{s=1}^{t}\left(\sum_{a=1}^{c} e_{h,a}\mathbf{v}'_w\boldsymbol{\Sigma}^{-1}_{qr}\mathbf{v}_a + \sum_{b=c+1}^{u} e^{q}_{h,b}\mathbf{v}'_w\boldsymbol{\Sigma}^{-1}_{qr}\mathbf{v}^{s}_b\right) + n_h\left(\sum_{q=1}^{t}\sum_{s=1}^{t}\sum_{a=1}^{c}(\bar{\mathbf{x}}^{s}_h - \boldsymbol{\mu}^{q}_0)'\boldsymbol{\Sigma}^{-1}_{qr}\mathbf{v}_w\right).$$

The estimating equations for $\mathbf{v}_i$ and $\mathbf{v}^{q}_i$ are the same as those in the previous section except that $e_{g,a}$ is substituted for $e^{q}_{g,a}$.

8.3.5 Obtaining Estimates

The estimating equations taken together with the constraints result in a system of non-linear equations which can be solved with a Gauss-Newton algorithm, implementing the Marquardt modification where appropriate. The estimates for all parameters are generally solved for simultaneously, as all the estimating equations must be simultaneously true in order to be at a solution. However, the estimation of $\boldsymbol{\mu}_0$ is an exception to this rule as its estimator is a function only of the data, not of any of the other parameter estimates; see equation (8.7).

The algorithm does not guarentee convergence to a local extremum or saddle point. The convergence of the algorithm to a globally optimal solution depends on good starting values. Reasonable starting values can be obtained for the common variates by performing a common principal components on the group means. Starting values for the unique variates can be obtained by performing separate canonical variate analysis or redundancy analyses at each occasion. For models that include both common and unique variates one has to use both methods to obtain starting values. In some cases it may be necessary to try more than one set of starting values. One can examine the matrix of second order partial derivatives of the likelihood function with to the parameters to determine whether a solution is a local maximum, minimum or saddle point. Unfortunately, one can never be certain one has achieved a global solution.

8.3.6 Statistical Inference

Maximum likelihood estimation has under regularity conditions (Wilks 1962) properties which allow one to perform various forms of statistical inference including hypothesis tests and confidence intervals for parameter estimates. This section discusses how such inference is obtained.

Firstly, define the composite hypothesis and alternative as follows:

$$H_0 = \boldsymbol{\theta} \in \Theta_0$$

$$H_1 = \boldsymbol{\theta} \in \Theta_1,$$

where $\Theta_0 = \Theta - \Theta_1$, $\boldsymbol{\theta}$ is an *r*-dimensional vector of unknown parameters to be estimated, Θ is an open region in *r*-dimensional Euclidean space and Θ_0 is *q*-dimensional, $q < r$. Then for such composite hypotheses likelihood ratio tests can be based on the asymptotic chi-square distribution of negative two times the log-likelihood ratio, which is denoted by $-2\log\lambda(\mathbf{X})$, where:

$$\lambda(\mathbf{X}) = \frac{\sup\{L(\mathbf{X},\theta): \theta \in \Theta_0\}}{\sup\{L(\mathbf{X},\theta): \theta \in \Theta_1\}},$$

and $\sup\{L(\mathbf{X},\theta): \theta \in \Theta_1\}$ is the maximum likelihood estimate given $\theta \in \Theta_1$. For large sample sizes the following is approximately true:

$$-2\log\lambda(\mathbf{X}) \sim X^2_{r-q}.$$

The form of an α-level test is straightforward: reject H_0 if

$$-2\log\lambda(\mathbf{X}) \geq X^2_{r-q}(1-\alpha).$$

One can perform a likelihood ratio test if the set of parameters of the null hypothesis is nested within the set of parameters of the alternative hypothesis. For example, the set of parameters of the common variate hypothesis is nested within the set of parameters of the unique variate hypothesis. Thus one can test the null hypothesis that a given number of variates are common versus the alternative that they are unique.

Next I point out that maximum likelihood estimates are consistent and asymptotically unbiased. Furthermore, they are asymptotically normal with a covariance matrix equal to the inverse of the information matrix. The information matrix is defined as:

$$\mathbf{I}(\boldsymbol{\theta}_0) = \left[-\frac{\delta^2 l(\mathbf{X},\boldsymbol{\theta})}{\delta\theta_i \delta\theta_\varphi}(\boldsymbol{\theta} = \boldsymbol{\theta}_0) \right],$$

though in practice one evaluates the information matrix at the parameter estimates based on the data. Estimates for variances of the parameter estimates can be obtained from $\mathbf{I}^{-1}(\boldsymbol{\theta}_0)$, enabling one to make confidence intervals and simple hypothesis tests for the parameters. Let θ_i be a parameter in the model, and $\hat{\theta}_i$ be its estimate. Then a $100 \times (1-\alpha)\%$ confidence interval for θ_i would be $\hat{\theta}_i \pm \sqrt{c}_{ii} Z_{(1-\alpha/2)}$, where c_{ii} is the i^{th} diagonal element of $\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}})$ and $Z_{(1-\alpha/2)}$ is the value of the standard normal variate corresponding to a cumulative probability of $1-\alpha/2$.

An α-level hypothesis test for testing the following hypothesis that a single parameter, θ_i, is zero

$$H_0: \theta_i = 0 \quad \text{vs.} \quad H_1: \theta_i \neq 0,$$

would be to reject H_0 if $\hat{\theta}_i \geq c_{ii} Z_{(1-\alpha/2)}$ or $\hat{\theta}_i \leq -c_{ii} Z_{(1-\alpha/2)}$. One can test more complex hypotheses of the form

$$H_0: \mathbf{C}\boldsymbol{\theta} = 0 \quad \text{vs.} \quad H_1: \mathbf{C}\boldsymbol{\theta} \neq 0$$

where $\boldsymbol{\theta}$ is a $q \times 1$ the vector of parameters and **C** is an $r \times q$ matrix of rank *r* with the Wald (Wald 1945) statistic W, where $W = (\mathbf{C}\hat{\boldsymbol{\theta}})'((\mathbf{C}\hat{\boldsymbol{\theta}})'\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{C}\hat{\boldsymbol{\theta}})^{-1}\mathbf{C}\hat{\boldsymbol{\theta}}$.

8.4 SIMULATIONS

In this section I present a simulated study of the methods developed in Section **8.3**. Simulated data have the advantage that one knows the true structure of the data, hence one can ascertain whether the method is successful in discerning that structure. In particular I attempt to answer the following questions: Is the method obtaining the true parameters? When the null hypothesis is true, is the likelihood ratio test rejecting at the specified alpha-level? Is the test statistic distributed as predicted in the theory? And, does the simulated variance-covariance matrix for the parameter estimates converge to the theoretical asymptotic variance-covariance based on the inverse of the information matrix?

These simulated data were generated using a pseudo-multivariate normal distribution in SAS's Proc IML. The generated multivariate normal data served as residuals which were added to a mean structure specified to be a one common variate model. The actual parameters were chosen arbitrarily. The simulated data consisted of measurements of three variables at three occasions for four groups. The SAS code used to generate the simulation and obtain the estimates is given in Appendix Six. There were three simulations performed. One simulation generated 1,000 datasets of a sample size of 100, or a sample of 25 for each of the four groups. The other two simulations generated 5,000 datasets each of samples of sizes of 400 and 4,000, or 100 and 1,000 for each of the four groups. There were fewer simulations for samples of size 100 (four groups of 25), because the algorithm to solve the estimating equations took prohibitively longer to converge.

The information matrix was calculated using the computer package "Mathematica" (Wolfram 1991). The Mathematica code is given in Appendix Seven. It is clear from equation (8.6) that the theoretical estimate for $\boldsymbol{\mu}_0$ based on the information matrix is independent of the estimates for the rest of the parameters because the derivative of (8.6) with respect to any parameter other than $\boldsymbol{\mu}_0$ will be zero. Further, it is straightforward to show that the derivatives of $\dfrac{\delta l(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})}{\delta\boldsymbol{\Sigma}}$ with respect to **v** and $\mathbf{e}_g$ will be zero. Thus $\hat{\boldsymbol{\Sigma}}$ is independent (asymptotically) of the estimates of **v** and $\mathbf{e}_g$. Hence **Table 8.2** and Appendix Eight only show the calculated variances of the parameters **v** and $\mathbf{e}_g$.

The model for the group means used to simulate the data is shown below. The terms are defined as in Section **8.3**:

$$\boldsymbol{\mu}_g = \boldsymbol{\mu}_0 + \mathbf{e}_g \otimes \mathbf{v},$$

where $g = 1,\ldots,m$ and $\boldsymbol{\mu}_g$ is a 9×1 vector. The parameter values for **v**, the common variate, are:

$$\mathbf{v} = [0.5, 0.5, 0.707]' .$$

Let **E** be the matrix whose columns are $\mathbf{e}_g$, $g = 1,\ldots,m$, then:

$$\mathbf{E} = \begin{bmatrix} 1 & 0.5 & -.5 & -1 \\ 0 & 1 & 0 & -1 \\ 1 & -.5 & .5 & -1 \end{bmatrix}.$$

The errors have the covariance matrix **C**:

$$\mathbf{C} = \begin{bmatrix} 4.8 & 2.1 & 1.0 & 2.4 & 1.05 & 0.5 & 1.2 & 0.525 & 0.25 \\ 2.1 & 3.3 & 1.4 & 1.05 & 1.65 & 0.7 & 0.525 & 0.825 & 0.35 \\ 1.0 & 1.4 & 2.9 & 0.5 & 0.7 & 1.45 & 0.25 & 0.35 & 0.725 \\ 2.4 & 1.05 & 0.5 & 4.8 & 2.1 & 1.0 & 2.4 & 1.05 & 0..5 \\ 1.05 & 1.65 & 0.7 & 2.1 & 3.3 & 1.4 & 1.05 & 1.65 & 0.7 \\ 0.5 & 0.7 & 1.45 & 1.0 & 1.4 & 2.9 & 0.5 & 0.7 & 1.45 \\ 1.2 & 0.525 & 0.25 & 2.4 & 1.05 & 0.5 & 4.8 & 2.1 & 1.0 \\ 0.525 & 0.825 & 0.35 & 1.05 & 1.65 & 0.7 & 2.1 & 3.3 & 1.4 \\ 0.25 & 0.35 & 0.725 & 0.5 & 0.7 & 1.45 & 1.0 & 1.4 & 2.9 \end{bmatrix}.$$

The one common variate model and the one unique variate model were fit for each dataset. The one unique variate model is as follows:

$$\boldsymbol{\mu}_g = \boldsymbol{\mu}_o + \begin{bmatrix} e_g^1 \mathbf{v}^1 \\ e_g^2 \mathbf{v}^2 \\ e_g^3 \mathbf{v}^3 \end{bmatrix},$$

where by definition $\sum_{g=1}^{m} n_g \mathbf{e}_g = 0$, $\mathbf{v}^{1\prime}\mathbf{v}^1 = 1$, $\mathbf{v}^{2\prime}\mathbf{v}^2 = 1$ and $\mathbf{v}^{3\prime}\mathbf{v}^3 = 1$. Note that for 108 of the 1,000 datasets with sample size of 100 that the algorithm for the unique variates converged to an estimate that was clearly not a global maximum. I determined that an estimate for the unique variates model was not a global maximum if it had a lower likelihood than the estimate of the common variate model. Since the unique variates model has more free parameters than the common variates model the likelihood of its estimate should be greater if it is at the global maximum. The 108 runs were not included in the tables. In only one of the 5,000 runs with the sample size of 400 did this problem occur, and in none of the runs with the sample size of 4,000.

Table 8.1 shows the means of the parameter estimates based on the simulations. At the left are the true parameter values. All three sets of estimates are in the area of the true parameter values. However, the estimates based on the larger samples are clearly closer to the true values.

