CHAPTER NINE

SCALING THE VARIABLES

Up to this point the scaling of the data and the (possible) scale invariance of the methods under consideration have been approached on an ad hoc basis in the examples. In this chapter I provide a more thorough discussion. To the best of my knowledge the only systematic treatment on the scale invariance of multivariate data is Jöreskog's (1989) discussion on scale invariance for the analysis of covariance structures. I try to extend his ideas to the methods considered in this dissertation. I show that my methods are often not scale invariant. Hence the issue of which scaling to employ is of obvious importance. I discuss several possible standardizations and when one would want to employ them.

9.1 AN EXAMPLE OF RESCALING THE DATA

Example 9.1

To emphasize the importance of the choice of scale, I begin with an example which shows how sensitive a solution may be to a rescaling of the data. I present a principal components analysis of the covariance matrix (a) before and after rescaling. For this example the rescaling involves multiplying the first variable by 10 and the second by five. The rescaling could be the result of changing units; for example, when one converts from millimeters to centimeters. The rescaling is equivalent to multiplying the data matrix by the diagonal matrix (b), or by pre-multiplying and post-multiplying the covariance matrix (a) by (b). (c) is the matrix whose columns are the principal components of (a), ordered by the size of their eigenvalues. Compare (c) with (d), the matrix of principal components after the data have been rescaled. There is no nontrivial way to relate the principal components derived from the unscaled data (c) and that of the rescaled data (d).

(a) $\begin{bmatrix} 15 & 9 & -7 \\ 9 & 23 & 0 \\ -7 & 0 & 6 \end{bmatrix}$

(b) $\begin{bmatrix} 10 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 0.578 & -0.598 & 0.554 \\ 0.798 & 0.558 & -0.228 \\ -0.172 & 0.573 & 0.801 \end{bmatrix}$

(d) $\begin{bmatrix} 0.926 & -0.372 & 0.061 \\ 0.375 & 0.926 & -0.048 \\ -0.039 & 0.067 & 0.996 \end{bmatrix}$

9.2 DEFINITIONS OF SCALE INVARIANCE

In point estimation for one location parameter, π, equivariance to scale is defined simply as

$$\hat{\pi}(cy_1, cy_2, \ldots, cy_n) = c\hat{\pi}(y_1, y_2, \ldots, y_n)$$

where $\hat{\pi}$ is the parameter estimate, c is a constant and y_i, $i = 1, \ldots, n$, are the data. Scale invariance needs to be defined more broadly for multivariate methods. For example, consider the relatively simple case of multiple regression. If one multiplies a regressor variable x_i by c, then $\hat{b}_i^* = \hat{b}_i / c$, where $\hat{b}_i$ and $\hat{b}_i^*$ are the estimates of the regression parameter associated with x_i before and after multiplication by c. This relationship of the parameter estimate to scale differs from that of the point estimation for a location parameter as seen above. Nevertheless, researchers consider multiple regression to be scale invariant for two reasons: because there is a simple relationship between each parameter estimate and the scale of its associated regressor variable, and because the parameter estimate for a given b_i is invariant to the scaling of those other regressors not associated with it.

Jöreskog (1989) gives a systematic discussion on scale invariance in covariance structure modeling (see Chapter Seven for a definition of covariance structure modeling). According to

Jöreskog one distinguishes between a model being scale invariant and a fit function being scale invariant. A covariance model $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scale invariant (Browne 1982) if for any diagonal matrix **D** of positive scalars and any parameter vector $\boldsymbol{\theta}$ there exists another parameter vector $\boldsymbol{\theta}^*$ such that

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}^*) = \mathbf{D}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D}.$$

