

Component Analysis  
of  
Multisubject Multivariate Longitudinal  
Data

© 2001, Marieke E. Timmerman

Printed by Stichting Drukkerij C. Regenboog, Groningen

The Netherlands Organization of Scientific Research (NWO) is gratefully acknowledged for funding this project (no. 575-30-008). The research presented in this thesis was carried out at the Heymans Institute of Psychology, University of Groningen, The Netherlands.

RIJKSUNIVERSITEIT GRONINGEN

Component Analysis  
of  
Multisubject Multivariate Longitudinal  
Data

Proefschrift  
ter verkrijging van het doctoraat in de  
Psychologische, Pedagogische en Sociologische Wetenschappen  
aan de Rijksuniversiteit Groningen  
op gezag van de  
Rector Magnificus, dr. D.F.J. Bosscher,  
in het openbaar te verdedigen op  
donderdag 5 juli 2001  
om 16.00 uur  
door  
Marieke Engelen Timmerman  
geboren op 2 mei 1972  
te Groningen

Promotores:

Prof. dr. H.A.L. Kiers

Prof. dr. J.M.F. ten Berge

Beoordelingscommissie:

Prof. dr. M.W. Browne

Prof. dr. W.J. Heiser

Prof. dr. A.K. Smilde

ISBN: 90-367-1402-8

# Contents

<b>1. Introduction</b>	<b>1</b>
1.1. Longitudinal research	1
1.2. Sampling from the longitudinal axis	1
1.3. Two types of multisubject longitudinal data: longitudinal two-way data and multiple time series	4
1.4. Two types of multivariate multisubject longitudinal data	5
<b>2. Component models for longitudinal three-way data</b>	<b>9</b>
2.1. Introduction	9
2.2. Notational issues and some matrix algebraic properties	9
2.3. Component models for three-way data	10
2.4. Fitting the component models for three-way data	13
2.5. Transformational freedom within the component models for three-way data	14
2.6. Issues in the application of the component models to three-way data	15
2.6.1. Preprocessing three-way data	15
2.6.2. Choice of a specific model and the numbers of components	15
2.7. Interpretation of three-way component models applied to longitudinal three-way data	18
<b>3. Occasion components as evaluations of latent curves: possibilities for constraints to the time mode</b>	<b>21</b>
<b>4. The CP and Tucker3 models with smoothness constraints</b>	<b>25</b>
4.1. Introduction	25
4.2. The choice of a smoother	25
4.3. How to smooth in the Tucker3 model and CP model?	28
4.4. Comparing constrained with unconstrained CP and Tucker3 models	31
4.4.1. Construction of the data for the simulation study	31
4.4.2. Analyses of simulation data	33
4.4.3. Criteria of interest	33
4.4.4. Results of the simulation studies	35
4.5. Empirical example: Learning to read study (I)	40
4.6. Discussion and conclusion	47

<b>5. Structured latent curve component models for longitudinal three-way data</b>	<b>49</b>
5.1. Introduction	49
5.2. Structured latent curve two-way component models for growth data	49
5.2.1. The SLC two-way component model for data measured at equal time points	49
5.2.2. Fitting the SLC two-way component model to data with equal measurements	53
5.2.3. The SLC two-way component model for growth data measured at unequal time points	57
5.3. The SLC Tucker3 model for longitudinal three-way data	58
5.3.1. Fitting the SLC Tucker3 model to data	59
5.3.2. Transformational freedom and interpretation in the SLC Tucker3 model	62
5.4. Empirical example: Learning to read study (II)	62
5.5. Discussion and conclusion	66
<b>6. Simultaneous Component Models of Multisubject Multivariate Time Series</b>	<b>69</b>
6.1. Introduction	69
6.2. Four models for simultaneous component analysis	70
6.2.1. Preprocessing of raw data before fitting the SCA model to data	70
6.2.2. SCA with invariant Pattern (SCA-P)	71
6.2.3. Constrained versions of SCA-P	72
6.2.4. SCA with PARAFAC2 constraints (SCA-PF2)	73
6.2.5. SCA with INDSCAL constraints (SCA-IND)	73
6.2.6. SCA with Equal average Cross-Products constraints (SCA-ECP)	74
6.2.7. Transformational freedom in the SCA-P, SCA-PF2, SCA-IND and SCA-ECP models	74
6.2.8. Model selection	75
6.3. Fitting the four SCA models to the data	77
6.3.1. Fitting the SCA-PF2 model	78
6.3.2. Fitting the SCA-IND model to data	78
6.3.3. Fitting the SCA-ECP model to data	79
6.3.4. Starting values of the parameters	79
6.4. Empirical examples of simultaneous component analyses	79
6.4.1. Empirical example 1: Mood in individuals with Parkinson's disease	80
6.4.2. Empirical example 2: The Big Five as states	86
6.5. Discussion and conclusion	91

<b>7. Lagged Simultaneous Component Models of Multisubject Multivariate Time Series</b>	<b>95</b>
7.1. Introduction	95
7.2. The dynamic factor model	95
7.3. Lagged SCA models	97
7.3.1. LSCA-P	98
7.3.2. LSCA-PF2, LSCA-IND and LSCA-ECP	99
7.3.3. Fitting the LSCA models to data	100
7.3.4. Transformational freedom in the LSCA models	105
7.3.5. Testing the LSCA-P and LSCA-IND algorithms	109
7.4. Empirical example: Mood in individuals with Parkinson's disease	114
7.5. Discussion and conclusion	115
<b>8. Conclusion</b>	<b>117</b>
8.1. Summary	117
8.2. Discussion and future work	118
<b>References</b>	<b>123</b>
<b>Summary in Dutch (Samenvatting)</b>	<b>131</b>
<b>Subject Index</b>	<b>135</b>
<b>Author Index</b>	<b>139</b>
<b>Acknowledgements</b>	<b>141</b>



# 1. Introduction

## 1.1. Longitudinal research

Longitudinal research is performed to study a phenomenon as it is evolving over time. The phenomenon will generally show changes over time, but it may also show stability. One can distinguish short-term changes and long-term changes. Short-term changes occur in relatively rapid succession, whereas long-term changes are characterized by more or less irreversible alterations. Short-term and long-term changes may also be confounded.

The process of interest can be studied for only one, or more than one observation units. The units can refer to human beings, households, countries, schools, animals, stock exchanges, chemicals, et cetera. We will indicate the observation units by 'subjects' or 'individuals' in the sequel.

One can distinguish three broad goals in studying a process (Bijleveld & Van der Kamp, 1998). In a research with the goal of description, one is interested in describing intraindividual and/or interindividual patterns of change. One may also want to explain patterns of change in terms of individual features. The features may be intrinsic to individuals (like age and socioeconomic status), but they may also be assigned by the researcher (like treatment). Furthermore, they can be stable over time or changing. Sometimes, one aims partly at description and partly at explanation in a single study. A third goal of longitudinal research is to forecast the future development of a process from previous stages. This type of forecasting is rarely aimed at in the social sciences. It is more often encountered in (business) economics, for example forecasting demand for a certain merchandise.

The key feature of longitudinal research designs is that a certain feature of a subject is observed repeatedly. Degree of depression, socioeconomic status, level of reading ability, quality of mother-child attachment, and speed of information processing are instances of features of interest in social sciences. Characteristics are measured by collecting scores on one or more variables, which are thought to be indicative for the feature. For example, the reaction time on a certain cognitive task could be used as an indicator for speed of information processing. Longitudinal data consist of scores on one or more variables from one or more individuals, measured at several time points.

## 1.2. Sampling from the longitudinal axis

The length of the observational period, the number of measurement occasions and the associated points in time should be chosen carefully. From a practical point of view, it is desirable to sample only a small number of occasions in a short observational

period. Every additional measurement occasion increases the research costs, simply because more data are collected. Moreover, the repeated measurement of the *same* subjects may lead to difficulties, because it may be hard to trace subjects or to motivate subjects to continue participating. From a theoretical point of view, on the other hand, it is desirable to measure the phenomenon under study frequently. One should choose the sampling points cleverly to find a balance between practical and theoretical interests. Three considerations are important in this respect. First, one should consider the aim of the research. If a researcher wants to gain a highly detailed insight into the process, a large number of samples is needed. In the most extreme case, one needs continuous sampling. In practice, researchers usually content themselves with a rough approximation of the process. A second point of concern is the (expected) pattern of change over time. This pattern will be denoted by the *functional form* in the sequel. An intricate functional form needs more sampling points than a simple one (e.g., a linear trend). If the functional form changes quickly in a certain period, one should sample more frequently in that period. Thirdly, if one is willing to assume a certain functional form, the expected size of (random) measurement error should play a role in determining the number of sampling points, as will be discussed later. To illustrate the first and second point, the true scores (i.e., without measurement error) of one subject on two fictitious variables during a certain period are plotted in Figures 1.1 and 1.2.

The solid line in Figure 1.1 might, for example, indicate the degree of depression during seven consecutive days. A short-term and a long-term change can be distinguished. Per day, the degree of depression follows a sinusoidal function, whereas a linear trend can be seen over days. Sampling daily at exactly the same hour (e.g., at the time points indicated by the circles in Figure 1.1) would reveal the linearly increasing trend in depression over days. If one were interested only in daily fluctuations of depression, it would be sufficient to collect scores within one day to discover the sinusoidal form.

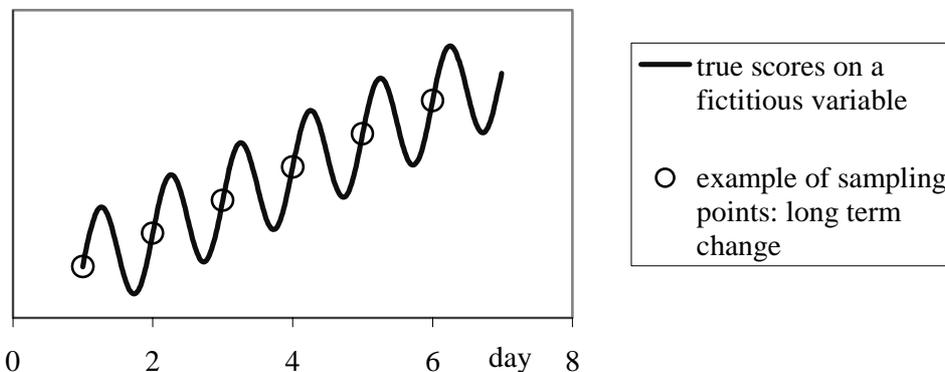


Figure 1.1. Example of true scores (i.e., without measurement error) of one subject on a fictitious variable over time: a sinusoidal function combined with a linear trend.

The number of sampling points that is needed depends on the degree of detail the researcher is interested in, and on the assumptions that can reasonably be made about the functional form of the process under study. The latter assumptions serve two purposes, namely as a guide in making a rational choice of the sampling points, and as a guide in interpolating between the sampled time points. To give an extreme and unrealistic example: if a researcher assumes linear growth in depression over days, and the interest focuses on the degree of long-term increase only, it would be sufficient to measure only twice in the whole process at exactly the same hour to estimate the ‘correct’ trend coefficient. If a research is focussed on short term fluctuations only, and the researcher is willing to assume that daily fluctuations are equal across days, it is sufficient to sample within one day. If no a priori knowledge about the functional form of the process is available, the degree of detail the researcher is interested in is the only remaining criterion. However, usually in practice, there is reason to assume a certain functional form.

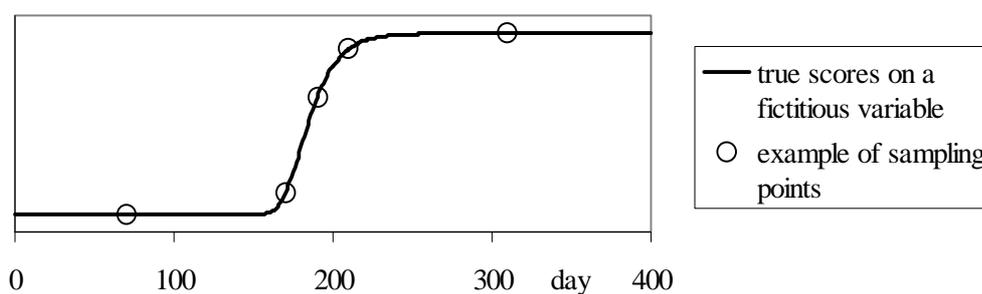


Figure 1.2. Example of true scores (i.e., without measurement error) of one subject on a fictitious variable over time: a Gompertz function.

Figure 1.2 could represent the true scores of a subject on reading ability, that follows a Gompertz curve (Richards, 1959). The plot clearly shows that taking five equidistant samples between day 0 and 400 would be inefficient, because then the steeply increasing part of the curve from day 175 to day 225 is hardly covered and the almost constant scores between days 0 and 175, and 225 and 400 are overrepresented. The number of samples needed to estimate the functional form reasonably well again depends on the aim of the research, and on the assumptions concerning the functional form a researcher can reasonably make. A researcher will usually not be satisfied with a simple approximation of the learning process, for example by measuring at days 1, 200 and 400, and then estimating the best fitting linear trend. A researcher would probably be reasonably satisfied with the scores obtained from sampling at the time points indicated by circles in Figure 1.2.

A third consideration in deciding about the sampling points should be the (expected) degree of measurement error. Measurement error hampers the study of the

process of change. Naturally, one should try to reduce the degree of measurement error, and use a reliable measurement instrument. Additionally, by assuming a certain functional form, one could *estimate* the true scores, and thus try to filter out the measurement error. This functional form can be strongly restricted (e.g., a linear trend), but also weakly restricted (e.g., only a small degree of smoothness). Based on the functional form assumptions, a curve that (according to a certain criterion) optimally represents the observed scores can be estimated.

### **1.3. Two types of multisubject longitudinal data: longitudinal two-way data and multiple time series**

In Section 1.2, we dealt with intraindividual change over time. Repeatedly collecting scores that index the same feature, from more than one subject, enables the study of interindividual variability in level and change. Based on the comparability of the measurement occasions of the different subjects, we distinguish two types of multisubject longitudinal data, namely longitudinal two-way data and multiple time series.

In longitudinal two-way data, the measurement occasions at which the scores are obtained, are comparable across subjects. The measurement occasions are comparable if they take place at similar points in the process that one intends to measure for all subjects. The scores obtained on the successive  $K$  occasions from the  $I$  subjects can be usefully arranged in a two-way  $K \times I$  matrix, hence the name 'longitudinal two-way data'. In such a  $K \times I$  matrix, the scores considered columnwise pertain to the same subject, and the scores considered rowwise pertain to the same occasion.

An empirical example of longitudinal two-way data is data provided by research on the effect of a therapy on the degree of social phobia in socially phobic patients, where the measurements are made at the start of the therapy, and one, five, nine, thirteen and 25 weeks after the start (Scholing, 1993). If the measurements had taken place in the first, second, sixth, tenth, fourteenth and 26<sup>th</sup> weeks of the year 2000, while the patients had started therapy at different weeks, the measurement occasions would not have been comparable. A specific process dependent marker is required for the comparison of measurement occasions across subjects. In the above example, it is the start of the therapy. This marker usually follows naturally from the research design. A second example is a study of the degree of cognitive recovery after brain injury, measured at one, three, six and twelve months after injury (Spikman, Timmerman, Deelman & Van Zomeren, 1999). The occurrence of the brain injury is the marker, here. A third example is a study of the degree of visual attention in babies, which is measured at two-week intervals from six to 26 weeks after birth (Butcher, Kalverboer & Geuze, in press). Here, the date of birth is taken as a marker. In practice, a marker may not be available, or may be unsatisfactory. That is, shifted or transformed time axes of the univariate series would make the data more

comparable. Ramsay and Silverman (1997, Ch. 5) discuss a number of methods for dealing with this problem.

Generally, in analyzing longitudinal two-way data, the average level and development, and the interindividual variability are the matters of interest. The research question is usually focussed on whether predictors of the level, degree and form of long-term changes can be found. The development and variability of the separate subjects is usually of marginal importance.

In multiple time series, the same feature is measured repeatedly for a number of subjects, but measurement occasions are not comparable across subjects. The focus is usually on intraindividual variability rather than interindividual variability, and on short-term changes rather than long-term changes. An example is a study into the variability in aspects of mood of a group of patients suffering from Parkinson's disease (Shifren, Hooker, Wood & Nesselroade, 1997; see also Chapter 6). Note that if such mood data are collected at, for instance, the same date and time, the measurements are not comparable across subjects, as no time point can be indicated as a marker. A large number of measurement occasions are needed to study the intraindividual variability in detail. Presumably because it is difficult to collect time series for many subjects, multiple time series are usually only gathered for a small number of subjects.

#### **1.4. Two types of multivariate multisubject longitudinal data**

In most longitudinal studies, multivariate data are collected. The variables themselves can be composite variables. One may be interested only in the longitudinal behavior of the distinct variables. It is then sufficient to perform several univariate analyses. Sometimes, one is interested in phenomena described by several variables jointly. It is then useful to study the structure of the multivariate data. This usually implies studying the mutual linear relationships between the variables, and looking for a smaller set of (latent) variables that summarizes the data well.

Traditionally, the structure of multivariate data collected at one time point is examined using factor analysis methods. One can distinguish two approaches, namely the common factor analysis approach and the component analysis approach. In component analysis, the components (factors) are linear combinations of the variables. The models give an approximation of the (usually standardized) data. In common factor analysis, the scores on a variable are conceived of as consisting of a part that is common to the variables (in terms of being correlated), and a unique part. The factors are linear combinations of the common parts of the variables, which, however, cannot be solved for explicitly within the factor analysis model. The common factor model gives an approximation of the covariances or correlations between the observed variables.

Longitudinal multivariate data collected at comparable time points from a number of subjects are a kind of three-way data, which will be denoted by 'longitudinal three-

way data' in the sequel. Three-way data are data that can be classified in three ways, hence using three indices (Kroonenberg, 1983). Other examples of three-way data are multivariate data from a number of subjects collected under a number of conditions, or from a number of observers. Such different sets of entities are denoted as 'modes' of the data array. To apply factor analysis to three-way data, extensions of common factor analysis (Bloxom, 1968; Bentler & Lee, 1978, 1979; Oort, 1999, 2001), as well as component analysis (Tucker, 1966a; Carroll & Chang, 1970; Harshman, 1970; Kroonenberg, 1983) have been proposed. The application of three-way factor analysis techniques to longitudinal three-way data is a special topic, as this offers specific opportunities for modeling the data. Oort (2001) discusses a number of common factor analysis models specifically suitable for three-way data with a longitudinal mode. In the present study, component models for longitudinal three-way data will be provided.

In three-mode common factor analysis, one of the modes, usually the subject mode, is considered random, and the other modes are considered fixed. Usually, distributional assumptions are made to estimate the model parameters in the factor analysis model. Furthermore, it is usually necessary to specify the model so that a unique solution can be obtained. The process of model identification can be a difficult task, which is sometimes managed by imposing ad hoc constraints that are not necessarily realistic given the content of the study.

In three-way component analysis, all three modes are considered fixed. Hence, they are placed on the same footing in the model. No distributional assumptions are made in estimating the model parameters. One aims at finding a model which minimizes the part of the data that is not described by the model. A source of inaccuracy in the component model is the failure to take the unique variance of every variable into account. However, this is usually not problematic in practice, as comparable factor loadings are obtained using the two approaches, at least in the two-way case (Harshman & Lundy, 1984a, pp. 142-144; Velicer & Jackson, 1990a & 1990b; Jackson, 1991, p. 397). Kroonenberg and Oort (2001) offer a theoretical and empirical comparison of a three-way component model and a three-way factor model. If the sample size is small, and/or the necessary distributional assumptions appear to be violated, three-way component techniques are favored over three-way common factor analysis. On the other hand, if the subjects form a random sample and the other requirements are met, three-way factor analysis approaches appear preferable.

Factor analysis methods for analyzing multivariate time series collected from a single subject have been treated by several authors. Molenaar (1985) and Immink (1986) used a common factor approach. This approach requires relatively long time series, and makes some, rather strict distributional assumptions. Bijleveld (1989) and Van Buuren (1990) used component techniques. The simultaneous analysis of multivariate time series obtained from more than one subject has received some attention (Van Buuren, 1990; Bijleveld & Van der Kamp, 1998; Nesselroade & Molenaar, 1999). However, interindividual differences are covered poorly or not at all in those models.

The present study deals with component analysis approaches for the analysis of multisubject multivariate longitudinal data. Extensions of existing three-way component models for longitudinal three-way data are proposed. These extensions take advantage of the fact that the data are obtained at successive time points. Furthermore, models are proposed for multisubject multivariate time series which take intraindividual and interindividual differences into account.

This thesis is organized as follows. Chapters 2 through 5 are devoted to component models for longitudinal three-way data. Chapter 2 discusses existing three-way component models that can be applied to longitudinal three-way data. Chapter 3 gives an overview of possibilities for modeling the longitudinal mode. In Chapters 4 and 5, the ideas presented in Chapter 3 are elaborated, and applied to three-way component models. Chapters 6 and 7 are devoted to component models for multisubject multivariate time series. In Chapter 6, a class of methods for simultaneously modeling multisubject multivariate time series is discussed. In Chapter 7, the class of methods from Chapter 6 is extended to model so-called lagged influences as well. Chapter 8 concludes the thesis with a discussion, and recommendations for future research.



## 2. Component models for longitudinal three-way data

### 2.1. Introduction

In this chapter, a class of component models, that can be applied to longitudinal three-way data is discussed. Longitudinal three-way data are defined as multisubject multivariate data measured at various occasions that are comparable across all subjects (see Section 1.3). The models aim at providing a meaningful summary of the data. For that purpose, the data are decomposed into a number of interpretable matrices, so that the data are described parsimoniously. Before the models and some issues in empirical applications are discussed (Sections 2.3 through 2.6), some terminology and notation used throughout this study will be introduced. The interpretation of longitudinal three-way models is treated in Section 2.7.

### 2.2. Notational issues and some matrix algebraic properties

Element  $i,j$  of the  $I \times J$  two-way datamatrix  $\mathbf{X}$  is denoted by  $x_{ij}$ . The  $j^{\text{th}}$  column of  $\mathbf{X}$  is denoted by  $\mathbf{x}_j$ , and the  $i^{\text{th}}$  row by  $\mathbf{x}_i'$ . The symbol  $\mathbf{I}_Q$  denotes the  $Q \times Q$  identity matrix. The  $P \times Q$  matrix having each element equal to zero is denoted by  $\mathbf{0}_{P \times Q}$ . The  $Q \times 1$  vectors with each element equal to zero and one are denoted by  $\mathbf{0}_Q$  and  $\mathbf{1}_Q$ , respectively, or just  $\mathbf{0}$  and  $\mathbf{1}$  if the size of the vector is clear from the context.

Longitudinal three-way data consist of scores on  $J$  variables ( $j=1, \dots, J$ ) collected from  $I$  subjects ( $i=1, \dots, I$ ) at  $K$  measurement occasions ( $k=1, \dots, K$ ). The data can be collected in a three-way data array  $\underline{\mathbf{X}}$  ( $I \times J \times K$ ). A graphical representation of such a data array  $\underline{\mathbf{X}}$  is presented in Figure 2.1.

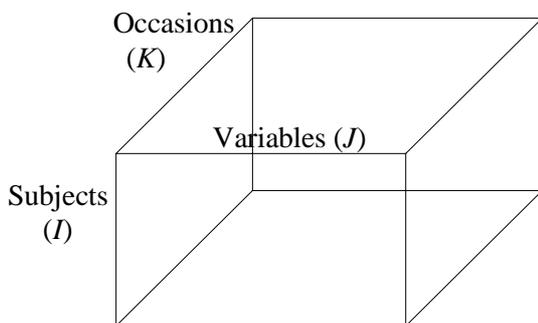


Figure 2.1. Graphical representation of a three-way data array  $\underline{\mathbf{X}}$  of size  $I \times J \times K$ .

This type of three-way data is also called *three-mode* data (Carroll & Arabie, 1980). The number of ways refers to the geometrical dimension of the array, whereas the number of modes refers to the number of different sets of entities. Hence, a set of covariance matrices obtained at different occasions from scores on the same variables would make up a two-mode three-way data array.

The data in Figure 2.1 can also be presented in  $I$ ,  $J$  or  $K$  separate submatrices. The  $k^{\text{th}}$  frontal slab of  $\underline{\mathbf{X}}$  is denoted by  $\mathbf{X}_k$  ( $I \times J$ ). Analogously,  $\mathbf{X}_i$  ( $J \times K$ ) is the  $i^{\text{th}}$  horizontal slab, and  $\mathbf{X}_j$  ( $K \times I$ ) the  $j^{\text{th}}$  lateral slab of  $\underline{\mathbf{X}}$ . Element  $i,j,k$  of the array  $\underline{\mathbf{X}}$  ( $I \times J \times K$ ) is denoted by  $x_{ijk}$ . Sometimes it is convenient to present the full three-way data set in ‘matricized’ form (see Kiers, 2000), that is, as a supermatrix with frontal, lateral or horizontal slabs next to each other. The size of the supermatrix and the subscript of the matrix symbol indicate the positioning of the submatrices in the sequel. Specifically, the matrix  $\mathbf{X}_a$  ( $I \times JK$ ) contains  $K$  frontal slabs  $\mathbf{X}_k$  ( $I \times J$ ),  $k=1, \dots, K$ , positioned next to each other. Analogously, the matrix  $\mathbf{X}_b$  ( $J \times KI$ ) denotes a supermatrix in which the  $I$  horizontal slabs  $\mathbf{X}_i$  ( $J \times K$ ),  $i=1, \dots, I$ , are positioned next to each other, and the matrix  $\mathbf{X}_c$  ( $K \times IJ$ ) is the supermatrix that contains the  $J$  lateral slabs  $\mathbf{X}_j$  ( $K \times I$ ) positioned next to each other. Another option, though rarely used, is to present the data in  $\underline{\mathbf{X}}$  as a vectorized version of  $\mathbf{X}_a$ . The vectorized version of  $\mathbf{X}_a$  ( $I \times JK$ ) is denoted by  $\text{Vec}(\mathbf{X}_a)$ , and it is obtained by stringing out  $\mathbf{X}_a$  column-wise to a column vector of length  $IJK$ .

The Khatri-Rao product, or the column-wise Kronecker product, is denoted by  $\vee$  (Rao & Mitra, 1971). The Khatri-Rao product of two matrices  $\mathbf{A}$  and  $\mathbf{B}$ , each with  $R$  columns, is defined as

$$\mathbf{A} \vee \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \quad \dots \quad \mathbf{a}_R \otimes \mathbf{b}_R], \quad (2.1)$$

where  $\otimes$  denotes the Kronecker product.

The elements of a matrix product of the form  $\mathbf{ABC}$  can be expressed as a function of  $\text{Vec}(\mathbf{B})$  as

$$\text{Vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A}) \text{Vec}(\mathbf{B}), \quad (2.2)$$

see, for instance, Magnus and Neudecker (1991).

### 2.3. Component models for three-way data

As the component models described below are generalizations of the well-know principal component model, this model for two-way data is discussed first. A principal component analysis (PCA) aims at describing the scores on a number of variables as a weighted sum of scores on a smaller number of components. Specifically, in PCA one decomposes a two-way data matrix  $\mathbf{X}$  ( $I \times J$ ) as

$$\mathbf{X} = \mathbf{A}\mathbf{B}' + \mathbf{E}, \quad (2.3)$$

where  $\mathbf{A}$  ( $I \times Q$ ) denotes the subject component matrix,  $\mathbf{B}$  ( $J \times Q$ ) the matrix with so-called variable loadings,  $\mathbf{E}$  ( $I \times J$ ) the matrix with residuals and  $Q$ ,  $q=1, \dots, Q$ , the number of components. The matrix with residuals is that part of the data that is not covered by the model. The data matrix is decomposed in a way that minimizes the sum of squared residuals. The part of the data that is covered by the structural part of the model ( $\mathbf{A}\mathbf{B}'$  in the case of a PCA), hence without the residual part, is denoted by  $\hat{\mathbf{X}}$  in the sequel.

Standard two-way principal component analysis leads to solutions that can be rotated without loss of fit. To avoid the rotational indeterminacy, Cattell (1944), and Cattell and Cattell (1955) proposed looking for 'Parallel Proportional Profiles' (PPP) in two PCA solutions of two two-way data matrices, where the same variables on the same observation units have been measured twice. Harshman (1970) generalized Cattell's idea from two to  $K$  'parallel occasions', and showed that PPP entail the 'PARAFAC' model, which is a three-mode generalization of the principal component model. In the PARAFAC model, only proportional differences exist between the subjects, variables and occasions with respect to each component. Carroll and Chang (1970) independently proposed their CANDECOMP model, which is equivalent to the PARAFAC model. We will refer to the CANDECOMP/PARAFAC model as the CP model. The CP model is defined as

$$x_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr} + e_{ijk} \quad (2.4)$$

where  $x_{ijk}$  denotes element  $i, j, k$  of a three-way data array  $\underline{\mathbf{X}}$ ,  $a_{ir}$ ,  $b_{jr}$ ,  $c_{kr}$  denote elements of the component matrices  $\mathbf{A}$  ( $I \times R$ ),  $\mathbf{B}$  ( $J \times R$ ),  $\mathbf{C}$  ( $K \times R$ ), respectively, and  $e_{ijk}$  denotes element  $i, j, k$  of the error array  $\underline{\mathbf{E}}$  ( $I \times J \times K$ ),  $i=1, \dots, I$ ,  $j=1, \dots, J$ ,  $k=1, \dots, K$ ,  $r=1, \dots, R$ .

The CP model can be written in matrix notation as

$$\mathbf{X}_a = \mathbf{A}(\mathbf{C} \cup \mathbf{B})' + \mathbf{E}_a, \quad (2.5)$$

where  $\mathbf{X}_a$  is the  $I \times JK$  matricized data array  $\underline{\mathbf{X}}$ ,  $\cup$  denotes the Khatri-Rao product, the matrix  $\mathbf{C}$  denotes the occasion component matrix, and, analogously to the PCA model, the matrix  $\mathbf{A}$  denotes the subject component matrix and  $\mathbf{B}$  the variable component matrix, and  $\mathbf{E}_a$  ( $I \times JK$ ) is the matricized error array  $\underline{\mathbf{E}}$ . Note that the variable component matrix  $\mathbf{B}$  has an equivalent role and interpretation as the loading matrix  $\mathbf{B}$  has in the PCA model (see (2.3)). To stress the fact that the CP model is completely symmetric, we name the matrices  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  all 'component matrices', instead of using different names for the three.

An alternative three-way generalization of PCA is the Tucker3 model (Tucker, 1966a; Kroonenberg & De Leeuw, 1980; Kroonenberg, 1983). Just as in the CP

model, the three-way array is decomposed into three component matrices, and the model is symmetric. However, in the Tucker3 model, all components of  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  are related to each other via the so-called core array. Also, the numbers of components of  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  may differ. The Tucker3 model is defined as

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk}, \quad (2.6)$$

where  $x_{ijk}$  denotes element  $i,j,k$  of the  $I \times J \times K$  three-way array  $\underline{\mathbf{X}}$ ,  $a_{ip}$ ,  $b_{jq}$ , and  $c_{kr}$  denote the elements of the component matrices  $\mathbf{A}$  ( $I \times P$ ),  $\mathbf{B}$  ( $J \times Q$ ), and  $\mathbf{C}$  ( $K \times R$ ), respectively;  $g_{pqr}$  denotes the elements of the core array  $\underline{\mathbf{G}}$  ( $P \times Q \times R$ ), and  $e_{ijk}$  denotes element  $i,j,k$  of the error array  $\underline{\mathbf{E}}$  ( $I \times J \times K$ ),  $i=1, \dots, I$ ,  $j=1, \dots, J$ ,  $k=1, \dots, K$ ;  $p=1, \dots, P$ ,  $q=1, \dots, Q$ ,  $r=1, \dots, R$ . The elements of the core array describe the weights of the interactions of the components in  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$ .

The Tucker3 model can be written in matrix notation as

$$\mathbf{X}_a = \mathbf{A} \mathbf{G}_a (\mathbf{C}' \otimes \mathbf{B}') + \mathbf{E}_a, \quad (2.7)$$

where  $\mathbf{X}_a$  is the  $I \times JK$  matricized data array  $\underline{\mathbf{X}}$ ,  $\mathbf{G}_a$  ( $P \times QR$ ) denotes the supermatrix containing the frontal slices of the core array  $\underline{\mathbf{G}}$  ( $P \times Q \times R$ ),  $\otimes$  the Kronecker product, and  $\mathbf{E}_a$  ( $I \times JK$ ) the matricized error array  $\underline{\mathbf{E}}$ .

The CP model can be notated equivalently to the Tucker3 model (2.7), namely as

$$\mathbf{X}_a = \mathbf{A} \mathbf{H} (\mathbf{C}' \otimes \mathbf{B}') + \mathbf{E}_a, \quad (2.8)$$

where  $\mathbf{X}_a$  is  $I \times JK$ , the fixed matrix  $\mathbf{H}$  is the  $Q \times Q^2$  two-way version of the 'superidentity' three-way array  $\underline{\mathbf{H}}$ , that is, an array with  $h_{pqr}=1$  if  $p=q=r$ , and  $h_{pqr}=0$  otherwise. This notation shows at once that the CP model is a constrained version of the Tucker3 model, a result given earlier by Carroll and Chang (1970, p. 312).

The Tucker2 model (Kroonenberg & De Leeuw, 1980) is less restricted than the Tucker3 and CP models. In the Tucker3 and CP models, each of the three modes are reduced into components. In the Tucker2 model, two modes are reduced and one is left unreduced. Consequently, the Tucker2 model is not symmetric, whereas the Tucker3 and CP models are. Here, we present the Tucker2 model for the situation where the subject mode is unreduced, but one can equally well choose to leave the variable or occasion mode unreduced. The Tucker2 model is given by

$$\mathbf{X}_a = \tilde{\mathbf{G}}_a (\mathbf{C}' \otimes \mathbf{B}') + \mathbf{E}_a, \quad (2.9)$$

where  $\tilde{\mathbf{G}}_a$  ( $I \times QR$ ) denotes the ‘extended core matrix’. The elements of the extended core matrix describe the weights of the interaction between the combinations of the  $Q$  variable components and the  $R$  occasion components for each of the  $I$  subjects.

The relationship between the Tucker2 and Tucker3 models can be seen as follows. If in the Tucker3 model (2.7) the number of subject components is chosen to be  $I$ , then the estimated component matrix  $\mathbf{A}$  is a basis for the full subject space. In this case, the Tucker3 model boils down to the Tucker2 model, because then  $\tilde{\mathbf{G}}_a$  can be written as  $\tilde{\mathbf{G}}_a = \mathbf{A}\mathbf{G}_a$ .

The Tucker1 model is the most weakly constrained of the models for longitudinal three-way data that are discussed here. In the Tucker1 model, only one mode is reduced. Here, we treat the case for which the subject mode is reduced, and the variable and occasion modes are left unreduced. The Tucker1 model with reduced subject mode is given by

$$\mathbf{X}_a = \mathbf{A}\mathbf{F}' + \mathbf{E}_a, \quad (2.10)$$

where  $\mathbf{F}$  ( $JK \times P$ ) denotes a component matrix containing the loadings of all combinations of variables and occasions on the  $P$  components. Note that the Tucker1 model equals a PCA model of a matricized three-way array. In practical applications of the Tucker1 model, it is often natural to reduce the variable mode, and to leave the subject and occasion modes unreduced. This can be done by applying the Tucker1 model to the matrix  $\mathbf{X}_b$  ( $J \times KI$ ). With the exception of the Tucker2 model, the hierarchical relations between the three-way models mentioned above are also discussed by Kiers (1991). The hierarchical relation between the Tucker3 and Tucker2 models is treated by Kroonenberg and de Leeuw (1980).

## 2.4. Fitting the component models for three-way data

The Tucker1, Tucker2, Tucker3, and CP models are fitted to observed data by minimizing the sum of squared residuals  $\|\mathbf{X}_a - \hat{\mathbf{X}}_a\|^2$ , where  $\hat{\mathbf{X}}_a$  denotes the part of  $\mathbf{X}_a$  that is covered by the structural part of the model, and  $\|\cdot\|^2$  denotes the squared Euclidean norm. The degree to which the estimated model describes the data is expressed by the proportion of sum of squares explained by the model, which we call the ‘fit’ in the sequel. The fit is defined as

$$1 - \frac{\|\mathbf{X}_a - \hat{\mathbf{X}}_a\|^2}{\|\mathbf{X}_a\|^2}. \quad (2.11)$$

The fit is often expressed as a percentage by multiplying it by 100.

Because the Tucker1 model equals a PCA model of a matricized three-way array, estimates of the parameters of the model are obtained by applying a standard PCA procedure. Kroonenberg and de Leeuw (1980) gave algorithms for fitting the Tucker2 and Tucker3 models to data. Other ways of fitting the Tucker3 model to data have been proposed by Weesie and Van Houwelingen (1983), Kiers, Kroonenberg and Ten Berge (1992), Andersson and Bro (1998), and Paatero and Andersson (1999). Harshman (1970), Carroll and Chang (1970), Kiers (1998a) and Paatero (1999) gave algorithms to estimate the parameters of the CP model. Typically, the algorithms are Alternating Least Squares (ALS) algorithms. In ALS algorithms, the matrices over which the function has to be minimized are alternately updated until convergence. The algorithm is said to have converged, if from one cycle (i.e., update of all parameters) to another, the residual sum of squares decreases less than a prespecified tolerance value. The latter tolerance value is sometimes denoted by the ‘convergence criterion’. As the residual sum of squares decreases monotonically, and is bounded below by zero, convergence in terms of this criterion is guaranteed. However, it is not guaranteed that the global minimum of the function is reached, hence the algorithm may land in a local minimum. In practice, it is recommended to use several differently started runs in order to decrease the chance of missing the global minimum of the function.