Table 8.1 Parameter Estimates

Para-meters	True Parameter Values	Estimates for Sample of 4000	Estimates for Sample of 400	Estimates for Sample of 100
v_1	0.5	0.4994	0.4967	0.4911
v_2	0.5	0.4997	0.4981	0.4910
v_3	0.7071	0.7073	0.7058	0.6940
e_1^1	1.0	0.9995	1.0048	1.0368
e_1^2	0.0	0.0006	0.0001	0.0290
e_1^3	1.0	0.9989	1.0018	1.0359
e_2^1	0.5	0.4999	0.4990	0.4892
e_2^2	1.0	0.9989	1.0024	1.0078
e_2^3	-0.5	-0.5000	-0.5057	-0.5181
e_3^1	-0.5	-0.4997	-0.5029	-0.4901
e_3^2	0.0	0.0003	0.0016	0.0039
e_3^3	0.5	0.5007	0.5033	0.5209
e_4^1	-1.0	-0.9997	-1.0009	-1.0360
e_4^2	-1.0	-0.9998	-1.0007	-1.0407
e_4^3	-1.0	-0.9996	-0.9994	-1.0389

Table 8.2 presents the observed variances based on the simulations for comparison with the theoretical variances based on the inverse of the information matrix. The theoretical variances are calculated assuming a sample of size 400. Not presented in **Table 8.2** are the theoretical variances for a sample of size 4,000, which are $\frac{1}{10}^{\text{th}}$ that of the variance for $n = 400$, and the theoretical variances for a sample size of 100, which are four times those of a sample size of 400. One sees that the estimated variances of the parameter estimates are close to the theoretical variances for the sample sizes of 400 and 4,000, but not for a sample size of 100. Examination of the full variance-covariance matrices would reveal the same pattern. The full theoretical variance-covariance matrix is presented in Appendix Eight, while the full variance-covariance matrices based on the estimates from the simulations is presented in Appendix Nine.

Table 8.2 Theoretical and Observed Variances for the Parameter Estimates

Para-meters	Theoretical Values For Sample of 400	Observed for Sample of 4000	Observed for Sample of 400	Observed for Sample of 100
v_1	0.002998	0.000314	0.00326	0.046457
v_2	0.001589	0.000158	0.001802	0.037871
v_3	0.001928	0.0001964	0.00215	0.06856
e_1^1	0.040908	0.00427	0.042776	0.325786
e_1^2	0.039628	0.003878	0.040881	0.185901
e_1^3	0.040908	0.00409	0.042394	0.306227
e_2^1	0.039948	0.004037	0.041292	0.204403
e_2^2	0.040908	0.004177	0.041389	0.324521
e_2^3	0.039948	0.004072	0.041348	0.218552
e_3^1	0.039948	0.004059	0.041273	0.219736
e_3^2	0.039628	0.003866	0.042012	0.197513
e_3^3	0.039948	0.003925	0.041697	0.208513
e_4^1	0.040908	0.004207	0.04335	0.30024
e_4^2	0.040908	0.004036	0.041934	0.308235
e_4^3	0.040908	0.004189	0.04264	0.324294

The next results of interest are the distributions of the likelihood ratio test statistics. **Table 8.3** shows the mean and variance of the test statistics, and the proportion that are greater than the 90th, 95th and 99th percentile of the cumulative distribution of a chi-square with four degrees of freedom. One sees for sample sizes of 4,000 that the mean and variance of the likelihood ratio test statistic are very close to the theoretical values. Also, the proportions of the observed test statistics that are above the $\mathbf{X}^2_{(1-\alpha)}$ are close to what is predicted by the theory. For the datasets with sizes of 400, the variance of the test statistic is larger at about 9, though the proportions are roughly correct. For the data with sample sizes of 100 the distribution of the test statistic deviates more noticeably from the theoretical. This is not surprising as one has only 100 observations with which to estimate a total of 61 parameters (one must also estimate the elements of $\mathbf{\Sigma}$).

Table 8.3 Theoretical and Observed Values of the Likelihood Ratio Test Statistic

Labels	Mean	Variance	Proportion over 90^{th} Percentile	Proportion over 95^{th} Percentile	Proportion over 99^{th} Percentile
Theoretical	4.0	8.0	0.1	0.05	0.01
Sample of 4000	4.005	8.036	0.099	0.0506	0.0099
Sample of 400	4.212	9.076	0.117	0.0614	0.0146
Sample of 100	3.904	7.000	0.084	0.0426	0.0034

In summary, these simulations confirm the basic methodology. First, they confirm that the algorithm and estimating equations yield correct estimates. The estimates are correct in the sense that they are approaching the true parameter values and that the observed variances of the parameter estimates are close to the theoretical variances, at least for the larger sample sizes. Second, they confirm the correctness of the hypothesis tests for the larger samples. With samples sizes of 400 or 4,000 the test statistic is distributed close to the theoretical distribution under the null hypothesis.

8.5 CVA/TIME - UNCORRELATED VARIATES

In this section I develop an alternative model for analyzing group structure with longitudinal multivariate data, CVA/time with uncorrelated canonical variates. Where CVA/time (orth.) hypothesizes that the group means lie in the space of orthogonal variates, CVA/time (unc.) hypothesizes that the group means lie in the space of uncorrelated canonical variates. CVA/time (unc.), unlike CVA/time (orth.), represents a true generalization of CVA. Furthermore, it is equivalent to Campbell and Tomenson's model under the special circumstance that the covariances of the variables between different occasions are zero.

CVA/time (unc.) follows much of the logic of that for CVA/time (orth.). The CVA/time (unc.) model hypothesizes common variates, unique variates, and group positions. It differs from CVA/time (orth.) in that the model for the means now involves the covariance matrix. Furthermore, as is seen in equation (8.1), a common within-groups covariance structure for each occasion, $\boldsymbol{\Sigma}_w$, is required, which will necessitate a certain structure for $\boldsymbol{\Sigma}$. The estimation of the structure of $\boldsymbol{\Sigma}$ will encompass a good part of this section. On the other hand, the development of the estimation for $\boldsymbol{\mu}_0$ is identical to that in Section **8.3.3**, as is the discussion about obtaining estimates in Section **8.3.5** and that of statistical inference in Section **8.3.6**. Hence these topics need not be further addressed with respect to CVA/time (unc.).

It is useful to give a simple example of a common variate model. Assume the positions of two group means can be plotted on a variate in the transformed space which is common over two occasions, and assume the positions of the group means change over time. **Figure 8.3** shows the positions at the first occasion, and **Figure 8.4** shows the positions at the second occasion.

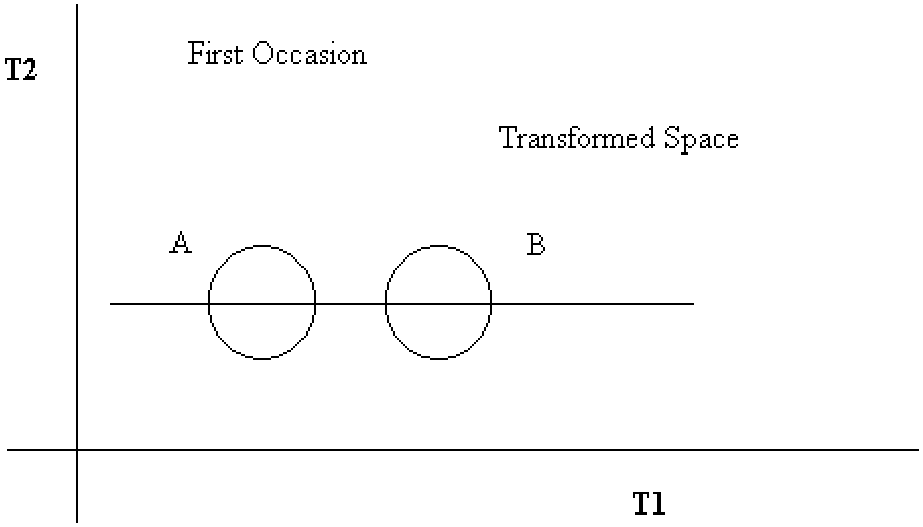

Figure 8.3

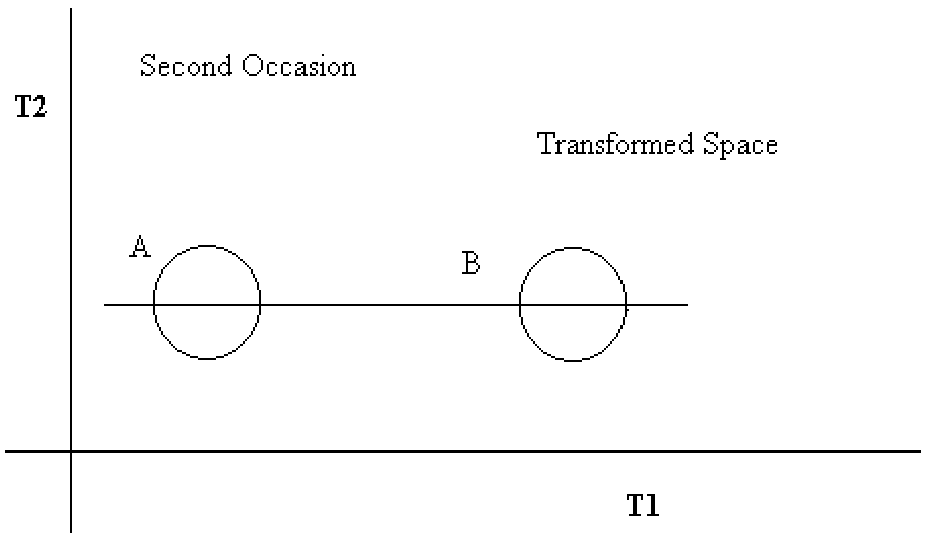

Figure 8.4

8.5.1 The CVA/Time Model with Uncorrelated Variates

CVA/time (unc.) model is as follows:

$$\boldsymbol{\mu}_{\mathrm{g}} = \boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_{\mathrm{w}} \mathbf{v}_1 \otimes \mathbf{e}_{\mathrm{g},1} + \ldots + \boldsymbol{\Sigma}_{\mathrm{w}} \mathbf{v}_c \otimes \mathbf{e}_{\mathrm{g},c} + \left(\boldsymbol{\Sigma}_{\mathrm{w}} \otimes \mathbf{I}_{t\times t}\right)\begin{bmatrix} \mathrm{e}^1_{\mathrm{g},c+1}\mathbf{v}^1_{c+1} \\ \vdots \\ \mathrm{e}^t_{\mathrm{g},c+1}\mathbf{v}^t_{c+1} \end{bmatrix} + \cdots + \left(\boldsymbol{\Sigma}_{\mathrm{w}} \otimes \mathbf{I}_{t\times t}\right)\begin{bmatrix} \mathrm{e}^1_{\mathrm{g},u}\mathbf{v}^1_{u} \\ \vdots \\ \mathrm{e}^t_{\mathrm{g},u}\mathbf{v}^t_{u} \end{bmatrix}, \quad (8.11)$$

where $\boldsymbol{\mu}_{\mathrm{g}}$ is a $pt \times 1$ vector of means for the g^{th} group, $\boldsymbol{\mu}_0$ is a $pt \times 1$ vector of overall means, $\mathbf{v}_{\mathrm{i}}$ are c $p \times 1$ vectors of common variates, $\mathbf{v}^{\mathrm{k}}_{\mathrm{j}}$ are $t(u - c)$ $p \times 1$ vectors of unique variates, $\mathrm{e}^t_{\mathrm{g,i}}$

is the score for the g^{th} group mean on the i^{th} canonical variate at the q^{th} occasion, and $\mathbf{e}_{\mathrm{g,i}}$ is the $t \times 1$ vector whose elements are $\mathrm{e}_{\mathrm{g,i}}^{\mathrm{q}}$.

Note the constraints on the parameters. Those for the positions of the group means are the same as for CVA/time (orth.), $\sum_{\mathrm{g}=1}^{m} \mathrm{e}_{\mathrm{g,i}}^{\mathrm{q}} = 0$ for all q,i. This constraint reflects the centering by $\boldsymbol{\mu}_0$. The constraints on the variates differ from those for CVA/time (orth.) as the variates are constrained to be mutually uncorrelated, not orthogonal:

$$\mathbf{V}_{\mathrm{com}}' \boldsymbol{\Sigma}_{\mathrm{w}} \mathbf{V}_{\mathrm{com}} = \mathbf{I}_{c \times c},$$

where $\mathbf{V}_{\mathrm{com}}$ is the matrix whose c columns are the common variates. Furthermore, within each set of unique variates for each occasion the variates are mutually uncorrelated, that is:

$$\mathbf{V}^{\mathrm{q}\prime} \boldsymbol{\Sigma}_{\mathrm{w}} \mathbf{V}^{\mathrm{q}} = \mathbf{I}_{(u\text{-}c) \times (u\text{-}c)},$$

where $\mathbf{V}^{\mathrm{q}}$ is the matrix whose columns are the unique variates for the q^{th} occasion, $\mathrm{q} = 1, \ldots, t$. Finally, each variate in each set of unique variates is orthogonal with each common variate. Thus:

$$\mathbf{V}_{\mathrm{com}}' \boldsymbol{\Sigma}_{\mathrm{w}} \mathbf{V}^{\mathrm{q}} = [0]_{c \times (u\text{-}c)},$$

for $\mathrm{q} = 1, \ldots, t$.