Denote a fit function as $\mathrm{F}(\mathbf{S},\boldsymbol{\Sigma})$, where **S** is an observed covariance matrix and $\boldsymbol{\Sigma}$ is a predicted covariance matrix. Then $\mathrm{F}(\mathbf{S},\boldsymbol{\Sigma})$ is scale invariant (Jöreskog 1989) if for any diagonal matrix **D** of positive scalars the following is true:

$$\mathrm{F}(\mathbf{DSD},\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}) = \mathrm{F}(\mathbf{S},\boldsymbol{\Sigma})\,. \tag{9.1}$$

Maximum likelihood and generalized least squares are scale invariant fit functions; least squares is not (Jöreskog 1989). To see that generalized least squares is a scale invariant fit function, note that it minimizes the sum of squares of the deviations weighted by the inverse of the sample covariance matrix, **S**. Generalized least squares satisfies (8.1) as

$$\mathrm{Tr}\left[\mathbf{D}^{-1}\mathbf{S}^{-1}\mathbf{D}^{-1}(\mathbf{DSD}-\mathbf{D}\boldsymbol{\Sigma}\mathbf{D})\right]^2 = \mathrm{Tr}\left[\mathbf{S}^{-1}(\mathbf{S}-\boldsymbol{\Sigma})\right]^2.$$

If both the model and the fit function are scale invariant, then the analysis of the same variables in different scales yields results which are properly related; i.e. one can obtain $\hat{\boldsymbol{\theta}}^*$ from $\hat{\boldsymbol{\theta}}$ and **D**. This is because scale invariance for the fit function implies that the global optima is the same for all scalings. Scale invariance for the model then implies that the parameter estimates under the various scalings can be related.

An additional point is that a parameter estimate, $\hat{\pi}$, is defined to be scale-free if for all **D** the following holds:

$$\hat{\pi}(\mathbf{DSD}) = \hat{\pi}(\mathbf{S})\,.$$

9.3 EXAMPLES OF SCALE INVARIANT METHODS

Multiple regression is an example of a scale invariant method where there is a simple relationship between the estimates of parameters under different choices of scale. However, scale invariance as defined in the previous section does not in itself imply a simple or easily interpretable relationship between the parameter estimates before and after scaling. For example, principal components analysis is scale invariant as defined above, although there is no simple relationship between the parameter estimates before and after rescaling as seen with multiple regression in Example **8.1**. The fit function for the full principal components model is scale invariant for either least squares or maximum likelihood estimation as the fit is always perfect. The principal components model, $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{PLP}'$, where **P** is orthogonal and **L** diagonal, is scale invariant, as pre-multiplication and post-multiplication by **D** yields the following:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}^*) = \mathbf{DPLP}'\mathbf{D} = \mathbf{P}^*\mathbf{L}^*\mathbf{P}^{*\prime},$$

where $\mathbf{P}^*\mathbf{L}^*\mathbf{P}^{*\prime}$ is the singular value decomposition of $\mathbf{DPLP}'\mathbf{D}$. However, this relationship is trivial and not useful.

One can identify a class of models which will have a simple relationship between parameter estimates based on different choices of scale. If the matrices $\mathbf{F}_i$ of the model $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sum_i \mathbf{F}_i \mathbf{M}_i \mathbf{F}_i'$ are unrestricted in their column space, e.g., not constrained to be orthogonal or of unit length, then clearly, $\mathbf{F}_i^* = \mathbf{D}\mathbf{F}_i$. Furthermore, $\mathbf{M}_i$ are scale-free. Factor analysis and multiple regression can both be put into this framework.

Canonical correlation analysis is also scale invariant. Clearly its estimation by maximum likelihood is scale invariant. Further, it can be expressed as a covariance structures model as is shown below, where **W**, **E** and **V** are defined as in Section **2.2.1**:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{S}_{XX} & \mathbf{S}_{XY} \\ \mathbf{S}_{YX} & \mathbf{S}_{YY} \end{bmatrix} = \begin{bmatrix} \mathbf{W}^{-1\prime}\mathbf{W}^{-1} & \mathbf{W}^{-1\prime}\mathbf{E}\mathbf{V}^{-1} \\ \mathbf{V}^{-1\prime}\mathbf{E}\mathbf{W}^{-1} & \mathbf{V}^{-1\prime}\mathbf{V}^{-1} \end{bmatrix}.$$