### 2.5. Transformational freedom within the component models for three-way data

The Tucker3, Tucker2 and Tucker1 models have transformational freedom. In the Tucker3 model, this means that estimates of the data array  $\underline{\mathbf{X}}$  are insensitive to orthogonal and oblique transformations of the component matrices  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$ , provided that such transformations are compensated in the core array  $\underline{\mathbf{G}}$  (Tucker, 1966a). The same holds for the two component matrices in the Tucker2 model, provided that the transformations are compensated in the extended core matrix. The component matrix of the reduced mode in the Tucker1 model can also be transformed, provided that this transformation is compensated in the component matrix for the two unreduced modes. This case is identical to ordinary two-way PCA, whose transformational freedom is well known.

Standard rotational procedures aiming at a simple structure (e.g., Varimax; Kaiser, 1958) can be used to obtain more easily interpretable solutions for the subject and variable related component matrices. Those procedures are generally not useful in transforming the occasion component matrix, as will be illustrated later. In the Tucker2 and Tucker3 models, one can also transform the (extended) core array (over all two and three modes, respectively), provided that the transformations are compensated in the component matrices. Kiers (1997a) proposed a procedure that aims at obtaining simplicity of the core array in the Tucker3 model. Kiers (1998b)

discusses a method for *jointly* transforming the core, and the component matrices to simplicity in the Tucker3 model.

As Harshman (1972) proved, the parameter estimates of a CP model are ‘essentially unique’, if there is at least one pair  $\mathbf{D}_k, \mathbf{D}_l$  ( $k \neq l; k, l = 1, \dots, K$ ), so that  $\mathbf{D}_k \mathbf{D}_l^{-1}$  has no equal pair of diagonal elements. Essential uniqueness of a CP solution means that CP estimates are unique up to trivial permutation, reflection and/or rescaling. An even more relaxed sufficient condition for the uniqueness of the estimates of the CP model was given by Kruskal (1977).

## 2.6. Issues in the application of the component models to three-way data

### 2.6.1. Preprocessing three-way data

In applying a PCA to a standard two-way matrix, one usually decomposes the centered and standardized data matrix, rather than the raw data. Centering and scaling serve different purposes, in two-way as well as multi-way component analysis (Harshman & Lundy, 1984b; Bro & Smilde, in preparation). Centering aims at removing constant terms in the data in order to make the data compatible with the model. That is, it causes interval scale data to behave as ratio scale data, which is required in component models. Scaling aims at eliminating artificial scale differences, and does not affect the model. It only influences the weights attached to certain elements in the least squares loss function.

Preprocessing in multi-way data is more complicated than it is in the two-way case. Harshman and Lundy (1984b) and Bro and Smilde (in preparation) extensively discuss centering and scaling of multi-way data. A centering is considered proper, if constants are eliminated from the data without introducing artificial variation. That is, if data (without error) consist of a structural part that is covered by the model, plus constant(s), then the centering should indeed remove those constants, without introducing new ones. Harshman and Lundy, and Bro and Smilde conclude that proper centering is obtained by centering *across one* mode. Several such centerings across one mode may be performed sequentially.

Harshman and Lundy (1984b), and Bro and Smilde (in preparation) showed that scaling is unproblematic only when performed *within one* mode (for example, per variable over occasions and subjects jointly). Combinations of centering and scaling do not conflict if scaling within one mode is combined with centering across (one of) the other modes.

### 2.6.2. Choice of a specific model and the numbers of components

In Section 2.3, four component models for three-way data were discussed. The models were ordered from severely constrained (CP model) to weakly constrained (Tucker1 model). A constrained version of a certain model implies that a restriction is imposed on certain parameters of that model. As a result, the fit of a constrained

model to the observed data set is at most equal to, but generally lower than the fit of its unconstrained version, all other things being equal. If one wishes to select a particular model and the number of components, one should take the interpretability and the degree of modeling error into account.

The interpretability of a model is important because in fitting models one aims at understanding underlying processes. The selection of an interpretable model is by no means a simple task. The interpretation of a model depends partly on the theoretical ideas and knowledge a researcher has. As a result, the criterion of interpretability adds a certain degree of subjectivity to the model selection process, and can therefore be subject to debate.

The second point of concern in model selection is the degree of modeling error. In fitting a model to data, one usually aims at obtaining a model that is more generally applicable than just to the current data. In the current context, this would imply that one has a three-way data set of which the individuals are thought to be drawn from a certain population. Then, the estimated model parameters are wished to be valid for the whole population, and not just for the current sample. In this context, it is useful to distinguish three types of error involved in fitting a particular model, namely the error of approximation, the error of estimation and the overall error (Linhart & Zucchini, 1986; Browne & Cudeck, 1992). Suppose one has chosen a particular model. The *error of approximation* is the lack of fit of the current model to the population data, if the parameters were optimally chosen. The error of approximation is not dependent on the data set at hand, on the sample size or the method of estimation. The *error of estimation* is the lack of fit of the current model fitted to the current data set, to the model of the population data with optimally chosen parameters. The error of estimation depends on the sample size, the current data set, and the method of estimation. Both types of error cannot be computed directly, because the population data, and the optimally chosen parameters of the model for the population data are unknown. The *overall error* refers to the lack of fit of the current model that is fitted to the current data set, to the population data. Usually, the overall error is the sum of the error of approximation and the error of estimation. Generally, as the number of parameters of a model increases, the error of approximation reduces, whereas the error of estimation increases.

In choosing a particular model, one should aim at finding an *interpretable* model with a small degree of overall error. A way to investigate the degree of overall error is using cross-validation, which assesses the predictive validity of the estimated model parameters. A cross-validation method, the ‘expectation maximization cross-validation’ (EM-CV), is proposed by Louwerse, Smilde and Kiers (1999) for application in the three-way context. The degree of overall error can also be examined using split-half analysis. Examples of split-half analysis applied to three-way models are presented by Harshman and Lundy (1984a), and Kiers and Van Mechelen (2001). In the references, the methods are applied to either the CP or Tucker3 model, but they can be adapted easily for application to other component models as well. In Chapter 6, the EM-CV method (Louwerse, Smilde & Kiers, 1999) and a variant of split-half

analysis (Harshman & Lundy, 1984a) are adapted, and applied to models for multivariate multisubject time series.

Besides modeling error, the interpretability of the model is important, because one aims at understanding processes. The requirement of interpretability may conflict with the desire to reduce the error of approximation, which is one part of the overall error. It is not useful to add parameters which are not interpretable at all to a model, only to reduce the error of approximation. Hence, it is more precise to say that it is desirable to minimize the error of approximation, provided that the fitted model is interpretable, and reflects underlying processes. Therefore, the decision on which model to take cannot be based solely on indices of the degree of overall error.

In certain cases, one may assume that the population three-way data consist of 'true scores', that follow the model concerned, and measurement error. In the framework discussed above, the measurement error is the error of approximation. The parameters of the models discussed here are obtained by least squares fitting of the model to observed three-way data. The estimated three-way array obtained this way will usually not be exactly equal to the 'true' three-way array, and the fitted model is partially based on measurement error. This phenomenon will be indicated by 'error fitting' in the sequel. Obviously, a high degree of error fitting is undesirable, since it may lead to errors in interpretation.

A special case, which is sometimes encountered in component modeling, occurs if one aims at meaningfully summarizing the current three-way data set only. One can view this case as having available the population data. Then, the overall error is fully or almost fully (depending on the discrepancy measure and the estimation procedure) determined by the error of approximation. Therefore, the fit of the estimated model to the observed data is a good indicator of the degree of error of approximation. One should find a balance between interpretability and degree of modeling error. One could try to select a parsimonious model that still covers the most important or most salient aspects of the data, thereby ignoring aspects of little importance (e.g., because they pertain to a small number of subjects, variables, or occasions) by using the fit to compare different estimated models. It is hoped that the model that covers the most important aspects is also a well interpretable model. Using this approach requires to choose the most constrained model with the smallest numbers of components that explains relatively much of the sum of squares. Specifically, for choosing the numbers of components in the Tucker3 model, Timmerman and Kiers (2000) proposed a method that aims at finding an optimal balance between the fit of solutions for the Tucker3 model and the numbers of components. They compared the results obtained by this method, called DIFFIT, with results from two other methods, by means of a simulation study. It was found that DIFFIT performed considerably better than the other methods in indicating the numbers of components.

## 2.7. Interpretation of three-way component models applied to longitudinal three-way data

First, the interpretation of the three parameter matrices of a CP model of longitudinal three-way data is discussed. The CP model (2.4) is the most constrained symmetrical three-way component model. The three modes are reduced into three component matrices that summarize the linear relationships between subjects, variables and occasions. As an empirical example, one could think of the scores on a number of variables that measure arithmetic and language skills. The scores are collected from a number of pupils receiving education in arithmetic and language at several time points. Suppose the variables are related to only two variable components, namely an arithmetic and a language component. The variable components can be viewed as referring to two 'latent variables', hence to variables underlying the observed variables. The relative size of a variable component score indicates the degree to which that (observed) variable measures the particular latent variable. In the CP model, the component matrices are usually not (columnwise) orthogonal, which implies that the components are linearly related to each other, to a smaller or larger extent. From a theoretical point of view, it is conceivable that latent variables are linearly related. For example, the degree of arithmetic and language skills are likely to be partly influenced by a general degree of intelligence. In the example, the two subject component scores indicate the relative position of the pupil to that of the other pupils on arithmetic and language. The occasion component scores at the time points are indicators of the overall level in arithmetic and language for the whole group at each particular time point. The scores at successive time points are interpreted as indicating the development over time. It may be helpful to plot the scores against time to get a notion of the development of the level over time. One interprets those scores as if they are evaluations of a certain latent curve. Note that the relative position of the subjects on each component is constant across time in the CP model. Supposing the time component scores to be positive, the model implies, for example, that if a pupil scores poorly in arithmetic at one time point, he will continue to score poorly at all subsequent time points.

A graphical representation of the CP model is given in Figure 2.2. The figure shows that the three-way array  $\hat{\underline{\mathbf{X}}}$  is a sum of  $R$  ( $r=1, \dots, R$ ) rank one three-way arrays, of which each element  $x_{ijk}$  is the triple product  $a_{ir}b_{jr}c_{kr}$ . The vectors with component scores can be collected into three component matrices for each of the three modes.

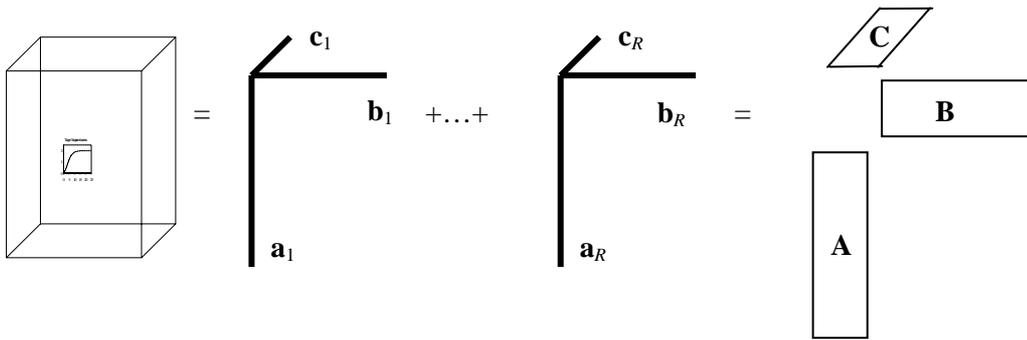


Figure 2.2. Graphical representation of the CP model.

As the Tucker3 model (2.7) is more flexible than the CP model, the interpretation is usually somewhat more complicated. The three component matrices summarize the properties of the subjects, variables and occasions. The weights of the combinations of the components are reflected in the core array. The variable component matrix is interpreted in the same way as in the CP model, that is in terms of latent variables. The subject components can be interpreted as 'idealized' or 'prototype' subjects (Kroonenberg, 1983). Real subjects are considered weighted combinations of the idealized subjects. The occasion components can be viewed as 'idealized' or 'prototype' occasions. The occasion component scores at the various time points reflect the level of the prototype occasion at the time points, and just as in the CP model, they can be interpreted as evaluations of a latent curve. Simplicity of the component matrices as well as the core array is needed to facilitate interpretability. A graphical representation of the Tucker3 model is given in Figure 2.3, which shows the decomposition of the three-way array into the three component matrices and the core array.

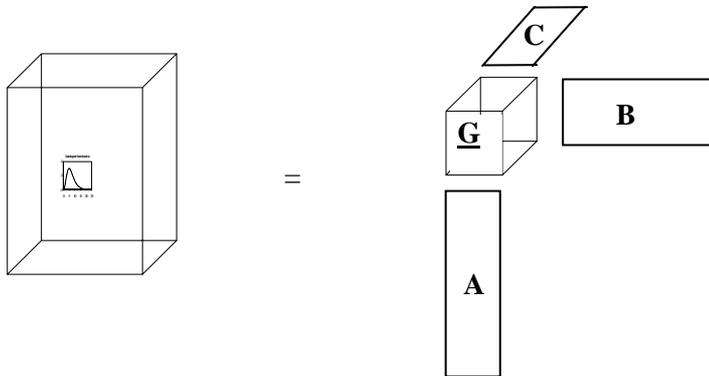


Figure 2.3. Graphical representation of the Tucker3 model.

The two component matrices of the two reduced modes in the Tucker2 model (2.9) are interpreted in the same way as their counterparts in the Tucker3 model. The extended core matrix provides the weights for all component combinations (of the reduced modes) per entity of the unreduced mode.

The component matrix of the reduced mode in the Tucker1 model (2.10) is interpreted in the same way as its counterpart in the Tucker2 or Tucker3 model. The component matrix contains the loadings for the components of the reduced mode for all combinations of entities of the two unreduced modes.

### **3. Occasion components as evaluations of latent curves: possibilities for constraints to the time mode**

In Chapter 2, four component models that are potentially useful for modeling longitudinal three-way data were discussed. Although the models can be used in a completely exploratory way, it is often advantageous to use existing knowledge of the processes generating the data to constrain the model. Potential advantages of model constraints are better estimates of the model parameters, reduction of numerical problems and time required for computation, reduction of transformational freedom in a substantively sensible way, and a reduction of the chance to end in a local minimum, especially in ill-defined models. Possible constraints are unimodality and non-negativity of components (Bro & Sidiropoulos, 1998), equality, symmetry, and orthogonality of components (Bro, 1998), and certain patterns of zero values in one (or more) of the component matrices and the core (Kiers, 1991; Kiers, Ten Berge & Rocci, 1997; Kiers & Smilde, 1998). In this chapter, the rationale for imposing constraints on component models for longitudinal data is introduced.

As discussed in Section 1.2, the expected functional form of the process under study is an important factor in selecting the sampling time points in longitudinal research. The successive scores obtained by repeatedly collecting the score on a variable from the same subject are a kind of 'functional data' (Ramsay & Silverman, 1997). Functional data may appear in various ways, but they are all smooth to a larger or smaller extent, in the sense of being repeatedly differentiable (Ramsay & Silverman, 1997). The smoothness property of functional data does not imply that the observed data are evaluations of a smooth function, since observational error or noise may disturb smoothness. An approach to modeling univariate functional data is trying to filter out the error term by imposing smoothness restrictions. In this way, no rigid parametric assumptions about the dependence of scores and time are imposed. Several smoothing techniques are available (Wahba, 1990; Hastie & Tibshirani, 1990; Ramsay & Silverman, 1997), like kernel smoothing and polynomial smoothing. Another approach is to impose a certain functional relationship between scores and time, and to estimate parameters of the function from the observed data. The form of the functional relationship is ideally determined on the basis of the mechanism producing the data. If the latter is not possible, one could base the decision on the form of the observed data. Interpretability of the parameters of the functions should play a role. Smoothing techniques and imposing functional relationships are widely used in the univariate case.

The earliest approach for identifying the structure in *two-way* functional data was principal component analysis (Tucker, 1958, 1966b; Rao, 1958). Functional principal component analysis (FPCA), as it is called by Ramsay and Silverman (1997), aims at describing the dominant modes of variation of a functional data set. This exploratory method is similar to PCA applied to a multisubject or multivariate longitudinal data matrix. A functional PCA reveals a number of component scores at the different time points, and subject or variable weights for these components. The component scores at the time points can be viewed as estimated evaluations of so-called ‘basis functions’. Hence, an observed score at a certain time point is a weighted sum of the basis functions that are evaluated in that particular time point.

Meredith and Tisak (1990) proposed a factor analysis counterpart of the functional PCA model, the latent curve model. They treated the individual weights for the latent curves as unobserved latent variables, and made distributional assumptions about the error term. The latent curve model differs from a standard factor analysis model in that the expected value of the observed variable is the average growth curve, and hence is (usually) non-zero. Also, the expected values of the weights for the latent curves are non-zero. The latent curves can be left unconstrained, but they can also be partly or fully prespecified.

Unconstrained latent curves, as well as components representing basis functions resulting from a functional PCA, have transformational indeterminacy. Common transformational procedures aiming at simple structure (e.g., Varimax; Kaiser, 1958) are generally not useful in transforming latent curves or basis functions. Tucker (1966b) defined some criteria for transforming the solution of his basis functions, like non-negativity of the component scores, and non-negativity of the slopes. However, in practice, empirical latent curve analyses with unconstrained latent curves seem to use only one latent curve (Meredith & Tisak, 1990; Jones & Meredith, 1996), possibly to avoid transformational indeterminacy, as Browne (1993) suggested. As in practice more than one latent curve could be needed to model empirical data satisfactorily, other approaches are needed. For example, transformational indeterminacy can be removed by fully fixing the basis functions or latent curves. Meredith and Tisak (1990) mentioned the use of orthogonal polynomials as a possibility. An alternative is to specify the basis functions or latent curves only partly, which reduces or even removes the transformational indeterminacy. In analyzing ‘growth data’, Browne and Du Toit (1991) and Browne (1993) used latent curves that summarize basis features of a certain (nonlinear) function that represents growth. The component models for longitudinal three-way data discussed in Chapter 2 can be viewed as three-way extensions of functional PCA. It can be useful to constrain the basis functions in the component models for longitudinal three-way data in similar ways.

The specification of the latent curves offers the opportunity to estimate evaluations of basis functions at unobserved time points. This is useful, for example, if the scores on variables are collected at different sets of time points. Standard procedures to estimate the component models for longitudinal three-way data require

all elements of the data array to be observed. The ‘missing elements’, which result from the unequal measurement occasions for the variables and/or subjects, could be estimated, and imputed in the three-way array. The thus obtained full three-way array can be analyzed using standard estimation procedures. To use this approach, it is necessary that the missing data can be considered to be missing completely at random (Little & Rubin, 1987). This implies that missingness is unrelated to the (observed or unobserved) scores themselves.

The possibilities for model constraints as discussed for two-way functional data can be used in the component models for longitudinal three-way data as well. The first one to explore further is the use of smoothness constraints on the occasion components. In Chapter 4, a method for imposing smoothness constraints in the Tucker3 and CP models is proposed. The smoothness constraints can be combined with monotonicity constraints. The method can also easily be applied to the Tucker2 and Tucker1 models, but this option is not treated explicitly.

A second approach is to impose a certain functional form on the occasion components. This is particularly attractive if substantive considerations point to a specific functional relationship. Browne and Du Toit (1991) and Browne (1993) elaborated structured latent curve models for learning data, in which the columns of the occasion component scores matrix are parameterized parsimoniously. This approach is extended to modeling longitudinal three-way data, and learning data in particular, in Chapter 5.



## 4. The CP and Tucker3 models with smoothness constraints

### 4.1. Introduction

In Chapter 3, the rationale for applying constraints to models for longitudinal data was introduced. In this context, imposing smoothness is a flexible type of constraint. In the present chapter, a method for applying smoothness constraints, possibly combined with monotonicity constraints, to the Tucker3 and CP models will be presented. The Tucker3 and CP models were introduced in Section 2.3. The smoothness constraints can easily be used in the Tucker2 and Tucker1 models as well, but this is not explicitly discussed here.

In the Tucker3 or CP models, smoothness constraints can be applied in two ways. One could smooth the raw data, that is the various univariate series per subject and variable. In doing so, one aims at (partly) eliminating measurement error from the data. Subsequently, an unconstrained Tucker3 or CP analysis can be performed on the smoothed data. Alternatively, one could constrain the occasion components in the Tucker3 or CP models to be smooth and thus fit a constrained model to the raw data. In this case, it is hoped that less measurement error is fitted by the model. In the case of growth data, it might be useful to combine smoothness constraints with monotonicity constraints. Fortunately, the question of which of the two approaches to take is trivialized in an important class of cases. As will be shown, fitting the smooth descriptions of the observed data by the unconstrained Tucker3 or CP models is equivalent to fitting the Tucker3 or CP models with smoothness constraints in a particular class of cases. The usefulness of smoothing in the Tucker3 and CP models will be examined under different conditions in a simulation study. An empirical example illustrates the use of smoothness constraints within the Tucker3 model.

### 4.2. The choice of a smoother

A smoother is used to describe a response measurement as a smooth function of one or more predictor measurements (Hastie & Tibshirani, 1990), usually by so-called local averaging. In our applications, the predictor will simply be the time point at which each measurement is made, to be denoted as  $t_k=t_1, \dots, t_K$ . Local averaging aims at averaging the observed measurements associated with predictor values close to each other (i.e., in each other's neighborhood). The different types of smoothers differ mainly in their method of averaging. The size of the neighborhood influences the smoothness, and the accuracy: a large neighborhood leads to an estimate with low variance (i.e., high smoothness) but high potential bias (i.e., low accuracy), whereas

the opposite holds for small neighborhoods: there is a trade-off between bias and variance, as discussed by Hastie and Tibshirani (1990).

Hastie and Tibshirani (1990) and Ramsay and Silverman (1997) offer overviews of different smoothers and their properties. Polynomial regression splines, which form a class of smoothers that is computationally convenient, will be used here. Polynomial regression splines are constructed from different polynomial pieces, which are joined at certain predictor values, the knots. B-splines are a popular type of polynomial splines (De Boor, 1978). B-splines can easily be used for smoothing the data before analysis (see Alsberg & Kvalheim (1993) for an example involving three-way data). Monotonicity restrictions on the solutions can be useful in certain longitudinal applications. They can be imposed by using I-splines (Ramsay, 1988).

B-splines (Basis splines) are non-negative basis functions. The degree ( $d$ ) of a B-spline is the degree of the polynomial pieces on which it is based, and any degree of polynomial can be chosen. Each B-spline is determined fully by its degree and by its knot sequence. The knots are positioned in the domain  $t_1$  through  $t_K$  of the predictor. Given the degree and the location of the knots, B-splines can be computed by a recursive formula (De Boor, 1978). If they are of equal degree and they are positioned equidistantly, the basis functions are equal in size and shape. An example of a set of seven third degree B-splines, evaluated on the interval 1 to 2 is presented in Figure 4.1.

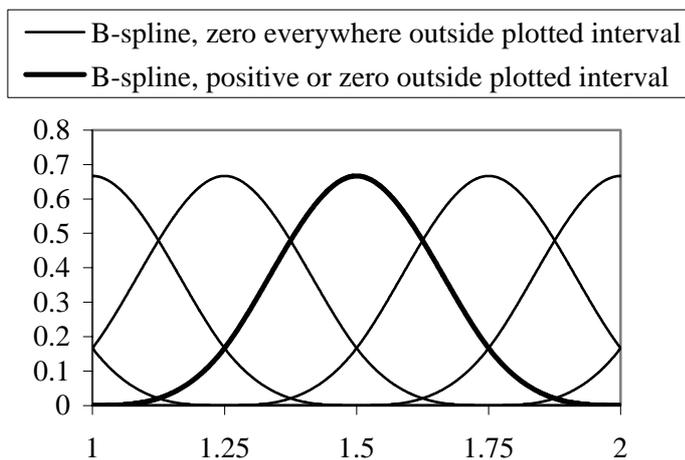


Figure 4.1. Example of seven non-zero third degree B-splines, evaluated on the interval 1 to 2.

In Figure 4.1, the five equidistantly placed knots are indicated by tick marks outside the x-axis. In general, the polynomial pieces join at  $d$  inner knots, and at these joining points, the derivative up to order  $d-1$  is continuous. The number of non-zero B-splines ( $N$ ) on the domain  $t_1$  through  $t_K$  is equal to the total number of knots on this

domain plus the degree of the polynomial minus one. In the example, this boils down to  $5+3-1=7$  non-zero B-splines in Figure 4.1. A B-spline is positive on a domain spanned by  $d+2$  knots; everywhere else it is zero. The B-spline that is zero everywhere outside the interval 1 to 2 in Figure 4.1 is indicated by the bold line.

Usually, a set of response measurements collected in  $\mathbf{y}$  ( $K \times 1$ ) is to be approximated by a linear combination of the B-splines that are evaluated in the values of the predictor  $\mathbf{t}$ . Let  $\mathbf{B}^s$  denote a  $K \times N$  B-spline matrix, in which the  $n^{\text{th}}$  column contains the values of the  $n^{\text{th}}$  B-spline that is evaluated for all values of the predictor  $\mathbf{t}$  ( $K \times 1$ ), and where  $K \geq N$ ; let  $\mathbf{w}$  ( $N \times 1$ ) denote the vector with weights for the  $N$  B-splines, and  $\hat{\mathbf{y}}$  the vector with estimated response measurements, which is called the *smooth* in the sequel, then

$$\hat{\mathbf{y}} = \mathbf{B}^s \mathbf{w}. \quad (4.1)$$

Since B-splines are always non-negative, the estimated response measurement  $\hat{\mathbf{y}}$  can be restricted to be non-negative by restricting the weights  $\mathbf{w}$  to be non-negative.

I-splines (Integrated splines; Ramsay, 1988) are monotonically increasing basis functions. They are based on integrated M-splines, which are proportional to B-splines. Because M-splines are non-negative everywhere, the integrated M-splines are a natural basis for monotone splines. Since bases for I-splines are monotonically increasing, a non-negativity constraint on the set of coefficients of the I-splines leads to monotonically non-decreasing estimated response variables.

The use of B-splines and I-splines requires the selection of the ‘smoothing parameters’, that is the degree of the splines and the number and the position of the knots. The degree of the spline is commonly fixed. For B-splines, a popular choice is a third degree B-spline (Hastie & Tibshirani, 1990, p. 22); smoothers based on higher degree splines tend to oscillate wildly (Van Rijckevorsel, 1988). Ramsay (1988) claims that low (e.g., second) degree I-splines generally suffice. The number and location of knots influences the smooth: more knots in a region lead to a greater flexibility of estimation in that region, whereas fewer knots lead to a greater rigidity in that region. Using too many knots leads to overfitting, and using too few knots leads to underfitting. As to choosing the position of the knots, a simple approach is to position them uniformly over the domain  $t_1$  through  $t_K$ . Another approach is to place them at appropriate quantiles of the predictor variable.

The smoothing parameters can be selected by subjective comparisons of several estimated response variables and the observed variable. Automatic selection methods for the smoothing parameters are also available (see Hastie & Tibshirani, 1990, pp. 42-52). Although the usefulness of these methods is debatable, they can be helpful in deciding on the number of knots. A commonly used procedure is cross-validation by means of the leave-one-out approach. Hastie and Tibshirani (1990, pp. 46-48) showed that the cross-validation sum of squares for linear smoothers can be computed by

$$\text{CV}(\boldsymbol{\lambda}) = \frac{1}{K} \sum_{k=1}^K \left( \frac{y_k - \hat{y}_k}{1 - \mathbf{S}(\boldsymbol{\lambda})_{kk}} \right)^2, \quad (4.2)$$

where  $\boldsymbol{\lambda}$  denotes the smoothing parameters (i.e., the degree, and the number and the positions of the knots), and  $\mathbf{S}(\boldsymbol{\lambda})_{kk}$  are the diagonal elements of the projection-matrix  $\mathbf{S}(\boldsymbol{\lambda})$ , which relates  $\hat{\mathbf{y}}$  to  $\mathbf{y}$ . To select the smoothing parameters, one may search for those that minimize  $\text{CV}(\boldsymbol{\lambda})$ .

### 4.3. How to smooth in the Tucker3 model and CP model?

Smoothers are generally used to describe *observed* scores by a function. As discussed above, smoothing in the Tucker3 or CP models can be performed by smoothing the *raw data* before analysis by the unconstrained Tucker3 or CP model, or by constraining the *component scores* in the Tucker3 or CP models to be smooth and thus fitting a constrained model to the raw data. The latter approach is used in functional PCA by Ramsay and Silverman (1997, Ch. 7), by applying a roughness penalty to prevent the roughness of the estimated principal components from being too large. We prefer a different approach, which has the advantage that smoothing the raw data and smoothing the components lead to the same estimated model parameters. We propose imposing a smoothness constraint on the occasion component matrix  $\mathbf{C}$  by constraining  $\mathbf{C}$  ( $K \times R$ ) so that it can be written as  $\mathbf{B}^s \mathbf{U}$ , for a B-spline matrix  $\mathbf{B}^s$  ( $K \times N$ ) and a particular weight matrix  $\mathbf{U}$  ( $N \times R$ ), and where  $N \geq R$ . As a result, the Tucker3 and CP models with smoothness constraints on the occasion component matrix can be written as

$$\mathbf{X}_c = \mathbf{B}^s \mathbf{U} \mathbf{G}_c (\mathbf{B}' \otimes \mathbf{A}') + \mathbf{E}_c, \quad (4.3)$$

where  $\mathbf{X}_c$  denotes the  $K \times IJ$  matricized data array  $\underline{\mathbf{X}}$ ,  $\mathbf{B}^s$  ( $K \times N, K \geq N$ ) a B-spline matrix,  $\mathbf{U}$  ( $N \times R, N \geq R$ ) a weight matrix,  $\mathbf{A}$  ( $I \times P$ ) and  $\mathbf{B}$  ( $J \times Q$ ) component matrices,  $\mathbf{G}_c$  ( $R \times PQ$ ) the supermatrix containing the frontal slices of the core array  $\underline{\mathbf{G}}$  ( $P \times Q \times R$ ), and  $\mathbf{E}_c$  ( $K \times IJ$ ) the matricized error array  $\underline{\mathbf{E}}$ ; in the case of the CP model with smoothness constraints the core array is fixed at superidentity. The B-spline bases are computed using 'time' as predictor. Note that the same basis is used for all components. In fact, formula (4.1) is used repeatedly for  $r=1, \dots, R$  as  $\mathbf{c}_r = \mathbf{B}^s \mathbf{u}_r$ . If monotonicity restrictions are required, it is proposed that the B-splines basis matrix be replaced by an I-splines basis matrix, and that non-negativity constraints be imposed on the weights.

The Tucker3 and CP models with smoothness constraints, like their unconstrained counterparts, are fitted to data by minimizing the sum of squared residuals. Now, it will be shown that restricting the component matrix  $\mathbf{C}$  to be in the column space of the B-spline matrix  $\mathbf{B}^s$  in the Tucker3 or CP models is equivalent to analyzing the projection of the data matrix  $\mathbf{X}_c$  on  $\mathbf{B}^s$  by the unrestricted Tucker3 or CP

model, which in turn comes down to a Tucker3 or CP analysis of the B-spline smoothed data. To show this, we replace  $\mathbf{B}^s$  by the QR-factorization  $\mathbf{B}^s = \mathbf{QR}$ , with  $\mathbf{Q}$  column-wise orthonormal, and  $\mathbf{R}$  a square upper triangular matrix. Note that since  $\mathbf{B}^s$  is of full column rank,  $\mathbf{R}$  is non-singular. Then, the function to be minimized is

$$f_1(\mathbf{U}, \mathbf{A}, \mathbf{B}, \mathbf{G}_c) = \|\mathbf{X}_c - \mathbf{QRUG}_c(\mathbf{B}' \otimes \mathbf{A}')\|^2, \quad (4.4)$$

with  $\mathbf{X}_c$  ( $K \times IJ$ ) the matricized data array  $\underline{\mathbf{X}}$ , and  $\mathbf{G}_c$  ( $R \times PQ$ ) the matricized core in the Tucker3 model, or the matricized superidentity array in the case of the CP model. As already noted by Carroll, Pruzansky and Kruskal (1980, p. 7), minimization of (4.4) is equivalent to minimizing

$$f_2(\tilde{\mathbf{U}}, \mathbf{A}, \mathbf{B}, \mathbf{G}_c) = \|\mathbf{Q}'\mathbf{X}_c - \tilde{\mathbf{U}}\mathbf{G}_c(\mathbf{B}' \otimes \mathbf{A}')\|^2, \quad (4.5)$$

with  $\tilde{\mathbf{U}}$  written for  $\mathbf{RU}$ .

It will now be shown that minimizing (4.5) is equivalent to analyzing the projection of  $\mathbf{X}$  on the B-spline matrix  $\mathbf{B}^s$  by the unrestricted Tucker3 or CP model. This in turn comes down to smoothing the data matrix  $\mathbf{X}_c$  by means of B-splines before Tucker3 or CP analysis, which is achieved by minimizing

$$f_3(\mathbf{W}) = \|\mathbf{X}_c - \mathbf{B}^s \mathbf{W}\|^2. \quad (4.6)$$

The optimal weights  $\mathbf{W}$  are given by  $(\mathbf{B}^s \mathbf{B}^s)^{-1} \mathbf{B}^s \mathbf{X}_c$ , hence the smooth of  $\mathbf{X}_c$  is  $\hat{\mathbf{X}}_c = \mathbf{B}^s (\mathbf{B}^s \mathbf{B}^s)^{-1} \mathbf{B}^s \mathbf{X}_c$ , the projection of  $\mathbf{X}_c$  on  $\mathbf{B}^s$ . Analyzing this projection by Tucker3 or CP comes down to minimizing

$$f_4(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{G}_c) = \|\hat{\mathbf{X}}_c - \mathbf{CG}_c(\mathbf{B}' \otimes \mathbf{A}')\|^2 = \|\mathbf{B}^s (\mathbf{B}^s \mathbf{B}^s)^{-1} \mathbf{B}^s \mathbf{X}_c - \mathbf{CG}_c(\mathbf{B}' \otimes \mathbf{A}')\|^2. \quad (4.7)$$

Let  $\mathbf{B}^s$  be replaced by the QR-factorization as  $\mathbf{B}^s = \mathbf{QR}$ . Note that  $\mathbf{R}$  is nonsingular. Minimization of (4.7) comes down to minimizing

$$f_4(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{G}_c) = \|\mathbf{QQ}'\mathbf{X}_c - \mathbf{CG}_c(\mathbf{B}' \otimes \mathbf{A}')\|^2. \quad (4.8)$$

The optimal  $\mathbf{C}$  will be in the column space of  $\mathbf{Q}$ , hence  $\mathbf{C}$  can be written as  $\mathbf{Q}\tilde{\mathbf{C}}$ , and minimizing (4.8) is equivalent to minimizing

$$\begin{aligned}
f_5(\mathbf{A}, \mathbf{B}, \tilde{\mathbf{C}}, \mathbf{G}) &= \|\mathbf{Q}\mathbf{Q}'\mathbf{X}_c - \mathbf{Q}\tilde{\mathbf{C}}\mathbf{G}_c(\mathbf{B}'\otimes\mathbf{A}')\|^2 \\
&= \|\mathbf{Q}(\mathbf{Q}'\mathbf{X}_c - \tilde{\mathbf{C}}\mathbf{G}_c(\mathbf{B}'\otimes\mathbf{A}'))\|^2 \\
&= \|\mathbf{Q}'\mathbf{X}_c - \tilde{\mathbf{C}}\mathbf{G}_c(\mathbf{B}'\otimes\mathbf{A}')\|^2, \tag{4.9}
\end{aligned}$$

(Kiers & Harshman, 1997). Clearly, minimizing (4.9) is equivalent to minimizing (4.5); the solutions for  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{G}_c$  of (4.5), and of (4.9) are equivalent; the solution for  $\tilde{\mathbf{U}}$  in (4.5) is equivalent to that for  $\tilde{\mathbf{C}}$  in (4.9). Because  $\tilde{\mathbf{U}}$  leads to  $\mathbf{C}$  by  $\mathbf{C}=\mathbf{B}^s\mathbf{U}=\mathbf{Q}\mathbf{R}\mathbf{U}=\mathbf{Q}\tilde{\mathbf{U}}$ , and  $\tilde{\mathbf{C}}$  leads to  $\mathbf{C}$  by  $\mathbf{C}=\mathbf{Q}\tilde{\mathbf{C}}$ , we see that both methods give the same solution for  $\mathbf{C}$  as well. It has thus been shown that analyzing the original data by means of a smooth constrained Tucker3 or CP models is equivalent to analyzing smoothed data by the unconstrained Tucker3 or CP model, as long as smoothness is defined in terms of unrestricted linear combinations of B-splines.

The matrix  $\mathbf{Q}'\mathbf{X}_c$  in (4.9) is a compressed version of the smooth matrix  $\mathbf{B}^s(\mathbf{B}^s'\mathbf{B}^s)^{-1}\mathbf{B}^s'\mathbf{X}_c$  (see Kiers & Harshman, 1997). Since the matrix  $\mathbf{Q}'\mathbf{X}_c$  is (much) smaller than the matrix  $\hat{\mathbf{X}}_c$ , minimization of (4.9) over  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\tilde{\mathbf{C}}$ , and  $\mathbf{G}_c$  can be considerably faster than minimization of (4.7) over  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$ , and  $\mathbf{G}_c$ . On the other hand, the use of (unconstrained) splines on the data rather than on the components may be easier to handle, since standard software can be used to obtain the smooths, and subsequently analyze them by the Tucker3 or the CP model.