To conclude this section I point out that a model that hypothesizes unique variates at each occasion is not equivalent to performing separate canonical variate analyses at each occasion. A unique variates model estimates the variates at each occasion as a part of a larger model. It may be superior to a separate analysis of each occasion because it models the covariances between the measurements made at different occasions.

8.5.2 Estimating the Within-Groups Covariance Matrix

CVA/time (unc.) necessitates a particular structure to $\boldsymbol{\Sigma}$. The estimation of this structure is discussed in this section. Also briefly discussed at the conclusion is how to estimate this same structure for $\boldsymbol{\Sigma}$ if one should hypothesize orthogonal canonical variates.

As stated previously, the logic of CVA/time (unc.) requires within-groups covariance matrices that are equal over time or at minimum proportional over time. However, given this assumption it is reasonable to make the further assumption that the covariances matrices between measurements at different occasions are proportional. Indeed, this additional assumption proves to be necessary to obtain workable estimating equations. Hence the structure given in (8.12) is assumed:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{a}_{11}\boldsymbol{\Sigma}_{\mathrm{w}} & \mathrm{a}_{12}\boldsymbol{\Sigma}_{\mathrm{w}} & \cdots & \mathrm{a}_{1t}\boldsymbol{\Sigma}_{\mathrm{w}} \\ \mathrm{a}_{21}\boldsymbol{\Sigma}_{\mathrm{w}} & \mathrm{a}_{22}\boldsymbol{\Sigma}_{\mathrm{w}} & \cdots & \mathrm{a}_{2t}\boldsymbol{\Sigma}_{\mathrm{w}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{a}_{t1}\boldsymbol{\Sigma}_{\mathrm{w}} & \mathrm{a}_{t2}\boldsymbol{\Sigma}_{\mathrm{w}} & \cdots & \mathrm{a}_{tt}\boldsymbol{\Sigma}_{\mathrm{w}} \end{bmatrix} = \mathbf{A} \otimes \boldsymbol{\Sigma}_{\mathrm{w}}, \tag{8.12}$$

where $\mathbf{A} = [a_{ij}]$, is a $t \times t$ positive definite matrix, and $\boldsymbol{\Sigma}_w$ is a $p \times p$ matrix that is proportional to the within-groups variance-covariance matrix at each occasion. The matrix **A** will be referred to as the matrix of proportionality constants.

The estimating equations for **A** are derived in Section **8.5.3**. I proceed by deriving the estimating equations for $\boldsymbol{\Sigma}_w$. The estimation of $\boldsymbol{\Sigma}_w$ is complicated by the fact that the group means, $\boldsymbol{\mu}_g$, are now functions of $\boldsymbol{\Sigma}_w$, which is seen in (8.11). The log-likelihood for the model is:

$$l(\mathbf{X}|\mu,\Sigma) = \frac{-np}{2}\log(2\pi) - \frac{nt}{2}\log|\boldsymbol{\Sigma}_w| - \frac{np}{2}\log|\mathbf{A}|$$
$$-\tfrac{1}{2}\left(\sum_{g=1}^{m}\sum_{i=1}^{n_g}\sum_{q=1}^{t}\sum_{s=1}^{t}\mathbf{A}^{-1}[q,s](\mathbf{x}_{iq}-\bar{\mathbf{x}}_{gq})'\boldsymbol{\Sigma}_w^{-1}(\mathbf{x}_{is}-\bar{\mathbf{x}}_{gs}) + \sum_{g=1}^{m} n_g \sum_{q=1}^{t}\sum_{s=1}^{t}\mathbf{A}^{-1}[q,s](\bar{\mathbf{x}}_{gq}-\boldsymbol{\Sigma}_w\boldsymbol{\mu}_{gq})'\boldsymbol{\Sigma}_w^{-1}(\bar{\mathbf{x}}_{gs}-\boldsymbol{\Sigma}_w\boldsymbol{\mu}_{gs})\right).$$

Taking the derivative of the log-likelihood with respect to $\boldsymbol{\Sigma}_w$ yields:

$$\frac{\delta l(\mathbf{X}|\mu,\Sigma)}{\delta\boldsymbol{\Sigma}_w} = -nt\boldsymbol{\Sigma}_w^{-1} + \frac{nt}{2}\operatorname{diag}(\boldsymbol{\Sigma}_w^{-1}) + \sum_{g=1}^{m}\sum_{i=1}^{n_g}\sum_{q=1}^{t}\sum_{s=1}^{t}\mathbf{A}^{-1}[q,s]\boldsymbol{\Sigma}_w^{-1}(\mathbf{x}_{is}-\bar{\mathbf{x}}_{gs})(\mathbf{x}_{iq}-\bar{\mathbf{x}}_{gq})'\boldsymbol{\Sigma}_w^{-1}$$
$$-\tfrac{1}{2}\operatorname{diag}\left(\sum_{g=1}^{m}\sum_{i=1}^{n_g}\sum_{q=1}^{t}\sum_{s=1}^{t}\mathbf{A}^{-1}[q,s]\boldsymbol{\Sigma}_w^{-1}(\mathbf{x}_{is}-\bar{\mathbf{x}}_{gs})(\mathbf{x}_{iq}-\bar{\mathbf{x}}_{gq})'\boldsymbol{\Sigma}_w^{-1}\right)$$
$$+\sum_{g=1}^{m} n_g \sum_{q=1}^{t}\sum_{s=1}^{t}\mathbf{A}^{-1}[q,s]\boldsymbol{\Sigma}_w^{-1}\bar{\mathbf{x}}_{gs}\bar{\mathbf{x}}_{gq}'\boldsymbol{\Sigma}^{-1} - \tfrac{1}{2}\operatorname{diag}\left(\sum_{g=1}^{m} n_g \sum_{q=1}^{t}\sum_{r=1}^{t}\mathbf{A}^{-1}[q,s]\boldsymbol{\Sigma}_w^{-1}\bar{\mathbf{x}}_{gs}\bar{\mathbf{x}}_{gq}'\boldsymbol{\Sigma}_w^{-1}\right)$$
$$-\tfrac{1}{2}\sum_{g=1}^{m} n_g \sum_{q=1}^{t}\sum_{s=1}^{t}\mathbf{A}^{-1}[q,s]\boldsymbol{\mu}_{gr}\boldsymbol{\mu}_{gq}'\ .$$

To obtain the estimating equations one sets these derivatives equal to a matrix of zeros. To put the equations in a simpler form first multiply through by $\boldsymbol{\Sigma}_w$. Then note that the last term above equals the following expression:

$$-\tfrac{1}{2}\sum_{g=1}^{m} n_g \sum_{q=1}^{t}\sum_{s=1}^{t}\mathbf{A}^{-1}[q,s]\boldsymbol{\mu}_{gs}\boldsymbol{\mu}_{gq}' =$$
$$-\tfrac{1}{2}\left(\sum_{g=1}^{m} n_g \sum_{q=1}^{t}\sum_{s=1}^{t}\mathbf{A}^{-1}[q,s]\boldsymbol{\Sigma}_w\boldsymbol{\mu}_{gs}\boldsymbol{\mu}_{gq}'\boldsymbol{\Sigma}_w + \operatorname{diag}\left(\sum_{g=1}^{m} n_g \sum_{q=1}^{t}\sum_{s=1}^{t}\mathbf{A}^{-1}[q,s]\boldsymbol{\Sigma}_w\boldsymbol{\mu}_{gs}\boldsymbol{\mu}_{gq}'\boldsymbol{\Sigma}_w\right)\right)$$
$$+\tfrac{1}{4}\operatorname{diag}\left(\sum_{g=1}^{m} n_g \sum_{q=1}^{t}\sum_{s=1}^{t}\mathbf{A}^{-1}[q,s]\boldsymbol{\Sigma}_w\boldsymbol{\mu}_{gs}\boldsymbol{\mu}_{gq}'\boldsymbol{\Sigma}_w + \operatorname{diag}\left(\sum_{g=1}^{m} n_g \sum_{q=1}^{t}\sum_{s=1}^{t}\mathbf{A}^{-1}[q,s]\boldsymbol{\Sigma}_w\boldsymbol{\mu}_{gs}\boldsymbol{\mu}_{gq}'\boldsymbol{\Sigma}_w\right)\right).$$

Then the estimating equations are solved when equation (8.13) below is solved. Note that $\boldsymbol{\Sigma}_w$ is on both sides of the equations.

$$nt\boldsymbol{\Sigma}_{\mathrm{w}} = \sum_{\mathrm{g}=1}^{m}\sum_{\mathrm{i}=1}^{n_{\mathrm{g}}}\sum_{\mathrm{q}=1}^{t}\sum_{\mathrm{s}=1}^{t}\mathbf{A}^{-1}[\mathrm{q,s}](\mathbf{x}_{\mathrm{is}}-\bar{\mathbf{x}}_{\mathrm{gs}})(\mathbf{x}_{\mathrm{iq}}-\bar{\mathbf{x}}_{\mathrm{gq}})' + \sum_{\mathrm{g}=1}^{m} n_{\mathrm{g}}\sum_{\mathrm{q}=1}^{t}\sum_{\mathrm{s}=1}^{t}\mathbf{A}^{-1}[\mathrm{q,s}]\bar{\mathbf{x}}_{\mathrm{gr}}\bar{\mathbf{x}}_{\mathrm{gq}}'$$
$$-\tfrac{1}{2}\sum_{\mathrm{g}=1}^{m} n_{\mathrm{g}}\sum_{\mathrm{q}=1}^{t}\sum_{\mathrm{s}=1}^{t}\mathbf{A}^{-1}[\mathrm{q,s}]\boldsymbol{\Sigma}_{\mathrm{w}}\boldsymbol{\mu}_{\mathrm{gs}}\boldsymbol{\mu}_{\mathrm{gq}}'\boldsymbol{\Sigma}_{\mathrm{w}} - \tfrac{1}{2}\mathrm{diag}\left(\sum_{\mathrm{g}=1}^{m} n_{\mathrm{g}}\sum_{\mathrm{q}=1}^{t}\sum_{\mathrm{s}=1}^{t}\mathbf{A}^{-1}[\mathrm{q,s}]\boldsymbol{\Sigma}_{\mathrm{w}}\boldsymbol{\mu}_{\mathrm{gs}}\boldsymbol{\mu}_{\mathrm{gq}}'\boldsymbol{\Sigma}_{\mathrm{w}}\right). \quad (8.13)$$

If one wants to estimate the covariance structure assuming orthogonal variates, that is, assuming the CVA/time (orth.) model, then the estimating equations are solved when $\boldsymbol{\Sigma}_{\mathrm{w}}$ equals the expression below. The derivation of this equation involves a modification of the derivation of equation (8.13).

$$\boldsymbol{\Sigma}_{\mathrm{w}} = (nt)^{-1}\sum_{\mathrm{g}=1}^{m}\sum_{\mathrm{i}=1}^{n_{\mathrm{g}}}\sum_{\mathrm{q}=1}^{t}\sum_{\mathrm{s}=1}^{t}\mathrm{ai}_{\mathrm{qr}}(\mathbf{x}_{\mathrm{is}}-\bar{\mathbf{x}}_{\mathrm{gs}})(\mathbf{x}_{\mathrm{iq}}-\bar{\mathbf{x}}_{\mathrm{gq}})' + \sum_{\mathrm{g}=1}^{m} n_{\mathrm{g}}\sum_{\mathrm{q}=1}^{t}\sum_{\mathrm{s}=1}^{t}\mathrm{ai}_{\mathrm{qs}}(\bar{\mathbf{x}}_{\mathrm{gs}}-\boldsymbol{\mu}_{\mathrm{gs}})(\bar{\mathbf{x}}_{\mathrm{gq}}-\boldsymbol{\mu}_{\mathrm{gq}})'.$$