Let the diagonal matrix of scale terms be $\mathbf{D} = \begin{bmatrix} \mathbf{K} & \\ & \mathbf{F} \end{bmatrix}$. Then when one pre-multiplies and post-multiplies $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ by **D** one has the following matrix:

$$\mathbf{D}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D} = \begin{bmatrix} \mathbf{K}'\mathbf{W}^{-1\prime}\mathbf{W}^{-1}\mathbf{K} & \mathbf{K}'\mathbf{W}^{-1\prime}\mathbf{E}\mathbf{V}^{-1}\mathbf{F} \\ \mathbf{F}'\mathbf{V}^{-1\prime}\mathbf{E}\mathbf{W}^{-1}\mathbf{K} & \mathbf{F}'\mathbf{V}^{-1\prime}\mathbf{V}^{-1}\mathbf{F} \end{bmatrix} = \begin{bmatrix} \mathbf{W}^{*-1\prime}\mathbf{W}^{*-1} & \mathbf{W}^{*-1\prime}\mathbf{E}^*\mathbf{V}^{*-1} \\ \mathbf{V}^{*-1\prime}\mathbf{E}^*\mathbf{W}^{*-1} & \mathbf{V}^{*-1\prime}\mathbf{V}^{*-1} \end{bmatrix}.$$

Thus $\mathbf{W}^* = \mathbf{W}\mathbf{K}^{-1}$, $\mathbf{V}^* = \mathbf{V}\mathbf{F}^{-1}$ and $\mathbf{E}^* = \mathbf{E}$. The relationship between the parameter estimates for the canonical variates before and after rescaling is the same as that of multiple regression. Note also that $\mathbf{E}$ is scale-free.

Estimating the full redundancy analysis (RA) model by least squares yields a perfect fit regardless of multiplication by **D**. Hence it is scale invariant with respect to fit function. RA can be expressed as a covariance structure model as follows:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{S}_{XX} & \mathbf{S}_{XY} \\ \mathbf{S}_{YX} & \mathbf{S}_{YY} \end{bmatrix} = \begin{bmatrix} \mathbf{W}^{-1\prime}\mathbf{W}^{-1} & \mathbf{W}^{-1\prime}\mathbf{E}\mathbf{V}' \\ \mathbf{V}\mathbf{E}\mathbf{W}^{-1} & \mathbf{V}\mathbf{E}^2\mathbf{V}' + \mathbf{J} \end{bmatrix},$$

where **W**, **R** and **V** are defined as in Section **2.2.2** and $\mathbf{J} = \mathbf{S}_{YY} - \mathbf{V}\mathbf{E}^2\mathbf{V}'$. Pre-multiply and post-multiply $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ by **D** as defined above. Then one has the following matrix:

$$\mathbf{D}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D} = \begin{bmatrix} \mathbf{K}'\mathbf{W}^{-1\prime}\mathbf{W}^{-1}\mathbf{K} & \mathbf{K}'\mathbf{W}^{-1\prime}\mathbf{E}\mathbf{V}'\mathbf{F} \\ \mathbf{F}'\mathbf{V}\mathbf{E}\mathbf{W}^{-1}\mathbf{K} & \mathbf{F}'\mathbf{V}\mathbf{E}^2\mathbf{V}'\mathbf{F} + \mathbf{F}'\mathbf{J}\mathbf{F} \end{bmatrix} = \begin{bmatrix} \mathbf{W}^{*-1\prime}\mathbf{W}^{*-1} & \mathbf{W}^{*-1\prime}\mathbf{E}^*\mathbf{V}^{*\prime} \\ \mathbf{V}^*\mathbf{E}^*\mathbf{W}^{*-1} & \mathbf{V}^*\mathbf{E}^{*2}\mathbf{V}^{*\prime} + \mathbf{J}^* \end{bmatrix}.$$