By constraining the B-spline or I-spline weights, one can impose a constraint on the solution in addition to the smoothness constraint. If I-spline weights are restricted to non-negativity, the smooth is monotonically increasing (and non-negative as well). The smooth can be restricted to be non-negative by requiring the B-splines weights to be non-negative. The problem of finding non-negative weights for the splines can be solved by treating the problem as a non-negative least squares problem (Lawson & Hanson; 1974, pp. 158-164; Bro & De Jong, 1997). Note that if spline weights are constrained, imposing a spline basis on a component matrix will have a different effect than imposing a spline basis on the data matrix. If a spline basis is imposed on a *component matrix* with constrained weights, we have to minimize (4.4) over  $\mathbf{U}$ ,  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{G}_c$ , subject to appropriate constraints. If a spline basis is imposed on the *data matrix* with constrained weights  $\mathbf{W}$ , (4.6) has to be minimized, subject to the appropriate constraints, before analyzing the restricted projection by Tucker3 or CP.

It is expected that applying smoothness constraints will help recover the underlying true curves when applying the Tucker3 or CP models to data with smooth underlying structure. In fitting the unconstrained CP model, local minima are often encountered when the underlying curves are smooth. In this case the associated component matrix often has a high degree of multicollinearity, which may cause the model to be ill-defined, and thus lead to local minimum problems. It is expected that

the algorithm to fit CP will land in a local minimum less frequently if smoothness constraints are used, especially in the case of high multicollinearity of the component matrices.

#### 4.4. Comparing constrained with unconstrained CP and Tucker3 models

To test the usefulness of smoothing in the CP and Tucker3 models and compare it with the unrestricted CP and Tucker3 models, respectively, we performed a simulation study on the basis of 480 data sets for the CP model, and 960 data sets for the Tucker3 model. The algorithms were programmed in MATLAB5 (1998), and the analyses were carried out on a Pentium 333Mhz 32 Mb RAM personal computer in a Windows 95 environment.

##### 4.4.1. Construction of the data for the simulation study

###### *CP data for the simulation study*

For the CP simulation study, 480 data sets were constructed with known CP structure with smooth components in one mode ( $\mathbf{C}$ ), and various data sizes, numbers of components, degrees of multicollinearity in  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$ , and error levels. The data matrices  $\mathbf{X}_c$  ( $K \times IJ$ ) were constructed according to

$$\mathbf{X}_c = \mathbf{C}_o \mathbf{H}(\mathbf{B}_o \hat{\otimes} \mathbf{A}_o) + \varepsilon \mathbf{N}_c, \quad (4.10)$$

where  $\mathbf{A}_o$  ( $I \times Q$ ),  $\mathbf{B}_o$  ( $J \times Q$ ), and  $\mathbf{C}_o$  ( $K \times Q$ ) are 'true' component matrices for the respective three modes,  $\mathbf{H}$  ( $Q \times Q^2$ ) is the matrix version of the superdiagonal three-way array  $\underline{\mathbf{H}}$  ( $Q \times Q \times Q$ ),  $\varepsilon$  is a scalar, and  $\mathbf{N}_c$  ( $K \times IJ$ ) denotes the matrix expression of the three-way error array  $\underline{\mathbf{N}}$  ( $I \times J \times K$ ).

The data sizes  $I, J, K$  were 10,10,20; 10,10,50; 10,50,20 and 10,50,50. The numbers of components were two and four. The elements of the matrices  $\mathbf{A}_o$  and  $\mathbf{B}_o$  were drawn randomly from the uniform [0,1] distribution (mild multicollinearity condition), and from the uniform [0.5,1.5] distribution (severe multicollinearity condition). To ensure smoothness of the components of  $\mathbf{C}_o$ , every component of  $\mathbf{C}_o$  followed a smooth function evaluated at  $K$  equidistant points (to be denoted by  $t_1, \dots, t_K$ ) in every condition. Half of the components of  $\mathbf{C}_o$  followed an exponential function, and half of the components followed a logistic function, both of which are often employed as growth curves (Browne, 1993). The parameters were varied so that in the mild multicollinearity condition of  $\mathbf{C}_o$  the condition numbers for two and four components of  $\mathbf{C}_o$  were two and six, respectively, whereas in the severe multicollinearity condition the condition numbers were six and 42. The values of  $\mathbf{N}_c$  were drawn randomly from the standard normal distribution and multiplied by a scalar  $\varepsilon$  chosen so that the expected percentages of error in  $\underline{\mathbf{X}}$  were 2%, 26%, or 50% (in terms of sums of squares). The number of replications was five. The design was

fully crossed, leading to a total of four (data sizes)  $\times$  two (numbers of components)  $\times$  two (degrees of multicollinearity of  $\mathbf{A}_o$  and  $\mathbf{B}_o$ )  $\times$  two (degrees of multicollinearity of  $\mathbf{C}_o$ )  $\times$  three (error levels)  $\times$  five (replications) = 480 matrices.

#### *Tucker3 data for the simulation study*

For the Tucker3 simulation study, 960 data sets were constructed with known Tucker3 model structure with smooth components in one mode ( $\mathbf{C}$ ), and various data sizes, numbers of components, degrees of multicollinearity in the core, and error levels. The data matrices were constructed as

$$\mathbf{X}_c = \mathbf{C}_o \mathbf{G}_o (\mathbf{B}_o' \otimes \mathbf{A}_o') + \varepsilon \mathbf{N}_c, \quad (4.11)$$

where  $\mathbf{A}_o$  ( $I \times P$ ),  $\mathbf{B}_o$  ( $J \times Q$ ), and  $\mathbf{C}_o$  ( $K \times R$ ) are ‘true’ component matrices for the three modes,  $\mathbf{G}_o$  ( $R \times P \times Q$ ) is the matricized version of the three-way core array  $\underline{\mathbf{G}}_o$  (the subscript ‘c’ is omitted for notational simplicity),  $\varepsilon$  is a scalar, and  $\mathbf{N}_c$  ( $K \times I \times J$ ) denotes the matrix expression of the three-way error array  $\underline{\mathbf{N}}_c$ .

The sizes of the data array  $\underline{\mathbf{X}}$   $I, J, K$  were 10,10,20; 10,20,20; 10,10,50; 10,50,20; 10,20,50; 30,20,20; 10,50,50 and 30,20,50. The numbers of components  $P, Q, R$  for the three modes were 2,2,2; 2,4,2; 2,2,4 and 4,4,4. The component matrices  $\mathbf{A}_o$ ,  $\mathbf{B}_o$ , and  $\mathbf{C}_o$  were chosen column-wise orthonormal. The components of the smooth  $\mathbf{C}_o$  followed the same functions as in the CP simulation study, but now the orthonormal bases of the matrices used in the CP simulation study were used. The matrices  $\mathbf{A}_o$  and  $\mathbf{B}_o$  were obtained by taking the orthonormal bases of a matrix with equal size as  $\mathbf{A}_o$  and  $\mathbf{B}_o$  with elements drawn randomly from the uniform [0,1] distribution. These choices do not place severe limitations on the simulation study, since the component matrices in the Tucker3 solution, and hence any set of ‘true’ component matrices of a Tucker3 model in a simulation study, can be transformed to orthonormality, provided that this transformation is compensated in the core. However, transformation of a multicollinear true component matrix to orthonormality and compensation for this in the core array would lead to a multicollinear core. For example, suppose we have a matrix  $\mathbf{C}$  and  $\mathbf{G}$ , where  $\text{cond}(\mathbf{C})=100$ , and  $\mathbf{G}$  is row-wise orthonormal so that  $\text{cond}(\mathbf{G})$  is 1, where  $\text{cond}(\cdot)$  means the condition number. Orthonormalization of  $\mathbf{C}$  into  $\tilde{\mathbf{C}}$ , and compensation for the orthonormalization in  $\mathbf{G}$  by transforming  $\mathbf{G}$  into  $\tilde{\mathbf{G}}$  results in  $\text{cond}(\tilde{\mathbf{C}})=1$ . This can be achieved by taking the QR-decomposition of  $\mathbf{C}=\mathbf{Q}\mathbf{R}$ , defining  $\tilde{\mathbf{C}}=\mathbf{Q}=\mathbf{C}\mathbf{R}^{-1}$ , and  $\tilde{\mathbf{G}}=\mathbf{R}\mathbf{G}$ , and, as a result,  $\text{cond}(\tilde{\mathbf{G}})=\text{cond}(\mathbf{R}\mathbf{G})=100$ . Therefore, to represent a reasonable range of possible data matrices, the degree of multicollinearity of the core is varied in this study. The elements of  $\mathbf{G}_o$  were drawn randomly from the uniform [0,1] distribution in the low multicollinearity condition, and from the uniform [.5,1.5] distribution in the high multicollinearity condition. The error level was varied in the same way as in the CP simulation study, that is the expected percentages of error sum of squares of  $\underline{\mathbf{X}}$  were 2%, 26%, and 50%. The

number of replications in each condition was five. The design was fully crossed, leading to a total of eight (data sizes)  $\times$  four (numbers of components)  $\times$  two (degrees of multicollinearity of  $\mathbf{G}_0$ )  $\times$  three (error levels)  $\times$  five (replications) = 960 matrices.

#### 4.4.2. Analyses of simulation data

The simulated data sets  $\mathbf{X}_c$  were all analyzed by one unconstrained CP or Tucker3 analysis, and by two CP or Tucker3 analyses with smoothness constraints. Specifically, in the analyses with smoothness constraints, the estimated component matrix  $\mathbf{C}$  was restricted to be in the column space of a set of B-splines  $\mathbf{B}^s$  of degree three. The knots were equidistantly placed on the time interval  $t_1, \dots, t_K$ , with a knot at  $t_1$  and one at  $t_K$ . The CP or Tucker3 analyses with smoothness constraints were performed on the compressed data array (see (4.9)) instead of the full data array to reduce computation time. In one of the analyses with smoothness constraints, the numbers of knots were chosen so that the sum of the cross-validation sum of squares,  $\text{CV}(\boldsymbol{\lambda})$ , see (4.2), over columns of  $\mathbf{X}_c$  was minimized. That is, for a fixed number of knots, the  $\text{CV}(\boldsymbol{\lambda})$  was computed for each column of  $\mathbf{X}_c$ , and then the sum of the  $\text{CV}(\boldsymbol{\lambda})$ 's obtained in this way was computed. The sum of the  $\text{CV}(\boldsymbol{\lambda})$ 's was computed successively for solutions based on  $2, 3, \dots, t_K$  knots, and the number of knots that goes with the minimal sum of  $\text{CV}(\boldsymbol{\lambda})$ 's was chosen. The CP and Tucker3 analyses with these restrictions are referred to as CP-Bs(CV) and Tucker3-Bs(CV), respectively. In the other analysis with smoothness constraints, the number of knots of the B-splines was fixed at three. A small number of knots was chosen, since it is known that using too many knots leads to overfitting. However, the number three was somewhat arbitrary. The CP and Tucker3 analysis with this restriction are referred to as CP-Bs(3) and Tucker3-Bs(3), respectively. The estimates of CP were obtained by means of the CP algorithm of Harshman (1970), and Carroll and Chang (1970). Each Tucker3 analysis was performed using the efficient algorithm by Andersson and Bro (1998). In each analysis, the CP and Tucker3 algorithms were run from five different starts, one started rationally and four randomly, to reduce the chance of missing the global minimum. The rationally started runs were started with the parameters resulting from Tucker's Method I (Tucker, 1966a). The convergence criterion was set at  $10^{-6}$ .

#### 4.4.3. Criteria of interest

The main interest in this study was to determine how well the original component matrices (and core matrix in the case of the Tucker3 model) were recovered by each of the methods. Since the CP analysis yields unique estimates of the component matrices (up to permutation and scaling), while the solution of the Tucker3 analysis is not uniquely defined, different comparison criteria are used for the CP and the Tucker3 analyses.

*CP analyses: criteria of interest*

In the analysis of the CP data, a comparison of the estimated component matrices  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{B}}$ , and  $\hat{\mathbf{C}}$  and the original component matrices  $\mathbf{A}_o$ ,  $\mathbf{B}_o$ , and  $\mathbf{C}_o$  has to take into account possible permutations, rescalings, and sign reversions of the estimated component matrices. Following Kiers (1998a), and Mitchell and Burdick (1994), we compared the CP solutions by computing the cosines between the tensor products  $\mathbf{a}_r^o \otimes \mathbf{b}_r^o \otimes \mathbf{c}_r^o$ ,  $r=1, \dots, R$ , for the original component matrices and  $\hat{\mathbf{a}}_r \otimes \hat{\mathbf{b}}_r \otimes \hat{\mathbf{c}}_r$ ,  $r=1, \dots, R$ , for the estimated component matrices, where the subscript  $r$  denotes the  $r^{\text{th}}$  column of the matrix at hand. Given a data array  $\underline{\mathbf{X}}$  that is represented by a set of  $R$  tensor products of components, which are collected in component matrices  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$ , other sets of component matrices that yield the same representation of  $\underline{\mathbf{X}}$  are composed of the same such tensor products, although possibly in a different order. Therefore, a useful comparison measure of the original and the estimated component matrices is the mean of the  $R$  cosines between the tensor products of the original components and the tensor products of the estimated components, with the latter tensor products ordered so that they lead to the highest mean of cosines. The cosines are known as Tucker's coefficient of congruence (Tucker, 1951). The coefficient of congruence between two columns  $\mathbf{x}$  and  $\mathbf{y}$ ,  $\phi_{\mathbf{xy}}$ , is defined as the normalized inner product between the columns  $\mathbf{x}$  and  $\mathbf{y}$ , namely as

$$\phi_{\mathbf{xy}} = \frac{\mathbf{x}'\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{x}}\sqrt{\mathbf{y}'\mathbf{y}}}. \quad (4.12)$$

One rationally started and four randomly started runs of the CP analysis were carried out. The runs which led to a sub-optimal solution (defined here as a solution with a function value higher than 1.001 times the fit of the optimal solution, out of the five runs) were counted to get an impression of the sensitivity to local minima of the constrained and unconstrained analyses.

*Tucker3 analyses: criteria of interest*

To investigate how well the original matrices of the Tucker3 model are recovered, two aspects are of importance, namely the recovery of the column spaces of the component matrices, and the recovery of the weights of the interactions of the components. The column spaces of the component matrices are compared as follows: A comparison of the component matrices  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{B}}$ , and  $\hat{\mathbf{C}}$  as estimated by Tucker3 to the underlying  $\mathbf{A}_o$ ,  $\mathbf{B}_o$ , and  $\mathbf{C}_o$  has to take into account the fact that  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{B}}$ , and  $\hat{\mathbf{C}}$  can be transformed without loss of fit, provided that such transformations are compensated in the core. Therefore, the estimates of  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{B}}$ , and  $\hat{\mathbf{C}}$  are transformed towards  $\mathbf{A}_o$ ,  $\mathbf{B}_o$ , and  $\mathbf{C}_o$ , respectively, by postmultiplying  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{B}}$ , and  $\hat{\mathbf{C}}$  by the

matrices  $\mathbf{S}$ ,  $\mathbf{T}$ , and  $\mathbf{V}$ , respectively. The transformation matrices  $\mathbf{S}$ ,  $\mathbf{T}$ , and  $\mathbf{V}$  are found by minimizing the Euclidean distance between the original component matrices  $\mathbf{A}_o$ ,  $\mathbf{B}_o$ , and  $\mathbf{C}_o$  and the transformed component matrices  $\hat{\mathbf{A}}\mathbf{S}$ ,  $\hat{\mathbf{B}}\mathbf{T}$ , and  $\hat{\mathbf{C}}\mathbf{V}$ , respectively. The transformations are compensated in the estimated core matrix  $\hat{\mathbf{G}}$  by computing the transformed core array  $\tilde{\mathbf{G}} = \mathbf{V}^{-1}\hat{\mathbf{G}}((\mathbf{T}')^{-1} \otimes (\mathbf{S}')^{-1})$ . The component matrices  $\hat{\mathbf{A}}\mathbf{S}$ ,  $\hat{\mathbf{B}}\mathbf{T}$ , and  $\hat{\mathbf{C}}\mathbf{V}$  are compared to the original component matrices  $\mathbf{A}_o$ ,  $\mathbf{B}_o$ , and  $\mathbf{C}_o$  by computing the Proportion of Agreement ( $PA_A$ ,  $PA_B$ , and  $PA_C$ , respectively) as

$$PA_A = 1 - \frac{\|\mathbf{A}_o - \hat{\mathbf{A}}\mathbf{S}\|^2}{\|\mathbf{A}_o\|^2}; PA_B = 1 - \frac{\|\mathbf{B}_o - \hat{\mathbf{B}}\mathbf{T}\|^2}{\|\mathbf{B}_o\|^2}; PA_C = 1 - \frac{\|\mathbf{C}_o - \hat{\mathbf{C}}\mathbf{V}\|^2}{\|\mathbf{C}_o\|^2}. \quad (4.13)$$

To be able to compare the Proportion of Agreement of the component matrices simultaneously, the average of  $PA_A$ ,  $PA_B$ , and  $PA_C$ , denoted as  $PA_{ABC}$ , is used as the measure of agreement between the original and the estimated component matrices.

The recovery of the weights of the interactions of the components is examined by comparing the transformed core matrix  $\tilde{\mathbf{G}}$  to the original core matrix  $\mathbf{G}_o$  by computing the Proportion of Agreement ( $PA_G$ ) as

$$PA_G = 1 - \frac{\|\mathbf{G}_o - \tilde{\mathbf{G}}\|^2}{\|\mathbf{G}_o\|^2}. \quad (4.14)$$

Note that the transformed component matrices are optimally transformed towards the original component matrices, whereas the associated core matrix is not optimally transformed towards the original core matrix. Hence, it can be expected that the  $PA_G$  is smaller than the  $PA_{ABC}$  in the case of a Tucker3 solution deviating from the original matrices.

#### 4.4.4. Results of the simulation studies

##### *Results of the CP simulation study*

The original component matrices and the estimated component matrices, as obtained by unconstrained CP analysis (CP) and CP with smoothness constraints (CP-Bs(CV) and CP-Bs(3)), are compared by inspecting the  $\phi$ -values. The  $\phi$ -values have a negatively skewed distribution over the replications within each condition. The median  $\phi$ -values of the three analysis methods are plotted overall as well as per main condition in Figure 4.2.

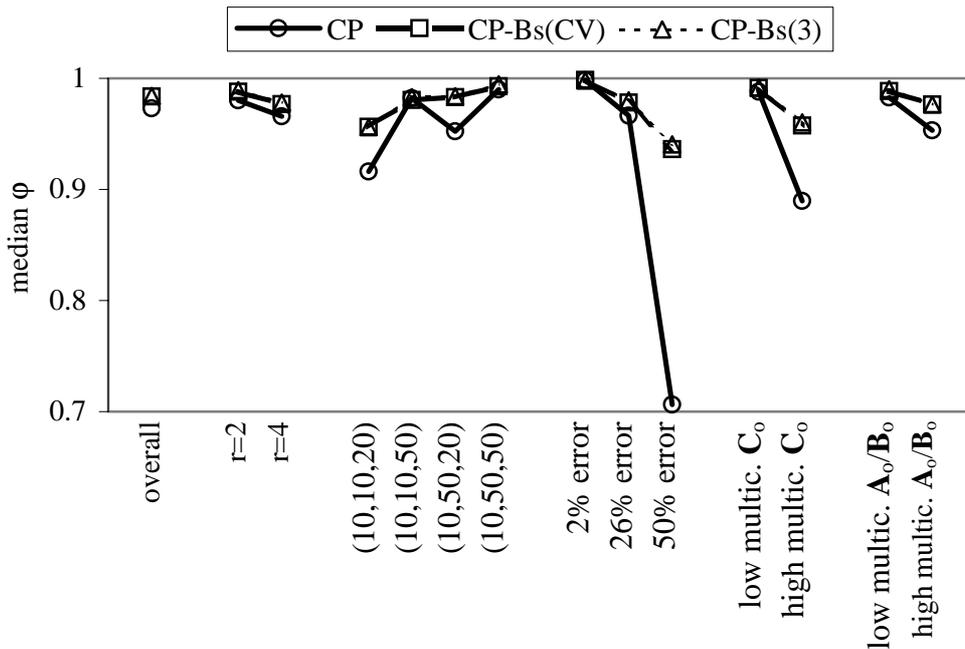


Figure 4.2. Median  $\phi$ -values of CP, CP-Bs(CV) and CP-Bs(3) per condition. 'Multic.' denotes 'multicollinearity'.

The following observations can be made in Figure 4.2. The median coefficient of congruence of the constrained CP solutions is larger than the median coefficient of congruence of the unconstrained CP solutions, whereas virtually no difference was found between the median coefficients of congruence of CP-Bs(CV) and CP-Bs(3). Furthermore, the difference between the unconstrained and the constrained CP solutions gets clearly larger with increasing condition numbers of  $C_o$ , and with increasing error level, and varies in a more complicated manner with data size (see Figure 4.2).

A repeated measurement ANOVA was performed to test whether the observed effects of type of analysis and of the interactions of analysis method with the various manipulated factors could be distinguished from random fluctuations. To correct for the deviation from normality for the repeated measurement ANOVA, the  $\phi$ -values were transformed into  $\tilde{\phi} = \log(\phi/(1-\phi))$  before analysis, where the two observed negative  $\phi$ -values were excluded from the analysis. The transformation of negatively skewed  $\phi$ -values on the interval  $[0,1]$  results in approximately normally distributed  $\tilde{\phi}$ -values on the interval  $[-\infty, \infty]$ . The effects which were described in the previous paragraph, were all found to be significant at  $\alpha=0.001$  in the repeated measurement ANOVA of the  $\tilde{\phi}$ -values.

In addition to the coefficients of congruence of the three analysis methods in the different conditions, the number of cases in which the unconstrained CP leads to a ‘good’ solution, and the constrained CP to a ‘bad’ solution is of interest. On the basis of inspection of a number of plots of original and estimated components and the accompanying coefficient of congruence, solutions with a coefficient of congruence smaller than 0.75 were considered to be bad. The resulting frequencies according to this criterion are presented in Table 4.1.

Table 4.1. Frequencies of good ( $\phi \geq 0.75$ ) and bad ( $\phi < 0.75$ ) solutions per analysis method (CP with CP-Bs(CV) and CP with CP-Bs(3)).

		CP-Bs(CV)		CP-Bs(3)	
		good	bad	good	bad
CP	good	348	3	348	3
	bad	43	86	40	89

In a large number of cases, both the constrained and the unconstrained CP model lead to a good solution. The most important finding is that if the unconstrained CP leads to a bad solution, the constrained CP model leads to a good solution in about 33% of the cases. Furthermore, it is rarely found that the unconstrained CP model leads to a good solution and the constrained CP model to a bad solution. The proportion of bad solutions, as well as the differences between the constrained and unconstrained CP model increases with error level, and condition number of  $C_0$ . Thus, on the basis of these results we can conclude that, if there is a smooth underlying structure, B-spline constrained CP is helpful in a fair number of cases, and that there is very little risk in replacing unconstrained CP by CP with smoothness constraints. Moreover, the choice for the number of knots does not seem crucial.

Differences between the constrained and unconstrained CP analyses in sensitivity to local minima were also studied. The constrained CP analyses led to a sub-optimal solution a little less frequently (both 0.10 out of five starts on average) than the unconstrained analyses (0.16 on average). No difference in average local minima has been found between the rationally and the randomly started runs. The number of local minima increased with increasing error level, whereas no substantial interaction between any other of the manipulated factors and type of analysis was found.

#### Results of the Tucker3 simulation study

The original component matrices and the original core matrix were compared to the estimated component matrices and the estimated core matrix by means of the Proportion of Agreement of the component matrices and the core matrix, the  $PA_{ABC}$ , which is based on the average of the expressions in (4.13), and the  $PA_G$ , (4.14), respectively. The average  $PA_{ABC}$  values per analysis method give a good impression of the condition effects, and they are plotted per condition in Figure 4.3.

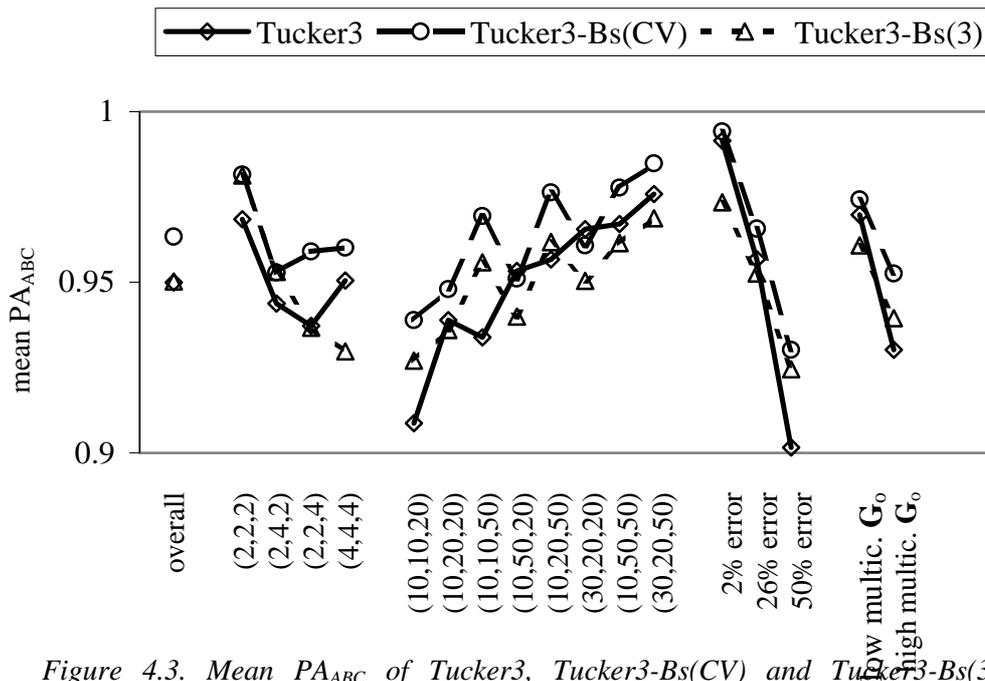


Figure 4.3. Mean  $PA_{ABC}$  of Tucker3, Tucker3-Bs(CV) and Tucker3-Bs(3) per condition. 'Multic.' denotes 'multicollinearity'.

As can be seen in Figure 4.3, the  $PA_{ABC}$  of Tucker3-Bs(CV) is generally higher than that of Tucker3 and Tucker3-Bs(3), whereas almost no difference was found between the  $PA_{ABC}$  of Tucker3 and Tucker3-Bs(3), over all conditions. The difference in  $PA_{ABC}$  between the three methods of analysis increases with increasing core size, error percentage, and degree of multicollinearity of the core, and varies with data size. The gain of the smooth Tucker3 over the unconstrained Tucker3 is largest in the case of a relatively large size of the smooth mode, and relatively small sizes of the non-smooth modes, for example, data size 10,10,50. If the size of the smooth mode is smaller than the size of one of the non-smooth modes, the performance of Tucker3 is better than the smooth Tucker3's (e.g., data sizes 30,20,20 and 10,50,20). Tucker3 clearly outperforms Tucker3-Bs(3) in the case of low error level (2%) and low multicollinearity of the core, whereas Tucker3-Bs(CV) performs best of the three. In high error level and high multicollinearity conditions, Tucker3-Bs(3) performs better than Tucker3, but Tucker3-Bs(CV) gives best recovery of the component matrices. This finding suggests that the smoothness restricted Tucker3 is sensitive to the choice of number of knots, and that in 'easy conditions' an unconstrained Tucker3 model performs even better than a smoothness constrained Tucker3 model with a non-optimal number of knots.

A repeated measurement ANOVA was performed to test whether the observed effects of type of analysis and of the interactions of analysis method with the various

manipulated factors could be distinguished from random fluctuations. For the repeated measurement ANOVA, the  $PA_{ABC}$ -values were transformed to correct for the observed heterogeneity of variances for the groups by computing  $\tilde{P}\tilde{A}_{ABC} = \arcsin(PA_{ABC})^{1/2}$  (Stevens, 1992). The effects that were explicitly described in the previous paragraph, were found to be significant at  $\alpha=0.001$  in the repeated measurement ANOVA of the  $\tilde{P}\tilde{A}_{ABC}$ -values.

The estimated component matrices  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{B}}$ , and  $\hat{\mathbf{C}}$  are optimally transformed to the original component matrices  $\mathbf{A}_o$ ,  $\mathbf{B}_o$ , and  $\mathbf{C}_o$ , whereas the transformation of the estimated core matrix  $\hat{\mathbf{G}}$  is so that the transformations of the original component matrices are compensated. Therefore, a non-optimal recovery will be expressed in a low  $PA_G$  value, and possibly in a low  $PA_{ABC}$  value. The  $PA_G$  values appear highly negatively skewed, with some extremely low values, hence the median  $PA_G$  values give a better insight into the condition effects than the mean  $PA_G$  values. The median  $PA_G$  values per condition appeared to be high ( $>0.985$ ), and they hardly differ from each other, neither between type of analysis nor between conditions. The extremely low values all occurred in the ‘more difficult’ conditions, namely large core size, small data size, high condition number of the core  $\mathbf{G}_o$ , and high error level. The Tucker3 analysis showed more extremely low  $PA_G$  values than the Tucker3-Bs(CV) and Tucker3-Bs(3) analyses, as is indicated by, for example, the percentages of the cases with  $PA_G$  values lower than 0.5 of 4.7%, 1.1% and 1.3%, respectively.

A second way of comparing the achievement of the three methods of analysis is to inspect the number of cases that are recovered well by the different methods. On the basis of inspection of a number of original and estimated components, cores and associated  $PA_{ABC}$  and  $PA_G$  solutions with a  $PA_{ABC}$  or a  $PA_G$  smaller than 0.9 were considered to be bad. The resulting frequencies of good and bad solutions are presented in Table 4.2. It can be seen in this table that if Tucker3 leads to a bad solution, Tucker3-Bs(CV) leads to a good solution in 44% of the cases. In only 1% of the cases, the Tucker3-Bs(CV) is bad, while the Tucker3 solution is good. According to the frequencies in Table 4.2, Tucker3-Bs(3) performs almost as well as Tucker3-Bs(CV).

Table 4.2. Frequencies of good ( $PA_{ABC}>0.9$  and  $PA_G>0.9$ ) and bad solution per analysis method (Tucker3 with Tucker3-Bs(CV) and Tucker3 with Tucker3-Bs(3)).

		Tucker3-Bs(CV)		Tucker3-Bs(3)	
		good	bad	good	bad
Tucker3	good	811	7	808	10
	bad	63	79	62	80

Although the number of solutions that were reasonably recovered by Tucker3-Bs(3) does not deviate much from the number of reasonable recoveries using Tucker3-Bs(CV), the smoothing technique is sensitive to the choice for the number of knots, as

indicated by the better recovery of the underlying component structure by Tucker3-Bs(CV) than of Tucker3-Bs(3). Thus, we can conclude on the basis of these results, that if there is a smooth underlying structure, a smoothness constrained Tucker3 model is helpful in a reasonable number of cases, and that, conversely, there is very little risk in using smoothness constrained instead of unconstrained Tucker3. Tucker3-Bs(CV), the method with optimal knot selection, performed best, and is therefore preferable to Tucker3-Bs(3).

#### 4.5. Empirical example: Learning to read study (I)

In this section, an empirical example is presented to illustrate the use of smoothness constraints in the Tucker3 model, and the use of monotonicity constraints on the data before analyzing the data by the Tucker3 model. The ‘Learning to read study’ (Jansen & Bus, 1982; Bus & Kroonenberg, 1982) investigates the learning process of reading. Seven pupils were tested weekly (except for holidays) on 37 occasions using five different tests intended to measure different aspects of reading ability. The primary research questions focused on whether the performance of the pupils per test and over tests was equal over time. This data set has been analyzed using a Tucker3 model by Kroonenberg (1983). However, mainly because our preprocessing procedure is different from the one applied by Kroonenberg, the results are not directly comparable.

The data elements  $x_{ijk}$ , where  $i=1,\dots,7$  denote the subjects,  $j=1,\dots,5$  the variables, and  $k=1,\dots,37$  the occasions, were collected in the data array  $\underline{\mathbf{X}}$ . Since the tests have different score ranges, the data array  $\underline{\mathbf{X}}$  was rescaled so that the scores of all tests ranged from 0 to 1. In this way, all the differences in variation were maintained in the data, while the test scores were comparable. The data have a meaningful zero point, as a score of zero on a test indicates that the pupil has not mastered the corresponding ability at all. We consider the data to be approximately ratio scale. We therefore did not center the data. The rescaled scores were collected in the data array  $\underline{\mathbf{Y}}$  and first analyzed by the unconstrained Tucker3 model.

The scores are viewed as evaluations of growth curves, which are assumed to follow some smooth curves in the course of time. Therefore, in the second analysis, the components of the occasion mode (C) are constrained to be smooth, and this analysis will be referred to as ‘T3-Bs’. A smoothness constrained Tucker3 model is fitted to  $\underline{\mathbf{Y}}_c$  by minimizing (4.9), which is equivalent to minimizing (4.4). The degree of the B-spline was fixed at three. The knots were placed equidistantly, and the number of knots was chosen so that the sum of cross-validation sum of squares ( $CV(\boldsymbol{\lambda})$ ; see (4.2)) of the columns of  $\underline{\mathbf{Y}}_c$  was minimized, by computing the sum of  $CV(\boldsymbol{\lambda})$ ’s related to B-splines with 2,3,...,10 knots, and choosing the number of knots that goes with the minimal sum of  $CV(\boldsymbol{\lambda})$ ’s.

Because the data pertain to learning data, it seems reasonable to assume that the true scores per variable per subject are non-decreasing on subsequent occasions, that

is that the reading ability of the child never decreases in the course of time. To model non-decreasing true scores, a smoothed data matrix  $\tilde{\mathbf{Y}}_c = \mathbf{B}^i \mathbf{W}$  is obtained by minimizing  $\|\mathbf{Y}_c - \mathbf{B}^i \mathbf{W}\|^2$  over  $\mathbf{W}$ , where  $\mathbf{B}^i$  is a fixed I-spline matrix and  $\mathbf{W}$  the matrix of weights for the I-splines that are restricted to non-negativity. As a result,  $\tilde{\mathbf{Y}}_c$  is restricted in the sense that  $\tilde{y}_{ijk} \leq \tilde{y}_{ij(k+1)}$  for all  $i=1, \dots, I; j=1, \dots, J$ , and  $k=1, \dots, K-1$ . An unconstrained Tucker3 analysis is applied to the smoothed data. This third analysis of the ‘Learning to read data’ will be referred to as ‘T3-Bi’. The degree of the I-spline matrix was fixed at two. The number of knots was selected by subjective comparison of the observed variables and several estimated response variables.

In all three analyses, the numbers of components were chosen to be (2,1,2) for the subject, variable and occasion modes, respectively. As will be discussed below, the models with this number of components fits the data well, and is interpretable, parsimonious, and stable.

The fit of the unconstrained Tucker3 model applied to  $\mathbf{Y}$  is 96.26%. The estimated core matrix  $\mathbf{G}_a$  of the model positioned in principal axes orientation was diagonal. The core matrix was transformed to identity (which can always be done in case  $P=QR$  (Murakami, Ten Berge & Kiers, 1998, p. 256)), and this rescaling was compensated in the subject component matrix. The columns of the component matrices were rescaled so that the solution was easy to interpret: the maximum values of the second subject component, of the variable component, and of the first occasion component were rescaled to 1, and these rescalings were compensated in the first subject component and the second occasion component. The component matrices for the subjects ( $\mathbf{A}$ ) and the variables ( $\mathbf{B}$ ) of the unconstrained Tucker3 analysis of  $\mathbf{Y}$  are presented in Tables 4.4 and 4.5, respectively. The occasion component scores are plotted in Figure 4.4.

Table 4.3. Subject component scores of the unconstrained Tucker3 solution.

$\mathbf{A}$ (subjects)	1 <sup>st</sup> component	2 <sup>nd</sup> component
1	1.06	-0.42
2	0.96	-0.30
3	0.99	-0.38
4	1.28	1.00
5	1.16	0.19
6	1.09	-0.01
7	0.89	-0.42

Table 4.4. Variable component scores of the unconstrained Tucker3 solution.

<b>B</b> (variables)	
Letter Knowledge	0.91
Regular Orthographic Short Words	1.00
Regular Orthographic Long Words	0.87
Regular Orthographic Long and Short Words within Context	0.99
Irregular Orthographic Long and Short Words	0.58

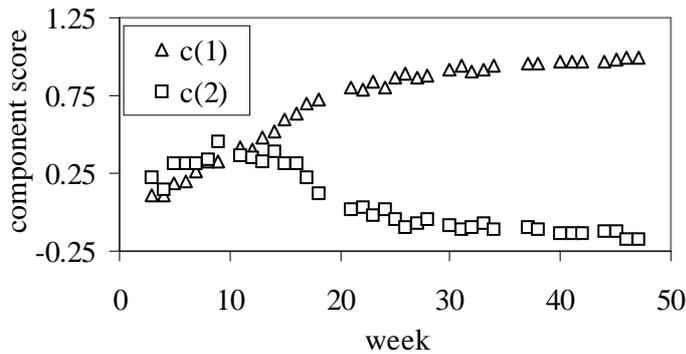


Figure 4.4. Occasion component scores resulting from the unconstrained Tucker3 model;  $c(1)$  denotes the scores on the first component,  $c(2)$  the scores on the second component.