8.5.3 Estimating the Matrix of Proportionality Constants (A)

To solve for **A** it will be necessary to define some new notation. Let $\mathbf{x}_{\mathrm{i(b)}}$ be the $t \times 1$ vector whose r^{th} element is the b^{th} variable of the i^{th} observation at the r^{th} occasion. In other words, $\mathbf{x}_{\mathrm{i(b)}}$ is composed of the measurements made over the *t* occasions of the b^{th} variable for the i^{th} subject. Let $\bar{\mathbf{x}}_{\mathrm{g(b)}}$ be the analogous for the g^{th} group mean. Then the log-likelihood becomes as follows:

$$l(\mathbf{X}|\mu,\Sigma) = \frac{-npt}{2}\log(2\pi) - \frac{nt}{2}\log|\boldsymbol{\Sigma}_{\mathrm{w}}| - \frac{np}{2}\log|\mathbf{A}|$$
$$-\tfrac{1}{2}\left(\sum_{\mathrm{g}=1}^{m}\sum_{\mathrm{i}=1}^{n_{\mathrm{g}}}\sum_{\mathrm{j}=1}^{p}\sum_{\mathrm{v}=1}^{p}\boldsymbol{\Sigma}_{\mathrm{w}}^{-1}[\mathrm{j,v}](\mathbf{x}_{\mathrm{i(j)}}-\bar{\mathbf{x}}_{\mathrm{g(j)}})'\mathbf{A}^{-1}(\mathbf{x}_{\mathrm{i(v)}}-\bar{\mathbf{x}}_{\mathrm{g(v)}})' + \sum_{\mathrm{g}=1}^{m} n_{\mathrm{g}}\sum_{\mathrm{j}=1}^{p}\sum_{\mathrm{v}=1}^{p}\boldsymbol{\Sigma}_{\mathrm{w}}^{-1}[\mathrm{j,v}](\bar{\mathbf{x}}_{\mathrm{g(j)}}-\boldsymbol{\mu}_{\mathrm{g(j)}})'\mathbf{A}^{-1}(\bar{\mathbf{x}}_{\mathrm{g(v)}}-\boldsymbol{\mu}_{\mathrm{g(v)}})\right)$$

The **A** matrix needs to be restrained for scale. The simplest way to do this is to set

$$\mathrm{trace}(\mathbf{A}) - p = 0,$$

which constrains the within-groups covariance matrix to be proportional over time. The constraint with Lagrangian multiplier is $\lambda(\mathrm{trace}(\mathbf{A}) - p)$, where λ is the Lagrangian multiplier. To obtain the estimating equations for **A** take the derivative with respect to **A** of the log-likelihood modified by the constraint with the Lagrangian multipliers yielding (8.14). Set this equal to a $t \times t$ matrix of zeros to get the estimating equations.

$$\frac{\delta l^*(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})}{\delta\mathbf{A}} = -np\mathbf{A}^{-1} + np\,\mathrm{diag}(\mathbf{A}^{-1}) + \sum_{g=1}^{m}\sum_{i=1}^{n_g}\sum_{j=1}^{p}\sum_{v=1}^{p}\boldsymbol{\Sigma}_w^{-1}[j,v]\mathbf{A}^{-1}(\mathbf{x}_{i(j)} - \bar{\mathbf{x}}_{g(j)})(\mathbf{x}_{i(v)} - \bar{\mathbf{x}}_{g(v)})'\mathbf{A}^{-1}$$
$$-\tfrac{1}{2}\mathrm{diag}\left(\sum_{g=1}^{m}\sum_{i=1}^{n_g}\sum_{j=1}^{p}\sum_{v=1}^{p}\boldsymbol{\Sigma}_w^{-1}[j,v]\mathbf{A}^{-1}(\mathbf{x}_{i(j)} - \bar{\mathbf{x}}_{g(j)})(\mathbf{x}_{i(v)} - \bar{\mathbf{x}}_{g(v)})'\mathbf{A}^{-1}\right)$$
$$+\sum_{g=1}^{m} n_g \sum_{j=1}^{p}\sum_{v=1}^{p}\boldsymbol{\Sigma}_w^{-1}[j,v]\mathbf{A}^{-1}(\bar{\mathbf{x}}_{g(j)} - \boldsymbol{\mu}_{g(j)})(\bar{\mathbf{x}}_{g(v)} - \boldsymbol{\mu}_{g(v)})'\mathbf{A}^{-1}$$
$$-\tfrac{1}{2}\mathrm{diag}\left(\sum_{g=1}^{m} n_g \sum_{j=1}^{p}\sum_{v=1}^{p}\boldsymbol{\Sigma}_w^{-1}[j,v]\mathbf{A}^{-1}(\bar{\mathbf{x}}_{g(j)} - \boldsymbol{\mu}_{g(j)})(\bar{\mathbf{x}}_{g(v)} - \boldsymbol{\mu}_{g(v)})'\mathbf{A}^{-1}\right) - \lambda\mathbf{I}_{p\times p}. \qquad (8.14)$$

If one wants to hypothesize that the within-groups covariance matrices are constant at each occasion, then one would restrict **A** to have ones as its diagonal elements. Let $\mathbf{h}_{(i)}$ denote a vector that has a one as its i^{th} element and zeros as the rest; i.e. $\mathbf{h}_{(i)}[j] = 1$ if $j = i$, else $\mathbf{h}_{(i)}[j] = 0$ if $j \neq i$. Then this restriction is equivalent to requiring:

$$\mathbf{h}'_{(i)}\mathbf{A}\mathbf{h}_{(i)} = 1 \text{ for } i = 1,\ldots,p.$$

The p constraints with p Lagrangian multipliers are as follows:

$$\sum_{i=1}^{p}\lambda_i\left(\mathbf{h}'_{(i)}\mathbf{A}\mathbf{h}_{(i)} - 1\right) = 0.$$

When differentiated with respect to **A** this expression yields a diagonal matrix **T** whose i^{th} diagonal element is λ_i. Hence the estimating equations for solving for **A** with this method are the same as in (8.14), except that one substitutes **T** for $\lambda\mathbf{I}_{p\times p}$.

8.5.4 Hypothesis Test for the Simple Structure of the Covariance Matrix (Σ)

One may wish to test the hypothesis of a simple structure for $\boldsymbol{\Sigma}$, that is $H_0: \boldsymbol{\Sigma} = \mathbf{A}\otimes\boldsymbol{\Sigma}_w$, versus the alternative: $H_1: \boldsymbol{\Sigma} \neq \mathbf{A}\otimes\boldsymbol{\Sigma}_w$. Such a test can be performed in two contexts. The first is more consistent with the rest of Section **8.5** but may not be practical. One assumes a specific mean structure, obtains estimates for both the structured and unstructured $\boldsymbol{\Sigma}$, and then performs a likelihood ratio test based on those estimates. The difficulty with this approach is that one usually does not know the means model. This is particularly troublesome since the estimate of $\boldsymbol{\Sigma}_w$ is sensitive to misspecification of the means model, as is seen in equations (8.8) and (8.13). Indeed, a more robust estimate of $\boldsymbol{\Sigma}_w$ may be desirable even if one does not intend to perform the hypothesis test for simple structure.

One can obtain such a robust approach by assuming a saturated model for $\boldsymbol{\mu}_g$; i.e., let $\bar{\mathbf{x}}_g$ be the estimate for $\boldsymbol{\mu}_g$. The log-likelihood for such a model is obtained by substituting the model for the structure of $\boldsymbol{\Sigma}$ (8.12) into the likelihood equation in its general form (8.4), yielding:

$$l(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{-np}{2}\log(2\pi) - \frac{nt}{2}\log|\boldsymbol{\Sigma}_{\mathrm{w}}| - \frac{np}{2}\log|\mathbf{A}|$$
$$-\tfrac{1}{2}\left(\sum_{\mathrm{g=1}}^{m}\sum_{\mathrm{i=1}}^{n_{\mathrm{g}}}\sum_{\mathrm{q=1}}^{t}\sum_{\mathrm{s=1}}^{t}\mathbf{A}^{-1}[\mathrm{q,s}](\mathbf{x}_{\mathrm{iq}}-\bar{\mathbf{x}}_{\mathrm{gq}})'\boldsymbol{\Sigma}_{\mathrm{w}}^{-1}(\mathbf{x}_{\mathrm{is}}-\bar{\mathbf{x}}_{\mathrm{gs}}) + \sum_{\mathrm{g=1}}^{m} n_{\mathrm{g}}\sum_{\mathrm{q=1}}^{t}\sum_{\mathrm{s=1}}^{t}\mathbf{A}^{-1}[\mathrm{q,s}](\bar{\mathbf{x}}_{\mathrm{gq}}-\boldsymbol{\mu}_{\mathrm{gq}})'\boldsymbol{\Sigma}_{\mathrm{w}}^{-1}(\bar{\mathbf{x}}_{\mathrm{gs}}-\boldsymbol{\mu}_{\mathrm{gs}})\right).$$

The estimating equations that result are seen to be reasonable. The estimate of the unconstrained variance-covariance reduces to **S**; see (8.8). The derivation of the estimate for the structured variance-covariance matrix is similar to, though simpler than, the derivation in Section **8.5.2**, and yields:

$$\boldsymbol{\Sigma}_{\mathrm{w}} = (nt)^{-1}\sum_{\mathrm{g=1}}^{m}\sum_{\mathrm{i=1}}^{n_{\mathrm{g}}}\sum_{\mathrm{q=1}}^{t}\sum_{\mathrm{s=1}}^{t}\mathbf{A}^{-1}[\mathrm{q,s}](\mathbf{x}_{\mathrm{is}}-\bar{\mathbf{x}}_{\mathrm{gs}})(\mathbf{x}_{\mathrm{iq}}-\bar{\mathbf{x}}_{\mathrm{gq}})',$$

where the matrix **A** is estimated as in Section **8.5.3**. This estimate for $\boldsymbol{\Sigma}_{\mathrm{w}}$ is just the sum of the $p \times p$ submatrices of $\mathbf{S}$ weighted by the appropriate element of $\mathbf{A}^{-1}$.

A reservation needs to be made about this hypothesis test. It is possible that despite rejecting the null hypothesis of structure for $\boldsymbol{\Sigma}$, a researcher may conclude the deviations from this structure are not of practical significance. Instead, the researcher may prefer to assume a simple structure for $\boldsymbol{\Sigma}$ to obtain a (crude) scale invariance or to reduce the number of parameters that need to be estimated.