One sees RA is model scale-invariant as defined in Section **9.2** as $\mathbf{W}^*$, $\mathbf{V}^*$, $\mathbf{E}^*$ and $\mathbf{J}^*$ can be found from **W**, $\mathbf{E}$, **V** and $\mathbf{J}$. However, their relationship to **W**, **V**, $\mathbf{E}$, $\mathbf{J}$ is not simple. Nevertheless, if one rescales only the X-variables one has simple relationships between the parameter estimates as $\mathbf{W}^* = \mathbf{W}\mathbf{K}^{-1}$, $\mathbf{E}^* = \mathbf{E}$, $\mathbf{V}^* = \mathbf{V}$ and $\mathbf{J}^* = \mathbf{J}$.

The CVA/time model (7.3) presented in Chapter Seven is not scale invariant. This model can be expressed in the form $\mathbf{V}\mathbf{Q}_{ij}\mathbf{V}' = \mathbf{P}_{ij}$, where $\mathbf{P}_{ij}$ is the submatrix of the between-group covariance matrix corresponding to the covariance between the i^{th} and j^{th} occasions, and $\mathbf{Q}_{ij}$ is a diagonal matrix. If this model were to have scale invariance then the following must be true: $\mathbf{DVQ}_{\mathrm{ij}}\mathbf{V}'\mathbf{D} = \mathbf{V}^*\mathbf{Q}^*_{\mathrm{ij}}\mathbf{V}^{*\prime}$ and $\mathbf{DVQ}_{\mathrm{i'j'}}\mathbf{V}'\mathbf{D} = \mathbf{V}^*\mathbf{Q}^*_{\mathrm{i'j'}}\mathbf{V}^{*\prime}$, where $\mathrm{i, j} \neq \mathrm{i}', \mathrm{j}'$. This implies

$$\mathbf{V}^*\mathbf{Q}^*_{\mathrm{ij}}\mathbf{Q}^*_{\mathrm{i'j'}}\mathbf{V}^{*\prime} = \mathbf{DVQ}_{\mathrm{ij}}\mathbf{V}'\mathbf{DDVQ}_{\mathrm{i'j'}}\mathbf{V}'\mathbf{D}. \tag{9.2}$$

But (9.2) can be seen to be generally untrue. The term on the left is symmetric since $\mathbf{Q}_{\mathrm{ij}}$ and $\mathbf{Q}_{\mathrm{i'j'}}$ are diagonal. However, the term on the right is symmetric only if $\mathbf{Q}_{\mathrm{ij}} = \mathbf{Q}_{\mathrm{i'j'}}$ or if $\mathbf{V}'\mathbf{D}^2\mathbf{V}$ is diagonal. The former is generally not true and the latter is true only if **V** is the diagonal. Hence there is no way to relate the parameters of the model before and after a rescaling.

9.4 SCALE INVARIANCE FOR THREE-MODE PRINCIPAL COMPONENTS ANALYSIS

Kroonenberg (1983) and Harshman and Lundy (1984) have discussed the choice of scale for three-mode data. However, they never discuss scale invariance. Indeed, it seems the researchers who develop and use three-mode methods implicitly assume that scale invariance is unattainable for three-mode data. In this section I show that for certain three-mode models that a type of scale invariance or approximate scale invariance exists. The discussion of how to scale three-mode data when scale invariance does not exist is deferred until Section **9.5**.