To interpret the component scores, we start with the occasion components. In Figure 4.4, it can be seen that the scores on the first component ( $c(1)$  in the figure) gradually increase from week 3 to week 20, and then levels off to an asymptote of one. The scores on the second component show a steady increase to week 10, a steady decrease from week 10 to 20, and then level off to slightly below zero. We would interpret the first component as indicating general performance level, and the second component as approximately reflecting learning rate. The latter component is not entirely interpretable as learning rate, because of the negative component scores, which are due to the estimated model parameters, and do not indicate that the performance decreases in the end.

The core matrix is identity, implying that the first component of the subject component matrix (**A**) is only related to the first component of the occasion component matrix (**C**), and the same holds for the second components of **A** and **C**.

Now, the subject component matrices can be interpreted. Recall that the general performance level of a subject (thus apart from specific variable effects) is a weighted sum of the two occasion components, which reflect general performance level and learning rate. A relatively high score on the first component means that the subject

concerned performs above the general performance level. A relatively high score on the second component implies that the subject shows a relatively fast growth in the learning. Hence, for example, Subject 4 performs by far best, as (s)he has a high performance level and a high learning rate. Subject 4 is followed at some distance by Subject 5, as (s)he has second position for performance level, and learning rate. Subjects 1 and 6 show approximately the same weighting for general performance level, but Subject 1 has a lower weight for the second component. Hence, their asymptote scores are more or less equal, but Subject 1 develops much more slowly than Subject 6. The performance order between Subjects 1 to 3 is somewhat difficult to see at once, as the performance level of a subject is a weighted sum of the two occasion components, and the weights are close to each other. One could plot the weighted sum of the occasion component scores for the three subjects (hence, rows 1, 2 and 3 of  $\mathbf{AG}_a\mathbf{C}'$  in this particular case), and this would reveal that Subject 1 performs best of the three, and Subject 2 worst.

As there is only one variable component, the relative sizes of the variable component scores denote the difficulties of the items. The variable 'Irregular Orthographic Long and Short words' is by far the most difficult variable, as indicated by the lowest variable component score. Hence, the scores on 'Irregular Orthographic Long and Short Words' develop slowly in the course of time, compared to the other variables. The variable component scores of the 'Regular Orthographic Short Words' and the 'Regular Orthographic Long and Short Words within Context' are the highest variable component scores, showing that these scores develop fastest in the course of time. The variable component score of the 'Letter Knowledge' is slightly larger, and thus develops slightly faster, than the 'Regular Orthographic Long Words'.

The stability of the model just discussed was investigated via a split-half analysis, following the guidelines by Kiers and Van Mechelen (2001). That is, the data were split in two halves over the subject mode, resulting in one data set of three subjects, and one of four subjects, to be denoted as  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$ . A Tucker3 analysis was performed for each of the two data sets. The solutions for  $\mathbf{B}$  and  $\mathbf{C}$  for each of the data sets were optimally transformed (in the least squares sense) to the solutions for  $\mathbf{B}$  and  $\mathbf{C}$  of the full data set (as presented in

Table 4.4 and Figure 4.4, respectively). The two transformed occasion component matrices obtained in this way were compared by computing the coefficients of congruence (Tucker, 1951; see also Section 4.4.3) between the columns of the matrices. The two variable component matrices were compared analogously. The subject component matrices were compared as follows: The two solutions for  $\mathbf{A}$  for each of the splits were collected in one matrix  $\mathbf{A}_{12}$ , in which the rows pertain to the same subjects as in  $\mathbf{A}$  of the full data set (as presented in Table 4.3). The matrix  $\mathbf{A}_{12}$  was regressed on  $\mathbf{A}$ , and the resulting transformed component matrix was compared to  $\mathbf{A}$  by computing the coefficients of congruence between the two columns of the matrices. For each of the splits, the transformations of  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  were compensated in the core array. The separate split-half core arrays and the full data set core array were compared by computing the mean absolute difference between the split-half core array and the full data set core array.

The split-half procedure was repeated for every possible combination of the seven subjects split into two groups of three and four subjects, resulting in 35 split-half analyses. The mean coefficients of congruence for the subject component matrices over the 35 analyses was 1.000 and 0.997. The mean coefficient of congruence for the variable components was 0.998. The mean coefficient of congruence for the occasion components was 0.996 and 0.713 for the first and second occasion component, respectively. This implies that the stability of the subject components, the variable component and the first occasion component is high, whereas the stability of the second occasion component is moderate. The mean absolute difference between the split-half core array and the full data set core array averaged over the 35 analyses was 0.000. On the basis of these results, we conclude that the current Tucker3 model is sufficiently stable, given the small sample size at hand. We now turn to the results of the smoothness constrained analyses (T3-Bs and the T3-Bi), which will be discussed successively.

In the smooth Tucker3 analysis with B-splines, the T3-Bs, an unconstrained Tucker3 analysis was performed on the smoothed data array. The number of knots as indicated by the (minimal) cross-validation sum of squares was five. (The CV values for two through seven knots were 0.0096, 0.0089, 0.0089, 0.0087, 0.0089, 0.0097, respectively.) The fit of this constrained model to the data array  $\mathbf{Y}$  was 96.18%, which is only 0.08% less than the fit of unconstrained Tucker3 model.

Just as in the unconstrained Tucker3 model, the estimated core matrix of the solution in principal axes orientation was diagonal, and was rescaled to identity. The component score matrices of the T3-Bs model were rescaled in the same way as was done with the unconstrained Tucker3 model. The estimated component matrices  $\mathbf{A}$  and  $\mathbf{B}$  of T3-Bs are compared to the solutions of the unconstrained Tucker3 model by computing the coefficient of congruence between the pairs of components concerned. This coefficient was large ( $>0.999$ ) for all pairs, and therefore the solutions of  $\mathbf{A}$  and  $\mathbf{B}$  for T3-Bs can be interpreted in the same way as the corresponding solutions for the unconstrained Tucker3. The component scores for the occasions for T3-Bs are plotted in Figure 4.5.

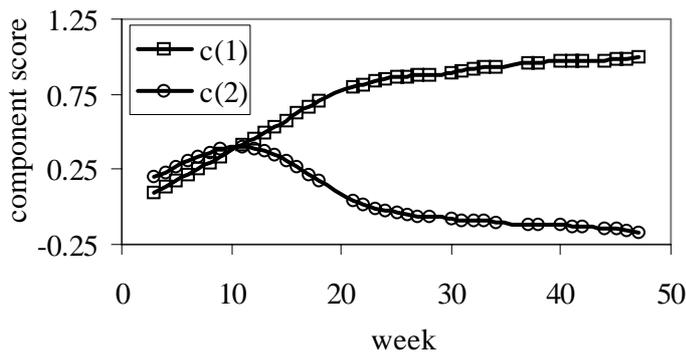


Figure 4.5. Occasion component scores of the T3-Bs model, where  $c(1)$  denotes the scores on the first component,  $c(2)$  the scores on the second component.

Not surprisingly, the component scores of the occasions follow more or less the same curve as in Figure 4.4. However, the wiggles have disappeared, and the overall trend in the component scores of the occasions is more clear.

It is interesting to investigate whether the stability of the occasion component matrices of the T3-Bs has been improved on the unconstrained Tucker3 model. Additionally, it is important to check whether the subject and variable component matrices, and the core array of the T3-Bs model have a high stability, just as their counterparts in the Tucker3 model. The stability of the T3-Bs model was investigated using the split-half procedure, as discussed before for the unconstrained Tucker3 model, where the procedure was applied to the smoothed data. The mean coefficients of congruence of the occasion component matrices were 0.998 and 0.792, which is higher than the coefficients of congruence found in the unconstrained analyses of 0.996 and 0.713, respectively. The mean coefficients of congruence for the subject component matrices and the variable component matrices were equal to the ones found for the unconstrained Tucker3 model (1.000 and 0.997 for the subject component matrices, and 0.998 for the variable component matrices). Also, the mean absolute difference between the split-half core array and the full data set core array averaged over the 35 analyses was 0.000. On the basis of this results, one can conclude that the stability of the second occasion component of the T3-Bs model has indeed been improved somewhat compared to the unconstrained counterpart. The subject and variable component matrices, the first occasion component and the core array of the T3-Bs model are highly stable, just as their unconstrained counterparts.

In the T3-Bi analysis, the subsequent scores per variable and per subject were restricted to be non-decreasing in the course of time before analysis. The number of knots for the second degree I-spline matrix was chosen to be seven, on the basis of subjective comparison of the observed variables and several estimated response variables. The I-splines were defined on the interval from week 0 to 50. The fit of the

resulting estimates of the Tucker3 model to the unconstrained data array  $\underline{\mathbf{Y}}$  was 96.17%. Just as in the unconstrained Tucker3 and the Tucker3-Bs models, the core matrix of the solution in principal axis direction was diagonal. The estimated core and component matrices were rescaled in the same way as in the Tucker3 and Tucker3-Bs models. The estimated solutions for T3-Bi of  $\mathbf{A}$  and  $\mathbf{B}$  were compared to the associated solutions for the unconstrained T3 by the coefficient of congruence. The coefficients were high ( $>.999$ ) for all pairs concerned, and  $\mathbf{A}$  and  $\mathbf{B}$  are interpreted in the same way as  $\mathbf{A}$  and  $\mathbf{B}$  of the unconstrained Tucker3. The occasion component scores for T3-Bi resemble the occasion component scores for T3-Bs closely, as indicated by the coefficients of congruence (1.000 and 0.999, respectively). Therefore, our interpretation of this model is identical to the interpretation of the T3-Bs model.

#### 4.6. Discussion and conclusion

The results from the CP and Tucker3 simulation study demonstrate that, if smooth underlying components are present, applying smoothness constraints in the CP and Tucker3 models is generally useful for providing a better estimate of the (underlying) components of the CP and Tucker3 models (and the core of the Tucker3 model). The gain in estimation accuracy of constrained estimation is more salient in the case of larger numbers of components, high condition numbers of the component matrices and high error levels.

In the simulation study, the smoothness constraints were imposed by requiring that the smooth component matrix lies in the column space of a B-spline matrix. The performance of the constrained Tucker3 model appeared considerably better if the number of knots of the B-splines was optimized according to the cross-validation criterion compared to the fixed knots choice (of three knots). On the other hand, the performance of CP did not appear to be influenced by the method of choosing the number of knots. This finding suggests that the performance of the smoothness constrained Tucker3 model is more sensitive to the choice of the number of knots than the smoothness constrained CP model. This might be due to the intrinsic axis property of the CP model.

The empirical example demonstrates the use of smoothing in Tucker3 analysis, and the use of monotonicity and smoothness constraints on the data before a Tucker3 analysis. The interpretation of the component matrices of the variables and the subjects, and the core did not alter if a constrained Tucker3 model was used instead of an unconstrained Tucker3 model. The T3-Bi model, in which the analyzed data are constrained to be non-decreasing in the course of time, appears to be reasonable for the data concerned. However, adding the monotonicity constraint to the smoothness constraint did not alter the interpretation of the solution at all, and therefore the simpler T3-Bs model can be preferred here. The interpretation of the time component scores of the smooth constrained T3-Bs solution is clearer than the interpretation of the unconstrained Tucker3 solution, as it is hard to judge whether certain wiggles in the plot of the time component scores of the unconstrained Tucker3 model should be considered important. Additionally, the stability of the T3-Bs solution is higher than of the unconstrained Tucker3 solution. Therefore, in the case of (presumed) smooth components, it appears to be useful to use smoothness constraints on the Tucker3 and CP model.

The commonly used procedures to estimate the Tucker3 or CP models require all elements of the three-way data box to be observed. In the case of data with a smooth mode, the use of the proposed procedures for smoothing the data can be helpful in estimating missing data elements. In longitudinal data, this procedure can be particularly useful if all measurements take place in the same time span, but at different sets of time points for different variables and/or occasions, where the missing data can be assumed to be missing completely at random (Little & Rubin, 1987; see also Chapter 3). Note that the B-spline matrix  $\mathbf{B}^s$  ( $K \times N, K \geq N$ ) is a matrix

with  $N$  B-splines which are evaluated in all  $K$  values of the predictor. In the present chapter, the  $K$  measurements of the predictor, which simply represents the time points in the case of longitudinal data, were assumed to be equal for all  $i$  ( $i=1, \dots, I$ ), and  $j$  ( $j=1, \dots, J$ ), thus the B-spline matrix  $\mathbf{B}^s$  is equal for all  $i$  and all  $j$ , and (4.4) can be used. However, if there are different measurements of the predictor for different  $(i, j)$  combinations, a B-spline matrix  $\mathbf{B}_{ij}^s$  must be defined for every combination of  $i$  and  $j$ . Now, provided that  $\mathbf{B}_{ij}^s$  is of full column rank,  $\mathbf{x}_{ij}$  can be projected on  $\mathbf{B}_{ij}^s$  by minimizing

$$f_6(\mathbf{w}_{ij}) = \left\| \mathbf{x}_{ij} - \mathbf{B}_{ij}^s \mathbf{w}_{ij} \right\|^2. \quad (4.15)$$

The weights  $\mathbf{w}_{ij}$  can be used to estimate  $\hat{\mathbf{x}}_{ij}$  on the same time points for all  $i$  and  $j$ , namely by defining  $\hat{\mathbf{x}}_{ij} = \mathbf{B}_{ij}^s \mathbf{w}_{ij}$ . If the vectors  $\hat{\mathbf{x}}_{ij}$  are collected in  $\hat{\mathbf{X}}$  ( $K \times IJ$ ),  $\hat{\mathbf{X}}$  can be analyzed by unrestricted Tucker3 or CP procedures. Hence, the use of smoothness constraints in the Tucker3 and CP models is not only useful in improving the estimation accuracy, but also in dealing with data measured at unequal sets of time points.

# 5. Structured latent curve component models for longitudinal three-way data

## 5.1. Introduction

In Chapter 4, the use of smoothness constraints in the Tucker3 and CP models was discussed. A different approach to constraining the occasion component matrix is to impose a particular functional form. This approach is attractive if theoretical considerations suggest a certain functional relationship. Then, potential advantages of imposing such a functional form are reduction of error fitting and an increase in the interpretability of the estimated model.

Browne and Du Toit (1991) and Browne (1993) proposed structured latent curve models for learning data, in which the occasion component scores are parameterized parsimoniously in terms of a small number of parameters. In the next sections, the principles of this factor analysis approach are elaborated for use with the Tucker3 model for longitudinal three-way data. One could also apply the approach with the usually more restricted CP model and the less restricted Tucker2 and Tucker1 models, as will be explained briefly in the discussion section.

## 5.2. Structured latent curve two-way component models for growth data

The idea of structured latent curve three-way component models for growth data will be introduced by discussing a structured latent curve *two-way* component model for growth data. The two-way structured latent curve component model is a straightforward modification of Browne and Du Toit's (1991) model, which was further elaborated by Browne (1993). In the present model, it is assumed that all individual growth curves are a weighted sum of certain basis functions, while the average growth curve as estimated by the model follows a particular function. As will be shown, the imposed constraints can be generalized fairly easily to component models for longitudinal three-way data.

### 5.2.1. The SLC two-way component model for data measured at equal time points

Let  $\mathbf{X}$  ( $K \times I$ ) denote the matrix of scores of  $I$  subjects ( $i=1, \dots, I$ ) on one variable collected at  $K$  occasions. The  $k^{\text{th}}$  occasion takes place at time point  $t_k$  ( $k=1, \dots, K$ ), and  $t_1$  is usually 0. To facilitate the explanation and notation, it is assumed here that the scores of the subjects are collected at the same time points. At a later stage (Section 5.2.3) the case involving different time points for different subjects is discussed.

The model imposed in structured latent curve (SLC) two-way component analysis is given by

$$\mathbf{X}=\mathbf{C}\mathbf{A}'+\mathbf{E}, \quad (5.1)$$

where  $\mathbf{C}$  ( $K \times R$ ) denotes the occasion component scores matrix,  $\mathbf{A}$  ( $I \times R$ ) denotes the subject component scores matrix, and  $\mathbf{E}$  ( $K \times I$ ) denotes the matrix of residuals; the  $R$  components in the occasion component scores matrix represent  $R$  basis functions evaluated at each of the  $K$  time points. The estimated series of a particular subject is thus a weighted sum of the  $R$  basis functions.

A method for specifying the basis curves in structured latent curve analysis, as suggested by Browne and Du Toit (1991) and Browne (1993), is followed broadly here in defining the SLC two-way component model. It is assumed that the average scores across subjects at successive occasions ( $\bar{x}_k$ ,  $k=1, \dots, K$ ), as estimated by the model, follow a particular target function, which is evaluated in the time points  $t_1, \dots, t_K$ . Browne (1993) discusses the Gompertz function (Richards, 1959), the exponential function and the logistic function as target functions. These are particularly useful functions for modeling growth data. The exponential function has no point of inflection, whereas the logistic and Gompertz functions do have such a point. One should choose the target function on the basis of the characteristics of the data. We will use the Gompertz curve as target function. It is given by

$$g(t_k, \boldsymbol{\tau}) = \alpha \exp \left\{ \ln \left( \frac{\beta}{\alpha} \right) \exp(-t_k \gamma) \right\}, \quad (5.2)$$

where  $\boldsymbol{\tau}=(\alpha, \beta, \gamma)$ ,  $\alpha$  is the asymptote and represents potential performance,  $\beta$  is the function value at  $t_k=0$  and represents previously acquired skills, and  $\gamma$  determines the rate of change, which reflects the learning speed. The Gompertz curve is asymmetrical around its point of inflection, which occurs for  $t_k = \frac{1}{\gamma} \ln \left( -\ln \frac{\beta}{\alpha} \right)$ , with

function value  $g(t_k, \boldsymbol{\tau}) = \frac{\alpha}{e}$ .

For any subject  $i$ , the scores at successive time points  $t_k=t_1, \dots, t_K$ , as estimated by the model, follow some function that is not necessarily monotonic. The first-order Taylor polynomial about  $\boldsymbol{\tau}$  is used to model the score of subject  $i$  at time point  $t_k$  (Browne & Du Toit, 1991), namely as

$$x_{ki} = g(t_k, \boldsymbol{\tau}) + \tilde{a}_{i1} g'_1(t_k, \boldsymbol{\tau}) + \tilde{a}_{i2} g'_2(t_k, \boldsymbol{\tau}) + a_{i3} g'_3(t_k, \boldsymbol{\tau}) + e_{ki}, \quad (5.3)$$

where  $x_{ki}$  denotes the observed score of subject  $i$  at time point  $t_k$ , and  $g'_m(t_k, \boldsymbol{\tau}) = (\delta/\delta\tau_m)g(t_k, \boldsymbol{\tau})$ , with  $\tau_1 = \alpha$ ,  $\tau_2 = \beta$ ,  $\tau_3 = \gamma$ . The first order derivatives  $g'_m(t_k, \boldsymbol{\tau})$  are given in (5.6). As the Gompertz curve (5.2) has the property

$$g(t_k, \boldsymbol{\tau}) = \alpha g'_1(t_k, \boldsymbol{\tau}) + \beta g'_2(t_k, \boldsymbol{\tau}), \quad (5.4)$$

the observed score of subject  $i$  at occasion  $k$  (=time point  $t_k$ ) can be written as

$$x_{ki} = a_{i1} g'_1(t_k, \boldsymbol{\tau}) + a_{i2} g'_2(t_k, \boldsymbol{\tau}) + a_{i3} g'_3(t_k, \boldsymbol{\tau}) + e_{ki}, \quad (5.5)$$

where  $a_{i1} = (\alpha + \tilde{a}_{i1})$ ,  $a_{i2} = (\beta + \tilde{a}_{i2})$ , and  $a_{ir}$  denotes element  $(i, r)$  of the subject component scores matrix  $\mathbf{A}$ . To ensure that the average estimated curve follows a Gompertz curve with parameters  $\alpha$ ,  $\beta$ , and  $\gamma$ , the average component score (across subjects) is required to be  $\alpha$  and  $\beta$  for the first and second components respectively, and 0 for the third component. The elements of the matrix  $\mathbf{C}$  ( $K \times 3$ ) are now defined as

$$\begin{aligned} c_{k1} &= g'_1(t_k, \boldsymbol{\tau}) = \{1 - \exp(-t_k \gamma)\} \exp\left\{\ln\left(\frac{\beta}{\alpha}\right) \exp(-t_k \gamma)\right\} \\ c_{k2} &= g'_2(t_k, \boldsymbol{\tau}) = \left(\frac{\alpha}{\beta}\right) \exp\left\{-t_k \gamma + \ln\left(\frac{\beta}{\alpha}\right) \exp(-t_k \gamma)\right\}, \\ c_{k3} &= g'_3(t_k, \boldsymbol{\tau}) = -\alpha \ln\left(\frac{\beta}{\alpha}\right) t_k \exp\left\{-t_k \gamma + \ln\left(\frac{\beta}{\alpha}\right) \exp(-t_k \gamma)\right\}, \end{aligned} \quad (5.6)$$

$k=1, \dots, K$ . The first function in (5.6) is called the ‘asymptote basis function’, the second the ‘initial value basis function’ and the third the ‘learning rate basis function’. Examples of the three basis curves, and the associated target functions are shown in Figure 5.1.

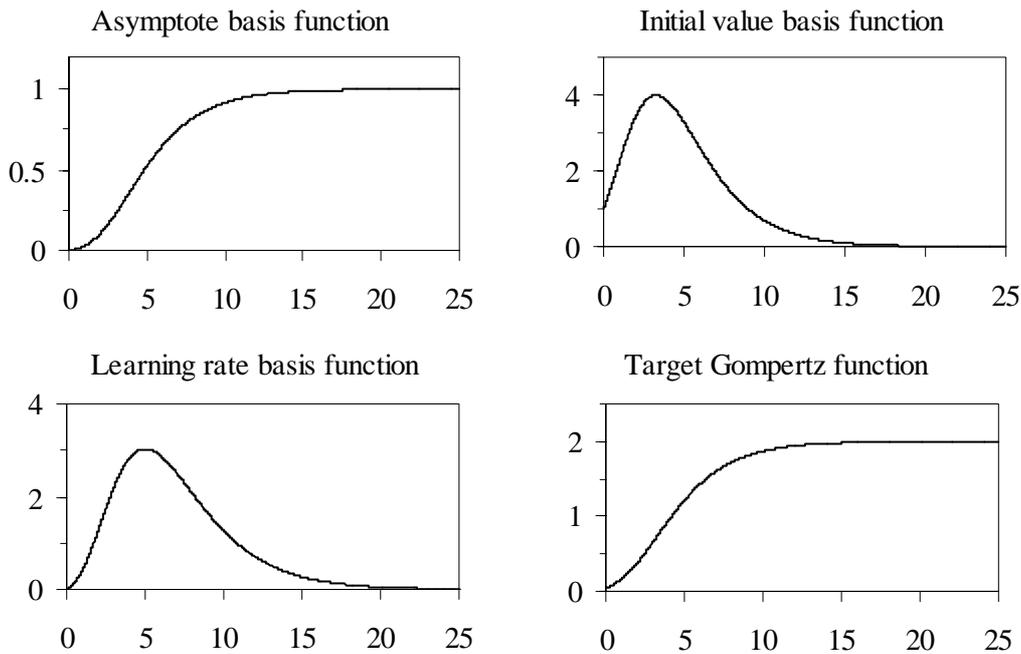


Figure 5.1. Examples of the three basis curves and the associated target functions.

The ‘asymptote basis function’ increases monotonically from zero at time point 0 towards an asymptote of one. A relatively large weight for this function for a particular subject denotes that the subject’s estimated growth curve has a relatively large asymptotic value compared to the other subjects. The ‘initial value basis function’ starts at 1, increases up to the time point where the inflection point of the target function occurs (here at  $t=3.26$ ), and then decreases towards an asymptote of zero. Unfortunately, this function not only reflects the initial value, but also the fact that the rate of learning first increases and then decreases. However, if the basis functions are evaluated only at time points after the inflection point, because of the particular choice of the measurement occasions, then the function decreases monotonically and the ‘initial value basis function’ can be interpreted well as only reflecting the initial value. If a particular subject has a relatively large weight for this function, this denotes that this subject’s estimated growth curve starts relatively high compared to the other subjects. The third basis function can be viewed as the ‘rate of learning basis function’. A relatively large weight for a particular subject for this function denotes that the subject concerned shows a relatively high rate of learning compared to the other subjects.

### 5.2.2. Fitting the SLC two-way component model to data with equal measurements

So far we have followed the approach proposed by Browne and Du Toit (1991) and Browne (1993). In their factor model, they make certain assumptions concerning the error structure. To fit the model to data, maximum likelihood estimation is used under the assumption of multivariate normality. Here, however, we use a component approach, and we estimate the parameters of the model via least squares estimation. As a result, to fit the SLC two-way component model to data, we propose to minimize the sum of squared residuals by minimizing

$$f_1(\boldsymbol{\tau}, \mathbf{A}) = \|\mathbf{X} - \mathbf{C}\mathbf{A}'\|^2, \quad (5.7)$$

where  $\boldsymbol{\tau} = (\alpha, \beta, \gamma)$ , the matrix  $\mathbf{C}$  ( $K \times 3$ ) consists of elements given by (5.6), and hence is a function of  $\alpha$ ,  $\beta$ , and  $\gamma$  only, and the matrix  $\mathbf{A}$  ( $I \times 3$ ) is restricted so that  $\frac{1}{I} \mathbf{1}' \mathbf{A} = [\alpha \ \beta \ 0]$ . In fitting the SLC two-way component model, using algorithms programmed in MATLAB5 (1998), we found that estimating the parameters of a reparametrization of the Gompertz function given in (5.2), and the corresponding reparametrizations of the basis functions in (5.6), which were given by Browne and du Toit (1991), led to far fewer computational problems than using the function in (5.2) itself as a target function. Therefore, the (reparametrized) occasion component scores matrix  $\tilde{\mathbf{C}}$  ( $K \times R$ ) and the subject component scores matrix  $\tilde{\mathbf{A}}$  ( $I \times R$ ), which can be transformed to  $\mathbf{C}$  and  $\mathbf{A}$  without altering the model estimates, as already noted by Browne (1993), are estimated. The parametrization of the Gompertz function used by Browne and du Toit (1991) is

$$\tilde{g}(t_k, \boldsymbol{\theta}) = \theta_1 \exp\{-\theta_2 \exp(-t_k \theta_3)\}. \quad (5.8)$$

The parameters  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$  in (5.8) can be transformed into the parameters of the Gompertz function in (5.2) by taking  $\alpha = \theta_1$ ;  $\beta = \theta_1 \exp(-\theta_2)$  and  $\gamma = \theta_3$ . The associated basis functions and hence the elements of the matrix  $\tilde{\mathbf{C}}$  ( $K \times 3$ ) are defined as

$$\begin{aligned} \tilde{c}_{k1} &= \tilde{g}'_1(t_k, \boldsymbol{\theta}) = \exp\{-\theta_2 \exp(-t_k \theta_3)\}, \\ \tilde{c}_{k2} &= \tilde{g}'_2(t_k, \boldsymbol{\theta}) = -\theta_1 \{\exp(-t_k \theta_3)\} \tilde{c}_{k1}, \\ \tilde{c}_{k3} &= \tilde{g}'_3(t_k, \boldsymbol{\theta}) = -t_k \theta_2 \tilde{c}_{k2}. \end{aligned} \quad (5.9)$$

The elements of  $\mathbf{C}$  and  $\tilde{\mathbf{C}}$  are related to each other as follows:

$$\begin{aligned}
c_{k1} &= \tilde{c}_{k1} + \frac{1}{\theta_1} \tilde{c}_{k2} \\
c_{k2} &= \frac{-1}{\theta_1 \exp(-\theta_2)} \tilde{c}_{k2} \\
c_{k3} &= \tilde{c}_{k3}.
\end{aligned} \tag{5.10}$$

Now, the model can be fitted to data by minimizing

$$f_1(\boldsymbol{\theta}, \tilde{\mathbf{A}}) = \|\mathbf{X} - \tilde{\mathbf{C}}\tilde{\mathbf{A}}'\|^2, \tag{5.11}$$

with  $\tilde{\mathbf{A}}$  ( $I \times 3$ ) restricted so that  $\frac{1}{T}\mathbf{1}'_I \tilde{\mathbf{A}} = [\theta_1 \ 0 \ 0]$ , because then the average estimated curve follows the Gompertz function as given in (5.8), with parameters  $\boldsymbol{\theta}$ . Finding an optimal solution for  $\tilde{\mathbf{A}}$  with unconstrained  $\tilde{\mathbf{a}}_1$ , which is the first column of  $\tilde{\mathbf{A}}$ , is easier than finding an optimal solution with  $\frac{1}{T}\mathbf{1}'_I \tilde{\mathbf{a}}_1 = \theta_1$ . Because the latter constraint can be satisfied afterwards without affecting the fit, (5.11) can be minimized subject to constraints on the second and third columns of  $\tilde{\mathbf{A}}$  only. When (5.11) has been minimized subject to the constraint on the second and third columns of  $\tilde{\mathbf{A}}$ ,  $\mathbf{1}'_I \tilde{\mathbf{a}}_2 = \mathbf{1}'_I \tilde{\mathbf{a}}_3 = 0$ , the constraint on the first column of  $\tilde{\mathbf{A}}$ ,  $\frac{1}{T}\mathbf{1}'_I \tilde{\mathbf{a}}_1 = \theta_1$ , can be satisfied as follows: Let  $\theta_1^c$ ,  $\mathbf{a}_2^c$  and  $\mathbf{a}_3^c$  be the estimates of  $\theta_1$ ,  $\tilde{\mathbf{a}}_2$  and  $\tilde{\mathbf{a}}_3$ , respectively, after convergence. Then defining  $\theta_1 = \frac{1}{T}\mathbf{1}'_I \tilde{\mathbf{a}}_1$ , rescaling  $\tilde{\mathbf{a}}_2$  and  $\tilde{\mathbf{a}}_3$  as  $\tilde{\mathbf{a}}_2 = \frac{\theta_1^c}{\theta_1} \mathbf{a}_2^c$  and

$\tilde{\mathbf{a}}_3 = \frac{\theta_1^c}{\theta_1} \mathbf{a}_3^c$ , and recomputing the elements of  $\tilde{\mathbf{C}}$  using (5.9), the product  $\tilde{\mathbf{C}}\tilde{\mathbf{A}}'$  remains

equal, and hence the fit is unaffected.

After having estimated the parameters  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ , one can obtain the parameters of the target function in (5.2) by taking  $\alpha = \theta_1$ ;  $\beta = \theta_1 \exp(-\theta_2)$  and  $\gamma = \theta_3$ . The matrices  $\mathbf{C}$  and  $\mathbf{A}$  can be obtained from  $\tilde{\mathbf{C}}$  and  $\tilde{\mathbf{A}}$  as  $\mathbf{C} = \tilde{\mathbf{C}}\mathbf{T}$ , and  $\mathbf{A} = \tilde{\mathbf{A}}(\mathbf{T}^{-1})'$ , where  $\mathbf{T}$  is given by

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{\theta_1} & \frac{-1}{\theta_1 \exp(-\theta_2)} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (5.12)$$

Estimates of  $\boldsymbol{\theta}$  and  $\tilde{\mathbf{A}}$  that minimize (5.11) subject to the constraint  $\frac{1}{I} \mathbf{1}'_I \tilde{\mathbf{a}}_2 = \frac{1}{I} \mathbf{1}'_I \tilde{\mathbf{a}}_3 = 0$  can be obtained via an alternating least squares (ALS) algorithm. The steps in the algorithm are as follows.

#### Initial estimates of $\boldsymbol{\theta}$ and $\tilde{\mathbf{A}}$ for the ALS algorithm

To start an ALS algorithm, initial estimates of the parameters are needed. An initial estimate of  $\boldsymbol{\theta}$ , and hence of  $\tilde{\mathbf{C}}$  as well, is obtained here by minimizing

$$f_2(\boldsymbol{\theta}) = \|\bar{\mathbf{x}} - \theta_1 \exp\{-\theta_2 \exp(-\mathbf{t}\theta_3)\}\|^2, \quad (5.13)$$

where  $\bar{\mathbf{x}}$  denotes the  $K \times 1$  vector containing the average scores over subjects at time points  $t_1, \dots, t_K$ , and  $\mathbf{t}$  is the vector with time points  $t_1, \dots, t_K$ . Least squares estimates of  $\boldsymbol{\theta}$  are obtained using the Levenberg-Marquardt algorithm (Seber & Wild, 1989). This algorithm needs starting values. As was discussed in Section 5.2.1, the parameters of the Gompertz curve have a physical interpretation. In the reparametrized form, the parameter  $\theta_1$  is the asymptote,  $\theta_2$  governs the distance from the asymptote, and  $\theta_3$  governs the learning rate. The point of inflection occurs at time point  $t_k = \log \theta_2 (\theta_3)^{-1}$ . Rational starting values can be obtained from a plot of the averaged scores across time, perhaps with a freehand smooth curve added (Seber & Wild, 1989). One should guess the asymptote value (denoted by  $\bar{x}_{max}$ ), the value at time point zero (denoted by  $\bar{x}_0$ ) and the time point at which the inflection point occurs (denoted by  $t_{infl}$ ). Then, as starting values, one takes  $\bar{x}_{max}$  for  $\theta_1$ ,  $-\log(\bar{x}_0 / \bar{x}_{max})$  for  $\theta_2$ , and subsequently  $(\log \theta_2 / t_{infl})$  for  $\theta_3$ .

As an initial component scores matrix  $\tilde{\mathbf{A}}$ , an unconstrained least squares estimate of  $\tilde{\mathbf{A}}$ , considering  $\tilde{\mathbf{C}}$  fixed, is taken as  $\tilde{\mathbf{A}} = \mathbf{X}'\tilde{\mathbf{C}}(\tilde{\mathbf{C}}'\tilde{\mathbf{C}})^{-1}$ . To satisfy the constraint  $\frac{1}{I} \mathbf{1}'_I \tilde{\mathbf{a}}_2 = \frac{1}{I} \mathbf{1}'_I \tilde{\mathbf{a}}_3 = 0$ , the second and third columns of the estimate of  $\tilde{\mathbf{A}}$  are centered.

*Finding an update of  $\tilde{\mathbf{A}}$* 

To find an update of  $\tilde{\mathbf{A}}$ , it is proposed that the first, and the second and third columns of  $\tilde{\mathbf{A}}$  be updated separately. To update the first column of  $\tilde{\mathbf{A}}$ , we minimize (5.11) considering  $\tilde{\mathbf{C}}$ , and  $\tilde{\mathbf{a}}_2$  and  $\tilde{\mathbf{a}}_3$  fixed. The latter is equivalent to minimizing

$$f_3(\tilde{\mathbf{a}}_1) = \|\mathbf{X} - \tilde{\mathbf{c}}_2 \tilde{\mathbf{a}}_2' - \tilde{\mathbf{c}}_3 \tilde{\mathbf{a}}_3' - \tilde{\mathbf{c}}_1 \tilde{\mathbf{a}}_1'\|^2 = \|\mathbf{X}_{-23} - \tilde{\mathbf{c}}_1 \tilde{\mathbf{a}}_1'\|^2, \quad (5.14)$$

where  $\mathbf{X}_{-23} \equiv \mathbf{X} - \tilde{\mathbf{c}}_2 \tilde{\mathbf{a}}_2' - \tilde{\mathbf{c}}_3 \tilde{\mathbf{a}}_3'$ . An update of  $\tilde{\mathbf{a}}_1$  can be obtained by taking  $\tilde{\mathbf{a}}_1 = \mathbf{X}'_{-23} \tilde{\mathbf{c}}_1' (\tilde{\mathbf{c}}_1' \tilde{\mathbf{c}}_1)^{-1}$ . An update for  $\tilde{\mathbf{a}}_2$  and  $\tilde{\mathbf{a}}_3$  subject to the constraint  $\mathbf{1}'_I \tilde{\mathbf{a}}_2 = \mathbf{1}'_I \tilde{\mathbf{a}}_3 = 0$  can be obtained by minimizing

$$f_4(\tilde{\mathbf{a}}_2, \tilde{\mathbf{a}}_3) = \|\mathbf{X} - \tilde{\mathbf{c}}_1 \tilde{\mathbf{a}}_1' - (\tilde{\mathbf{c}}_2 \tilde{\mathbf{a}}_2' + \tilde{\mathbf{c}}_3 \tilde{\mathbf{a}}_3')\|^2, \quad (5.15)$$

subject to  $\mathbf{1}'_I \tilde{\mathbf{a}}_2 = \mathbf{1}'_I \tilde{\mathbf{a}}_3 = 0$ . This is equivalent to minimizing

$$f_5(\hat{\mathbf{A}}) = \|\mathbf{X}_{-1} - \hat{\mathbf{C}} \hat{\mathbf{A}}' \mathbf{N}'\|^2, \quad (5.16)$$

over arbitrary  $\hat{\mathbf{A}}$   $((I-1) \times 2)$ , where  $\mathbf{X}_{-1} \equiv \mathbf{X} - \tilde{\mathbf{c}}_1 \tilde{\mathbf{a}}_1'$ ,  $\hat{\mathbf{C}}$   $(K \times 2)$  denotes a matrix containing the second and third columns of  $\tilde{\mathbf{C}}$ ,  $\mathbf{N}$   $(I \times (I-1))$  is a basis for the null space of  $\mathbf{1}'_I$ , and  $[\tilde{\mathbf{a}}_2 | \tilde{\mathbf{a}}_3] = \mathbf{N} \hat{\mathbf{A}}$ . A least squares update for  $\hat{\mathbf{A}}$  can be obtained by taking  $\hat{\mathbf{A}} = (\mathbf{N}' \mathbf{N})^{-1} \mathbf{N}' \mathbf{X}_{-1} \hat{\mathbf{C}}' (\hat{\mathbf{C}}' \hat{\mathbf{C}})^{-1}$  (Penrose, 1956). Hence,  $[\tilde{\mathbf{a}}_2 | \tilde{\mathbf{a}}_3] = \mathbf{N} \hat{\mathbf{A}} = \mathbf{J} \mathbf{X}_{-1}' \hat{\mathbf{C}}' (\hat{\mathbf{C}}' \hat{\mathbf{C}})^{-1}$ , where  $\mathbf{J}$  is the centering operator  $\mathbf{N}(\mathbf{N}' \mathbf{N})^{-1} \mathbf{N}'$ .