8.5.5 Estimating the Canonical Variates and the Group Scores

Next I develop the estimating equations for the canonical variates, $\mathbf{v}_{\mathrm{i}}$ and $\mathbf{v}_{\mathrm{i}}^{\mathrm{q}}$, and the group scores $\mathbf{e}_{\mathrm{g,i}}$. $\boldsymbol{\mu}_{\mathrm{o}}$ is estimated as in equation (8.8). Denote the log-likelihood by $l(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ and the terms which include neither $\mathbf{v}_{\mathrm{i}}$, $\mathbf{v}_{\mathrm{i}}^{\mathrm{q}}$ nor $\mathbf{e}_{\mathrm{g,i}}$ by C. Then:

$$l(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \mathrm{C} - \tfrac{1}{2}\sum_{\mathrm{g=1}}^{m} n_{\mathrm{g}}\sum_{\mathrm{q=1}}^{t}\sum_{\mathrm{s=1}}^{t}\left(\sum_{\mathrm{a=1}}^{c}\sum_{\mathrm{b=1}}^{c}\mathbf{A}^{-1}[\mathrm{q,s}]\mathrm{e}_{\mathrm{g,a}}^{\mathrm{q}}\mathbf{v}_{\mathrm{a}}'\boldsymbol{\Sigma}_{\mathrm{w}}\mathbf{v}_{\mathrm{b}}\mathrm{e}_{\mathrm{g,b}}^{\mathrm{s}} + 2\sum_{\mathrm{a=1}}^{c}\sum_{\mathrm{b}=c+1}^{u}\mathbf{A}^{-1}[\mathrm{q,s}]\mathrm{e}_{\mathrm{g,a}}^{\mathrm{q}}\mathbf{v}_{\mathrm{a}}'\boldsymbol{\Sigma}_{\mathrm{w}}\mathbf{v}_{\mathrm{b}}^{\mathrm{s}}\mathrm{e}_{\mathrm{g,b}}^{\mathrm{s}}\right.$$
$$\left. + \sum_{\mathrm{a}=c+1}^{u}\sum_{\mathrm{b}=c+1}^{u}\mathbf{A}^{-1}[\mathrm{q,s}]\mathrm{e}_{\mathrm{g,a}}^{\mathrm{q}}\mathbf{v}_{\mathrm{a}}^{\mathrm{q}}\boldsymbol{\Sigma}_{\mathrm{w}}\mathbf{v}_{\mathrm{b}}^{\mathrm{s}}\mathrm{e}_{\mathrm{g,b}}^{\mathrm{s}}\right)$$
$$+\tfrac{1}{2}\sum_{\mathrm{g=1}}^{m} n_{\mathrm{g}}\sum_{\mathrm{q=1}}^{t}\sum_{\mathrm{s=1}}^{t}\left(2\mathbf{A}^{-1}[\mathrm{q,s}]\sum_{\mathrm{b=1}}^{c}(\bar{\mathbf{x}}_{\mathrm{g}}^{\mathrm{q}}-\boldsymbol{\mu}_{0}^{\mathrm{q}})'\mathbf{v}_{\mathrm{b}}\mathrm{e}_{\mathrm{g,b}}^{\mathrm{s}} + 2\mathbf{A}^{-1}[\mathrm{q,s}]\sum_{\mathrm{b}=c+1}^{\mathrm{k}}(\bar{\mathbf{x}}_{\mathrm{g}}^{\mathrm{q}}-\boldsymbol{\mu}_{0}^{\mathrm{q}})'\mathbf{v}_{\mathrm{b}}^{\mathrm{s}}\mathrm{e}_{\mathrm{g,b}}^{\mathrm{s}}\right), \quad (8.15)$$

where $\boldsymbol{\mu}_0^{\mathrm{q}}$ is the $p \times 1$ vector of overall means for the q^{th} occasion and $\bar{\mathbf{x}}_{\mathrm{g}}^{\mathrm{q}}$ is the $p \times 1$ vector of sample means for the g^{th} group.

I start by deriving the equations for the common variates. The constraints are incorporated by the method of Lagrangian multipliers. Note that these and subsequent constraints with Lagrangian multipliers are implicitly set to zero. The constraints with Lagrangian multipliers for the unit length of the common variates are as follows below.

$$\sum_{\mathrm{a=1}}^{\mathrm{c}}\frac{\gamma_{\mathrm{a}}}{2}\left(\mathbf{v}_{\mathrm{a}}'\boldsymbol{\Sigma}_{\mathrm{w}}\mathbf{v}_{\mathrm{a}}-1\right),$$

where γ_a are c Lagrangian multipliers. The constraints with Lagrangian multipliers for the orthogonality of the common variates are:

$$\sum_{a=1}^{c}\sum_{b=1}^{a-1}\gamma_{ab}\mathbf{v}_a'\boldsymbol{\Sigma}_w\mathbf{v}_b\ ,$$

where γ_{ab} are $c(c-1)/2$ Lagrangian multipliers. The constraints with Lagrangian multipliers for the orthogonality of each common variate with all of the unique variates are:

$$\sum_{q=1}^{t}\sum_{a=1}^{c}\sum_{b=c+1}^{u}\gamma_{abq}\mathbf{v}_a'\boldsymbol{\Sigma}_w\mathbf{v}_b^q\ , \tag{8.16}$$

where γ_{abq} are $ct(u-c)$ Lagrangian multipliers. The constraints with Lagrangian multipliers for the restriction to unit length of the unique variates are:

$$\sum_{a=c+1}^{u}\sum_{q=1}^{t}\frac{\gamma_{aq}}{2}\left(\mathbf{v}_a^{q\prime}\boldsymbol{\Sigma}_w\mathbf{v}_a^q-1\right),$$

where γ_{aq} are $(u-c)t$ Lagrangian multipliers. The constraints with Lagrangian multipliers for the mutual orthogonality of each unique variate with the other unique variates of the same occasion are:

$$\sum_{q=1}^{t}\sum_{a=c+1}^{u}\sum_{b=c+1}^{a-1}\gamma_{abq}\mathbf{v}_a^{q\prime}\boldsymbol{\Sigma}_w\mathbf{v}_b^q\ ,$$

where γ_{abq} are $t(u-c)(u-c-1)/2$ Lagrangian multipliers.

Denote the log-likelihood modified by constraints with Lagrangian multipliers by $l^*(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})$. Take the derivative of $l^*(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ with respect to $\mathbf{v}_f$:

$$\frac{\delta l^*(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})}{\delta\mathbf{v}_f}=\sum_{g=1}^{m}n_g\sum_{q=1}^{t}\sum_{s=1}^{t}\mathrm{ai}_{qs}\left(-\sum_{a=1}^{c}e_{g,a}^q\boldsymbol{\Sigma}_w\mathbf{v}_a e_{g,f}^s-\sum_{b=c+1}^{u}e_{g,f}^q\boldsymbol{\Sigma}_w\mathbf{v}_b^s e_{g,b}^s+(\bar{\mathbf{x}}_g^q-\boldsymbol{\mu}_0^q)e_{g,f}^s\right)$$
$$+\gamma_f\boldsymbol{\Sigma}_w\mathbf{v}_f+\sum_{\substack{a=1\\a\neq f}}^{c}\gamma_{af}\boldsymbol{\Sigma}_w\mathbf{v}_a+\sum_{q=1}^{t}\sum_{b=c+1}^{u}\gamma_{fbq}\boldsymbol{\Sigma}_w\mathbf{v}_b^q\ .$$

Set this equal to a zero vector to yield the estimating equations for $\mathbf{v}_f$.

Next I derive the estimating equations for the unique variates. Take the derivative of $l^*(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ with respect to $\mathbf{v}_f^r$:

$$\frac{\delta l^*(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})}{\delta\mathbf{v}_f^r}=\sum_{g=1}^{m}n_g\sum_{q=1}^{t}\mathbf{A}^{-1}[q,r]\left(-\sum_{a=1}^{c}e_{g,a}^q\boldsymbol{\Sigma}_w\mathbf{v}_a e_{g,f}^r-\sum_{b=c+1}^{u}e_{g,b}^q\boldsymbol{\Sigma}_w\mathbf{v}_f^r e_{g,f}^r+(\bar{\mathbf{x}}_{gq}-\boldsymbol{\mu}_0^q)e_{g,f}^r\right)$$
$$+\gamma_{rf}\boldsymbol{\Sigma}_w\mathbf{v}_f^r+\sum_{s=1}^{c}\gamma_{afr}\boldsymbol{\Sigma}_w\mathbf{v}_a+\sum_{b=c+1}^{u}\gamma_{fbr}\boldsymbol{\Sigma}_w\mathbf{v}_b^r\ .$$

Set this equal to a vector of zeros to yield the estimating equations for solving for $\mathbf{v}_f^r$.

Lastly I derive estimating equations for the e_{gb}^s terms, beginning with those for the group positions corresponding to the common variates. The constraints here are $\sum_{g=1}^{m}n_g e_{g,b}^s=0$ for

$s = 1,\ldots,t$ and $b = 1,\ldots,c$. They are handled by letting $e^{s}_{m,b} = -\sum_{h=1}^{m-1} e^{s}_{h,b}$ for $s = 1,\ldots,t$, and then taking the derivative of the log-likelihood with respect to $e^{r}_{h,w}$ for $h \neq m$. Then the estimating equations for the e^{s}_{gb} are obtained by setting this derivative equal to zero:

$$\frac{\delta \boldsymbol{l}(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})}{\delta e^{r}_{h,f}} = \sum_{q=1}^{t} \mathbf{A}^{-1}[q,r]\left(-\sum_{a=1}^{c} e^{q}_{h,a}\mathbf{v}'_{f}\boldsymbol{\Sigma}\mathbf{v}_{a} - \sum_{b=c+1}^{u} e^{q}_{h,b}\mathbf{v}'_{f}\boldsymbol{\Sigma}\mathbf{v}^{s}_{b} + (\bar{\mathbf{x}}^{s\,\prime}_{h} - \boldsymbol{\mu}^{q}_{0})\mathbf{v}_{f} \right).$$

The estimating equations for the group positions corresponding to the unique variates are handled similarly to those corresponding to the common variates. The constraints and the manner of incorporating them into the estimating equations are the same as the for the common variates: $\sum_{g=1}^{m} n_g e^{s}_{g,b} = 0$ for $s = 1,\ldots,t$ and $b = c+1,\ldots,u$. Let $e^{s}_{m,b} = -\sum_{h=1}^{m-1} e^{s}_{h,b}$, for $s = 1,\ldots,t$ and $b = c+1,\ldots,u$, and take the derivative of $l(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ with respect to $e^{r}_{h,w}$ for $h \neq m$:

$$\frac{\delta l(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})}{\delta e^{r}_{h,f}} = \sum_{q=1}^{t} \mathbf{A}^{-1}[q,r]\left(2\sum_{a=1}^{c} e^{q}_{h,a}\mathbf{v}'_{a}\boldsymbol{\Sigma}_{w}\mathbf{v}^{r}_{f} + 2\sum_{b=c+1}^{u} \mathbf{v}'_{f}\boldsymbol{\Sigma}_{w}\mathbf{v}^{s}_{b}e^{s}_{h,b} + 2\bar{\mathbf{x}}^{q\,\prime}_{h}\mathbf{v}^{r}_{f} \right).$$

Set these derivatives equal to zero to obtain the estimating equations.

8.5.6 Estimating Unchanging Group Positions

It may be of interest to hypothesize that the scores for the group means on the common variates, $e^{q}_{g,a}$, do not change, i.e., are equal over occasion, and to estimate these stable scores. The unchanging score of the g^{th} group for the b^{th} common variate shall be denoted by $e_{g,b}$. The likelihood equation is obtained by substituting $e_{g,b}$ for $e^{q}_{g,b}$ in the likelihood equation for CVA/time (unc.) (8.15). The constraints and the manner of incorporating them are the same as for the $e^{q}_{g,b}$ terms in Section **8.5.5**. Taking the derivative of $l(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ with respect to $e_{h,w}$ and setting it equal to zero yields the following estimating equation for $e_{h,w}$:

$$\frac{\delta l(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})}{\delta e_{h,w}} = -n_h \sum_{q=1}^{t}\sum_{s=1}^{t}\left(\sum_{a=1}^{c} e_{h,a}\mathbf{v}'_{w}\boldsymbol{\Sigma}_{qr}\mathbf{v}_{a} + \sum_{b=c+1}^{u} e^{q}_{h,b}\mathbf{v}'_{w}\boldsymbol{\Sigma}_{qr}\mathbf{v}^{s}_{b} \right)$$
$$+ n_h \left(\sum_{q=1}^{t}\sum_{s=1}^{t}\sum_{a=1}^{c} (\bar{\mathbf{x}}^{s}_{h} - \boldsymbol{\mu}^{q}_{0})'\mathbf{v}_{w} \right).$$

The estimating equations for $\mathbf{v}_i$ and $\mathbf{v}^{s}_{i}$ are the same as those in the previous section except that $e_{g,b}$ is substituted for $e^{q}_{g,b}$.

8.6 EXAMPLE FOR CVA/TIME WITH UNCORRELATED VARIATES - SEX DIFFERENCES IN MATH ANXIETY BEFORE AND AFTER INTRODUCTORY CALCULUS

In this section I present a real data example of modeling the group means over time in the space of uncorrelated canonical variates. The example is a relatively simple one. 423 male college students and 118 female college students enrolled in an introductory calculus course at Virginia Tech were given a questionnaire at the beginning and end of the course. The questionnaire included 19 questions pertaining to "math anxiety". Math anxiety is generally construed to be a particular apprehension some students have about mathematics. The groups of interest are men and women. Thus $p = 19$, $t = 2$ and $m = 2$. The 19 questions are presented in **Table 8.4**. The responses to these questions followed the ordinal scale, as seen in **Table 8.5**. Thus the data are not normal and the inferential techniques used in the analysis are at best approximate.