I will provide a definition of scale invariance for three-mode PCA that is analogous to Jöreskog's definition of scale invariance for covariance structures. As in Jöreskog's development I will need to define both scale invariance for the fit function and scale invariance for the model. First of all, however, I must define what is meant by rescaling in the context of three-mode PCA. In the analysis of covariance structures there are two modes, a variables mode and an observations mode, and by definition one rescales only the variables mode. In three-mode PCA all three modes are candidates for rescaling. However, (approximate) scale invariance can exist only when one rescales one mode at a time. Hence I restrict myself to considering the rescaling of just one mode. The rescaling of three-way data is defined as follows: if one has g slices of the three-way array, $\mathbf{Z}_{\mathrm{i}}, \ldots, \mathbf{Z}_g$, $\mathrm{i} = 1, \ldots, g$, one can post-multiply them by **B**, a diagonal matrix of positive scalars. Because of symmetry the following arguments generalize to any one of the three modes being rescaled.

Having defined choice of scale in the context of three-mode PCA, I can address the issue of scale invariance for the fit function. The least squares fit function is generally not invariant to scale. However, the Tucker2 and Tucker3 models can decompose a three-mode matrix exactly, which yields scale invariance; i.e., regardless of the scaling, the fit function is zero. If one chooses a solution with less than the full complement of components one can get approximate scale invariance if the fit is good. How good the fit must be remains to be determined.

Next I consider model scale invariance for various three-mode PCA models. The Tucker2 model can be shown to be invariant to column scaling. The Tucker2 is expressed in matrix form as follows, where $\mathbf{G}$, $\mathbf{C}_{\mathrm{i}}$, $\mathbf{H}$ are defined as in Section **2.32**

$$\mathbf{Z}_{\mathrm{i}} = \mathbf{GC}_{\mathrm{i}}\mathbf{H}', \quad \text{for} \quad \mathrm{i} = 1, \ldots, g.$$

Post-multiplying the $\mathbf{Z}_i$ terms by **B** yields the following:

$$\mathbf{Z}_i\mathbf{B} = \mathbf{G}\mathbf{C}_i\mathbf{H}'\mathbf{B} = \mathbf{G}^*\mathbf{C}_i^*\mathbf{H}^{*\prime}, \quad \text{for} \quad i = 1, \ldots, g. \tag{9.3}$$

To get $\mathbf{G}^*$, $\mathbf{C}_i^*$ and $\mathbf{H}^*$ in terms of **G**, $\mathbf{C}_i$, **H** and **B**, perform a singular value decomposition on $\mathbf{H}'\mathbf{B}$, yielding $\mathbf{H}'\mathbf{B} = \mathbf{M}\mathbf{N}\mathbf{P}'$. Then (9.3) becomes

$$\mathbf{G}\mathbf{C}_i\mathbf{M}\mathbf{N}\mathbf{P}' = \mathbf{G}^*\mathbf{C}_i^*\mathbf{H}^{*\prime}, \quad \text{for} \quad i = 1, \ldots, g. \tag{9.4}$$

Thus $\mathbf{G}^* = \mathbf{G}$, $\mathbf{C}_i^* = \mathbf{C}_i\mathbf{M}\mathbf{N}$ and $\mathbf{H}^{*\prime} = \mathbf{P}'$. The model is invariant to rescaling the column space of the $\mathbf{Z}_i$ in the sense that $\mathbf{G}^*$, $\mathbf{C}_i^*$ and $\mathbf{H}^*$ can be found for all **G**, $\mathbf{C}_i$, **H** and **B**. Further, **G** is $\mathbf{G}^*$. The core matrices, $\mathbf{C}_i^*$, and the components for the column space, $\mathbf{H}^*$, however, are not related to $\mathbf{C}_i$ and **H** in any simple or useful manner. With similar logic the Tucker3 can be shown to have scale invariance properties.