*Finding an update of  $\boldsymbol{\theta}$* 

To update  $\boldsymbol{\theta}$ , we minimize (5.11) considering  $\tilde{\mathbf{A}}$  fixed. Least squares estimates of  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$  can be obtained using the Levenberg-Marquardt algorithm (Seber & Wild, 1989). Note that after convergence the elements of  $\tilde{\mathbf{a}}_2$ ,  $\tilde{\mathbf{a}}_3$ ,  $\tilde{\mathbf{c}}_2$ , and  $\tilde{\mathbf{c}}_3$  have to be rescaled to satisfy the constraint  $\frac{1}{I} \mathbf{1}'_I \tilde{\mathbf{a}}_1 = \theta_1$  as described above. Furthermore, the solutions for  $\tilde{\mathbf{C}}$  and  $\tilde{\mathbf{A}}$  have to be transformed to  $\mathbf{C}$  and  $\mathbf{A}$  as described above.

### 5.2.3. The SLC two-way component model for growth data measured at unequal time points

The method discussed in Section 5.2.1 aims at decomposing a two-way matrix consisting of the scores of  $I$  subjects on one variable measured at  $K$  time points into an occasion component matrix  $\mathbf{C}$ , and a subject component matrix  $\mathbf{A}$ . The three components in the occasion component matrix represent three basis functions evaluated at each of the occasions  $t_1, \dots, t_K$ . In practice, the measurement time points are not necessarily equal for the different subjects. This can be easily covered in the model by allowing the evaluated time points of the basis functions to differ across subjects. To explore this idea further, the time point of the  $k_i^{\text{th}}$  measurement of subject  $i$  is indicated by  $t_{ik_i}$ , where  $k_i=1, \dots, K_i$  denotes the sequence number of measurements of subject  $i$ ,  $i=1, \dots, I$ , and the time points at which the scores of *any* subject are collected are denoted by  $t_1, \dots, t_K$ , where  $K$  is the total number of different time points at which measurements are available. A two-way data matrix  $\mathbf{X}^{\text{F}}$  ( $K \times I$ ) is constructed with rows corresponding to *all* time points for which measurements are encountered. The scores of subject  $i$ ,  $i=1, \dots, I$ , are positioned in the rows of  $\mathbf{X}^{\text{F}}$  that correspond to the time points at which the scores of subject  $i$  are collected ( $t_{i1}, \dots, t_{iK_i}$ ); the remaining  $K-K_i$  values are missing. A binary indicator matrix  $\mathbf{W}$  ( $K \times I$ ) is constructed, with zeros indicating missing values in  $\mathbf{X}^{\text{F}}$ ,  $k=1, \dots, K$ ,  $i=1, \dots, I$ . The SLC two-way component model for growth data can be fitted to the observed data by minimizing

$$f_6(\boldsymbol{\theta}, \mathbf{A}) = \|\mathbf{W} * (\mathbf{X}^{\text{F}} - \mathbf{C}\mathbf{A}')\|^2, \quad (5.17)$$

where  $*$  denotes the Hadamard (or elementwise) product,  $\mathbf{C}$  is restricted to be a function of  $\alpha$ ,  $\beta$  and  $\gamma$  according to (5.6), and  $\frac{1}{7}\mathbf{1}'_I \mathbf{A} = [\alpha \ \beta \ 0]$ . Analogously to the unweighted case (see Section 5.2.2), the minimization problem in (5.17) is treated using a reparametrized Gompertz function (see (5.8)) as target function, and associated reparametrized matrices  $\tilde{\mathbf{C}}$  and  $\tilde{\mathbf{A}}$ , with  $\tilde{\mathbf{A}}$  constrained so that  $\frac{1}{7}\mathbf{1}'_I \tilde{\mathbf{A}} = [\theta_1 \ 0 \ 0]$ . Again, the constraint on the first column of  $\tilde{\mathbf{A}}$  can be satisfied by proper rescaling after minimizing (5.17), and hence we will only discuss an ALS algorithm to minimize the weighted least squares loss function in (5.17) subject to the constraint  $\mathbf{1}'_I \tilde{\mathbf{a}}_2 = \mathbf{1}'_I \tilde{\mathbf{a}}_3 = 0$ . Weighted least squares estimates of  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$ , and hence of  $\tilde{\mathbf{C}}$ , can be obtained using the Levenberg-Marquardt algorithm (Seber & Wild, 1989). The first column of  $\tilde{\mathbf{A}}$  can be updated by means of row-wise weighted least squares regression (Gabriel & Zamir, 1979). The second and third columns of  $\tilde{\mathbf{A}}$ ,  $\tilde{\mathbf{a}}_2$  and  $\tilde{\mathbf{a}}_3$ , can be estimated using the procedure for weighted least squares fitting by Kiers (1997b). The latter method boils down to missing data imputation in  $\mathbf{X}^{\text{F}}$ , where the optimal least squares estimates of the missing elements are imputed at

each step, followed by performing the OLS step to find updates for  $\tilde{\mathbf{a}}_2$  and  $\tilde{\mathbf{a}}_3$ , which is discussed in Section 5.2.2.

The above approach is useful if the missing data are missing completely at random (Little & Rubin, 1987; see also Chapter 3). The degree of reliability of the estimates is influenced by the number of missing data as well as by the time points at which missing data occur. Generally, a large amount of missing data restricted to certain small time periods will decrease the reliability greatly. However, this also depends on the functional form of the true scores. For example, the occurrence of missing data at a certain time interval decreases the reliability more when the true scores fluctuate greatly, than when they fluctuate only slightly.

### 5.3. The SLC Tucker3 model for longitudinal three-way data

In this section, a three-way generalization of the SLC two-way component model, namely the SLC Tucker3 model, is elaborated. In the Tucker3 model, three-way data are decomposed into three component matrices (see Section 2.3). The components for the three modes are weighted via the core array. The SLC Tucker3 model, as elaborated here, is particularly useful if the longitudinal three-way data consist of scores on variables that are intended to measure a certain growth process. In the SLC two-way component model, it is assumed that the estimated average scores across subjects follow a Gompertz curve. In the SLC Tucker3 model, it is assumed that the estimated average scores across subjects *and* variables follow a Gompertz curve. Furthermore, in the models, the estimated score of a subject on a variable at a particular time point is a weighted sum of basis functions that are evaluated in that particular time point. Just as in the SLC two-way component model, the measurement time points are not necessarily identical for all subjects, and for all variables. However, to facilitate description and notation, the model is described here as if the measurements are collected at the same time points for all subjects and all variables.

The SLC Tucker3 model is given by

$$\mathbf{X}_c = \mathbf{C}\mathbf{G}_c(\mathbf{B}' \otimes \mathbf{A}') + \mathbf{E}_c, \quad (5.18)$$

where  $\mathbf{X}_c$  ( $K \times IJ$ ) denotes the matricized three-way array  $\underline{\mathbf{X}}$  ( $I \times J \times K$ ),  $\mathbf{C}$  ( $K \times 3$ ) the occasion component scores matrix,  $\mathbf{A}$  ( $I \times P$ ) the subject component matrix,  $\mathbf{B}$  ( $J \times Q$ ) the variable component matrix,  $\mathbf{G}_c$  ( $3 \times PQ$ ) the matricized core array  $\underline{\mathbf{G}}$  ( $3 \times P \times Q$ ), and  $\mathbf{E}_c$  ( $K \times IJ$ ) the matricized error array  $\underline{\mathbf{E}}$ ; the three components in the occasion component matrix  $\mathbf{C}$  represent three basis functions following (5.6), that are evaluated at each of the time points  $t_1, \dots, t_K$ ; the matrix  $\mathbf{G}_c(\mathbf{B}' \otimes \mathbf{A}')$  is restricted so that  $\frac{1}{IJ} \mathbf{1}'_{IJ} (\mathbf{B} \otimes \mathbf{A}) \mathbf{G}_c' = [\alpha \ \beta \ 0]$ . Because of the nature of  $\mathbf{C}$ , and the restriction on the weights for the basis functions, the estimated average scores across variables and subjects follow a Gompertz curve.

The parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  govern the estimated average scores across subjects and variables, and they are interpreted in the same way as in the two-way SLC model (see Section 5.2.1). The occasion component matrix consists of evaluations of three basis functions, which can be interpreted as the asymptote basis function, the initial basis function, and the rate of learning basis function. The subject and variable component matrices and the core array are interpreted just like their counterparts in the unconstrained Tucker3 model (see Section 2.7).

### 5.3.1. Fitting the SLC Tucker3 model to data

In this section, an algorithm to fit the SLC Tucker3 model to data will be discussed. The function to be minimized is the least squares loss function, for which an ALS algorithm is proposed.

The SLC Tucker3 algorithm aims at minimizing

$$f_7(\boldsymbol{\tau}, \mathbf{A}, \mathbf{B}, \mathbf{G}) = \|\mathbf{X}_c - \mathbf{C}\mathbf{G}_c(\mathbf{B}' \otimes \mathbf{A}')\|^2, \quad (5.19)$$

subject to the constraint  $\frac{1}{H} \mathbf{1}'_H (\mathbf{B} \otimes \mathbf{A}) \mathbf{G}_c' = [\alpha \ \beta \ 0]$ , and where  $\mathbf{C}$  consists of elements given by (5.6). Just as in the two-way case (see Section 5.2.2), a reparametrized version of the Gompertz function, given by (5.8), is used to arrive at an estimate of the parameters  $\alpha$ ,  $\beta$  and  $\gamma$ . Instead of  $\mathbf{C}$  and  $\mathbf{G}_c$ , the (reparametrized) occasion component matrix  $\tilde{\mathbf{C}}$  and the core array  $\tilde{\mathbf{G}}_c$  which can be transformed to  $\mathbf{C}$  and  $\mathbf{G}_c$  without altering the model estimates, are estimated, analogously to the two-way case.

The elements of  $\tilde{\mathbf{C}}$  are given by (5.9), and thus are a function of  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  only. The model can be fitted to data by minimizing

$$f_8(\boldsymbol{\theta}, \mathbf{A}, \mathbf{B}, \tilde{\mathbf{G}}_c) = \|\mathbf{X}_c - \tilde{\mathbf{C}}\tilde{\mathbf{G}}_c(\mathbf{B}' \otimes \mathbf{A}')\|^2, \quad (5.20)$$

with  $\tilde{\mathbf{C}}$  given by (5.9), subject to the constraint  $\frac{1}{H} \mathbf{1}'_H (\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{G}}_c' = [\theta_1 \ 0 \ 0]$ , to achieve the average estimated curve to follow (5.8), with parameters  $\theta_1$ ,  $\theta_2$  and  $\theta_3$ . Analogously to the algorithm to fit the SLC two-way component model (see Section 5.2.2), the constraint on the first column of  $(\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{G}}_c'$  can be satisfied by proper rescaling of  $\theta_1$  (and hence of  $\tilde{\mathbf{C}}$ ) after minimizing (5.20), subject to the constraints on the second and third columns of  $(\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{G}}_c'$  only. In the SLC Tucker3 model, the rescaling should be compensated in the core matrix  $\tilde{\mathbf{G}}_c$ . Estimates of  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\tilde{\mathbf{G}}_c$ , and  $\boldsymbol{\theta}$  can be obtained by alternating least squares, as follows.

*Initial estimates of  $\boldsymbol{\theta}$  and  $\mathbf{A}$  and  $\mathbf{B}$  for the ALS algorithm*

To start the ALS algorithm, initial estimates of the parameters are needed. Analogously to the two-way case, initial values of  $\boldsymbol{\theta}$  can be chosen on the basis of a plot of the average scores across subjects and variables (see Section 5.2.2). Initial estimates of  $\mathbf{A}$  and  $\mathbf{B}$  can be obtained by taking the first  $P$  and  $Q$  eigenvectors of  $\mathbf{X}_a\mathbf{X}_a'$  and  $\mathbf{X}_b\mathbf{X}_b'$ , respectively. Subsequently, an initial (unconstrained) estimate of  $\tilde{\mathbf{G}}_c$  is obtained by taking  $\tilde{\mathbf{G}}_c = (\tilde{\mathbf{C}}'\tilde{\mathbf{C}})^{-1}\tilde{\mathbf{C}}'\mathbf{X}_c(\mathbf{B} \otimes \mathbf{A})(\mathbf{B}'\mathbf{B} \otimes \mathbf{A}'\mathbf{A})^{-1}$  (Penrose, 1956). Those initial values for  $\boldsymbol{\theta}$ ,  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\tilde{\mathbf{G}}_c$  suffice to start the algorithm by finding an estimate for the second and third columns of  $\tilde{\mathbf{G}}_c$ .

*Finding an update of  $\boldsymbol{\theta}$*

An update of  $\boldsymbol{\theta}$ , considering  $\mathbf{A}$  and  $\mathbf{B}$  fixed, can be obtained the same way as finding an update of  $\boldsymbol{\theta}$  is obtained in the SLC two-way component model, as discussed in Section 5.2.2.

*Finding an update of the core array*

To find an update for  $\tilde{\mathbf{G}}_c$ , we propose updating the rows of  $\tilde{\mathbf{G}}_c$  successively. An update of the first row of  $\tilde{\mathbf{G}}_c$  can be obtained by minimizing

$$\begin{aligned} f_9(\tilde{\mathbf{g}}_{c,1}) &= \left\| (\mathbf{X}_c - \tilde{\mathbf{c}}_2\tilde{\mathbf{g}}'_{c,2}(\mathbf{B}' \otimes \mathbf{A}') - \tilde{\mathbf{c}}_3\tilde{\mathbf{g}}'_{c,3}(\mathbf{B}' \otimes \mathbf{A}')) - \tilde{\mathbf{c}}_1\tilde{\mathbf{g}}'_{c,1}(\mathbf{B}' \otimes \mathbf{A}') \right\|^2 \\ &= \left\| \mathbf{X}_{c,-23} - \tilde{\mathbf{c}}_1\tilde{\mathbf{g}}'_{c,1}(\mathbf{B}' \otimes \mathbf{A}') \right\|^2, \end{aligned} \quad (5.21)$$

where  $\tilde{\mathbf{g}}'_{c,m}$  ( $1 \times QP$ ) denotes the  $m^{\text{th}}$  row of  $\tilde{\mathbf{G}}_c$ , and  $\mathbf{X}_{c,-23} \equiv (\mathbf{X}_c - \tilde{\mathbf{c}}_2\tilde{\mathbf{g}}'_{c,2}(\mathbf{B}' \otimes \mathbf{A}') - \tilde{\mathbf{c}}_3\tilde{\mathbf{g}}'_{c,3}(\mathbf{B}' \otimes \mathbf{A}'))$ . An update of  $\tilde{\mathbf{g}}_{c,1}$  can be obtained by taking  $\tilde{\mathbf{g}}'_{c,1} = (\tilde{\mathbf{c}}_1'\tilde{\mathbf{c}}_1)^{-1}\tilde{\mathbf{c}}_1'\mathbf{X}_{c,-23}(\mathbf{B}' \otimes \mathbf{A})(\mathbf{B}'\mathbf{B} \otimes \mathbf{A}'\mathbf{A})^{-1}$  (Penrose, 1956).

An update of the second row of  $\tilde{\mathbf{G}}_c$  can be obtained by minimizing

$$\begin{aligned} f_{10}(\tilde{\mathbf{g}}_{c,2}) &= \left\| (\mathbf{X}_c - \tilde{\mathbf{c}}_1\tilde{\mathbf{g}}'_{c,1}(\mathbf{B}' \otimes \mathbf{A}') - \tilde{\mathbf{c}}_3\tilde{\mathbf{g}}'_{c,3}(\mathbf{B}' \otimes \mathbf{A}')) - \tilde{\mathbf{c}}_2\tilde{\mathbf{g}}'_{c,2}(\mathbf{B}' \otimes \mathbf{A}') \right\|^2 \\ &= \left\| \mathbf{X}_{c,-13} - \tilde{\mathbf{c}}_2\tilde{\mathbf{g}}'_{c,2}(\mathbf{B}' \otimes \mathbf{A}') \right\|^2, \end{aligned} \quad (5.22)$$

subject to the constraint  $\mathbf{1}'(\mathbf{B} \otimes \mathbf{A})\tilde{\mathbf{g}}_{c,2} = 0$ , where  $\mathbf{X}_{c,-13} \equiv (\mathbf{X}_c - \tilde{\mathbf{c}}_1\tilde{\mathbf{g}}'_{c,1}(\mathbf{B}' \otimes \mathbf{A}') - \tilde{\mathbf{c}}_3\tilde{\mathbf{g}}'_{c,3}(\mathbf{B}' \otimes \mathbf{A}'))$ . By requiring  $\tilde{\mathbf{g}}_{c,2} = \mathbf{N}\mathbf{g}_{c,2}$ , where  $\mathbf{N}$  ( $QP \times (QP-1)$ ) is a basis for the nullspace of  $\mathbf{1}'_U(\mathbf{B} \otimes \mathbf{A})$ , and minimizing (5.22) over

unconstrained  $\mathbf{g}_{c,2}$  ( $(QP-1) \times 1$ ), a least squares estimate of  $\tilde{\mathbf{g}}_{c,2} = \mathbf{N} \mathbf{g}_{c,2}$ , subject to the constraint  $\mathbf{1}'(\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{g}}_{c,2} = 0$ , can be obtained. An update of the unconstrained vector  $\mathbf{g}_{c,2}$  can be obtained as  $\mathbf{g}_{c,2}' = (\tilde{\mathbf{c}}_2' \tilde{\mathbf{c}}_2)^{-1} \tilde{\mathbf{c}}_2' \mathbf{X}_{c,-13} (\mathbf{B} \otimes \mathbf{A}) \mathbf{N} (\mathbf{N}' (\mathbf{B}' \mathbf{B} \otimes \mathbf{A}' \mathbf{A}) \mathbf{N})^{-1}'$ .

An update of the third row of  $\tilde{\mathbf{G}}_c$ , subject to the constraint  $\mathbf{1}'(\mathbf{B} \otimes \mathbf{A}) \mathbf{g}_{c,3} = 0$ , can be obtained the same way as an update of the second row of  $\tilde{\mathbf{G}}_c$  was obtained.

#### Finding an update of $\mathbf{A}$ and $\mathbf{B}$

An update of  $\mathbf{A}$  can be obtained by minimizing

$$f_{11}(\mathbf{A}) = \|\mathbf{X}_c - \tilde{\mathbf{C}} \tilde{\mathbf{G}}_c (\mathbf{B}' \otimes \mathbf{A}')\|^2, \quad (5.23)$$

subject to the constraint  $\mathbf{1}'_J (\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{G}}'_{23} = \mathbf{0}'$ , where  $\tilde{\mathbf{G}}'_{23} \equiv [\tilde{\mathbf{g}}_{c,2} | \tilde{\mathbf{g}}_{c,3}]$ . Upon defining  $\mathbf{V}$  as

$$\mathbf{V} \equiv \begin{bmatrix} \mathbf{1}'_J \mathbf{B} \tilde{\mathbf{G}}'_2 \otimes \mathbf{1}'_I \\ \mathbf{1}'_J \mathbf{B} \tilde{\mathbf{G}}'_3 \otimes \mathbf{1}'_I \end{bmatrix},$$

where  $\tilde{\mathbf{G}}'_m$  ( $P \times Q$ ) denotes the  $m^{\text{th}}$  horizontal slab of the core array  $\tilde{\mathbf{G}}$  ( $3 \times P \times Q$ ), the latter constraints are equivalent to requiring  $\mathbf{V} \text{Vec}(\mathbf{A}) = \mathbf{0}$ , as is shown in Appendix 5.1. Minimizing (5.23) over  $\mathbf{A}$  is equivalent to minimizing

$$f_{12}(\mathbf{A}) = \|\mathbf{X}_a - \mathbf{A} \tilde{\mathbf{G}}_a (\tilde{\mathbf{C}}' \otimes \mathbf{B}')\|^2. \quad (5.24)$$

An update for  $\mathbf{A}$  subject to the constraint  $\mathbf{V} \text{Vec}(\mathbf{A}) = \mathbf{0}$  can be obtained by minimizing

$$f_{13}(\mathbf{a}) = \|\text{Vec}(\mathbf{X}_a) - ((\tilde{\mathbf{C}} \otimes \mathbf{B}) \tilde{\mathbf{G}}'_a \otimes \mathbf{I}_I) \mathbf{N} \mathbf{a}\|^2, \quad (5.25)$$

where  $\mathbf{N}$  is a basis for the nullspace of  $\mathbf{V}$ , and  $\text{Vec}(\mathbf{A}) = \mathbf{N} \mathbf{a}$ . By defining  $\mathbf{Z} \equiv ((\tilde{\mathbf{C}} \otimes \mathbf{B}) \tilde{\mathbf{G}}'_a \otimes \mathbf{I}_I) \mathbf{N}$ , an update for  $\text{Vec}(\mathbf{A})$  can be obtained by taking  $\text{Vec}(\mathbf{A}) = \mathbf{N} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \text{Vec}(\mathbf{X}_a)$ . Rearranging the elements of  $\text{Vec}(\mathbf{A})$  into an  $I \times P$  matrix yields  $\mathbf{A}$ .

Upon interchanging the role of the A and B mode, an update of  $\mathbf{B}$  can be obtained the same way an update for  $\mathbf{A}$  was obtained in the SLC Tucker3 model, as was discussed above.

### Rescaling of the estimates after convergence of the ALS algorithm

After convergence, the rescaling of  $\theta_1$  so that  $\frac{1}{I'} \mathbf{1}'_{IJ} (\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{g}}_{c,1} = \theta_1$  can be obtained as follows: Let  $\theta_1^c$ ,  $\mathbf{g}_{c,2}^c$  and  $\mathbf{g}_{c,3}^c$  be the estimates of  $\theta_1$ ,  $\tilde{\mathbf{g}}_{c,2}$  and  $\tilde{\mathbf{g}}_{c,3}$ , respectively, after convergence. After defining  $\theta_1 = \frac{1}{I'} \mathbf{1}'_{IJ} (\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{g}}_{c,1}$ , the second and third rows of  $\tilde{\mathbf{G}}_c$  are defined as  $\tilde{\mathbf{g}}_{c,2} = \frac{\theta_1^c}{\theta_1} \mathbf{g}_{c,2}^c$  and  $\tilde{\mathbf{g}}_{c,3} = \frac{\theta_1^c}{\theta_1} \mathbf{g}_{c,3}^c$ , respectively. The rescaled estimates of the elements of  $\tilde{\mathbf{C}}$  are obtained using (5.9).

Subsequently, the parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  can be obtained by taking  $\alpha = \theta_1$ ,  $\beta = \theta_1 \exp(-\theta_2)$ , and  $\gamma = \theta_3$ . Then, the matrix  $\mathbf{C}$  is computed as  $\mathbf{C} = \tilde{\mathbf{C}}\mathbf{T}$ , and the matrix  $\mathbf{G}_c$  as  $\mathbf{G}_c = (\mathbf{T})^{-1} \tilde{\mathbf{G}}_c$ , where  $\mathbf{T}$  is given by (5.12).

### 5.3.2. Transformational freedom and interpretation in the SLC Tucker3 model

The SLC Tucker3 model has transformational freedom, as the subject and variable component matrices can be transformed provided that such transformations are compensated in the core array  $\underline{\mathbf{G}}$ . The occasion component matrix  $\mathbf{C}$  is given by (5.6), and may therefore not be transformed.

A rescaling of the occasion component matrix in the SLC Tucker3 model would imply that the elements of  $\mathbf{C}$  no longer satisfy (5.6). However, in practice, the interpretation of the model may be facilitated by rescaling the component matrix  $\mathbf{C}$  column-wise. Then the columns of the occasion component matrix follow (5.6) up to a multiplication by a constant. The rescaling should be compensated in the subject and/or variable component matrices, or in the core array, to preserve the fit. Note that after rescaling the first column of  $\mathbf{C}$ , the constraint on the first column of  $\mathbf{1}'(\mathbf{B} \otimes \mathbf{A})\mathbf{G}'_c$  in the SLC Tucker3 model is no longer satisfied. However, this does not pose a problem as the estimated average curve across subjects and variables, as given by  $\frac{1}{I'} \mathbf{C}\mathbf{G}_c(\mathbf{B} \otimes \mathbf{A})'\mathbf{1}_{IJ}$  in the SLC Tucker3 model, is still a Gompertz curve with parameters  $\alpha$ ,  $\beta$ , and  $\gamma$ . Such a rescaling of the occasion component matrix is illustrated in the empirical example (Section 5.4).

## 5.4. Empirical example: Learning to read study (II)

In this section, the results of an SLC Tucker3 analysis of the data of the Learning to Read study (Jansen & Bus, 1982; Bus & Kroonenberg, 1982) are discussed. The learning to read study was discussed in Section 4.5. The data contain the scores of seven subjects on five reading tests, collected at 37 time points during process of learning to read. This data set was analyzed by the unrestricted Tucker3 model, and two smoothness constrained Tucker3 models. The results are discussed in Section 4.5.

In this section, we present the results of an SLC Tucker3 analysis and compare them to the results of the previous analyses.

The raw scores were rescaled in the same way as in Section 4.5, that is, so that the scores on all variables ranged from 0 to 1. The SLC Tucker3 model (see (5.18)) was fitted to the rescaled data. Analogously to the unconstrained Tucker3 analysis (see Section 4.5), we chose two components for the subject mode, and one component for the variable mode.

The parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  were estimated as 0.92, 0.06, and 0.13, respectively. Hence, the asymptote of the mean curve (over all subjects and variables) is 0.92, and the estimated mean score at time point (=week) 0 is 0.06. The inflection point is

estimated to occur at week  $7.7 = \left\{ \frac{1}{\gamma} \ln \left( - \ln \frac{\beta}{\alpha} \right) \right\}$ . The observed mean scores across

variables and subjects, and the mean curve as estimated by the SLC Tucker3 model are plotted in Figure 5.2.

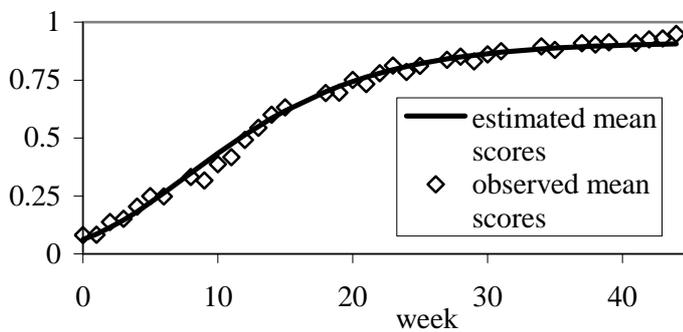


Figure 5.2. The observed mean scores across variables and subjects, and the mean curve as estimated by the SLC Tucker3 analysis of the data of the Learning to Read study.

The fit of the SLC Tucker3 model is 96.13%, which is 0.13% lower than the unconstrained Tucker3 model. Note that, although the unconstrained Tucker3 model has two occasion components, whereas the SLC Tucker3 model has three, the latter model uses fewer parameters: the Tucker3 model requires

$$(I \times P + J \times Q + K \times R + P \times Q \times R - P^2 - Q^2 - R^2) - (I \times P + J \times Q + 2 + P \times Q \times 3 - P^2 - Q^2 - 2) = 68$$

parameters more (see also Weesie & Van Houwelingen, 1983) than the SLC Tucker3 model in the current example.

The estimated subject and variable component matrices of the SLC Tucker3 model were regressed onto their counterparts found in the unconstrained Tucker3 analysis. This transformation was compensated in the core array. The transformed component matrices were compared to the solutions of the unconstrained Tucker3 model by computing the coefficient of congruence between the pairs of components

concerned. The large coefficients (minimally 0.9997) indicate that those pairs resembled each other strongly. The subject component matrix and the variable component matrix are presented in Tables 5.1 and 5.2, respectively.

*Table 5.1. Subject component scores of the SLC Tucker3 solution of the Learning to Read study.*

<b>A</b> (subjects)	1 <sup>st</sup> component	2 <sup>nd</sup> component
1	1.06	-0.41
2	0.96	-0.30
3	0.99	-0.37
4	1.28	1.01
5	1.16	0.17
6	1.09	-0.01
7	0.89	-0.43

*Table 5.2. Variable component scores of the SLC Tucker3 solution of the Learning to Read study.*

<b>B</b> (variables)	
Letter Knowledge	0.91
Regular Orthographic Short Words	1.00
Regular Orthographic Long Words	0.87
Regular Orthographic Long and Short Words within Context	0.99
Irregular Orthographic Long and Short Words	0.58

To facilitate the interpretation, the slabs of the core array pertaining to the three basis functions were rescaled so that the maximal weight for each of the three basis functions equals one, and this rescaling was compensated in the basis functions. As explained earlier, the estimated mean curve still is a Gompertz function with parameters  $\alpha$ ,  $\beta$ , and  $\gamma$ . The core array is presented in Table 5.3. The estimated and rescaled occasion basis functions are plotted in Figure 5.3.

*Table 5.3. Core of the SLC Tucker3 model of the Learning to Read data.*

<b>G</b> (core)	<b>b<sub>1</sub> c<sub>1</sub></b>	<b>b<sub>1</sub> c<sub>2</sub></b>	<b>b<sub>1</sub> c<sub>3</sub></b>
<b>a<sub>1</sub></b>	1.00	0.33	0.05
<b>a<sub>2</sub></b>	-0.18	1.00	1.00

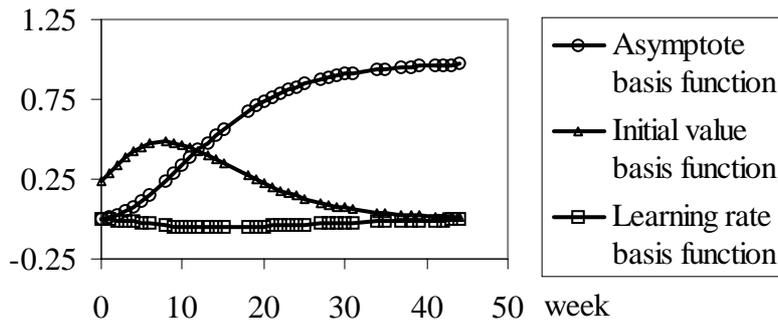


Figure 5.3. Estimated and rescaled basis functions of the SLC Tucker3 model of the Learning to Read data.

The sizes of the core elements indicate that the combinations of the first subject component and the first basis function, and the second subject component and the second and third basis functions weigh heavily in the final solution. However, the learning rate basis function (the third basis function) hardly influences the estimated solution, as the scores are almost zero, as can be seen in Figure 5.3. The initial value basis function reflects not only the initial value, but also the fact that the acceleration first increases and then decreases. In fact, one could interpret the initial value basis function as a second learning rate basis function, with its maximum occurring at the inflection point of the Gompertz curve.

The ordering in the size of the variable component scores indicates the relative difficulty of the variable, with a large score indicating that the skill measured by the variable concerned is mastered relatively quickly. As only one component is used for the variables, the ordering in difficulty among variables is equal across time for the subjects.

The subject component scores can be interpreted as follows: The subject component scores on the first component indicate the weight of the asymptote basis function, whereas the score on the second subject component indicates the weight of the initial value basis function and the learning rate basis function.

As the learning rate and initial basis functions approach zero at the last measurement occasion, the weight of the first subject component scores indicates the relative size of the maximal score of the subject. Thus, Subject 4 ends highest, successively followed by Subjects 5, 6, 1, 3, 2, 7.

The evaluated values of the learning rate basis function are close to zero, and therefore hardly play a role in the model. On the contrary, the initial value basis function influences the solution greatly. It reflects the initial score as well as the growth rate. Thus, on the basis of the ordering of the second subject component scores, the subjects can be ordered from fast growth rate combined with relatively high initial values, to slow learning rate combined with low initial values. Note that a subject can also start and end high, and show relatively slow learning rate due to

ceiling effects. This phenomenon is reflected in the subject component scores by a relatively high score on the first component and a relatively low score on the second component. This can be seen in Subject 6.

In order to compare subjects, it can be useful to estimate the growth curve(s) per subject. This can be done by computing  $\mathbf{G}_b(\mathbf{A}'\otimes\mathbf{C}')$ : the rows of this matrix refer to the weights for the  $Q$  variable components; columns 1 through  $K$  are the evaluations of the estimated growth curve at the  $k=1,\dots,K$  measurement occasions for the first subject,  $K+1$  through  $2K$  for the second subject, and so forth. Here, we have only one variable component, and thus each subject has only one estimated curve, which could be called the ‘general growth curve’. Per variable, the estimated curve per subject is just proportional to the general growth curve of the subject concerned. For illustration, the curves for Subjects 1, 4 and 6 are plotted in Figure 5.4. It can be seen in this figure that Subject 4 starts high, and approaches the asymptote quickly. Subject 6 starts somewhat higher than Subject 1, who shows a somewhat higher growth rate than Subject 6. Subjects 1, 4 and 6 end up at about the same asymptote level.

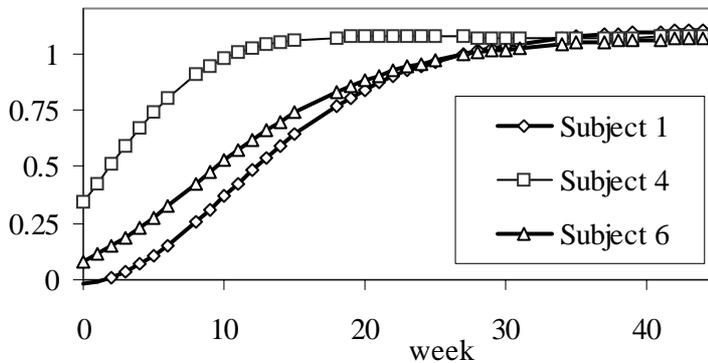


Figure 5.4. Estimated ‘general growth curves’ of Subjects 1, 4 and 6, of the SLC Tucker3 model of the learning to read data.

## 5.5. Discussion and conclusion

In this chapter, the Structured Latent Curve (SLC) two-way component and the SLC Tucker3 models were discussed. The empirical example showed the use of the SLC Tucker3 model in practice. In the SLC Tucker3 model of the ‘Learning to read’ data, it can be observed that the asymptote basis function and the initial value basis function (see Figure 5.3) resemble the two respective occasion component scores of the T3-Bs model (see Figure 4.5) closely. However, the initial value basis function can be interpreted somewhat better than the second occasion component of the T3-Bs analysis. In the latter analysis, the scores are negative after week 22, which may suggest decreasing scores on the variables after week 22. This is not the case, for

either the observed scores or the model, but this is difficult to see at once. (In problematic cases, plots of the estimated scores per subject per variable can be helpful.) The interpretation of the SLC Tucker3 model is therefore somewhat easier than the interpretation of the T3-Bs model.

In general, the SLC Tucker3 model is more parsimonious than its smoothness constrained counterpart. On the other hand, smoothing the component matrix is to be preferred over imposing a functional form if knowledge about the functional form is lacking. Moreover, if the functional form of the mean curve is intricate, the smoothness constrained model offers a simple approach to restricted modeling of the data.

The principles of the SLC Tucker3 model could also be applied to the CP model, which is more restricted, and to the Tucker2 and Tucker1 models that are less restricted than the Tucker3 model. That is, SLC CP, SLC Tucker 2 and SLC Tucker1 models can be defined completely analogously to the SLC Tucker3 model. An SLC CP model is usually heavily constrained, because the latent curves, as well as the CP model itself, are restricted. Note that, in a CP model, the number of components is equal for all three modes. As a result, when, for example, a Gompertz function is used as the target function, the number of basis functions is three, and hence the number of subject components and the number of variable components must always be three. In general, a CP model implies that the scores of the entities of one mode are proportional to each other, and hence no interactions across the modes are allowed for. To put it differently in the current context, the weights for the variables for each basis function are equal for all subjects, and per variable the subjects' curves over time are proportional to each other. Hence, the SLC CP model requires the data to have a rather special structure. Moreover, it is difficult to fit the model to data. For these two reasons, we did not elaborate the SLC CP model further.

The SLC Tucker2 model and SLC Tucker1 model are less restricted than the SLC Tucker3 model. The SLC Tucker2 model can be fitted to data using the SLC Tucker3 fitting procedure by taking the number of observed entities of the unreduced mode as the number of components for the unreduced mode in the SLC Tucker3 fitting procedure. Then, the extended core array is computed by multiplying the component matrix of the unreduced mode by the (appropriately reordered) estimated core array. The SLC Tucker1 model can be fitted to data by applying the fitting procedure for the SLC two-way component model to the matricized data array  $\mathbf{X}_c (K \times IJ)$ .

The SLC Tucker3 model was discussed for the case where the mean curve (across subjects and variables) could be described well by a Gompertz curve. The approach can also be used if the mean curve follows another target function. For an example, in the three-way SLC factor analysis context, we refer to Oort (2001), who applied a linear target function in an empirical example. Of course, one may think of more intricate target functions as well.