Table 8.4 The Math Anxiety Questions

1. Generally, I have felt secure about attempting mathematics.
2. The thought of a math test scares me.
3. I usually have been at ease in math classes.
4. It wouldn't bother me at all to take more math classes.
5. It would make me happy to be recognized as an excellent student in math.
6. Figuring out mathematical problems does not appeal to me.
7. Math is enjoyable and stimulating to me.
8. I get a sinking feeling when I think of trying hard math problems.
9. Winning a prize in mathematics would make me feel uncomfortably conspicuous.
10. Even though I study, math seems unusually hard for me.
11. I study mathematics because I know how useful it can be.
12. I wouldn't like people to think I'm smart in math.
13. I like math puzzles.
14. I memorize math formulas and techniques but often don't understand the underlying concepts.
15. I am sure I could do advanced work in math.
16. I'm not the type to do well in math.
17 Mathematics is a worthwhile and necessary subject.
18. I'd be proud to be a top student in math.
19. I would rather have someone give me the solution to a hard math problem than solve it myself.

Table 8.5 Possible Responses

1) Agree 2) Tend to agree 3) Tend to disagree 4)Disagree

The analysis pursues both statistical inference and the interpretation of the results. The first question for statistical inference is, is there a weighted sum of the variables that distinguishes between the sexes? Now, if there is such a weighted sum, is it interpretable as a "math anxiety" construct? Further inferential questions are, is such a variate stable over time? If yes, are the scores of the group means on the variate, that is, the positions of the group means on the variate, stable over time? How does one interpret the scores or positions of the group means?

It is believed from previous research that there are differences between the sexes in math anxiety. However, it is not known whether such a construct is stable over the course of taking introductory calculus. Since there are two groups the maximum number of canonical variates is one. First the one common canonical variate and one unique variate model will be estimated. Then a hypothesis test will be performed based on the likelihood ratio test statistics with the common variate hypothesis being the null hypothesis.

The one common variate model is:

$$\boldsymbol{\mu}_g = \boldsymbol{\mu}_0 + \begin{bmatrix} e_g^1 \boldsymbol{\Sigma}_w \mathbf{v} \\ e_g^2 \boldsymbol{\Sigma}_w \mathbf{v} \end{bmatrix}, \text{ for } g = 1,2 . \qquad (8.17)$$

And the one unique variate model is:

$$\boldsymbol{\mu}_g = \boldsymbol{\mu}_0 + \begin{bmatrix} e_g^1 \boldsymbol{\Sigma}_w \mathbf{v}^1 \\ e_g^2 \boldsymbol{\Sigma}_w \mathbf{v}^2 \end{bmatrix}, \text{ for } g = 1,2 .$$

The test for common versus unique variates is stated as:

$$H_0 : \mathbf{v}^1 = \mathbf{v}^2$$
$$H_1 : \mathbf{v}^1 \neq \mathbf{v}^2 .$$

There are a total of 19 weights to be estimated for each variate, with the constraint that each variate have a variance of one. Thus the difference between the number of parameters to be estimated in the null and in the alternative hypotheses is 18, and the test statistic under the null hypothesis is distributed approximately as a chi-square with 18 degrees of freedom. The parameter estimates were obtained by solving the estimating equations given in Sections **8.5.2**, **8.5.3** and **8.5.5**. The likelihood test statistic was determined as described in Section **8.3.6**. The observed value of the test statistic is 20.8, which is not significant, so one fails to reject H_0.

Given that one has failed to reject the null hypothesis of one common variate, the next question of interest is whether the positions of the group means over time on the common variate are changing. The null hypothesis is that they are unchanging. The test for equality of group positions is stated as:

$$H_0 : e_1^1 = e_1^2 , \quad e_2^1 = e_2^2$$
$$H_1 : \text{at least one of the above is untrue.}$$

The estimating equations for the unchanging group positions are given in Section **8.5.6**. The estimates are: $e_1 = 0.1427$ and $e_2 = -0.5117$. The estimates for the changing group positions are obtained as part of the estimation of the common variate model Section **8.5.5**. Those estimates are: $e_1^1 = 0.1909$, $e_1^2 = 0.0804$, $e_2^1 = -0.6844$ and $e_2^2 = -0.2882$. (Note that the group positions for any occasion sum to zero when weighted by sample size).

Under the null hypothesis the test statistic follows a chi-square distribution with one degree of freedom. The resulting p-value is p<0.002. Hence one rejects H_0 and concludes $e_g^1 \neq e_g^2$. In other words, one concludes that the differences between the group means change over the two occasions. In this case they move closer, see **Figure 8.5**. Note that rejecting $H_0: e_g^1 = e_g^2$ obviates any need to test $H_0: e_g^1 = e_g^2 = 0$, which is the test for the existence of treatment effects.

At this point it is appropriate to examine the common canonical variates and ascertain if they arguably comprise a math anxiety construct. The interpretation is clearer when examining the structural coefficients, which are the correlations of the variates with the variables (see Section **2.2.1**). The canonical variate weights and the structural coefficients are presented in **Table 8.6**. From inspection of the structural coefficients it is apparent that the variate is correlated positively with answers that seem to indicate low math anxiety. For example, a high score on Question #4, “It wouldn’t bother me at all to take more math classes”, is arguably indicative of low math anxiety. The signs of the correlations of all 19 variables with the canonical variate are all arguably consistent with low math anxiety, though these correlations vary in magnitude.

Table 8.6 Canonical Variate Weights and Structural Coefficients

Question #	Canonical Variates	Structural Coefficients
1	-0.2007	0.0459
2	-0.2257	-0.0810
3	0.0632	0.1246
4	0.8437	0.5578
5	-0.2397	0.2101
6	0.1774	-0.2092
7	0.6225	0.4772
8	0.2444	-0.0289
9	-0.6116	-0.4668
10	-0.3278	-0.1353
11	0.0473	0.1242
12	-0.1554	-0.2091
13	0.1338	0.2921
14	-0.6087	-0.2734
15	-0.0988	0.0996
16	0.2919	-0.0620
17	-0.2328	0.0957
18	-0.1069	-0.2529
19	0.1955	-0.0220

Next, consider the estimates of the positions of the group means. Men clearly score higher on this low math anxiety construct, though the difference between the sexes diminishes over time; see **Figure 8.5** below. Note that the axis in **Figure 8.5** is the canonical variate, which by definition has a variance of one, and also that the group means have been centered at zero.

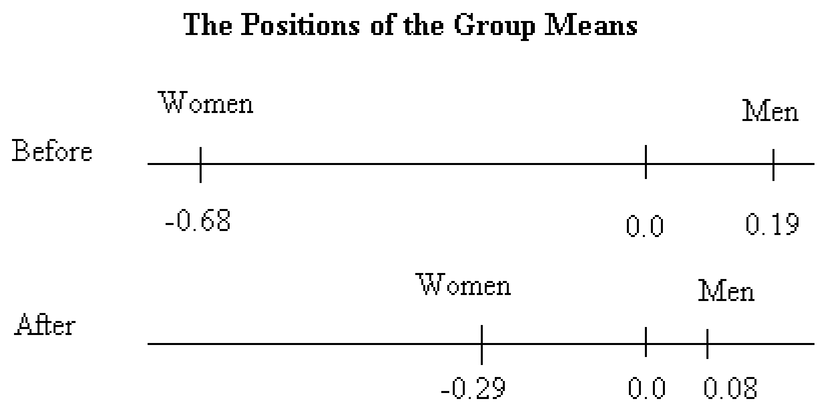

Figure 8.5

Figures 8.6, 8.7, 8.8 and **8.9** below show the group separation more clearly than does **Figure 8.5**. They are histograms of the scores for the 423 men and 118 women at both occasions. These historgrams are based on the uncentered data. As in **Figure 8.5**, the separation is greater for the first occasion.

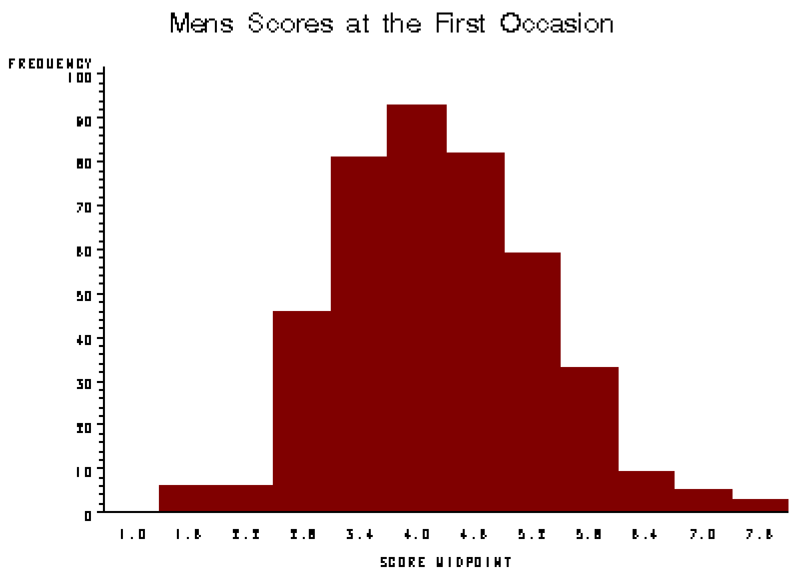

Figure 8.6

Figure 8.7

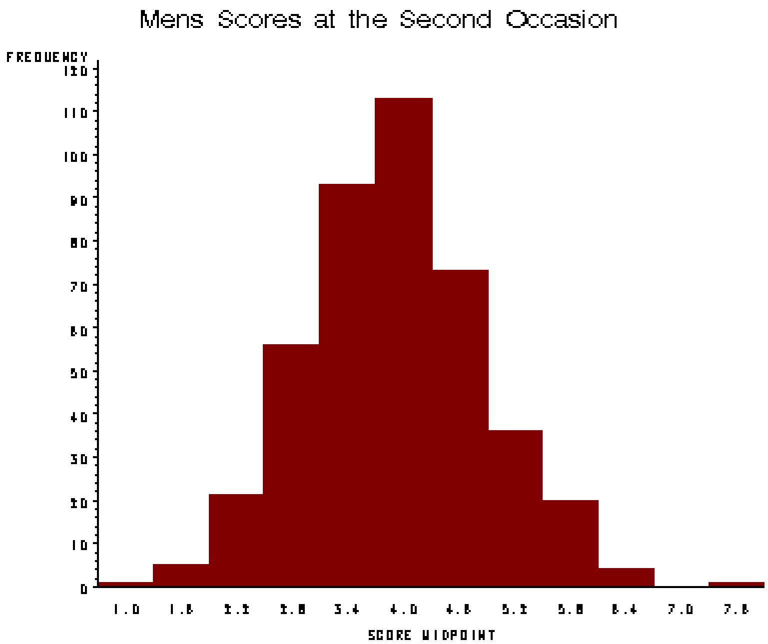

Figure 8.8

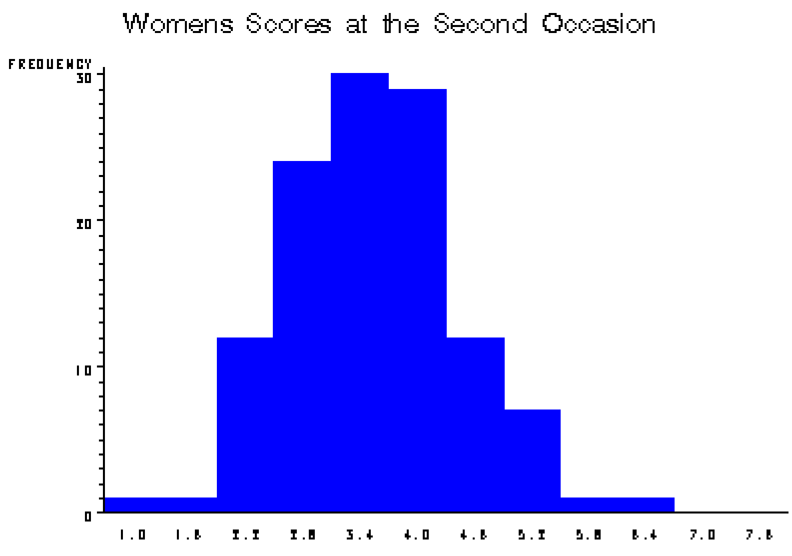

Figure 8.9

To summarize the analysis so far, one can conclude that men and women do differ on the construct of math anxiety, and that this construct is stable over time. Further, the women admit to more math anxiety, though the difference in admitted math anxiety between the sexes shrinks at the end of the course.