Now consider the PARAFAC model, first with two sets of components restricted to orthonormality. This is the model one gets when one requires diagonal $\mathbf{C}_i$ in (9.3). Clearly, in order for a solution to exist there must be orthonormal $\mathbf{G}^*$ and $\mathbf{H}^*$ that simultaneously diagonalize $\mathbf{G}\mathbf{C}_i\mathbf{H}'\mathbf{B}$ for $i = 1, \ldots, g$. As they generally do not exist the model is not scale invariant. However, the PARAFAC model without the orthonormality constraints is scale invariant. Again referring to (9.3), one sees that one can relate $\mathbf{G}^*$, $\mathbf{C}_i^*$ and $\mathbf{H}^*$ to **G**, $\mathbf{C}_i$, **H** and **B** by letting $\mathbf{G}^* = \mathbf{G}$, $\mathbf{H}^* = \mathbf{L}\mathbf{H}\mathbf{B}$ and $\mathbf{C}_i^* = \mathbf{C}_i\mathbf{L}^{-1}$, where $\mathbf{L} = \mathrm{Diag}\left((\mathbf{H}\mathbf{B})'\mathbf{H}\mathbf{B}\right)$. Not only is **G** unchanged by the rescaling, but there is an interpretable relationship between the coefficients of $\mathbf{H}^*$ and **H** and between $\mathbf{C}_i^*$ and $\mathbf{C}_i$.

Example 9.2:

This example illustrates the approximate scale invariance of the Tucker2 method. $\mathbf{X}_1$ and $\mathbf{X}_2$ are the data matrices. These data are purposely constructed to fit the one component Tucker2 relatively poorly but to fit the two component Tucker2 excellently, with 59% of the sums of squares being explained by the former model but 99% by the latter. $\mathbf{X}_1$ and $\mathbf{X}_2$ of (a) are the data before rescaling, $\mathbf{X}_1^*$ and $\mathbf{X}_2^*$ of (b) are the data after rescaling. Note the extent to which the scaled and unscaled solutions differ for the one-component solution (c). On the other hand, the scaled and unscaled solutions for two-component matrices are recognizably similar. If the fit were perfect, then the there would exist perfect scale invariance.

(a) $$\mathbf{X}_1 = \begin{bmatrix} -19.5 & 10.5 & -7 \\ -10.5 & 19.5 & 7 \\ 7 & 7 & 2 \end{bmatrix} \qquad \mathbf{X}_2 = \begin{bmatrix} -27.\overline{6} & 10.\overline{3} & -10.\overline{6} \\ -10.\overline{3} & 27.\overline{6} & 10.\overline{6} \\ 10.\overline{6} & 10.\overline{6} & 6.\overline{6} \end{bmatrix}$$

(b) $$\mathbf{X}_1 = \begin{bmatrix} -195 & 105 & -70 \\ -10.5 & 19.5 & 7 \\ 7 & 7 & 2 \end{bmatrix} \quad \mathbf{X}_2 = \begin{bmatrix} -276.\overline{6} & 103.\overline{3} & -106.\overline{6} \\ -10.\overline{3} & 27.\overline{6} & 10.\overline{6} \\ 10.\overline{6} & 10.\overline{6} & 6.\overline{6} \end{bmatrix}$$

(c)

Original	Rescaled
0.7071	0.8680
–0.7071	-0.3760
0.0000	0.3243

(d)

Original		Rescaled	
0.7071	0.5774	0.6688	0.5906
-0.7071	0.5774	-0.7398	0.5887
0.0000	0.5774	0.07335	0.5520

9.5 HOW TO SCALE THE DATA

Since the models I propose are often not scale invariant, the choice of scale of the variables is salient. There are several plausible ways to scale the data. The simplest is to analyze the raw data. This scaling is appropriate only if the measures in their unscaled form are comparable. Some examples of this could be species counts or company sales in dollars for particular types of goods. Because measures are often not comparable, one must choose a scaling that makes them so. In many multivariate applications one “standardizes”, or scales the variables to unit length. This standardization effectively gives each variable equal importance in the modeling. Another possibility is to apply the Mahalanobis transformation to the data by post-multiplying $\mathbf{Y}$ by $\mathbf{S}_{\mathrm{YY}}^{-1/2}$.