### Appendix 5.1.

The constraint  $\mathbf{1}'_I (\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{G}}'_{23} = \mathbf{0}'$  is equivalent to the constraint  $\mathbf{V}\text{Vec}(\mathbf{A}) = \mathbf{0}$ , where  $\mathbf{A}$  ( $I \times P$ ) and  $\mathbf{B}$  ( $J \times Q$ ) are component matrices,  $\tilde{\mathbf{G}}'_{23} \equiv [\tilde{\mathbf{g}}_{c,2} | \tilde{\mathbf{g}}_{c,3}]$ , with  $\tilde{\mathbf{g}}'_{c,m}$  ( $1 \times QP$ ) denoting the  $m^{\text{th}}$  row of  $\tilde{\mathbf{G}}_c$ , where  $\tilde{\mathbf{G}}_c$  ( $3 \times PQ$ ) denotes the matricized core array  $\tilde{\mathbf{G}}$  ( $3 \times P \times Q$ ),  $\text{Vec}(\mathbf{A})$  ( $IP \times 1$ ) denotes the vectorized version of matrix  $\mathbf{A}$ , and  $\mathbf{V}$  is defined as

$$\mathbf{V} \equiv \begin{bmatrix} \mathbf{1}'_J \mathbf{B} \tilde{\mathbf{G}}'_2 \otimes \mathbf{1}'_I \\ \mathbf{1}'_J \mathbf{B} \tilde{\mathbf{G}}'_3 \otimes \mathbf{1}'_I \end{bmatrix}.$$

This can be seen as follows. The constraint  $\mathbf{1}'_I (\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{G}}'_{23} = \mathbf{0}'$  is equivalent to the constraints  $\mathbf{1}'_I (\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{g}}_{c,2} = \mathbf{1}'_I (\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{g}}_{c,3} = 0$ . The constraint on the scalar  $\mathbf{1}'_I (\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{g}}_{c,2}$  can be written as  $\mathbf{1}'_I (\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{g}}_{c,2} = (\mathbf{1}'_J \otimes \mathbf{1}'_I) (\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{g}}_{c,2} = ((\mathbf{1}'_J \mathbf{B}) \otimes (\mathbf{1}'_I \mathbf{A})) \tilde{\mathbf{g}}_{c,2} = \text{Vec}(\mathbf{1}'_I \mathbf{A} \tilde{\mathbf{G}}_2 \mathbf{B}' \mathbf{1}_J) = ((\mathbf{1}'_J \mathbf{B} \tilde{\mathbf{G}}'_2) \otimes \mathbf{1}'_I) \text{Vec}(\mathbf{A}) = 0$ , where  $\tilde{\mathbf{G}}_2$  ( $P \times Q$ ) denotes the 2<sup>nd</sup> horizontal slab of the core array  $\tilde{\mathbf{G}}$  ( $3 \times P \times Q$ ). Analogously, the constraint on the scalar  $\mathbf{1}'_I (\mathbf{B} \otimes \mathbf{A}) \tilde{\mathbf{g}}_{c,3}$  can be written as  $((\mathbf{1}'_J \mathbf{B} \tilde{\mathbf{G}}'_3) \otimes \mathbf{1}'_I) \text{Vec}(\mathbf{A}) = 0$ . When  $\mathbf{V}$  is defined as above, the constraints  $((\mathbf{1}'_J \mathbf{B} \tilde{\mathbf{G}}'_2) \otimes \mathbf{1}'_I) \text{Vec}(\mathbf{A}) = ((\mathbf{1}'_J \mathbf{B} \tilde{\mathbf{G}}'_3) \otimes \mathbf{1}'_I) \text{Vec}(\mathbf{A}) = 0$  can be rewritten as  $\mathbf{V}\text{Vec}(\mathbf{A}) = \mathbf{0}$ .

# 6. Simultaneous Component Models of Multisubject Multivariate Time Series

## 6.1. Introduction

This chapter and the next chapter both deal with a class of models for the analysis of multivariate time series from more than one subject, for the case in which the measurement occasions are not comparable across the subjects. The multisubject multivariate time series are analyzed simultaneously. The models capture intraindividual as well as interindividual variability.

For the modeling of multivariate time series collected from a single observation unit, various models, in which the relationships between the observed variables are modeled using latent variables, have been proposed. In those models, it is assumed that the multivariate time series are generated by a latent uni- or multivariate time series, where the latent time series are of a lower order than the observed multivariate time series. One such model is the model underlying the 'P-technique', proposed by Cattell (1952, 1963), where a conventional cross-sectional factor analysis is applied to the multivariate time series of one subject. Anderson (1963) objected to this method since only simultaneous relations between variables are taken into account: possible relations between factor series at different times are not modeled in a P-technique analysis. Anderson (1963) proposed an alternative procedure, which has been elaborated by several authors. The elaboration is known under the name 'dynamic factor analysis' (Engle & Watson, 1981; Immink, 1986; Molenaar, 1985). The various dynamic factor models differ in the way the latent time series are related to observed time series, the model of the latent time series and the estimation procedure of the parameters of the model.

A different approach is to use component models, which are directed at fitting the data themselves, rather than their covariances. Bijleveld (1989) and Van Buuren (1990) offered dynamic component models. The linear dynamic system model (Bijleveld, 1989) is a model in which the autoregressive structure of successive component scores is explicitly modeled. Van Buuren (1990) offered a very general dynamic component model, which is denoted as the canonical class model.

The models mentioned above aim at modeling multivariate time series of a single subject, thus allowing intraindividual variability to be studied. Researchers are often interested in studying interindividual differences and similarities in intraindividual variability. One of the methods used so far is to compare the various fitted time series models obtained for each of the subjects. This approach is used quite often in P-technique factor analysis, as indicated in Jones and Nesselroade's review (1990) of studies in which multivariate time series were analyzed using P-technique, and also in

dynamic factor models (e.g., see Shifren, Hooker, Wood & Nesselroade, 1997). Another approach is to model the multisubject multivariate time series *simultaneously*. However, the multisubject extensions of the models for multivariate time series of a single subject so far proposed leave little or no room for interindividual differences. The dynamic factor model for time series collected from more than one subject (Nesselroade & Molenaar, 1999) assumes that there are no interindividual differences in intraindividual variability in the data. The extension of the linear dynamic system model (Bijleveld & Bijleveld, 1997) offers only limited possibilities for modeling interindividual differences. The extension of the canonical class model (Van Buuren, 1990) appears particularly useful for modeling data from a number of observers on the same subject, rather than modeling data of a number of subjects, and hence is not appropriate for modeling interindividual differences.

In this and the next chapter, we propose models for the exploratory analysis of multisubject multivariate time series which explicitly model interindividual differences. Specifically, in the present chapter, we propose a class of four simultaneous component analysis (SCA) models, two of which are new. In each of the four models, the multivariate time series of each subject is decomposed into a few series of component scores and a loading matrix. The loading matrix is common for all subjects. The four SCA models differ with respect to the constraints imposed on the cross-products (covariances) of the component scores. Apart from the cross-product constraints, the component scores themselves are estimated freely, and the approach deviates therefore from the usual time series analysis. The models can be ordered hierarchically from weakly to strongly constrained, thus allowing for large to small interindividual differences in the model. Which model is most appropriate depends on the degree of variability between subjects. After the models themselves have been described, the SCA models and alternating least squares algorithms to fit the models to data will be treated. Finally, the methods will be illustrated by means of two empirical examples. In the next chapter, the SCA models will be elaborated to include lagged effects.

## 6.2. Four models for simultaneous component analysis

In the next sections, steps in the preprocessing of the raw data before performing a simultaneous component analysis (SCA; Section 6.2.1), and the four SCA models will be discussed (Sections 6.2.2 through 6.2.6). The transformational freedom of each of the four models is treated in Section 6.2.7, and model selection issues are elaborated in Section 6.2.8.

### 6.2.1. Preprocessing of raw data before fitting the SCA model to data

A convenient way to organize the raw data is to let  $\mathbf{X}_i$  ( $K_i \times J$ ) denote the matrix containing the scores of the  $i^{\text{th}}$  subject ( $i=1, \dots, I$ ) on  $J$  variables measured on  $K_i$  occasions ( $k_i=1, \dots, K_i$ ). Before fitting one of the four SCA models, one has to decide

whether the raw data or some preprocessed version thereof should be analyzed. In most cases in practice, the data can be considered to have interval level. It is then advisable to center the raw scores across occasions per variable and per subject, hence to center  $\mathbf{X}_i$  column-wise across the  $K_i$  occasions,  $i=1, \dots, I$ . This approach eliminates constants from the data without introducing artificial variation (Harshman & Lundy, 1984b; Bro & Smilde, in preparation). In addition, the average estimated component score (over occasions) per component and per subject can be shown to be zero (see Appendix 6.1). This is useful because it allows the different restrictions on the component score matrices in each of the four models to be interpreted directly in terms of different restrictions on the covariances between components. This will become clear later.

Scaling aims at eliminating artificial scale differences. In the case of SCA, scaling to equalize the importance of the variables in the final solution appears to be most reasonable. We advise normalizing the (centered) scores ‘within variables’ (i.e., per variable over occasions and subjects jointly), so that the sum of squares per variable over occasions *and* subjects is equal to the sum of the number of measurements of all subjects ( $\sum_{i=1}^I K_i$ , where  $K_i$  denotes the number of occasions of subject  $i$ ,  $i=1, \dots, I$ ). As a consequence, differences in intraindividual variability are preserved. Furthermore, this type of scaling does not affect the form of the structural model (cf., Harshman & Lundy, 1984b; Bro & Smilde, in preparation).

### 6.2.2. SCA with invariant Pattern (SCA-P)

The model for SCA with invariant Pattern (SCA-P; Kiers & Ten Berge, 1994) was originally proposed for modeling multivariate data of a number of subjects drawn from more than one population. SCA-P can be used for modeling multivariate time series of a number of subjects as follows: Let  $\mathbf{X}_i$  ( $K_i \times J$ ) denote the matrix of (usually preprocessed) scores of the  $i^{\text{th}}$  subject ( $i=1, \dots, I$ ) on  $J$  variables measured at  $K_i$  occasions ( $k_i=1, \dots, K_i$ ). The SCA-P model is given by

$$\mathbf{X}_i = \mathbf{F}_i \mathbf{B}' + \mathbf{E}_i, \quad (6.1)$$

where  $\mathbf{F}_i$  ( $K_i \times Q$ ) denotes the  $Q$  component scores of subject  $i$  at time points  $1, \dots, K_i$ ,  $\mathbf{B}$  ( $J \times Q$ ) denotes the loading matrix, and  $\mathbf{E}_i$  ( $K_i \times J$ ) denotes the matrix of residuals. The component scores matrix  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , is unconstrained. It is assumed that the true variable scores (i.e., without error) at occasion  $k_i$  are a linear combination of the component scores at occasion  $k_i$ . The fact that the component scores matrices are unconstrained implies that the inner products between the components may vary across subjects. If the variables are centered across occasions for each subject, then the component scores are also centered (see Appendix 6.1), and the inner products of the component scores of subject  $i$  divided by  $K_i$  are covariances between the components of the subject concerned. If the analyzed data matrices were normalized

‘within variables’ (see Section 6.2.1), the differences in intraindividual variability are preserved. In the sequel, it will be assumed that the raw scores are preprocessed as discussed in Section 6.2.1.

The multivariate time series of each subject are decomposed into a number of time series of component scores and a loading matrix, which is common for all subjects and for all occasions. As a result, the loading matrix is assumed to be subject and time invariant. The component score on a certain occasion can be interpreted as the degree of the particular property measured by the particular component, present in that subject, on that occasion. The interpretation of the components is, as usual, based on the loadings. To investigate each individual’s behavior on the various components, the series of component scores for each component, and for each subject, can be plotted against the time axis. Possible trends and deviating scores on certain occasions can be seen immediately.

The variance of the component scores (over time) for each component and for each subject can be computed. Per component, differences in variances between subjects can be interpreted as differences in intraindividual variability with respect to that particular component. Differences between subjects in covariances between components (within subjects) are easiest to interpret in terms of differences in correlations. Hence, it is possible that certain components correlate highly for one subject, and almost zero for another.

### 6.2.3. Constrained versions of SCA-P

In the SCA-P model, the component scores matrices are unconstrained, implying that the variances of component scores and the covariances between component scores within subjects may vary across subjects. If no interindividual differences in covariation and/or variability are present in the data, a more parsimonious model than SCA-P should be used. By imposing proper constraints on the variances and covariances of the individual component scores, three models that are restricted variants of SCA-P are defined. The restrictions on the component scores of the four models are summarized in Table 6.1. The three restricted variants of SCA-P are described in Sections 6.2.4 to 6.2.6.

*Table 6.1. The restrictions of the four SCA models on the covariances between and the variances of the component scores.*

	Variances of components	Covariances between components
SCA-P	free	free
SCA-PF2	free	equal across subjects
SCA-IND	free	equal to 0
SCA-ECP	equal across subjects	equal across subjects

Depending on the extent of the interindividual differences, a weakly or strongly restricted SCA model can be chosen. The strength of the approach is that one can explicitly choose the most parsimonious model possible for the particular data set, without ignoring important aspects of the data. The choice for the most parsimonious model is not only important in terms of the interpretation of the model. Fitting a less parsimonious model than the one that is underlying the data usually leads to a considerable amount of ‘error fitting’, that is, a part of the error term is mistakenly fitted in the model. This usually leads to unstable parameter estimates.

#### 6.2.4. SCA with PARAFAC2 constraints (SCA-PF2)

The model for SCA with PARAFAC2 constraints (SCA-PF2; Kiers, Ten Berge & Bro, 1999) is a constrained version of the SCA-P model. Kiers, Ten Berge & Bro (1999) named this model ‘direct fitting PARAFAC2’, but we choose to denote it as SCA-PF2 to be consistent with the other SCA models. The SCA-PF2 model is given by (6.1) with  $\mathbf{F}_i' \mathbf{F}_i$  constrained to  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i \mathbf{\Phi} \mathbf{D}_i$ , with  $\mathbf{D}_i$  a diagonal  $Q \times Q$  matrix and  $\mathbf{\Phi}$  a positive definite  $Q \times Q$  matrix, and without loss of generality, we further require  $\mathbf{\Phi}$  to have unit diagonal elements. Thus, in SCA-PF2, the congruence coefficients (Tucker, 1951; see also Section 4.4.3) between columns of  $\mathbf{F}_i$  are invariant over all subjects  $i=1, \dots, I$ .

Here, the component scores matrices  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , contain centered scores, because the variables are centered across occasions for each subject (see Appendix 6.1). As a result, the restriction on  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , implies that the components have the same mutual correlations for all subjects, and that the variances of the components may vary across subjects. These variances are given by the diagonal elements of  $\mathbf{D}_i^2$ ,  $i=1, \dots, I$ . Thus, the SCA-PF2 model is suitable if the variables indicate concepts that are equally correlated for different subjects, and if the degree of intraindividual variability with respect to these concepts varies between subjects.

#### 6.2.5. SCA with INDSCAL constraints (SCA-IND)

The model for SCA with INDSCAL constraints (SCA-IND) is a constrained version of both SCA-P and SCA-PF2. That is, the SCA-IND model is given by (6.1) with  $\mathbf{F}_i' \mathbf{F}_i$  constrained to  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i^2$ , with  $\mathbf{D}_i$  a diagonal matrix. Thus, the inner products between the components are zero and the sums of squares of the components may differ across subjects. To the best of our knowledge, the SCA-IND model is new. If the component scores matrices contain centered scores, which is the case here, the restriction on  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , implies that the components are constrained to be uncorrelated, but the variances of the several components may vary across subjects. Again, the elements of  $\mathbf{D}_i^2$  contain the variances of the components of subject  $i$ .

The name ‘SCA-IND’ is obtained from a counterpart model for cross-product matrices. That is, taking the cross-products of the (error free parts of) the left and right hand side of (6.1), and imposing  $\frac{1}{k_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i^2$ ,  $i=1, \dots, I$ , we get

$$\frac{1}{k_i} \mathbf{X}_i' \mathbf{X}_i = \mathbf{B} \mathbf{D}_i^2 \mathbf{B}' \quad (6.2)$$

which equals the scalar products version of the INDSCAL (INDividual Differences SCALing) model (Carroll & Chang, 1970) applied to the matrices  $\frac{1}{k_i} \mathbf{X}_i' \mathbf{X}_i$ ,  $i=1, \dots, I$ .

The SCA-IND model should be used if there are interindividual differences in intraindividual variability (over time) for the separate components, but the separate components within subjects are uncorrelated for all subjects. The model should be used if the variables indicate several uncorrelated concepts, and if the subjects show differences in variability across time in the concepts indicated by the components.

### 6.2.6. SCA with Equal average Cross-Products constraints (SCA-ECP)

The model for SCA with Equal average Cross-Products constraints (SCA-ECP) is a constrained version of the SCA-P, SCA-PF2, and SCA-IND models. The name of this new method expresses the constraints on the component scores. That is, the SCA-ECP model is given by (6.1) with  $\mathbf{F}_i' \mathbf{F}_i$  constrained so that  $\frac{1}{k_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{\Phi}$ ,  $i=1, \dots, I$ , where  $\mathbf{\Phi}$  is a positive definite  $Q \times Q$  matrix. The restriction on the component scores implies that the inner products of the components are equal for all subjects. If the component scores matrix contains centered scores, the restriction implies that the correlations between, and the variances of the components are equal for all subjects. The estimates of the data matrices  $\mathbf{X}_i$ ,  $i=1, \dots, I$ , are insensitive to an orthogonal or oblique transformation of the component score matrices  $\mathbf{F}_i$ , provided that such a transformation is compensated in the loading matrix  $\mathbf{B}$ . Therefore, the sum of squares explained by the model does not alter by requiring  $\frac{1}{k_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{I}$  (instead of  $\frac{1}{k_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{\Phi}$ ), and thus it is clear that SCA-ECP is a constrained version of SCA-IND, and hence of the other models as well.

In the SCA-ECP model, the variances of components and covariances between components within subjects are equal for all subjects. This implies that the variable covariance matrix (computed over time) is equal for all subjects. This model is particularly useful if all subjects show equal variability over time on the separate components, and the correlations between the components are equal for all subjects.

### 6.2.7. Transformational freedom in the SCA-P, SCA-PF2, SCA-IND and SCA-ECP models

As will be discussed further in Section 6.3, the SCA models described above can all be fitted to a particular data set by least squares minimization of the residuals. In the

sequel it is assumed that parameter estimates of  $\mathbf{F}_i$  for at least one of the subjects, and the parameter estimates of  $\mathbf{B}$  have full rank. The estimates of the data matrices  $\mathbf{X}_i$ ,  $i=1, \dots, I$ , by the parameter matrices of the SCA-P, and SCA-ECP models are insensitive to orthogonal and oblique transformations of the loading matrix  $\mathbf{B}$ , provided that such a transformation is compensated in the component scores matrices  $\mathbf{F}_i$ ,  $i=1, \dots, I$ . Standard rotational procedures (e.g., Varimax; Kaiser, 1958) can be used to obtain solutions which are easier to interpret.

Kiers, Ten Berge and Bro (1999) have shown that, under some assumptions, SCA-PF2 solutions are ‘essentially unique’, which means that SCA-PF2 estimates are unique up to trivial permutation, reflection and/or rescaling. In the proof, it is required, among other things, that the number of subjects relative to the number of components is rather large ( $I \geq Q(Q+1)(Q+2)(Q+3)/24$ , where  $I$  denotes the number of subjects and  $Q$  the number of components). However, they report on the basis of simulations, that the uniqueness properties of PARAFAC2 appear to hold generally for  $I \geq 4$ .

Assuming that there is at least one pair  $\mathbf{D}_h, \mathbf{D}_i$  ( $h \neq i$ ;  $h, i=1, \dots, I$ ), so that  $\mathbf{D}_h \mathbf{D}_i^{-1}$  has no equal pair of diagonal elements, it can be proven that estimates of SCA-IND are essentially unique. The proof is largely based on the uniqueness proof for PARAFAC, as given by Harshman (1972).

#### 6.2.8. Model selection

In modeling data via an SCA-model, one aims at optimally separating observed data into a systematic part that is described by an interpretable model, and a residual part. As one does not know which part of the data is systematic, the choice of a particular SCA-model and number of components is a fairly complicated matter. The issues to take into consideration are the same as those discussed in Section 2.6.2. That is, the model should be interpretable and, have a small degree of overall error. The degree of overall error can be investigated using resampling techniques. We elaborated two such methods for use in the SCA-models. They will be discussed now. In practice, we apply the two methods to all SCA-models for a reasonable range of numbers of components.

Wold (1978) suggested the use of cross-validation for determining the number of components in PCA by assessing the predictive power of a model. Taking Wold’s method as starting point, Louwse, Smilde, and Kiers (1999) proposed the expectation maximization cross-validation (EM-CV) method to determine the numbers of components in multi-way component models. A variant of the EM-CV method can easily be applied to SCA-models. The expectation maximization cross-validation procedure for SCA (EM-SCA) can be summarized as follows. One element of the observed data is left out of the data, and the remaining part is preprocessed in the usual way (see Section 6.2.1). A sensible starting value is imputed for the missing value in the preprocessed data array, and this data array is then analyzed by the present SCA-model. The SCA estimation procedure is iterative, as will be discussed

in Section 6.3, and after each cycle the ‘missing element’ is estimated on the basis of the current model parameters, and it is imputed. The cycles are repeated until convergence, and the estimated value of the missing element is retained. The procedure is repeated by leaving out each element once until all elements have been eliminated once, and, as a result, each ‘missing element’ has been estimated once. Finally, the predictive residual error sum of squares (PRESS) is calculated as the sum of squared differences between estimated values and observed (preprocessed) values. A high PRESS value is indicative of a model with low predictive value, either because the model underfits or overfits the data. Underfitting may be caused by estimating too few components and/or using a too strongly restricted model, whereas the reverse leads to an overfitted model. Hence, only models with small PRESS values have a high predictive value. The size of the PRESS value depends partly on the degree of noise in the data. It is therefore not possible to give a generally valid bench-mark value for PRESS. We advise considering only those models that, or the data set at hand, have relatively low PRESS values compared to the PRESS values of the other models considered. A disadvantage of the EM-SCA cross-validation approach is that a huge number of analyses is needed, which may take much computing time. The process can be speeded up by leaving out several elements simultaneously to be handled as missing values. Still, the method can be impractical in the case of large data sets.

The second resampling based technique is split-half analysis. Split-half analysis can be performed in various ways. It is applied, for instance, in the PARAFAC model (Harshman & Lundy, 1984a) and the Tucker3 model (Kiers & Van Mechelen, 2001; see also Section 4.5), and it is used here for the SCA models. The observed data are randomly split into two parts over the mode that can be viewed as replications, thus the subject or the occasion mode. Then, the SCA model at issue is fitted to each of the (preprocessed) data halves. We propose comparing the two estimated loading matrices, after rotation to simple structure in the case of SCA-P and SCA-ECP (e.g., Varimax rotation; Kaiser, 1958). The rotation to simple structure hopefully leads to two similar interpretable solutions that are to be compared. If the solution is stable, the respective columns of the two loading matrices should be (approximately) equal up to permutation and/or reflection. Hence, permutation and reflection should be taken into account in a stability measure. As a stability measure, we propose using the average of the mean absolute difference of the columns  $\mathbf{b}_{q,1}$  and  $\mathbf{b}_{q,2}$ ,  $q=1,\dots,Q$ , where  $\mathbf{b}_{q,1}$  and  $\mathbf{b}_{q,2}$  denote the  $q^{\text{th}}$  column of the loading matrices  $\mathbf{B}_1$  and  $\mathbf{B}_2$ , and where the columns of  $\mathbf{B}_1$  are ordered and reflected so as to yield the lowest average mean of the (minimal) absolute differences. This stability measure will be denoted as the ‘split-half stability coefficient’. We advocate replicating the procedure a number of times (e.g., 50 times) and using the average split-half stability coefficients obtained this way, in order to increase the reliability of the stability study. A relatively low stability coefficient (e.g.,  $<0.10$ ) is indicative of a stable model.

Ideally, the above criteria point in the same direction as to the choice of a model. It is possible, however, that a few competing models each with its own

(dis)advantages, will remain. It might also happen that all criteria indicate different models, which might indicate that the SCA models are unsuitable for describing the data concerned.

### 6.3. Fitting the four SCA models to the data

To fit each of the four models for SCA to observed data, we propose minimizing the sum of squared residuals. Hence, we minimize

$$F(\mathbf{F}_i, \mathbf{B}) = \sum_{i=1}^I \|\mathbf{X}_i - \mathbf{F}_i \mathbf{B}'\|^2, \quad (6.3)$$

subject to the constraint imposed in the particular SCA. Hence, the total sum of squares that is explained by the model is maximized. To identify the solution partly,

we require for the four SCA models that  $\left(\sum_{i=1}^I K_i\right)^{-1} \sum_{i=1}^I \text{diag}(\mathbf{F}_i' \mathbf{F}_i) = \mathbf{I}_Q$ , where  $K_i$

denotes the number of measurements of subject  $i$ ,  $i=1, \dots, I$ , and  $\text{diag}(\mathbf{X})$  the diagonal of matrix  $\mathbf{X}$ . This identification constraint can always be obtained after estimates of  $\mathbf{F}_i$  and  $\mathbf{B}$  have been obtained, namely by a simple column-wise scaling transformation of  $\mathbf{B}$  and the  $\mathbf{F}_i$ 's. That is, the  $q^{\text{th}}$  column of the rescaled matrix  $\mathbf{F}_i^r$  is obtained by

$$\text{computing } \mathbf{f}_{qi}^r = \sqrt{\frac{\sum_{i=1}^I K_i}{\sum_{i=1}^I \mathbf{f}_{qi}' \mathbf{f}_{qi}}} \mathbf{f}_{qi}, \quad i=1, \dots, I; \quad \text{subsequently taking } \mathbf{b}_q^r = \sqrt{\frac{\sum_{i=1}^I \mathbf{f}_{qi}' \mathbf{f}_{qi}}{\sum_{i=1}^I K_i}} \mathbf{b}_q$$

compensates the rescaling in the  $q^{\text{th}}$  column of the loading matrix  $\mathbf{B}$ , where  $\mathbf{b}_q^r$  is the  $q^{\text{th}}$  column of the rescaled loading matrix. If the component scores matrices  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , are centered, this identification constraint implies that the variance per component over all subjects is one. In the sequel, we assume that this transformation is performed after the estimates of  $\mathbf{F}_i$  and  $\mathbf{B}$  have been obtained, and we will not explicitly discuss this transformation.

The degree to which the estimated model describes the data is expressed by the proportion of sum of squares explained by the model, which we call the 'fit' in the sequel. The fit is defined as

$$1 - \frac{\sum_{i=1}^I \|\mathbf{X}_i - \mathbf{F}_i \mathbf{B}'\|^2}{\sum_{i=1}^I \|\mathbf{X}_i\|^2}. \quad (6.4)$$

The fit is often multiplied by 100 and expressed as a percentage.

The SCA-P algorithm aims at minimizing (6.3) over arbitrary  $\mathbf{F}_i$ ,  $i=1,\dots,I$ , and  $\mathbf{B}$ . Kiers and Ten Berge (1994) give an explicit solution to this problem.

For SCA-PF2, SCA-IND and SCA-ECP we use alternating least squares (ALS) algorithms. The fitting of the SCA-PF2, SCA-IND, and SCA-ECP models to data will be treated successively.

### 6.3.1. Fitting the SCA-PF2 model

The SCA-PF2 algorithm aims at minimizing (6.3), subject to  $\frac{1}{k_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i \mathbf{\Phi} \mathbf{D}_i$ ,  $i=1,\dots,I$ , where  $\mathbf{D}_i$  is a  $Q \times Q$  diagonal matrix and  $\mathbf{\Phi}$  a positive definite  $Q \times Q$  matrix with unit diagonal elements. Kiers, Ten Berge and Bro (1999) proposed an ALS algorithm for the equivalent problem of minimizing (6.3) subject to  $\mathbf{F}_i' \mathbf{F}_i = \tilde{\mathbf{D}}_i \tilde{\mathbf{\Phi}} \tilde{\mathbf{D}}_i$  with  $\tilde{\mathbf{D}}_i$  a diagonal  $Q \times Q$  matrix and  $\tilde{\mathbf{\Phi}}$  an arbitrary positive definite  $Q \times Q$  matrix. Their algorithm is essentially based on the fact that every matrix  $\mathbf{F}_i$  that meets the constraint  $\mathbf{F}_i' \mathbf{F}_i = \tilde{\mathbf{D}}_i \tilde{\mathbf{\Phi}} \tilde{\mathbf{D}}_i$  can be written as  $\mathbf{F}_i = \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i$  provided that  $\mathbf{P}_i' \mathbf{P}_i = \mathbf{I}_Q$ ,  $\tilde{\mathbf{F}}$  is an arbitrary  $Q \times Q$  matrix, and  $\tilde{\mathbf{D}}_i$  a diagonal  $Q \times Q$  matrix,  $i=1,\dots,I$ . The SCA-PF2 algorithm as proposed by Kiers, Ten Berge and Bro (1999) boils down to minimizing

$$f_1(\mathbf{P}_i, \tilde{\mathbf{F}}, \tilde{\mathbf{D}}_i, \mathbf{B}) = \sum_{i=1}^I \left\| \mathbf{X}_i - \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i \mathbf{B}' \right\|^2. \quad (6.5)$$

subject to  $\mathbf{P}_i' \mathbf{P}_i = \mathbf{I}_Q$ , and  $\tilde{\mathbf{D}}_i$  a diagonal matrix,  $i=1,\dots,I$ . The function in (6.5) is minimized by updating  $\mathbf{B}$ ,  $\mathbf{P}_i$ ,  $\tilde{\mathbf{F}}$ , and  $\tilde{\mathbf{D}}_i$  alternately. We propose using this algorithm to find solutions for  $\mathbf{B}$ ,  $\mathbf{F}_i$ ,  $\mathbf{\Phi}$ , and  $\mathbf{D}_i$ . Solutions for  $\mathbf{D}_i$  and  $\mathbf{\Phi}$  so that  $\frac{1}{k_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i \mathbf{\Phi} \mathbf{D}_i$  can then be obtained by taking  $\mathbf{\Phi} = \text{diag}(\tilde{\mathbf{\Phi}})^{-\frac{1}{2}} \tilde{\mathbf{\Phi}} \text{diag}(\tilde{\mathbf{\Phi}})^{-\frac{1}{2}}$ , and  $\mathbf{D}_i = \frac{1}{\sqrt{k_i}} \tilde{\mathbf{D}}_i \text{diag}(\tilde{\mathbf{\Phi}})^{\frac{1}{2}}$ .

### 6.3.2. Fitting the SCA-IND model to data

The SCA-IND algorithm aims at minimizing (6.3) subject to  $\frac{1}{k_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i^2$ ,  $i=1,\dots,I$ , with  $\mathbf{D}_i$  a diagonal  $Q \times Q$  matrix. The ALS algorithm to find estimates of the parameters of the SCA-PF2 model (Kiers, Ten Berge & Bro, 1999) can be used to find estimates of  $\mathbf{F}_i$ ,  $i=1,\dots,I$ , subject to  $\mathbf{F}_i' \mathbf{F}_i = \tilde{\mathbf{D}}_i^2$  by keeping  $\tilde{\mathbf{\Phi}} = \mathbf{I}$ . In the SCA-PF2 algorithm this is obtained by keeping  $\tilde{\mathbf{F}}$  fixed as  $\tilde{\mathbf{F}} = \mathbf{I}$ , and only updating  $\mathbf{B}$ ,  $\mathbf{P}_i$ , and

$\tilde{\mathbf{D}}_i$ . With this algorithm we find solutions for  $\mathbf{B}$ ,  $\mathbf{F}_i$  and  $\tilde{\mathbf{D}}_i$ . Solutions for  $\mathbf{D}_i$  so that  $\frac{1}{k_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i^2$  can be obtained by taking  $\mathbf{D}_i = \frac{1}{\sqrt{k_i}} \tilde{\mathbf{D}}_i$ .

### 6.3.3. Fitting the SCA-ECP model to data

The SCA-ECP algorithm aims at minimizing (6.3) subject to  $\frac{1}{k_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{M}$ ,  $i=1, \dots, I$ , which is equivalent to (i.e., without affecting the model fit) requiring that  $\frac{1}{k_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{I}_Q$ . Updating  $\mathbf{B}$ , and  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , alternately can solve this problem. The problem of finding an update for  $\mathbf{B}$  is analogous to finding an update for  $\mathbf{B}$  in the SCA-PF2 algorithm (Kiers, Ten Berge & Bro, 1999). The next problem is to find, for every value of  $i$ , an update for  $\mathbf{F}_i$ , subject to  $\frac{1}{k_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{I}_Q$ . This can be done by minimizing

$$f_2(\mathbf{F}_i) = \|\mathbf{X}_i - \mathbf{F}_i \mathbf{B}'\|^2. \quad (6.6)$$

subject to  $\frac{1}{k_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{I}_Q$ . By defining  $\tilde{\mathbf{F}}_i = \frac{1}{\sqrt{k_i}} \mathbf{F}_i$ , this is equivalent to maximizing  $\text{tr}(\tilde{\mathbf{F}}_i' \mathbf{X}_i \mathbf{B})$ , subject to  $\tilde{\mathbf{F}}_i' \tilde{\mathbf{F}}_i = \mathbf{I}$ . Consider the singular value decomposition  $\mathbf{X}_i \mathbf{B} = \mathbf{U}_i \mathbf{\Delta}_i \mathbf{Q}_i'$ , with  $\mathbf{U}_i' \mathbf{U}_i = \mathbf{Q}_i' \mathbf{Q}_i = \mathbf{Q}_i \mathbf{Q}_i' = \mathbf{I}_Q$ , and  $\mathbf{\Delta}_i$  a diagonal matrix with nonnegative diagonal elements in weakly descending order. Then the maximum of  $\text{tr}(\tilde{\mathbf{F}}_i' \mathbf{X}_i \mathbf{B})$  is given by  $\tilde{\mathbf{F}}_i = \mathbf{U}_i \mathbf{Q}_i'$  (Cliff, 1966), hence an update of  $\mathbf{F}_i$  is given by  $\mathbf{F}_i = \sqrt{k_i} \mathbf{U}_i \mathbf{Q}_i'$ .

### 6.3.4. Starting values of the parameters

Each of the iterative algorithms (i.e., SCA-PF2, SCA-IND and SCA-ECP) has to be initialized with certain starting values. The starting matrices can be drawn randomly from, for example, a normal distribution. In practice, it is recommended that several differently started runs are used in order to decrease the chance of missing the global minimum of the function. As a so-called rational start, the (explicit) SCA-P solution for the matrix  $\mathbf{B}$  can be used in all algorithms. In the case of SCA-IND and SCA-PF2, the starting matrices for the diagonal matrices  $\mathbf{D}_i$ ,  $i=1, \dots, I$ , are set at  $Q \times Q$  identity matrices. In the case of SCA-PF2, the identity matrix ( $Q \times Q$ ) is also used as a starting matrix for  $\mathbf{\Phi}$ . These starting values suffice to start the iterative process (by updating  $\mathbf{P}_i$  or  $\mathbf{F}_i$ ).

## 6.4. Empirical examples of simultaneous component analyses

In this section two empirical examples will be provided to illustrate the application of the SCA models to multisubject multivariate time series.

#### 6.4.1. Empirical example 1: Mood in individuals with Parkinson's disease

In a study by Shifren, Hooker, Wood and Nesselroade (1997), mood structure among twelve individuals diagnosed with Parkinson's disease was examined. Positive and negative affect was measured with the Positive and Negative Affect Schedule (PANAS; Watson, Clark & Tellegen, 1988). This measure contains ten positive and ten negative affect items. Subjects were asked to rate the 20 adjectives on a five point scale (ranging from one (not at all) to five (all the time)) to indicate to what degree they experienced the particular affect on that day. Subjects scored the PANAS daily on successive days, for periods ranging from 53 to 71 days. For the twelve subjects together, scores on 817 days were obtained. The study investigated the intraindividual structure of mood, as well as the interindividual differences. Watson (1988) showed that interindividual differences in mood (of healthy subjects) can be described well by two relatively independent dimensions, namely positive and negative affect. Central questions were whether this structure can also be used in describing intraindividual differences in mood of subjects suffering from Parkinson's disease, and whether the degree of intraindividual variability differed across subjects.

The scores on the items on successive days were analyzed by Shifren et al. (1997) using a dynamic factor analysis. A dynamic factor model was estimated for each subject separately, where, based on issues regarding content, the maximal number of factors was two, and the maximal 'lag' was one. Items showing responses that were too stable over time (over 90% of the responses in the same category) were eliminated from the analyses. Also, any linear trend over time per variable per subject was removed from the data. Further information concerning the research, method of analyses and results can be found in Shifren et al. (1997).

Here, the data were analyzed by simultaneous component analyses. Prior to the analyses, the data of each subject per variable were centered over the time points, and normalized within variables (i.e., over occasions and subjects jointly), so that the sum of squares per variable over occasions *and* subjects was equal to the sum of the number of occasions for all subjects ( $\sum_{i=1}^I K_i$ ,  $i=1, \dots, I$ ) (see Section 6.2.1). In contrast to Shifren et al. (1997), to keep as much of the information in the data as possible, we did not remove trends from the data, nor did we exclude variables from the analyses.

We first assessed the sensitivity to sampling fluctuations and the fit of each of the four models (see Section 6.2.8) with one through five components, thus of 20 models in total. In each SCA analysis to assess the model fit, the SCA algorithm was run five times, using the rational start (see Section 6.3.4), and four random starts. In each SCA analysis performed to assess the stability, the SCA algorithm was run only one time, using the rational start, to reduce computing time. In each analysis the convergence criterion was set at  $10^{-6}$ .