Next consider what the parameters mean in a geometric sense. The weights for the canonical variate are a direction in the multi-dimensional variable space. The one canonical variate model hypothesizes that the group means, when centered, will line up on this canonical variate. The group positions indicate where on these variates the group means are centered.

Another way to interpret the results is in the spirit of a multivariate regression. That is, one predicts the mean response of each variable for each group. Such an interpretation will allow one to consider the canonical variate and the group positions in conjunction. To obtain the vector of predicted group means for any occasion, apply equation (8.17). **Table 8.7** shows the observed and predicted group means of men and women on the 19 questions at the first occasion. The predicted group means are generally similar to the observed. For example, in Question 18 they are almost the same.

For comparison **Table 8.8** shows the overall group means for each question and the standard deviations. Appendix Ten has the complete sample variance-covariance matrix and the maximum likelihood estimate of the variance-covariance matrix.

Table 8.7 Observed and Predicted Means at the First Occasion

	Men		Women	
Question #	Observed	Predicted	Observed	Predicted
1	1.284	1.281	1.254	1.264
2	2.071	2.073	2.136	2.129
3	1.655	1.652	1.576	1.585
4	1.993	1.988	1.576	1.593
5	1.317	1.317	1.220	1.219
6	1.837	1.837	1.720	1.719
7	2.135	2.133	1.831	1.835
8	1.993	1.993	1.983	1.982
9	1.976	1.976	1.678	1.678
10	1.773	1.772	1.678	1.680
11	1.600	1.597	1.525	1.537
12	1.548	1.549	1.424	1.420
13	2.035	2.036	1.839	1.838
14	2.778	2.777	2.585	2.588
15	1.882	1.877	1.805	1.822
16	1.485	1.484	1.458	1.462
17	1.248	1.247	1.212	1.215
18	1.317	1.317	1.195	1.195
19	1.716	1.712	1.703	1.717

Table 8.8 Group Means and Standard Deviations for each Question

Question #	Overall Mean	Sample Variance	MLE of Variance
1	1.34288	0.591816	0.585489
2	1.98429	0.835527	0.815629
3	1.61553	0.680435	0.660555
4	1.82348	0.837711	0.822719
5	1.34843	0.607680	0.583501
6	1.83087	0.721879	0.700720
7	2.04621	0.749102	0.720939
8	1.97412	0.768150	0.741683
9	1.87061	0.770398	0.744001
10	1.78466	0.726312	0.706184
11	1.64418	0.692900	0.669706
12	1.53789	0.749156	0.751014
13	2.06285	0.859188	0.815019
14	2.78928	0.800275	0.771214
15	1.85120	0.744667	0.731749
16	1.55638	0.624593	0.608705
17	1.31978	0.497792	0.492861
18	1.33272	0.599802	0.589608
19	1.77542	0.694748	0.679944

In concluding this example, **Figure 8.10** presents the matrix of proportionality constants, **A**. What is noteworthy here is that the weights for the within-groups covariance matrices for measurements at the beginning and at the end of the course are nearly equal.

$$\begin{bmatrix} 0.986 & 0.289 \\ 0.289 & 1.014 \end{bmatrix}$$

Figure 8.10 The Matrix of Proportionality Constants (**A)**

8.7 A COMPARISON TO ALTERNATIVE METHODS, INCLUDING DOUBLY MULTIVARIATE REPEATED MEASURES

In this section I make a comparison between CVA/time and alternative methods for longitudinal multivariate data with group structure. These alternative methods attempt to answer the same questions as the common variate hypothesis does; that is to determine what is and is not changing over time. However they attempt this without the clarity and efficacy achieved by explicitly modeling common variates. The interpretations of these models for data which have the common canonical variate structure will be illuminating.

I will begin by briefly considering two simple alternative approaches. Then I will discuss in greater depth two more ambitious approaches. The first of these involves performing a canonical variate analysis with the measurements at different occasions treated as different variables. The second, which is of particular interest, is an analysis that is loosely called "doubly multivariate repeated measures".

8.7.1 Two Simple Approaches

One simple approach to longitudinal multivariate data with group structure is to perform a separate canonical variate analysis at each occasion. It is easy to see that if common canonical variate structure exists over time that the common variates will be found by each analysis. The limitations of such an approach are that one has no means to test for the appropriateness of a common variate structure, nor for the number of common variates. Furthermore, due to chance variation one does not have a single estimate for a given common variate.

Another simple approach would be to pool the measurements over time. This approach raises the question of whether one centers by an overall mean or separately at each occasion. All of the other approaches discussed in this section either explicitly or implicitly assume that one centers at each occasion. This issue aside, it is clear that pooling the variables will estimate the common variates if they exist. However, inspecting the estimated variates yields neither a hint of which (if any) of the estimated variates are common, nor what is changing over time.

8.7.2 Measurements at Different Occasions Treated as Distinct Variables

An approach one may take is to treat the measurements taken at different occasions as distinct variables in a single canonical variate analysis. A failing of this approach is that if there is an effect, i.e., a statistically significant canonical variate and an associated non-zero canonical correlation, it cannot be attributed specifically to either treatment effects or to time-treatment interaction effects. Further, neither common variates, unique variates nor group positions are estimated. However, the canonical variates one obtains do have a particular structure. The $tp \times 1$ vector of weights of each canonical variate consists of t $p \times 1$ subvectors, each of which is a linear combination of the common and unique variates. Although this point is in itself of minor interest, it is illustrative to show it.

First determine **B**, where **B** is the matrix of between-groups sums of squares and crossproducts, under the assumption that the common variates hypothesis is true. **B** is generally defined as:

$$\mathbf{B} = \sum_{\mathrm{g}=1}^{m} n_{\mathrm{g}} (\boldsymbol{\mu}_{\mathrm{g}} - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_{\mathrm{g}} - \boldsymbol{\mu}_0)'.$$

Now consider the common variates model after the data are centered:

$$\boldsymbol{\mu}_{\mathrm{g}} - \boldsymbol{\mu}_0 = \boldsymbol{\Sigma}_{\mathrm{w}} \mathbf{v}_1 \otimes \mathbf{e}_{\mathrm{g},1} + \ldots + \boldsymbol{\Sigma}_{\mathrm{w}} \mathbf{v}_c \otimes \mathbf{e}_{\mathrm{g},c},$$

for $\mathrm{g} = 1, \ldots, c$. Then the $p \times p$ submatrix of **B** corresponding to the q^{th}, s^{th} occasions, denoted as $\mathbf{B}^{\mathrm{q,s}}$, is as follows:

$$\mathbf{B}^{\mathrm{q,s}} = \sum_{\mathrm{g}=1}^{m} n_{\mathrm{g}} \left(\boldsymbol{\Sigma}_{\mathrm{w}} \mathbf{v}_1 \mathrm{e}_{\mathrm{g},1}^{\mathrm{q}} + \ldots + \boldsymbol{\Sigma}_{\mathrm{w}} \mathbf{v}_c \mathrm{e}_{\mathrm{g},c}^{\mathrm{q}} \right) \left(\boldsymbol{\Sigma}_{\mathrm{w}} \mathbf{v}_1 \mathrm{e}_{\mathrm{g},1}^{\mathrm{s}} + \ldots + \boldsymbol{\Sigma}_{\mathrm{w}} \mathbf{v}_c \mathrm{e}_{\mathrm{g},c}^{\mathrm{s}} \right)'$$

$$\mathbf{B}^{q,s} = \sum_{i=1}^{r}\sum_{j=1}^{r}\mathbf{v}_i\mathbf{v}_j'\left(\sum_{g=1}^{m} n_g e_{g,i}^{q} e_{g,j}^{s}\right) = \mathbf{V}\left(\sum_{g=1}^{m}\mathbf{d}_g^{q}\mathbf{d}_g^{s\prime}\right)\mathbf{V}',$$

where **V** is the matrix whose i^{th} column is the i^{th} canonical variate, $\mathbf{v}_i$, for $i = 1,\ldots,r$, and $\mathbf{d}_g^q$ is the $r \times 1$ vector whose i^{th} element is $n_g^{1/2} e_{g,i}^{q}$, for $i = 1,\ldots,r$ and $q = 1,\ldots,t$. If $r < p$ then complement **V** with $p - r$ canonical variates which correspond to canonical correlations of zero, making **V** a $p \times p$ matrix, and likewise complement each $\mathbf{d}_g^q$ vector with $p - r$ zeros, making them $p \times 1$ vectors. Then the implied between-groups crossproducts matrix, **B**, is

$$\mathbf{B} = (\mathbf{\Sigma}_w \otimes \mathbf{I}_{t\times t})(\mathbf{V} \otimes \mathbf{I}_{t\times t})\mathbf{D}(\mathbf{V}' \otimes \mathbf{I}_{t\times t})(\mathbf{\Sigma}_w \otimes \mathbf{I}_{t\times t}),$$

where **D** is an $pt \times pt$ matrix, $\mathbf{D} = \sum_{g=1}^{m}\mathbf{d}_g\mathbf{d}_g'$, with $\mathbf{d}_g = [\mathbf{d}_g^{1\prime}, \mathbf{d}_g^{2\prime}, \ldots, \mathbf{d}_g^{t\prime}]'$.

The canonical variates are obtained as described in Section **2.2.1**, by performing a SVD on $\frac{1}{n-1}\mathbf{\Sigma}^{-1/2\prime}\mathbf{B}\mathbf{\Sigma}^{-1/2}$, where $n = \sum_{g=1}^{m} n_g$. The first step in determining this SVD is to express $\frac{1}{n-1}\mathbf{\Sigma}^{-1/2\prime}\mathbf{B}\mathbf{\Sigma}^{-1/2}$ as follows:

$$\tfrac{1}{n}\mathbf{\Sigma}^{-1/2\prime}\mathbf{B}\mathbf{\Sigma}^{-1/2} = \tfrac{1}{n}\mathbf{\Sigma}^{-1/2\prime}(\mathbf{\Sigma}_w \otimes \mathbf{I}_{t\times t})(\mathbf{V} \otimes \mathbf{I}_{t\times t})\mathbf{D}(\mathbf{V}' \otimes \mathbf{I}_{t\times t})(\mathbf{\Sigma}_w \otimes \mathbf{I}_{t\times t})\mathbf{\Sigma}^{-1/2}. \qquad (8.18)$$

Assume $\mathbf{\Sigma} = \mathbf{A} \otimes \mathbf{\Sigma}_w$ (8.12), so $\mathbf{\Sigma}^{-1/2} = \mathbf{\Sigma}_w^{-1/2} \otimes \mathbf{A}^{-1/2}$. Recall from Section **2.2.1** that $\mathbf{V} = \mathbf{\Sigma}_w^{-1/2}\mathbf{V}^*$, where $\mathbf{V}^*$ is orthogonal. Then $\mathbf{V} \otimes \mathbf{I}_{t\times t} = \left(\mathbf{\Sigma}_w^{-1/2} \otimes \mathbf{I}_{t\times t}\right)\left(\mathbf{V}^* \otimes \mathbf{I}_{t\times t}\right)$. Substituting the above expressions back into (8.18) gives:

$$\tfrac{1}{n}\mathbf{\Sigma}^{-1/2\prime}\mathbf{B}\mathbf{\Sigma}^{-1/2} = \tfrac{1}{n}\left(\mathbf{\Sigma}_w^{-1/2\prime} \otimes \mathbf{A}^{-1/2\prime}\right)(\mathbf{\Sigma}_w \otimes \mathbf{I}_{t\times t})\left(\mathbf{\Sigma}_w^{-1/2} \otimes \mathbf{I}_{t\times t}\right)\left(\mathbf{V}^* \otimes \mathbf{I}_{t\times t}\right)\mathbf{D}$$
$$\left(\mathbf{V}^{*\prime} \otimes \mathbf{I}_{t\times t}\right)\left(\mathbf{\Sigma}_w^{-1/2} \otimes \mathbf{I}_{t\times t}\right)(\mathbf{\Sigma}_w \otimes \mathbf{I}_{t\times t})\left(\mathbf{\Sigma}_w^{-1/2} \otimes \mathbf{A}^{-1/2}\right).$$