Choosing an appropriate scaling for data over time is complicated by two things. First, the covariances may be changing over time. Hence a scaling that is appropriate for one occasion may not be appropriate for other occasions. Second, the covariance of the Y-variables is assumed to be related to the X-variables either in a causal manner or in a correlational manner. Hence one must decide whether to, and perhaps how to, remove the effect of the X-variables on the Y-variables. This issue is simplified if the X-variables are group indicators, because then one can calculate a within-groups covariance.

I first discuss the issue of standardization for the situation where the X-variables are group indicators. The simplest scenario is that the within-group covariances are assumed not to vary over time. Then one could standardize by post-multiplying $\mathbf{Y}$ by $\mathbf{D} = \mathrm{Diag}^{-1/2}(\mathbf{S}_{\mathrm{YY}})$ or apply the Mahalanobis transformation by post-multiplying $\mathbf{Y}$ by $\mathbf{S}_{\mathrm{YY}}^{-1/2}$, where $\mathbf{S}_{\mathrm{YY}}$ is a pooled estimate over time.

If the within-group covariances are assumed to vary over time the situation is more complex. One possible way to standardize such data is to choose a baseline occasion and use its within-group covariance to standardize. Another possibility is to average the variances over occasions and standardize by the average variance. That is, one can post-multiply **Y** by $\mathbf{D} = \mathrm{Diag}^{-\frac{1}{2}}((\mathbf{S}_{\mathrm{YY1}} + \mathbf{S}_{\mathrm{YY2}} + \cdots + \mathbf{S}_{\mathrm{YY3}}) / g)$, where g is the number of occasions. Such a standardization is akin to what is done in the factor analysis of multiple groups (Loehlin 1992) and is also recommended for some situations in three-mode PCA (Kroonenberg 1983). It gives each variable the same weight in the overall analysis while allowing the variances to vary over occasion. Applying the Mahalanobis transformation to the data based on a baseline or averaged covariance matrix is problematic as the transformed data will not satisfy $\mathbf{S}_{\mathrm{X^*X^*k}} = \mathbf{I}$, for $\mathrm{k} = 1,\ldots, g$, where $\mathbf{x}^* = \mathbf{S}_{\mathrm{XX}}^{-\frac{1}{2}}\mathbf{x}$.

If the X-variables are not group indicators but continuous variables the standardization is further complicated because one does not have the elegant partitioning of the variation into between-group effects and within-group effects. One can express the matrix of the total sums of squares of the Y-variables as the sum of a regression sums of squares matrix and a residuals or error matrix. Then one could standardize by the residuals matrix. Such a standardization attempts to make the error terms equal for each observation, which is analogous to what is done when one standardizes by a within-group covariance matrix. For this standardization to be reasonable one should have X-variables that are controlled by the experimenter. If the X-variables are random variables in their own right, then the X-variables and Y-variables are correlated and one may prefer to standardize **Y** based on the total covariance matrix for the Y-variables.

The analysis of covariance structures and three-mode PCA are both methods based on the decomposition of matrices. In contrast, Campbell and Tomenson's model and the CVA/time model of Chapter Eight are means models. These means cannot be put into the framework used for evaluating the scale invariance of the analysis of covariance or three-mode PCA which begins by multiplying a mode by a diagonal matrix of positive scalars. However, there is another useful way to look at these models. Campbell and Tomenson's analysis is equivalent to plotting the group means in the space transformed by the Mahalanobis transformation and then finding a reduced space of common orthogonal variates in which the means approximately lay. This method effectively transforms the data by $\mathbf{S}_{\mathrm{YY}}^{-\frac{1}{2}}$ and thus is scale invariant. Likewise, CVA/time with uncorrelated variates plots the group means in the transformed space and is scale invariant. On the other hand CVA/time with orthogonal variates does not transform the data and is not scale invariant. Hence one must standardize the data by one of the methods described in the previous paragraphs of this section.

This discussion on scaling should make clear that the researcher needs to give thoughtful consideration to the standardization and scaling he uses.