For each of the 20 models, the PRESS value was computed following the procedure for EM-SCA cross-validation, as discussed in Section 6.2.8. The procedure was applied 150 times: 149 times a different set of 109 randomly chosen observations

was left out, and one time the remaining 99 observations were left out, so that each element was left out of the data set in one of the 150 analyses. As starting value, a zero was imputed for each removed observation. This is the average score per subject per variable (as a result of the preprocessing procedure). In this way PRESS values were obtained for each of the 20 models.

The split-half procedure was applied following the guidelines as discussed in Section 6.2.8. We repeated the split-half procedure 50 times, resulting in an average split-half stability coefficient over 50 replications. The average split-half stability coefficients (SHS), the PRESS values and the fit percentages of the SCA-P, SCA-PF2, SCA-IND, and SCA-ECP models with one through five components are reported in Table 6.2.

Table 6.2. PRESS values, split-half stability coefficients (SHS) and Fit (%) of the four SCA-models with one through five components for the 'Mood data'.

Model	Measure	$Q=1$	$Q=2$	$Q=3$	$Q=4$	$Q=5$
SCA-ECP	PRESS ( $\times 10^4$ )	1.39	1.49	1.76	2.02	2.72
	SHS	0.06	0.08	0.09	0.12	0.13
	Fit	24.0	31.8	37.1	41.0	44.0
SCA-IND	PRESS ( $\times 10^4$ )	1.24	1.16	1.18	1.27	1.51
	SHS	0.07	0.09	0.13	0.14	0.14
	Fit	30.8	42.8	49.5	55.3	60.1
SCA-PF2	PRESS ( $\times 10^4$ )	1.24	1.16	1.18	1.20	2.05
	SHS	0.07	0.58	1.03	2.12	2.00
	Fit	30.8	43.1	50.0	56.6	61.2
SCA-P	PRESS ( $\times 10^4$ )	1.24	1.14	1.12	1.13	1.25
	SHS	0.07	0.09	0.20	0.10	0.12
	Fit	30.8	43.5	51.2	58.4	63.2

\*Note that for  $Q=1$ , the models SCA-IND, SCA-PF2 and SCA-P are equivalent.

In Table 6.2, we see that using the SCA-IND model instead of the more constrained SCA-ECP model increases the fit considerably, for all models with one through five components. Almost no improvement in fit is gained by using the even less constrained SCA-PF2 or SCA-P model instead of the SCA-IND model. Thus, the fit percentages indicate that SCA-IND is preferable, but they do not clearly indicate the preferred number of components.

Low PRESS values and low split-half stability coefficients (SHS) indicate insensitivity to sampling fluctuations. On the basis of a comparison of the present PRESS and SHS values, we deem models sufficiently insensitive if their SHS values

do not exceed 0.10 and their PRESS values do not exceed  $1.20 \times 10^4$ . Thus, SCA-IND and SCA-P with two components, and SCA-P with four components will be considered. We start the discussion of the models with the most restricted model of the three, SCA-IND with two components. Note that using either SHS or PRESS to investigate the insensitivity would lead to the selection of different models.

The loading matrix **B** of the SCA-IND solution is presented on the left hand side of Table 6.3. The components can be interpreted as 'Introversion', and 'Emotional Instability'.

Table 6.3. Loading matrix of the SCA-IND model with two components (Columns 2 and 3), and Varimax rotated loading matrix of the SCA-P model with four components (Columns 4 through 7) for the 'Mood data'. Loadings  $\leq -0.40$  or  $\geq 0.40$  are printed in bold face.

	SCA-IND		SCA-P			
	Introver- sion	Emotional Instability	Arousal	Nervous- ness	Emotional Instability	Comp. IV
jittery	<b>0.59</b>	<b>0.43</b>	<b>0.45</b>	<b>0.50</b>	0.37	-0.15
distressed	-0.02	<b>0.65</b>	0.02	0.17	<b>0.66</b>	0.27
upset	-0.01	<b>0.74</b>	0.09	0.11	<b>0.75</b>	0.15
afraid	<b>0.61</b>	0.02	0.19	<b>0.75</b>	-0.00	0.11
scared	<b>0.63</b>	-0.04	0.18	<b>0.78</b>	-0.06	0.08
hostile	<b>0.44</b>	<b>0.52</b>	<b>0.41</b>	0.35	<b>0.46</b>	-0.13
irritable	-0.07	<b>0.76</b>	0.09	-0.01	<b>0.78</b>	0.09
guilty	-0.10	<b>0.66</b>	0.12	-0.18	<b>0.75</b>	-0.22
ashamed	<b>0.45</b>	0.22	0.26	<b>0.40</b>	0.29	<b>-0.40</b>
nervous	<b>0.55</b>	<b>0.40</b>	0.37	<b>0.53</b>	0.37	-0.12
inspired	<b>-0.71</b>	0.03	<b>-0.59</b>	<b>-0.41</b>	0.21	-0.12
excited	0.01	-0.24	-0.14	0.02	-0.07	<b>-0.69</b>
determined	-0.25	<b>-0.45</b>	<b>-0.58</b>	0.25	-0.29	-0.01
interested	<b>-0.75</b>	-0.24	<b>-0.78</b>	-0.27	-0.04	0.05
enthusiastic	<b>-0.68</b>	-0.29	<b>-0.72</b>	-0.28	-0.06	-0.08
attentive	<b>-0.69</b>	-0.25	<b>-0.77</b>	-0.19	-0.04	-0.07
proud	0.06	-0.23	0.04	-0.14	-0.12	<b>-0.68</b>
strong	-0.04	<b>-0.59</b>	<b>-0.45</b>	0.30	-0.37	-0.33
active	<b>-0.58</b>	<b>-0.46</b>	<b>-0.73</b>	-0.10	-0.29	0.08
alert	<b>-0.69</b>	-0.32	<b>-0.75</b>	-0.25	-0.11	-0.03

The size of the occasion component scores can be compared *between* subjects: extremely high (and low) component scores indicate a large degree of variability in scores over time compared to the other subjects. The degree of variability across persons can easily be compared on the basis of the variances of the component scores,

that equal the diagonal elements of  $\mathbf{D}_i^2$  here (see Section 6.2.5), and these values are presented in the left hand panel of Table 6.4. As can be seen in this table, for example, Subject 5 shows most, and Subject 10 shows least variability on the Introversion component.

Table 6.4. Variances ( $\mathbf{D}_i^2$ ) of and covariances and correlations between the component scores per subject of the mood data. 'In' denotes Introversion, 'EI' denotes Emotional Instability.

Subject	SCA-IND		SCA-P				SCA-IND compared to SCA-P	
	variances ( $\mathbf{D}_i^2$ )		variance		covariance	correlation	correlation	
	In	EI	In	EI	In/EI	In/EI	In/In	EI/EI
1	0.73	1.07	0.61	0.87	0.39	0.54	0.95	0.96
2	0.18	5.30	0.21	5.71	-0.43	-0.39	0.92	1.00
3	0.38	0.53	0.38	0.55	-0.06	-0.13	0.99	1.00
4	2.04	1.20	2.69	1.64	-1.01	-0.48	0.99	0.93
5	6.42	0.97	5.99	0.89	0.42	0.18	1.00	0.99
6	0.31	1.27	0.31	0.97	0.31	0.56	0.90	0.99
7	0.63	0.30	0.53	0.23	0.11	0.31	0.99	0.96
8	0.16	0.38	0.16	0.33	0.04	0.19	0.99	1.00
9	0.65	1.50	0.61	1.45	0.12	0.13	0.99	1.00
10	0.08	0.17	0.10	0.19	-0.05	-0.35	0.98	0.99
11	0.20	0.06	0.19	0.05	0.02	0.24	0.99	0.97
12	0.17	0.28	0.17	0.26	0.01	0.06	1.00	1.00

A comparison of the size of the component scores *within* subjects reveals the subject's degree of 'Introversion' or 'Emotional Instability' compared to the degree at the other days of the subject concerned. To give some insight into variation of the component scores over time, the Introversion scores of Subjects 5 and 10, and the Emotional Instability scores of Subjects 2 and 4, resulting from SCA-IND and SCA-P, are plotted in Figure 6.1.

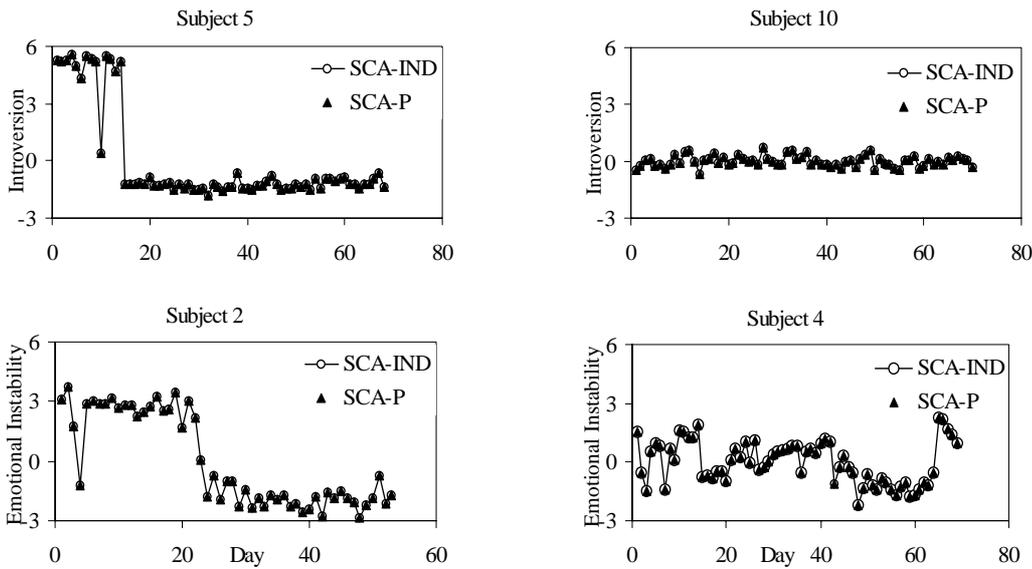


Figure 6.1. Introversion component scores of Subjects 5 and 10, and Emotional Instability component scores of Subjects 2 and 4.

Figure 6.1 not only illustrates differences between subjects in intraindividual variability, but also offers the possibility of identifying trends in component scores. For example, Subject 5 shows a remarkable shift towards extraversion on Day 15 (but, as has been verified, a similar change was not found in the Emotional Instability component scores of this subject). Unfortunately, additional information about the subjects to explain these changes in component scores is lacking.

The SCA-P model with two components explains 43.5% of the variance, which is only 0.7% more than the SCA-IND model with two components. To be able to compare the loading matrices of SCA-IND and SCA-P, we rotated the loading matrix of SCA-P orthonormally to the loading matrix of SCA-IND, and this rotation is compensated in the component score matrices of all subjects. The loading matrix obtained in this way (which is not presented here) strongly resembles the loading matrix of the SCA-IND solution, as is indicated by a maximal absolute difference in the loadings of 0.08, and an average absolute difference of 0.02. (Incidentally, the normalized *Varimax* rotated loading matrix of SCA-P is also rather similar to the SCA-IND loading matrix, with maximal and mean absolute differences in loadings 0.11, and 0.05, respectively.) Therefore, the components for the transformed SCA-P solution are interpreted as for the SCA-IND solution, namely as ‘Introversion’, and ‘Emotional Instability’.

The component scores of the transformed SCA-P solution with two components can be compared to the scores of the SCA-IND solution in different ways. For each subject we computed the average absolute difference between the component scores (both Introversion and Emotional Instability) of the SCA-P solution and the SCA-IND solution. On average, the SCA-P component scores deviate more than 0.10 from the scores of the SCA-IND solution for Subjects 1, 2, 4, 6, and 7. Not surprisingly, these are subjects with relatively high covariances between the two components in the transformed SCA-P solution, as reflected in correlations larger than 0.30 (see Table 6.4). The SCA-IND and SCA-P component scores can also be compared by investigating whether the score profiles (over time) of the two solutions are approximately equal. The correlation coefficients between the component scores of SCA-P and SCA-IND per component per subject are reported in Table 6.4. An impression of the differences in component scores between the two methods can be obtained from the plots in Figure 6.1. In Table 6.4, it can be seen that the correlation coefficients are rather high, thus the score profiles of Introversion and Emotional Instability as estimated by SCA-IND and SCA-P are approximately equal for all subjects. Even the estimated Emotional Instability scores of Subject 4, who shows the lowest correlation among all subjects (0.93), do not differ so much that the interpretation of the development over time would be different. Thus, in comparing the SCA-P to the SCA-IND solution, the presence of clearly non-zero correlations between the SCA-P components for a number of subjects is most striking.

The SCA-P model with *four* components explains 58.4% of the variance, which is as much as 14.9% more than the SCA-P model with two components. The Varimax rotated loading matrix is presented on the right hand side of Table 6.3. In comparison to the loading matrices of SCA-IND and SCA-P with two components, the Varimax rotated SCA-P loading matrix roughly shows a split of the Introversion component. That is, most of the positive affect items ('inspired' through 'alert') load high ( $<-0.40$ ) on a different component than the negative items ('jittery' through 'nervous') do. These two components can be labeled as 'Arousal' and 'Nervousness'. Furthermore, the items 'excited' and 'proud', that were not assigned to a component in the SCA-IND solution, form, together with 'ashamed', one component (comp. IV in Table 6.3). It is interesting to note that one subject shows a large positive correlation between 'Arousal' and 'Nervousness'. This is strange from the point of view of the interpretation, but it fully reflects the trend that is also perceptible in the raw data of this subject. The covariances between the components for the different subjects vary from large to small, but they are difficult to interpret.

The broad interpretations of the three models just discussed have much in common, but the models differ in the details. The less restricted the model is, the more details of the data (whether of the variables or the subjects) are captured in the model. The preference for a particular model then depends on the degree of interest for those details. The SCA-IND model with two components is fairly simple to interpret, and covers the main features of the data, as is indicated by a relatively high fit, at least compared to SCA-P with two components. In both SCA-P models (two

and four components) the components covary in a complicated manner for the different subjects. The SCA-P model with four components offers detailed insight into the structure of the variables, and fits the data much better than SCA-IND with two components. This model is preferred if a detailed description is desired.

On the basis of the estimated SCA-P models and SCA-IND model, one can conclude that the positive/negative dimensions, which were found when the PANAS scale was used to describe interindividual differences in healthy subjects, are not found in subjects suffering from Parkinson's disease. Instead, the intraindividual differences in mood in Parkinson patients mainly follow the 'Emotional Instability' and 'Introversion' dimensions. Furthermore, the three models show that the degree of intraindividual variability differed greatly across subjects, indicating that the state of mind stability varies across Parkinson patients.

#### 6.4.2. Empirical example 2: The Big Five as states

The paper by Borkenau and Ostendorf (1998) deals with the question whether a similarity exists between the factor structures of longitudinal variations in states and the factor structures of individual differences in traits. Individual differences in traits have been successfully described along five dimensions, which are denoted by the Big Five (see e.g., Tupes & Cristal, 1961; Goldberg, 1990). Generally, the research into the Big Five makes use of (self or peer) ratings on a number of adjectives, which are thought to be indicative of the five dimensions. Borkenau and Ostendorf investigated whether the same dimensions also underlie variations of states within persons.

Over a period of 90 days, 22 subjects rated 30 self-report items describing their behavior on that particular day. The 30 adjective items (see Table 6.6) were marker items for Neuroticism, Extraversion, Agreeableness, Conscientiousness, and Intellect, and every construct was supposed to be measured by six items. A comparison of the current research with earlier Big Five trait research was made on the basis of a so-called reference loading matrix. This matrix (reported by Borkenau & Ostendorf, 1998) was obtained from a study into individual differences in traits, which used the same variables as the current study.

To analyze the multivariate time series of the 22 subjects, Borkenau and Ostendorf (1998) computed a correlation matrix for each subject, that contained the correlations between variables over occasions. They performed a principal component analysis on the average variable correlation matrix, computed over subjects, a method proposed by Levin (1966). The loading matrix obtained this way was orthogonally Procrustes rotated to the reference loading matrix and compared to the reference loading matrix using coefficients of congruence (Tucker, 1951). The factor structure of the averaged correlations matched the factor structure of individual differences in traits quite well (congruence coefficients 0.94, 0.96, 0.88, 0.93 and 0.88). If in Levin's method the correlation matrices are based on the same sample size (which is the case here), this method gives equivalent estimates of the loading matrix obtained

from an SCA-P analysis (Kiers & Ten Berge, 1994) applied to separately standardized data matrices for each subject.

We performed a series of SCA analyses on these data to investigate whether interindividual differences in intraindividual variability exist in the data, to what extent the covariances between components differ across subjects, and whether trends could be discovered over time. Before analysis, we centered the data of each subject per variable over the time points, and normalized within variables (i.e., over occasions and subjects jointly), as described in Section 6.2.1. The normalization procedure applied this way implies that differences in intraindividual variability are kept in the data, which is not the case in Borkenau and Ostendorf's approach (1998). Because Borkenau and Ostendorf hypothesized that the structure of the data is covered by five dimensions we focus on SCA models with five components. However, we also considered models with numbers of components close to five to be able to compare the relative stability and fit.

We first assessed the stability and fit of the four SCA models with three through six components. The stability was investigated by applying the split-half procedure (see Section 6.2.8). The procedure was repeated 25 times, resulting in an average split-half stability coefficient over 25 replications. The SCA algorithm was run five times, using the rational start (see Section 6.3.4), and four random starts, in each SCA analysis to assess the model fit. In the split-half procedure, the SCA algorithm was run only once, using the rational start, to reduce computing time. In each analysis the convergence criterion was set at  $10^{-6}$ . The EM-SCA cross-validation procedure was not applied, as this would have lead to a prohibitively long computing time. The average split-half stability coefficients (SHS) and the fit percentages of the SCA-P, SCA-PF2, SCA-IND, and SCA-ECP models with three through six components are reported in Table 6.5.

Table 6.5. PRESS values, split-half stability coefficients (SHS) and fit (%) of the four SCA-models with three through six components for the 'Big Five data'.

Model	SCA-ECP		SCA-IND		SCA-PF2		SCA-P	
	SHS	Fit	SHS	Fit	SHS	Fit	SHS	Fit
$Q=3$	0.05	37.4	0.09	45.8	0.58	46.0	0.07	46.6
$Q=4$	0.05	41.2	0.15	49.9	1.28	50.4	0.05	51.2
$Q=5$	0.09	44.4	0.14	53.6	1.79	54.2	0.07	55.2
$Q=6$	0.05	47.4	0.14	56.8	2.26	57.5	0.05	58.6

Low SHS values indicate a stable model. On the basis of a comparison of the current SHS values, we deem models with an SHS value not exceeding 0.10 stable. According to this criterion, the SCA-ECP and SCA-P models with three through six components, and the SCA-IND model with three components are stable. A comparison of fit values of the stable models reveals that the SCA-P models investigated fit the data considerably better than the SCA-ECP models, thus an SCA-

P model is favored on the basis of fit. For the three component solutions, SCA-IND fits only slightly worse than SCA-P, but SCA-IND with three components appeared hard to interpret. The number of components cannot clearly be inferred from a comparison of the fit values. Because five dimensions were hypothesized, we treat the SCA-P solution with five components here.

The loading matrix of the SCA-P solution with five components was orthogonally Procrustes rotated to the reference loading matrix used by Borkenau and Ostendorf (1998). The loadings obtained this way are presented in Table 6.6. The Procrustes rotated loading matrix resembles the reference loading matrix reasonably well, as indicated by congruence coefficients of 0.95, 0.96, 0.91, 0.91 and 0.95 between the corresponding columns of the two loading matrices. The item 'changeable', that is assigned to 'Intellect' in trait research, is assigned here to 'Neuroticism' (loading >0.50). 'Reckless' and 'bad tempered' are not assigned here to any component (loadings between -0.50 and 0.50).

Table 6.6. Loading matrix of the SCA-P solution after orthogonal Procrustes rotation to the Reference loading matrix of the 'Big Five data'. Loadings  $\leq -0.50$  and  $\geq 0.50$  are indicated in bold face.

Trait adjective	Neuroticism (N)	Extra-version (E)	Agreeableness (A)	Conscientiousness (C)	Intellect (I)
irritable	<b>0.71</b>	-0.14	-0.04	-0.11	-0.09
bad-tempered	0.35	-0.04	-0.46	-0.14	-0.17
vulnerable emotionally	<b>0.72</b>	-0.13	-0.01	-0.15	-0.05
stable	<b>-0.69</b>	0.04	0.08	0.29	0.11
calm	<b>-0.63</b>	-0.04	0.09	0.10	0.02
resistant	<b>-0.69</b>	0.08	-0.06	0.11	0.03
dynamic	-0.07	<b>0.50</b>	0.08	0.26	0.08
sociable	-0.09	<b>0.66</b>	0.16	0.00	0.02
lively	0.02	<b>0.67</b>	-0.09	-0.10	0.10
shy	0.14	<b>-0.55</b>	0.02	0.10	-0.20
silent	0.06	<b>-0.64</b>	-0.05	0.10	-0.12
reserved	0.06	<b>-0.61</b>	0.04	0.08	-0.13
good-natured	-0.13	0.20	<b>0.54</b>	0.21	-0.11
helpful	0.08	0.26	<b>0.53</b>	0.38	-0.11
considerate	-0.03	0.08	<b>0.64</b>	0.25	-0.10
selfish	0.11	0.06	<b>-0.61</b>	-0.16	-0.03
domineering	0.05	0.23	<b>-0.61</b>	0.06	-0.06
obstinate	0.08	0.22	<b>-0.59</b>	0.04	-0.05
industrious	-0.02	0.05	0.12	<b>0.68</b>	0.15
persistent	-0.27	0.09	-0.08	<b>0.53</b>	0.01
responsible	-0.05	0.04	0.25	<b>0.62</b>	0.09
lazy	-0.05	-0.06	-0.21	<b>-0.65</b>	-0.11
reckless	-0.00	0.47	-0.29	-0.28	-0.13
changeable	<b>0.58</b>	0.01	-0.11	-0.35	-0.15
witty	-0.10	0.16	-0.06	0.09	<b>0.60</b>
knowledgeable	-0.11	-0.07	-0.17	0.26	<b>0.73</b>
prudent	-0.09	-0.01	-0.06	0.24	<b>0.70</b>
fanciness	-0.08	-0.40	-0.08	0.15	<b>-0.50</b>
uninformed	0.15	0.06	0.07	-0.16	<b>-0.71</b>
unimaginative	-0.04	-0.33	-0.04	0.00	<b>-0.56</b>

The variance/covariance matrices of the components per subject give insight into the variability over time, and the correlations between the components of the subjects.

The variance/covariance matrix for Subjects 1 and 2 is presented in Table 6.7 to give an impression of two of those matrices.

Table 6.7. Variance/covariance matrices of the components of Subjects 1 and 2 of the 'Big Five data'.

	Subject 1					Subject 2				
	N	E	A	C	I	N	E	A	C	I
N	0.90					0.37				
E	-0.27	0.75				0.14	0.80			
A	-0.25	-0.20	0.87			-0.24	0.01	0.87		
C	0.25	-0.09	0.19	0.69		0.14	0.09	-0.31	1.00	
I	-0.24	0.27	0.13	-0.08	0.83	-0.01	0.23	0.16	0.04	0.51

The degree of variability over time across persons can be compared on the basis of the variances of the component scores. The variability within subjects differs across components, as is indicated by variances of the component score variances of 0.75, 3.15, 0.32, 1.03 and 1.12 for components one through five, respectively. The component score variances also differ across subjects, but those variances are to a large extent linearly related, as reflected by generally high correlation coefficients ( $>0.50$ ) between the variances for the five components. Thus, in general, subjects with high (or low) component scores variance on one component tend to show high (or low) variances on the others as well. We next inspected correlations between component scores for the five components within subjects. This revealed that Neuroticism and Agreeableness are negatively correlated for almost all subjects with six out of 22 subjects showing a correlation even smaller than  $-0.50$ . Finally, plots of the component scores per component and per subject against the time axis were made. These did not reveal any trend in the data.

The SCA-ECP model with five components is also considered a stable model, but the fit is much lower than of the SCA-P model. The orthogonally Procrustes rotated loading matrix of this SCA-ECP model resembles the reference loading matrix to a lesser degree than that of the SCA-P model (congruence coefficients of 0.95, 0.95, 0.92, 0.89 and 0.89 for components one through five, respectively). The interindividual differences in variability are not captured in the model. Therefore, we prefer the SCA-P model, as discussed in the previous paragraph.

On the basis of the comparison of the loading matrix of the SCA-P model with five components and the reference loading matrix of the 'trait research', we come to the same conclusion as Borkenau and Ostendorf (1998): Not only individual differences in traits, but also intraindividual differences in states can be described along five dimensions. Additionally, the estimates of the SCA-P model show interindividual differences in variability across components as well as across subjects. The high negative correlation between two components ('Neuroticism' and 'Agreeableness') in the SCA-P model for a part of the sample suggests that four

instead of five dimensions would suffice to describe differences in states for some of the subjects.

## 6.5. Discussion and conclusion

Four variants of models for Simultaneous Component Analysis and the properties of the models were discussed in this chapter. The models for SCA are particularly useful for the exploratory analysis of multisubject multivariate time series. The four models can be ordered hierarchically. The most restricted model, the SCA-ECP model, does not allow for variation between subjects in terms of average cross-products of component scores (or covariances, if the component scores are centered), whereas the least restricted model, the SCA-P model, allows for most variation between subjects. A prerequisite for any chosen model is that it is interpretable and has a low degree of modeling error, implying that the overall error should be small. This can be assessed using resampling methods, like the EM-SCA cross-validation or split-half analysis. The methods evaluate different aspects of insensitivity, and therefore we advise using both methods simultaneously. However, because EM-SCA cross-validation may take very much computing time in the case of large data sets, this method is not always feasible. The final decision on which model to use should be based on the interpretability of the model. The empirical examples illustrated the use of the SCA models in practice. The examples offered nicely interpretable solutions, in which the intraindividual as well as the interindividual structure is covered.

The four models are not only suitable for modeling multivariate time series of two or more subjects simultaneously. SCA-P has been used to analyze scores of two or more groups of individuals on the same variables (e.g., Kiers & Ten Berge, 1994; Niesing, 1997). An application of SCA-PF2 to chemical data that does not involve time series is given in Bro, Andersson and Kiers (1999). Those types of data can be modeled by the other three methods as well. Depending on the extent of the differences between groups, samples, or individuals, SCA-P, SCA-PF2, or the new SCA-IND and SCA-ECP models can be chosen.

In the SCA models, the components are defined on the basis of the degree to which the variable scores are correlated over time, at least over all subjects simultaneously. Low correlations between the component scores over time certainly do not imply that the distinct series of component scores are unrelated: They could be associated non linearly. One could possibly extend the present models to models that incorporate non-linear relationships between component scores. Alternatively, a simple approach to reveal possible non-linear relationships between component scores is to inspect visually the various component scores time series simultaneously per subject. Furthermore, it is possible that components in SCA models have instantaneous as well as lagged influences on the variables. Such situations are discussed in the next chapter.

### Appendix 6.1.

Let  $\mathbf{X}_i$  ( $K_i \times J$ ) denote the matrix of scores of the  $i^{\text{th}}$  subject ( $i=1, \dots, I$ ) on  $J$  variables measured on  $K_i$  occasions ( $k_i=1, \dots, K_i$ ). To fit the SCA models to observed data, the following function is minimized

$$F(\mathbf{F}_i, \mathbf{B}) = \sum_{i=1}^I \|\mathbf{X}_i - \mathbf{F}_i \mathbf{B}'\|^2, \quad (6.7)$$

where  $\mathbf{F}_i$  ( $K_i \times Q$ ) contains the  $Q$  component scores of subject  $i$  at time points  $1, \dots, K_i$ ,  $\mathbf{B}$  ( $J \times Q$ ) denotes the loading matrix, and  $\mathbf{E}_i$  ( $K_i \times J$ ) denotes the matrix of residuals, and where  $\mathbf{F}_i$  is subject to the constraint in the SCA at issue.

*Assumption 1.*  $\text{rank}(\mathbf{X}_{\text{sup}}) \geq Q$ , where  $\mathbf{X}_{\text{sup}}$  contains the matrices  $\mathbf{X}_1, \dots, \mathbf{X}_I$  stacked below each other,  $\mathbf{X}_i$  ( $K_i \times J$ ) is the matrix with scores of the  $i^{\text{th}}$  subject ( $i=1, \dots, I$ ), and  $Q$  is the number of components as estimated in the SCA-P model.

*Theorem 1.* If  $\mathbf{X}_i$  ( $K_i \times J$ ), the matrix with scores of the  $i^{\text{th}}$  subject ( $i=1, \dots, I$ ), is centered column-wise, i.e.,  $\mathbf{1}'_{K_i} \mathbf{X}_i = \mathbf{0}'_J$ ,  $i=1, \dots, I$ , then, under Assumption 1, the component scores matrix  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , in the SCA-P model is centered column-wise as well, i.e.,  $\mathbf{1}'_{K_i} \mathbf{F}_i = \mathbf{0}'_Q$ ,  $i=1, \dots, I$ .

*Proof.* To fit the SCA-P model to observed data, the sum of squared residuals is minimized. The algorithm to find SCA-P estimates is essentially based on a singular value decomposition of the supermatrix  $\mathbf{X}_{\text{sup}} = \mathbf{U}_r \mathbf{\Delta}_r \mathbf{Q}_r'$ , where  $\mathbf{X}_{\text{sup}}$  contains the observed score matrices  $\mathbf{X}_1, \dots, \mathbf{X}_I$  positioned below each other, and  $\mathbf{U}_r' \mathbf{U}_r = \mathbf{Q}_r' \mathbf{Q}_r = \mathbf{Q}_r \mathbf{Q}_r' = \mathbf{I}_r$ ,  $\mathbf{\Delta}_r$  a diagonal matrix with positive diagonal elements in weakly descending order, and  $r$  the rank of  $\mathbf{X}_{\text{sup}}$ . The data matrix  $\mathbf{X}_i$  of subject  $i$ ,  $i=1, \dots, I$ , can be written as  $\mathbf{X}_i = \mathbf{U}_{ir} \mathbf{\Delta}_r \mathbf{Q}_r'$ , where  $\mathbf{U}_{ir}$  is obtained by selecting the rows of  $\mathbf{U}_r$  that correspond to the rows with the observed scores for subject  $i$  of  $\mathbf{X}_{\text{sup}}$ . From  $\mathbf{X}_i = \mathbf{U}_{ir} \mathbf{\Delta}_r \mathbf{Q}_r'$  it follows that  $\mathbf{U}_{ir} = \mathbf{X}_i \mathbf{Q}_r \mathbf{\Delta}_r^{-1}$ , which, given that  $\mathbf{1}' \mathbf{X}_i = \mathbf{0}'$ , implies that  $\mathbf{1}' \mathbf{U}_{ir} = \mathbf{0}'$ ,  $i=1, \dots, I$ , as well. The matrices  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , are obtained as the first  $Q$  columns of  $\mathbf{U}_{ir}$ , where  $Q \leq r$ , as follows from Assumption 1, and hence  $\mathbf{1}'_{K_i} \mathbf{F}_i = \mathbf{0}'_Q$ . ~

*Theorem 2.* If  $\mathbf{X}_i$  ( $K_i \times J$ ), the matrix with scores of the  $i^{\text{th}}$  subject ( $i=1, \dots, I$ ), is centered column-wise, i.e.,  $\mathbf{1}'_{K_i} \mathbf{X}_i = \mathbf{0}'_J$ ,  $i=1, \dots, I$ , then, under a mild assumption specified in the proof, the component score matrix  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , in the SCA-PF2, SCA-IND, and SCA-ECP models is centered column-wise as well, i.e.,  $\mathbf{1}'_{K_i} \mathbf{F}_i = \mathbf{0}'_Q$ ,  $i=1, \dots, I$ .

*Proof.* To fit the SCA-PF2, SCA-IND, or SCA-ECP models to observed data, the sum of squared residuals is minimized via an alternating least squares algorithm. The

algorithm to fit the SCA-PF2 model, requiring that  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i \boldsymbol{\Phi} \mathbf{D}_i$ , with  $\mathbf{D}_i$  a diagonal  $Q \times Q$  matrix and  $\boldsymbol{\Phi}$  a positive definite  $Q \times Q$  matrix with unit diagonal elements, uses an ALS algorithm for the equivalent problem of minimizing (6.3), subject to the constraint  $\mathbf{F}_i' \mathbf{F}_i = \tilde{\mathbf{D}}_i \tilde{\boldsymbol{\Phi}} \tilde{\mathbf{D}}_i$  with  $\tilde{\boldsymbol{\Phi}}$  an arbitrary positive definite  $Q \times Q$  matrix and  $\tilde{\mathbf{D}}_i$  a diagonal  $Q \times Q$  matrix. The algorithm is essentially based on the fact that every matrix  $\mathbf{F}_i$  that meets the constraint  $\mathbf{F}_i' \mathbf{F}_i = \tilde{\mathbf{D}}_i \tilde{\boldsymbol{\Phi}} \tilde{\mathbf{D}}_i$  can be written as  $\mathbf{F}_i = \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i$  provided that  $\mathbf{P}_i' \mathbf{P}_i = \mathbf{I}_Q$ ,  $i=1, \dots, I$ , and  $\tilde{\boldsymbol{\Phi}} = \tilde{\mathbf{F}}' \tilde{\mathbf{F}}$ . The matrices  $\mathbf{P}_i$ ,  $\tilde{\mathbf{F}}$ ,  $\tilde{\mathbf{D}}_i$ , and  $\mathbf{B}$  are alternately updated in the SCA-PF2 algorithm. The algorithm to fit the SCA-IND model uses the SCA-PF2 algorithm as follows: to find estimates of  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , in the SCA-IND algorithm, the SCA-PF2 algorithm is used, keeping  $\tilde{\mathbf{F}}$  fixed at identity. An update of  $\mathbf{P}_i$  in the SCA-PF2 algorithm, subject to  $\mathbf{P}_i' \mathbf{P}_i = \mathbf{I}_Q$ , can be found by first computing the singular value decomposition of  $\mathbf{X}_i \mathbf{B} \tilde{\mathbf{D}}_i \tilde{\mathbf{F}}' = \mathbf{U}_{iQ} \boldsymbol{\Delta}_{iQ} \mathbf{Q}_{iQ}'$ ,  $\mathbf{U}_{iQ}' \mathbf{U}_{iQ} = \mathbf{Q}_{iQ}' \mathbf{Q}_{iQ} = \mathbf{Q}_{iQ} \mathbf{Q}_{iQ}' = \mathbf{I}_Q$ ,  $\boldsymbol{\Delta}_{iQ}$  a diagonal matrix with nonnegative diagonal elements in weakly descending order, and then taking  $\mathbf{P}_i = \mathbf{U}_{iQ} \mathbf{Q}_{iQ}'$ , where  $Q$  is the number of components. Now assuming that  $\text{rank}(\mathbf{X}_i \mathbf{B} \tilde{\mathbf{D}}_i \tilde{\mathbf{F}}') = Q$ , which can be expected to be satisfied in practice, the equality  $\mathbf{X}_i \mathbf{B} \tilde{\mathbf{D}}_i \tilde{\mathbf{F}}' = \mathbf{U}_{iQ} \boldsymbol{\Delta}_{iQ} \mathbf{Q}_{iQ}'$  implies that  $\mathbf{U}_{iQ} = \mathbf{X}_i \mathbf{B} \tilde{\mathbf{D}}_i \tilde{\mathbf{F}}' \mathbf{Q}_{iQ} \boldsymbol{\Delta}_{iQ}^{-1}$ , which, given that  $\mathbf{1}' \mathbf{X}_i = \mathbf{0}'$ , implies that  $\mathbf{1}' \mathbf{U}_{iQ} = \mathbf{0}'$ ,  $i=1, \dots, I$ , as well. After convergence, for SCA-PF2,  $\mathbf{F}_i = \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i$ , whereas for SCA-IND  $\mathbf{F}_i = \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i$  with  $\tilde{\mathbf{F}}$  fixed at identity. Given that  $\mathbf{F}_i = \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i$  and that every update of  $\mathbf{P}_i$  is taken as  $\mathbf{P}_i = \mathbf{U}_{iQ} \mathbf{Q}_{iQ}'$ , it follows that if  $\mathbf{1}' \mathbf{X}_i = \mathbf{0}'$ , then  $\mathbf{1}' \mathbf{F}_i = \mathbf{1}' \mathbf{U}_{iQ} \mathbf{Q}_{iQ}' \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i = \mathbf{0}' \mathbf{Q}_{iQ}' \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i = \mathbf{0}'$ .