The above simplifies to:

$$\tfrac{1}{n}\mathbf{\Sigma}^{-1/2\prime}\mathbf{B}\mathbf{\Sigma}^{-1/2} = \left(\mathbf{V}^{*\prime} \otimes \mathbf{I}_{t\times t}\right)\left[\tfrac{1}{n}\left(\mathbf{I}_{p\times p} \otimes \mathbf{A}^{-1/2\prime}\right)\mathbf{D}\left(\mathbf{I}_{p\times p} \otimes \mathbf{A}^{-1/2}\right)\right]\left(\mathbf{V}^{*\prime} \otimes \mathbf{I}_{t\times t}\right). \qquad (8.19)$$

Note that $\left(\mathbf{V}^{*\prime} \otimes \mathbf{I}_{t\times t}\right)$ is orthogonal. Let the singular value decomposition of the square-bracketed part of (8.19) be $\frac{1}{n}\left(\mathbf{I}_{p\times p} \otimes \mathbf{A}^{-1/2\prime}\right)\mathbf{D}\left(\mathbf{I}_{p\times p} \otimes \mathbf{A}^{-1/2}\right) = \mathbf{D}^*\mathbf{M}\mathbf{D}^{*\prime}$. Then $\left(\mathbf{V}^{*\prime} \otimes \mathbf{I}_{t\times t}\right)\mathbf{D}^*$ is also orthonormal because $\mathbf{D}^*$ is. Hence the matrix of canonical variates obtained, denoted by **U**, will be:

$$\mathbf{U} = \left(\mathbf{\Sigma}_{\mathrm{w}}^{-1/2} \otimes \mathbf{I}_{t\times t}\right)\left(\mathbf{V}^{*'} \otimes \mathbf{I}_{t\times t}\right)\mathbf{D}^{*} = \left(\mathbf{V}' \otimes \mathbf{I}_{t\times t}\right)\mathbf{D}^{*}. \tag{8.20}$$

Now one sees that each column of $\mathbf{U}$ is a concatenation of t subvectors which are a linear compound of **V**: $\mathbf{U}_{\mathrm{i}}^{\mathrm{q}} = \mathbf{V}\mathbf{D}_{\mathrm{i}}^{\mathrm{q}}$, where $\mathbf{U}_{\mathrm{i}}^{\mathrm{q}}$ and $\mathbf{D}_{\mathrm{i}}^{\mathrm{q}}$ are the i^{th} column and q^{th} $p \times 1$ subvector of $\mathbf{U}$ and **D**.

Next consider that unique variates can be modeled as multiple common variates, a point which is touched on briefly at the end of Section **8.3.1**. The $(u - c)$ unique variates at each occasion can be viewed as up to $t(u - c)$ common variates with the appropriate group positions; if $t(u - c) \geq p$ then the unique variates model is equivalent to a common variates model with a full complement of *p* common variates. For a simple example, assume that one has one unique variate at each of *t* occasions and that these unique variates are mutually uncorrelated. Then these unique variates can be viewed as *t* common variates; a group position on a given common variate is either the position of the original unique variate or zero, depending on whether or not the common variate corresponds to an original unique variate at the given occasion. As a rule, unique variates can always be put into the form of common variates, with the group positions conveying the change over time.

Putting a model with unique variates in the form of a model with only common variates allows one to take advantage of the earlier results for common variates, in particular to generalize (8.20). Hence one can assert that the canonical variates one obtains consist of subvectors which are linear combinations of these “common” variates; that is, of the original common and unique variates.

8.7.3 Doubly Multivariate Repeated Measures

I am aware of only one method other than CVA/time that specifically deals with longitudinal multivariate data with group structure. This is the analysis which goes by the name of doubly multivariate repeated measures, an example of which is given in the SAS-STAT Users Guide (1990). In this section I compare doubly multivariate repeated measures with CVA/time. Doubly multivariate repeated measures is the most sophisticated of the alternatives I discuss. It will be seen that it answers some but not all of the questions CVA/time answers.

The method described in the SAS/STAT User’s Guide tests for time effects, treatment effects and time-treatment interactions by performing either a standard or a modified multivariate analyses of variance (MANOVA) on transformed variables. I will review each of these tests in the context of common and unique variate structure and compare their performance to CVA/time (unc.).

First I will discuss the test for time effects. The first step in the SAS approach for testing for simple time effects is to create time profile variates. One transforms the *tp* original measurements into $(t-1)p$ variates by taking the profiles of the measurements at different occasions with respect to a baseline occasion. If the t^{th} occasion is chosen to be the baseline occasion then one would have $\mathrm{X}_{\mathrm{ij}}^{*} = \mathrm{X}_{\mathrm{ij}} - \mathrm{X}_{\mathrm{i}t}$ for $\mathrm{i} = 1, \ldots, p$ and $\mathrm{j} = 1, \ldots, t-1$. Then, disregarding group membership (treatment effects), the vector of means of these variates is tested to be zero. In SAS one does this by using the "manova" statement with "$\text{H} = \text{intercept}$" in Proc GLM.

In comparison, the CVA/time model centers the data at each occasion. This centering removes the time effect that the SAS approach tests for. However, if the examination of a simple time effect is of interest, one can perform this same test that SAS does in addition to performing a CVA/time analysis. Note that the doubly multivariate repeated measures analysis test for treatment effects and the test for treatment-time interactions also remove the simple time effect from the analysis by the transformations of the variables that they use.

The first step in the test for treatment effects given by SAS is to create transformed variables by summing up the measurements over occasion. That is, one obtains p transformed variables, X_j^*, where $X_j^* = \sum_{q=1}^{t} X_j^q$ for $i = 1,\ldots,p$ and $j = 1,\ldots,t$. Then one performs a one-way MANOVA on the transformed data.

This test for treatment effects is powerful only in the following circumstances: if the common variate hypothesis is true or approximately true; i.e., the unique variates are nearly collinear; and the scores for the group means at different occasions are equal or similar over time. Indeed, if the common variate hypothesis holds exactly and the positions of the group means are completely stable over time ($e_{g,j}^q = e_{g,k}^q$, $j \neq k$), then the doubly multivariate repeated measures finds the common variates exactly. However, to the extent that variates or group positions change over time the effects will be muddled and the test rendered ineffective. To realize these assertions one can examine the vector of expected values for the transformed variables for each group, denoted by $\boldsymbol{\mu}_g^*$, under the assumption of the common variate structure. That is, for $g = 1,\ldots,m$:

$$E\left(\sum_{q=1}^{t} \mathbf{x}_g^q\right) = \boldsymbol{\mu}_g^* = \sum_{q=1}^{t} \boldsymbol{\mu}_0^q + \boldsymbol{\Sigma}_w \mathbf{v}_1 \left(\sum_{q=1}^{t} e_{g,1}^q\right) + \ldots + \boldsymbol{\Sigma}_w \mathbf{v}_c \left(\sum_{q=1}^{t} e_{g,c}^q\right)$$

$$\boldsymbol{\mu}_g^* = \boldsymbol{\mu}_0^* + \boldsymbol{\Sigma}_w \mathbf{v}_1 e_{g,1}^* + \ldots + \boldsymbol{\Sigma}_w \mathbf{v}_c e_{g,c}^* \tag{8.21}$$

where $\mathbf{x}_g^q$ is the $p \times 1$ vector of random variables at the q^{th} occasion for the g^{th} group. The form in (8.21) is identical to the form of a canonical variate analysis with the common variates as the canonical variates (see 2.1). (Recall that a one-way MANOVA is equivalent to a canonical variate analysis; to obtain the canonical variates from SAS one requests "canonical" in the "manova" statement). Hence the method estimates the common variates if the e_{gi}^* are not zero. However, if the group positions are not at least similar over time they tend to cancel each other out, resulting in a weaker or non-detectable effect; i.e., $e_{g,i}^*$ terms that are close to zero. Hence the treatment effects may not be detected.

On the other hand, if the variates change over time, i.e., one has unique variates, one can view the analysis as one with extra common variates where the changes in the group positions convey the change over time, as described in Section **8.7.2**. Since group positions on these common variates change, the treatment effects get muddled in the summation and again the resulting treatment effect will be weaker.

The SAS test for time-treatment interactions analyzes the $(t-1)p$ variables created by taking the profiles of the measurements at different occasions with respect to a baseline occasion, as was done to analyze the time effects. Here, however, one performs a usual one-way MANOVA on the transformed data.

The test for time-treatment interactions does indeed detect change over time in the group structure. But it provides no way to determine if that change is due to changing group positions on common variates or to unique variates at different occasions. Nor does it estimate the common or unique variates. (Recall that a one-way MANOVA is equivalent to a CVA). To see these points, consider the expected values of the transformed variables if the common variate structure exists. Then the expected value of the time profile variates is a function of the common variates and the group positions. That is,

$$\begin{aligned}\mathrm{E}\left(\mathbf{x}_{\mathrm{g}}^{\mathrm{q}}-\mathbf{x}_{\mathrm{g}}^{t}\right)=\boldsymbol{\mu}_{\mathrm{g}}^{\mathrm{q}}&=\boldsymbol{\mu}_{0}^{\mathrm{q}}+\boldsymbol{\Sigma}_{\mathrm{w}}\mathbf{v}_{1}\mathrm{e}_{\mathrm{g},1}^{\mathrm{q}}+\ldots+\boldsymbol{\Sigma}_{\mathrm{w}}\mathbf{v}_{c}\mathrm{e}_{\mathrm{g},c}^{\mathrm{q}}-\left(\boldsymbol{\mu}_{0p}+\boldsymbol{\Sigma}_{\mathrm{w}}\mathbf{v}_{1}\mathrm{e}_{\mathrm{g},1}^{t}+\ldots+\boldsymbol{\Sigma}_{\mathrm{w}}\mathbf{v}_{c}\mathrm{e}_{\mathrm{g},c}^{t}\right)\\&=\left(\boldsymbol{\mu}_{0}^{\mathrm{q}}-\boldsymbol{\mu}_{0}^{t}\right)+\boldsymbol{\Sigma}_{\mathrm{w}}\mathbf{v}_{1}\left(\mathrm{e}_{\mathrm{g},1}^{\mathrm{q}}-\mathrm{e}_{\mathrm{g},1}^{t}\right)+\ldots+\boldsymbol{\Sigma}_{\mathrm{w}}\mathbf{v}_{c}\left(\mathrm{e}_{\mathrm{g},c}^{\mathrm{q}}-\mathrm{e}_{\mathrm{g},c}^{t}\right),\end{aligned}$$

for $\mathrm{q}=1,\ldots,t-1$. Now let $\boldsymbol{\mu}_{\mathrm{g}}^{*}$ be the $p(t-1)$ vector of expected values of the transformed variables; in other words the concatenation of $\boldsymbol{\mu}_{\mathrm{g}}^{\mathrm{q}}$, $\mathrm{q}=1,\ldots,t-1$. Similarly let $\boldsymbol{\mu}_{0}^{*}$ be the $p(t-1)$ vector of overall means. Then

$$\boldsymbol{\mu}_{\mathrm{g}}^{*}-\boldsymbol{\mu}_{0}^{*}=\boldsymbol{\Sigma}_{\mathrm{w}}\mathbf{v}_{1}\otimes\mathbf{e}_{\mathrm{g},1}^{*}+\ldots+\boldsymbol{\Sigma}_{\mathrm{w}}\mathbf{v}_{c}\otimes\mathbf{e}_{\mathrm{g},c}^{*}\,. \tag{8.22}$$

Observe that (8.22) has the form of a common variate model. The first implication of this is that if the common variate structure exists then the test for the time-treatment interaction is equivalent to a test for the equality of the group positions at various occasions, because if the original group positions are equal, the group positions for the transformed data will be zero. A second implication is that when one performs a CVA on the $p(t-1)$ transformed variables one has a situation similar to that in Section **8.7.2**. As argued in Section **8.7.2**, the variates generated will consist of subvectors which are linear compounds of the common variates. Furthermore, the logic of these arguments can be extended to data with unique variates as was done in Section **8.7.2** by replacing each unique variate with up to *t* common variates. By definition the group positions corresponding to these (extra) common variates cannot be equal over time. Thus this test also detects changes due to changing or unique canonical variates, though one will not be able to determine if the observed effects are due to changes in group positions or changes in unique variates.

In summary, the doubly multivariate approach answers the questions of whether there are differences among the groups, and if these differences change over (interact with) time. But it does not determine what changes over time; i.e., whether it is the variates or the group scores that change.