An update of  $\mathbf{F}_i$  in the SCA-ECP algorithm, subject to  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{I}_Q$ , can be found by first computing the singular value decomposition of  $\mathbf{X}_i \mathbf{B} = \mathbf{U}_{iQ} \boldsymbol{\Delta}_{iQ} \mathbf{Q}_{iQ}'$ ,  $\mathbf{U}_{iQ}' \mathbf{U}_{iQ} = \mathbf{Q}_{iQ}' \mathbf{Q}_{iQ} = \mathbf{Q}_{iQ} \mathbf{Q}_{iQ}' = \mathbf{I}_Q$ ,  $\boldsymbol{\Delta}_{iQ}$  a diagonal matrix with nonnegative diagonal elements in weakly descending order, and then taking  $\mathbf{F}_i = \sqrt{K_i} \mathbf{U}_{iQ} \mathbf{Q}_{iQ}'$ , where  $Q$  is the number of components. Assuming that  $\text{rank}(\mathbf{X}_i \mathbf{B}) = Q$ , which can be expected to be satisfied in practice, the equality  $\mathbf{X}_i \mathbf{B} = \mathbf{U}_{iQ} \boldsymbol{\Delta}_{iQ} \mathbf{Q}_{iQ}'$  implies that  $\mathbf{U}_{iQ} = \mathbf{X}_i \mathbf{B} \mathbf{Q}_{iQ} \boldsymbol{\Delta}_{iQ}^{-1}$ , which, given that  $\mathbf{1}' \mathbf{X}_i = \mathbf{0}'$ , implies that  $\mathbf{1}' \mathbf{U}_{iQ} = \mathbf{0}'$ ,  $i=1, \dots, I$ , as well. Given that  $\mathbf{F}_i$  is taken as  $\mathbf{F}_i = \sqrt{K_i} \mathbf{U}_{iQ} \mathbf{Q}_{iQ}'$ , it follows that if  $\mathbf{1}' \mathbf{X}_i = \mathbf{0}'$ , then  $\mathbf{1}' \mathbf{F}_i = \mathbf{1}' \mathbf{U}_{iQ} \mathbf{Q}_{iQ}' \sqrt{K_i} = \mathbf{0}' \mathbf{Q}_{iQ}' \sqrt{K_i} = \mathbf{0}'$ . ~



# 7. Lagged Simultaneous Component Models of Multisubject Multivariate Time Series

## 7.1. Introduction

In the previous chapter, four models for the simultaneous component analysis (SCA) of multisubject multivariate time series were discussed. The models that were discussed covered intraindividual and interindividual variability, but possible time dependencies were not explicitly modeled. Therefore, in the current chapter, the four SCA models will be elaborated so that dependencies between successive measurement occasions are modeled. As will be explained later, the models are useful for data with a special time dependent structure. If data have such a structure, the elaborated SCA models offer a sparse model, that clearly reveals the time dependencies between component scores and observed scores.

The elaborated models to be proposed here are component analytic counterparts of one of the dynamic factor model variants (Engle & Watson, 1981; Geweke & Singleton, 1981; Molenaar, 1985; Immink, 1986). In dynamic factor models, observed multivariate time series are modeled by latent time series, which are of lower dimension than the observed ones. Time-dependencies are covered in the models. Several variants of dynamic factor models have been proposed. They differ in the way the latent time series are related to observed time series, the model of the latent time series and the procedure for estimating the parameters of the model. It should be noted that, in contrast to our SCA models, dynamic factor modeling is usually applied to the multivariate time series obtained from a single subject. Because our extension of the SCA models is based on Molenaar's dynamic factor model (Molenaar, 1985), this model is described first.

## 7.2. The dynamic factor model

In the dynamic factor model as proposed by Molenaar (1985), observed scores are related to factor scores at the same time point and at previous time points. Hence, simultaneous as well as lagged effects are modeled. The factors at time point  $k$  are denoted as the factors of lag zero, whereas the factors at  $k-u$  are denoted as the factors of lag  $u$ . The model is given by

$$\mathbf{x}_k = \sum_{u=0}^U \mathbf{\Lambda}_u \boldsymbol{\eta}_{k-u} + \boldsymbol{\varepsilon}_k, \quad (7.1)$$

where  $\mathbf{x}_k$  denotes the vector that contains the scores on  $J$  variables at time point  $k$  ( $k=1, \dots, K$ ) of a single subject,  $\Lambda_u$  denotes the loading matrix for lag  $u$  ( $u=0, \dots, U$ ),  $\boldsymbol{\eta}_{k-u}$  the  $Q$  factors at time point  $k-u$ ,  $\boldsymbol{\varepsilon}_k$  the residual at time point  $k$ , and  $U \geq 0$  the maximal lag. Thus, the observed series at time point  $k$  are represented by a weighted sum of factors at time points  $k, k-1, \dots, k-U$ . It is assumed that the time series  $\mathbf{x}_k$ ,  $k=1, \dots, K$  are weakly stationary, which implies that the first and second order moments are independent of the time. Hence, the time series contains no trend, and the lagged covariances are invariant under a translation along the time axis. Because the model relates the scores to time points separated from the present time point at most by a lag  $U$ , it is called a lag  $U$  dynamic factor model.

For illustration, we represent a lag one dynamic factor model graphically in Figure 7.1. Let  $\mathbf{x}_{k-2}, \dots, \mathbf{x}_{k+1}$  be the observed scores on all variables at time points  $k-2, \dots, k+1$ , hence these are the transposed rows  $k-2, \dots, k+1$  of matrix  $\mathbf{X}$ , for a single subject;  $\boldsymbol{\eta}_{k-2}, \dots, \boldsymbol{\eta}_{k+1}$  denote the vectors of factor scores,  $\Lambda_0$  and  $\Lambda_1$  the lag zero and lag one loading matrices, respectively, and  $\boldsymbol{\varepsilon}_{k-2}, \dots, \boldsymbol{\varepsilon}_{k+1}$  the residuals at time points  $k-2, \dots, k+1$ . Then, the model is given by

$$\mathbf{x}_k = \Lambda_0 \boldsymbol{\eta}_k + \Lambda_1 \boldsymbol{\eta}_{k-1} + \boldsymbol{\varepsilon}_k, \quad (7.2)$$

which can be depicted as follows.

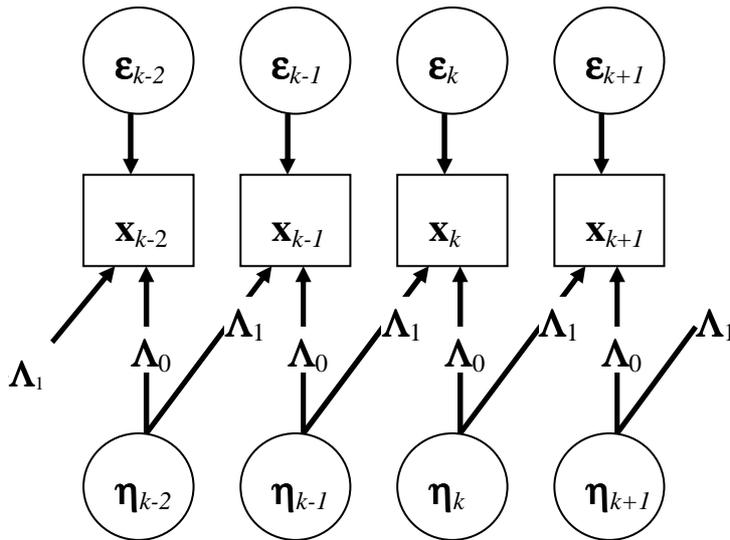


Figure 7.1. A graphical representation of a lag one dynamic factor model.

For example, it can be seen in Figure 7.1 that the estimated scores at time point  $k$  are given by the factor scores at time point  $k$  weighted by the lag zero loading matrix ( $\mathbf{\Lambda}_0$ ), and by the lag one factor scores (at  $k-1$ ) weighted by the lag one loading matrix ( $\mathbf{\Lambda}_1$ ). The dynamic factor model is suitable to model multivariate time series with a rather special structure, because, as shown above, a factor score at time point  $k$  influences the observed score not only at time point  $k$ , but also at time points  $k+1, \dots, k+U$ .

To obtain parameter estimates, Molenaar (1985) proposed fitting the model to (lagged) covariance matrices. This can be done by using pseudo-maximum likelihood or asymptotically distribution free estimation (Molenaar & Nesselrode, 1998). To obtain a structurally identified model, the (lagged) covariance matrix  $\Xi(u)$  of the  $Q$  factors in  $\eta_k$  at lag  $u, u=0,1,\dots$ , is usually constrained so that  $\Xi(u)=\delta(u)\mathbf{I}_Q$ , with  $\delta(u)=1$  if  $u=0$ , and  $\delta(u)=0$ , otherwise. Hence, in estimating the parameters of the model, the factor scores are assumed to be generated by a white noise process. All time dependencies in the data are modeled via the (lagged) loading matrices. Molenaar (1985) showed that the model specified this way can be transformed into a model with correlated factor scores, as long as an infinite number of lags is taken. Under the assumption of stationarity, the number of lags may be truncated in fitting the model, without altering the model estimates.

As an example of a dynamic factor model, one could think of a study of the level of anxiety over time in a single subject. Anxiety is measured by a series of psychological measures (e.g., a questionnaire) and a series of physiological measures (e.g., heart rate), both of which are indicative of anxiety. Suppose that the scores on the physiological measures at the same time point are highly correlated with each other, the scores on the psychological measures at the same time point are also highly correlated with each other, but that the psychological and the physiological measures at the same time point correlate zero. Suppose also that the physiological scores at any time point  $k$  are only highly correlated to the psychological scores at  $k+1$ . A lag zero dynamic factor model would reveal two factors, namely a physiological and a psychological anxiety factor. A lag one dynamic factor model would reveal just one factor, that can be labeled as anxiety, and the relationship between the ‘physiological anxiety’ variables at  $k$  and the ‘psychological anxiety’ variables at  $k+1$  is made clear via their direct and lag one relations, respectively, to this single factor. In this example, the model shows that the feeling of anxiety at time point  $k$  follows a high level of ‘physiological anxiety’ at  $k-1$ .

### 7.3. Lagged SCA models

In this section, we propose an extension of the SCA models, which will be called ‘lagged SCA models’. They are component analytic counterparts of Molenaar’s dynamic factor model (1985) for multisubject multivariate time series. The models cover interindividual and intraindividual differences, as well as lagged effects.

Component analytic counterparts of dynamic factor models that cover lagged effects have already been proposed by Brillinger (1975) and Van Buuren (1990). Brillinger (1975) defined a principal component model for multivariate time series, in which simultaneous as well as lagged effects are modeled, although he did not use the term ‘lag’. Van Buuren (1990) proposed the ‘canonical class model’, initially for multivariate time series. In modeling multisubject multivariate time series, Van Buuren assumes a *common* series of component scores for all subjects, whereas the loading matrices are allowed to differ across subjects. This model appears particularly useful for modeling time series data from a number of observers on a single subject, rather than for modeling multivariate time series of a number of subjects, as is considered in the SCA models. Unfortunately, a number of aspects of the modeling procedure remain unclear in the canonical class model, including the transformational freedom of the model parameters, the desirability to fix certain matrices, and the applicability of the model.

The lagged SCA (LSCA) models to be proposed here differ from the SCA models defined in Chapter 6 in that not only simultaneous, but also lagged effects of component scores on the observed scores are modeled in the same way as in the dynamic factor model (7.1). That is, in LSCA, a component score at time point  $t_k$  influences the observed score at  $t_k$  via the loading matrix, and the observed scores at  $t_{k+1}, t_{k+2}, \dots$  via lagged loading matrices.

In the LSCA models, it is assumed that the lagged loading matrices are subject and time invariant. This implies that the observed time series are related to the latent time series similarly across subjects and across time. This assumption is unlikely to be met if the intervals between measurement occasions differ within time series, or across subjects. Note that the interval between measurement occasions does not matter in the SCA models because only simultaneous effects are modeled in the SCA models.

The LSCA models will be introduced in the next sections. Alternating least squares algorithms to fit the models to data, and transformational freedom in each of the models are discussed. The results of a small simulation study performed to examine some estimation properties of the fitting procedures are presented. We briefly discuss the attempts to fit LSCA models to empirical data. The fitted models appeared to be difficult to interpret. Possible causes of interpretation difficulties and some ideas for possible solutions will be discussed in the discussion section.

### 7.3.1. LSCA-P

Before performing an LSCA, it is often needed to preprocess the observed scores. The considerations and the steps to be taken are the same as in SCA modeling, and therefore we refer to Section 6.2.1. In the sequel, we assume the observed scores to be preprocessed properly.

Let  $\mathbf{X}_i$  ( $K_i \times J$ ) denote the matrix of (preprocessed) scores of the  $i^{\text{th}}$  subject ( $i=1, \dots, I$ ) on  $J$  variables measured at  $K_i$  equidistant occasions ( $k_i=1, \dots, K_i$ ). The LSCA-P model is given by

$$\mathbf{X}_i = \sum_{u=0}^U \mathbf{S}_u^i \mathbf{F}_i \mathbf{B}_u' + \mathbf{E}_i, \quad (7.3)$$

where  $\mathbf{S}_u^i$  ( $K_i \times (K_i + U)$ ) denotes the so-called shift matrix (see below) of subject  $i$ ,  $i=1, \dots, I$ , for lag  $u$  ( $u=0, \dots, U$ ),  $\mathbf{F}_i$  ( $(K_i + U) \times Q$ ) denotes the  $Q$  component scores of subject  $i$  at time points  $-U + 1, \dots, K_i$ ,  $\mathbf{S}_u^i \mathbf{F}_i$  contains the  $Q$  component scores of subject  $i$  at time points  $1-u, \dots, K_i-u$  (i.e., at lag  $u$ ),  $\mathbf{B}_u$  ( $J \times Q$ ) denotes the loading matrix for lag  $u$ , and  $\mathbf{E}_i$  ( $K_i \times J$ ) denotes the matrix of residuals. To keep the notation as simple as possible, the  $K_i + U$  rows of matrix  $\mathbf{F}_i$  are indexed as rows  $1-U, \dots, K_i$ , rather than rows  $1, \dots, K_i + U$ , because, when indexed this way, row  $k_i$  of  $\mathbf{F}_i$  contains the component scores at time point  $k_i$ . The component scores matrix  $\mathbf{F}_i$  is unconstrained, implying that cross-products within and between the components may vary across subjects.

The shift matrix  $\mathbf{S}_u^i$  is defined as  $[\mathbf{0}_{K_i \times (U-u)} \mid \mathbf{I}_{K_i} \mid \mathbf{0}_{K_i \times u}]$ , where  $\mathbf{0}_{K_i \times U}$  denotes a  $K_i \times U$  zero matrix, and  $\mathbf{I}_{K_i}$  denotes the  $K_i \times K_i$  identity matrix. Premultiplying  $\mathbf{F}_i$  by the shift matrix  $\mathbf{S}_u^i$  gives the component scores of subject  $i$  at lag  $u$  (i.e., at time points  $1-u, \dots, K_i-u$ ). Hence, for example, the shift matrix  $\mathbf{S}_0^i$  is defined as  $[\mathbf{0}_{K_i \times U} \mid \mathbf{I}_{K_i}]$ , and  $\mathbf{S}_0^i \mathbf{F}_i$  is a matrix composed of the component scores of subject  $i$  at time points  $1, \dots, K_i$ , and, as a result,  $\mathbf{S}_0^i \mathbf{F}_i$  contains rows  $1, \dots, K_i$  of  $\mathbf{F}_i$ .

A lag  $U$  LSCA-P has  $U+1$  lagged loading matrices, namely  $\mathbf{B}_0, \dots, \mathbf{B}_U$ . An observed score at a certain time point is influenced by the component scores at that particular time point (via  $\mathbf{B}_0$ ), but also by component scores at  $U$  previous time points (via  $\mathbf{B}_1, \dots, \mathbf{B}_U$ ). Thus, simultaneous as well as lagged effects are modeled. The SCA-P model is a special case of the LSCA-P model, namely the case for which the number of lags is zero.

### 7.3.2. LSCA-PF2, LSCA-IND and LSCA-ECP

Analogously to the constrained version of the SCA-P model, we can define constrained versions of the LSCA-P model, namely the LSCA-PF2, LSCA-IND and LSCA-ECP models. They are defined analogously to the three constrained versions of the SCA-P model, that is by imposing particular constraints on the component scores.

The LSCA-PF2 model is given by (7.3), with  $\mathbf{F}_i$  constrained to  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i \Phi \mathbf{D}_i$ ,  $i=1, \dots, I$ , with  $\mathbf{D}_i$  a diagonal  $Q \times Q$  matrix and  $\Phi$  a positive definite  $Q \times Q$  matrix with

unit diagonal elements. This implies that the congruence coefficients (Tucker, 1951) between columns of  $\mathbf{F}_i$  are invariant over  $i$ .

The LSCA-IND model is given by (7.3) with  $\mathbf{F}_i$  constrained to  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i^2$ ,  $i=1, \dots, I$ , with  $\mathbf{D}_i$  a diagonal  $Q \times Q$  matrix. Thus, the inner products of the components are zero and the sum of squares of the components may differ across subjects in the LSCA-IND model.

The LSCA-ECP model is given by (7.3) with  $\mathbf{F}_i$  constrained so that  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \Phi$ ,  $i=1, \dots, I$ , implying that the sums of squares of the components and the inner products between the components are equal for all subjects.

Just as in the SCA model, the four LSCA models require that  $\left( \sum_{i=1}^I K_i \right)^{-1} \sum_{i=1}^I \text{diag}(\mathbf{F}_i' \mathbf{F}_i) = \mathbf{I}_Q$ , to identify the solution partly.

The constraints on the component scores in the four lagged models can be interpreted as in the SCA models. It is advised, that the scores per variable and per subject are centered, as one of the preprocessing steps (see Section 6.2.1). As a result, in the SCA models, the average component scores over occasions per subject are zero. This is a nice property, because then the different restrictions on the component score matrices in each of the four models can be interpreted directly in terms of different restrictions on the *covariances* between components. This property is lost if *lagged SCA* models (with  $U > 0$ ) are estimated. However, if the number of lags is small compared to the number of observations, the average component scores per subject per component will be close to zero in practice.

### 7.3.3. Fitting the LSCA models to data

As with fitting the SCA models, we fit the LSCA models to observed data by minimizing the sum of squared residuals. Hence, we minimize

$$F(\mathbf{F}_i, \mathbf{B}_0, \dots, \mathbf{B}_U) = \sum_{i=1}^I \left\| \mathbf{X}_i - \sum_{u=0}^U \mathbf{S}_u^i \mathbf{F}_i \mathbf{B}_u' \right\|^2, \quad (7.4)$$

subject to the constraints imposed in the particular LSCA. An Alternating Least Squares (ALS) algorithm is used to minimize this function.

#### 7.3.3.1. Fitting the LSCA-P model to data

The LSCA-P model can be fitted to data by minimizing (7.4) alternatingly over  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , and  $\mathbf{B}_u$ ,  $u=0, \dots, U$ . Let  $\mathbf{F}_i^S$  denote a supermatrix containing the component scores of lag zero to lag  $U$  of subject  $i$  ( $i=1, \dots, I$ ),  $\mathbf{F}_i^S = [\mathbf{S}_0^i \mathbf{F}_i | \dots | \mathbf{S}_U^i \mathbf{F}_i]$ , and let  $\mathbf{B}^S$  denote a supermatrix containing the loading matrices of lag zero to lag  $U$  positioned

next to each other,  $\mathbf{B}^S = [\mathbf{B}_0 | \dots | \mathbf{B}_U]$ . Let  $\mathbf{X}^S$  denote a supermatrix containing the matrices  $\mathbf{X}_1, \dots, \mathbf{X}_I$  with the observed scores of subjects  $1, \dots, I$ , stacked below each other, and let  $\mathbf{F}^S$  denote a supermatrix containing the matrices  $\mathbf{F}_1^S, \dots, \mathbf{F}_I^S$ , positioned below each other.

To update  $\mathbf{B}_0, \dots, \mathbf{B}_U$ , we minimize (7.4) considering  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , fixed. This comes down to minimizing

$$f_1(\mathbf{B}^S) = \|\mathbf{X}^S - \mathbf{F}^S \mathbf{B}^S\|^2. \quad (7.5)$$

This is a multiple regression problem, which is solved by taking  $\mathbf{B}^S = (\mathbf{F}^S \mathbf{F}^S)^{-1} \mathbf{F}^S \mathbf{X}^S$ . The matrix  $\mathbf{B}^S = [\mathbf{B}_0, \dots, \mathbf{B}_U]$ , so that updates for  $\mathbf{B}_0, \dots, \mathbf{B}_U$  can now easily be obtained.

An update for  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , can be obtained by minimizing

$$f_2(\mathbf{F}_i) = \left\| \mathbf{X}_i - \sum_{u=0}^U \mathbf{S}_u^i \mathbf{F}_i \mathbf{B}_u \right\|^2 = \left\| \text{Vec}(\mathbf{X}_i) - \sum_{u=0}^U (\mathbf{B}_u \otimes \mathbf{S}_u^i) \text{Vec}(\mathbf{F}_i) \right\|^2, \quad (7.6)$$

where  $\text{Vec}(\mathbf{X})$  denotes the vector containing all the elements of the matrix  $\mathbf{X}$  strung out column-wise into a column vector. Minimizing (7.6) over  $\text{Vec}(\mathbf{F}_i)$  is an ordinary regression problem, solved by taking

$\text{Vec}(\mathbf{F}_i) = \left( \sum_{u=0}^U (\mathbf{B}_u \otimes \mathbf{S}_u^i) \sum_{u=0}^U (\mathbf{B}_u \otimes \mathbf{S}_u^i)' \right)^{-1} \sum_{u=0}^U (\mathbf{B}_u \otimes \mathbf{S}_u^i)' \text{Vec}(\mathbf{X}_i)$ . By arranging the elements of  $\text{Vec}(\mathbf{F}_i)$  back into a matrix, the update for  $\mathbf{F}_i$  is obtained.

### 7.3.3.2. Fitting the LSCA-PF2 model to data

The LSCA-PF2 model can be fitted to data by minimizing (7.4) alternatingly over  $\mathbf{B}_u$ ,  $u=0, \dots, U$ , and  $\mathbf{F}_i$ , subject to  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i \mathbf{\Phi} \mathbf{D}_i$ ,  $i=1, \dots, I$ , where  $\mathbf{D}_i$ ,  $i=1, \dots, I$ , is a diagonal  $Q \times Q$  matrix and  $\mathbf{\Phi}$  a positive definite  $Q \times Q$  matrix with unit diagonal elements. We solve the equivalent minimization problem with  $\mathbf{F}_i$  subject to  $\mathbf{F}_i' \mathbf{F}_i = \tilde{\mathbf{D}}_i \tilde{\mathbf{\Phi}} \tilde{\mathbf{D}}_i$ , with  $\tilde{\mathbf{D}}_i$  a diagonal  $Q \times Q$  matrix and  $\tilde{\mathbf{\Phi}}$  an arbitrary positive definite  $Q \times Q$  matrix first, then discuss how to obtain  $\mathbf{F}_i$ , subject to  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i \mathbf{\Phi} \mathbf{D}_i$  from the obtained solutions for  $\tilde{\mathbf{D}}_i$  and  $\tilde{\mathbf{\Phi}}$ . As showed by Kiers, Ten Berge and Bro (1999), every matrix  $\mathbf{F}_i$  that meets the constraint  $\mathbf{F}_i' \mathbf{F}_i = \tilde{\mathbf{D}}_i \tilde{\mathbf{\Phi}} \tilde{\mathbf{D}}_i$  can be written as  $\mathbf{F}_i = \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i$ , provided that  $\mathbf{P}_i' \mathbf{P}_i = \mathbf{I}_Q$ ,  $\tilde{\mathbf{F}}$  is an arbitrary  $Q \times Q$  matrix, and  $\tilde{\mathbf{D}}_i$  a diagonal  $Q \times Q$  matrix,  $i=1, \dots, I$ . The problem to be solved boils down to minimizing

$$f_3(\mathbf{P}_i, \tilde{\mathbf{F}}, \tilde{\mathbf{D}}_i, \mathbf{B}_0, \dots, \mathbf{B}_U) = \sum_{i=1}^I \left\| \mathbf{X}_i - \sum_{u=0}^U \mathbf{S}_u^i \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i \mathbf{B}_u' \right\|^2, \quad (7.7)$$

$i=1, \dots, I$ , subject to  $\mathbf{P}_i' \mathbf{P}_i = \mathbf{I}_Q$ , and  $\tilde{\mathbf{D}}_i$  a diagonal matrix. The function in (7.7) can be minimized by updating  $\mathbf{B}_0, \dots, \mathbf{B}_U$ ,  $\mathbf{P}_i$ ,  $\tilde{\mathbf{F}}$ , and  $\tilde{\mathbf{D}}_i$  alternately. The problem of minimizing (7.7) over  $\mathbf{B}_0, \dots, \mathbf{B}_U$  is the same as minimizing (7.4) over  $\mathbf{B}_0, \dots, \mathbf{B}_U$ . The solution to this problem has been treated in Section 7.3.3.1. .

The next problem is to find, for every value of  $i$ , an update of  $\mathbf{P}_i$ , where  $\mathbf{P}_i$  is constrained to  $\mathbf{P}_i' \mathbf{P}_i = \mathbf{I}_Q$ . An update for  $\mathbf{P}_i$  can be obtained by minimizing (7.7) over  $\mathbf{P}_i$ , subject to  $\mathbf{P}_i' \mathbf{P}_i = \mathbf{I}_Q$ , which can be reformulated as minimizing the function

$$f_4(\mathbf{P}_i) = \text{tr} \mathbf{X}_i' \mathbf{X}_i + \text{tr} \left( \left( -2 \sum_{u=0}^U \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i \mathbf{B}_u' \mathbf{X}_i \mathbf{S}_u^i \right) \mathbf{P}_i \right) + \text{tr} \left( \sum_{u=0}^U \sum_{v=0}^U \mathbf{S}_u^i{}' \mathbf{S}_v^i \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i \mathbf{B}_v' \mathbf{B}_u \tilde{\mathbf{D}}_i \tilde{\mathbf{F}}' \mathbf{P}_i' \right). \quad (7.8)$$

An update of  $\mathbf{P}_i$  can be obtained by majorization (Kiers, 1990). Consider the singular value decomposition

$$\mathbf{P}_i - \frac{1}{2 \sum_{u=0}^U \sum_{v=0}^U \alpha_{uv}} \left( \left( -2 \sum_{u=0}^U \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i \mathbf{B}_u' \mathbf{X}_i \mathbf{S}_u^i \right)' + \sum_{u=0}^U \sum_{v=0}^U \mathbf{S}_u^i{}' \mathbf{S}_v^i \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i \mathbf{B}_v' \mathbf{B}_u \tilde{\mathbf{D}}_i \tilde{\mathbf{F}}' + \sum_{u=0}^U \sum_{v=0}^U \mathbf{S}_v^i{}' \mathbf{S}_u^i \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i \mathbf{B}_u' \mathbf{B}_v \tilde{\mathbf{D}}_i \tilde{\mathbf{F}}' \right) = \mathbf{U}_i \mathbf{\Delta}_i \mathbf{Q}_i', \quad (7.9)$$

with  $\mathbf{U}_i' \mathbf{U}_i = \mathbf{Q}_i' \mathbf{Q}_i = \mathbf{Q}_i \mathbf{Q}_i' = \mathbf{I}_Q$ , and  $\mathbf{\Delta}_i$  a diagonal matrix with nonnegative diagonal elements in weakly descending order;  $\alpha_{uv}$  has to be chosen as a scalar greater than or equal to the largest eigenvalue of the symmetric part of  $(\mathbf{S}_u^i{}' \mathbf{S}_v^i \otimes \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i \mathbf{B}_u' \mathbf{B}_v \tilde{\mathbf{D}}_i \tilde{\mathbf{F}}')$ . Now, the update of  $\mathbf{P}_i$  can be obtained as  $\mathbf{U}_i \mathbf{Q}_i'$ .

An update of  $\tilde{\mathbf{D}}_i$ , for every  $i$ , is obtained by minimizing (7.7) over  $\tilde{\mathbf{D}}_i$ , which comes down to minimizing

$$f_5(\mathbf{d}_i) = \left\| \text{Vec}(\mathbf{X}_i) - \sum_{u=0}^U (\mathbf{B}_u \vee \mathbf{S}_u^i \mathbf{P}_i \tilde{\mathbf{F}}) \mathbf{d}_i \right\|^2, \quad (7.10)$$

where  $\mathbf{d}_i$  denotes the vector containing the diagonal elements of  $\tilde{\mathbf{D}}_i$ , and  $\mathfrak{v}$  denotes the Khatri-Rao product (Rao & Mitra, 1971; see also Section 2.2). Minimizing (7.10) over  $\mathbf{d}_i$  is an ordinary regression problem. The minimum is obtained by taking

$$\mathbf{d}_i = \left( \sum_{u=0}^U (\mathbf{B}_u \mathfrak{v} \mathbf{S}_u^i \mathbf{P}_i \tilde{\mathbf{F}})' \sum_{u=0}^U (\mathbf{B}_u \mathfrak{v} \mathbf{S}_u^i \mathbf{P}_i \tilde{\mathbf{F}}) \right)^{-1} \sum_{u=0}^U (\mathbf{B}_u \mathfrak{v} \mathbf{S}_u^i \mathbf{P}_i \tilde{\mathbf{F}})' \text{Vec}(\mathbf{X}_i). \quad (7.11)$$

By positioning the elements of the optimal  $\mathbf{d}_i$  on the diagonal of a diagonal  $Q \times Q$  matrix, an update for  $\tilde{\mathbf{D}}_i$  is obtained.

The update of  $\tilde{\mathbf{F}}$  is obtained by minimizing (7.7) over  $\tilde{\mathbf{F}}$  ( $Q \times Q$ ), which comes down to minimizing

$$f_6(\tilde{\mathbf{F}}) = \sum_{i=1}^I \left\| \text{Vec}(\mathbf{X}_i) - \sum_{u=0}^U (\mathbf{B}_u \tilde{\mathbf{D}}_i \otimes \mathbf{S}_u^i \mathbf{P}_i) \text{Vec}(\tilde{\mathbf{F}}) \right\|^2. \quad (7.12)$$

Let  $\mathbf{x}^S$  denote a supervector containing the vectors  $\text{Vec}(\mathbf{X}_1), \dots, \text{Vec}(\mathbf{X}_I)$  with the observed scores of subjects  $1, \dots, I$ , positioned below each other, and let  $\mathbf{V}^S$  denote a supermatrix containing the matrices  $\left( \sum_{u=0}^U (\mathbf{B}_u \tilde{\mathbf{D}}_1 \otimes \mathbf{S}_u^1 \mathbf{P}_1) \right), \dots, \left( \sum_{u=0}^U (\mathbf{B}_u \tilde{\mathbf{D}}_I \otimes \mathbf{S}_u^I \mathbf{P}_I) \right)$  positioned below each other. Now, minimizing (7.12) boils down to a regression problem that can be solved by taking  $\text{Vec}(\tilde{\mathbf{F}}) = (\mathbf{V}^S \mathbf{V}^S)^{-1} \mathbf{V}^S \mathbf{x}^S$ . By arranging the elements of  $\text{Vec}(\tilde{\mathbf{F}})$  back into a matrix, the update for  $\tilde{\mathbf{F}}$  is obtained, and  $\mathbf{F}_i = \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i$  now gives the solution for  $\mathbf{F}_i$ ,  $i=1, \dots, I$ .

After convergence, solutions for  $\mathbf{D}_i$  and  $\Phi$  so that  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i \Phi \mathbf{D}_i$ , with  $\Phi$  unit diagonal elements, can be obtained by taking  $\Phi = \text{diag}(\tilde{\Phi})^{\frac{1}{2}} \tilde{\Phi} \text{diag}(\tilde{\Phi})^{\frac{1}{2}}$ , and  $\mathbf{D}_i = \frac{1}{\sqrt{K_i}} \tilde{\mathbf{D}}_i \text{diag}(\tilde{\Phi})^{\frac{1}{2}}$ , where  $\tilde{\Phi} = \tilde{\mathbf{F}}' \tilde{\mathbf{F}}$ .

### 7.3.3.3. Fitting the LSCA-IND model to data

The LSCA-IND model can be fitted to data by minimizing (7.4) alternately over  $\mathbf{B}_u$ ,  $u=0, \dots, U$ , and  $\mathbf{F}_i$ , subject to  $\frac{1}{\sqrt{K_i}} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i^2$  with  $\mathbf{D}_i$ ,  $i=1, \dots, I$ , a diagonal  $Q \times Q$  matrix.

The ALS algorithm to find estimates of the parameters of the LSCA-PF2 model can be used to find estimates of  $\mathbf{F}_i$ ,  $i=1, \dots, I$  subject to  $\mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i^2$ , by keeping  $\tilde{\Phi} = \mathbf{I}$ . In the LSCA-PF2 algorithm this is obtained by keeping  $\tilde{\mathbf{F}}$  fixed as  $\tilde{\mathbf{F}} = \mathbf{I}$ , and only updating  $\mathbf{B}_u$ ,  $\mathbf{P}_i$ , and  $\tilde{\mathbf{D}}_i$ . With this algorithm we find solutions for  $\mathbf{B}_u$ ,  $\mathbf{P}_i$  and  $\tilde{\mathbf{D}}_i$ , and hence

$\mathbf{F}_i = \mathbf{P}_i \tilde{\mathbf{F}} \tilde{\mathbf{D}}_i = \mathbf{P}_i \mathbf{F} \tilde{\mathbf{D}}_i$ . Solutions for  $\mathbf{D}_i$  so that  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{D}_i^2$  can be obtained by taking  $\mathbf{D}_i = \frac{1}{\sqrt{K_i}} \tilde{\mathbf{D}}_i$ .

### 7.3.3.4. Fitting the LSCA-ECP model to data

The LSCA-ECP model can be fitted to data by minimizing (7.4) alternately over  $\mathbf{B}_u$ ,  $u=0, \dots, U$ , and  $\mathbf{F}_i$ , subject to  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{M}$ ,  $i=1, \dots, I$ , which is equivalent to (i.e., without affecting the model fit) imposing the constraint that  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{I}_Q$ . Updating  $\mathbf{B}_0, \dots, \mathbf{B}_U$ , and  $\mathbf{F}_i$  alternately can solve this problem. The problem of finding an update for  $\mathbf{B}_0, \dots, \mathbf{B}_U$  is analogous to finding an update of  $\mathbf{B}_0, \dots, \mathbf{B}_U$  in the LSCA-PF2 algorithm. The next problem is to find an update for  $\mathbf{F}_i$  subject to  $\frac{1}{K_i} \mathbf{F}_i' \mathbf{F}_i = \mathbf{I}_Q$ . Upon substitution of  $\tilde{\mathbf{F}}_i = \frac{1}{\sqrt{K_i}} \mathbf{F}_i$ , this is equivalent to finding an update for  $\tilde{\mathbf{F}}_i$  subject  $\tilde{\mathbf{F}}_i' \tilde{\mathbf{F}}_i = \mathbf{I}$ . An update of  $\tilde{\mathbf{F}}_i$  can be obtained by majorization (Kiers, 1990). Consider the singular value decomposition

$$\tilde{\mathbf{F}}_i = \frac{1}{2 \sum_{u=0}^U \sum_{v=0}^U \alpha_{uv}} \left( \left( -2 \sum_{u=0}^U \mathbf{B}'_u \mathbf{X}'_i \mathbf{S}_u^i \right)' + \sum_{u=0}^U \sum_{v=0}^U \mathbf{S}_u^i{}' \mathbf{S}_v^i \tilde{\mathbf{F}}_i \mathbf{B}'_v \mathbf{B}_u + \sum_{u=0}^U \sum_{v=0}^U \mathbf{S}_v^i{}' \mathbf{S}_u^i \tilde{\mathbf{F}}_i \mathbf{B}'_u \mathbf{B}_v \right) = \mathbf{U}_i \mathbf{\Delta}_i \mathbf{Q}'_i, \quad (7.13)$$

with  $\mathbf{U}'_i \mathbf{U}_i = \mathbf{Q}'_i \mathbf{Q}_i = \mathbf{Q}_i \mathbf{Q}'_i = \mathbf{I}_Q$ , and  $\mathbf{\Delta}_i$  a diagonal matrix with nonnegative diagonal elements in weakly descending order;  $\alpha_{uv}$  has to be chosen as a scalar greater than or equal to the largest eigenvalue of the symmetric part of  $(\mathbf{S}_u^i{}' \mathbf{S}_v^i \otimes \mathbf{B}'_u \mathbf{B}_v)$ . Now, the update of  $\tilde{\mathbf{F}}_i$  can be obtained as  $\mathbf{U}_i \mathbf{Q}'_i$ , hence an update of  $\mathbf{F}_i$  is given by  $\mathbf{F}_i = \sqrt{K_i} \mathbf{U}_i \mathbf{Q}'_i$ .

### 7.3.3.5. Starting values of the parameters

Alternating least squares algorithms have to be initialized with certain starting values, chosen randomly or rationally. For rational starting values of  $\mathbf{B}_0$ , and, if applicable, of  $\mathbf{D}_i$  and  $\mathbf{F}$ , we advise taking the rational starting values for  $\mathbf{B}$ ,  $\mathbf{D}_i$  and  $\mathbf{F}$ , respectively, as discussed in Section 6.3.4. The starting matrices of the loading matrices  $\mathbf{B}_1$  through  $\mathbf{B}_U$  are set to zero. These starting values suffice to start the iterative process (by updating  $\mathbf{P}_i$  or  $\mathbf{F}_i$ ).

### 7.3.4. Transformational freedom in the LSCA models

In this section, some results will be presented concerning the transformational freedom in each of the four LSCA models. The transformational freedom of the solutions of LSCA-P and LSCA-ECP models will be shown to be fairly constrained provided that certain, usually mild, assumptions hold. Under some assumptions, the LSCA-PF2 and LSCA-IND solutions are even unique up to permutation, rescaling and sign permutation of the parameter matrices.

First, the transformational freedom in the LSCA-P model will be examined. The transformational freedom in this model will be shown to be fairly constrained under the conditions specified below. Specifically, postmultiplication of  $\mathbf{F}_i$  by any non-singular matrix  $\mathbf{T}_0$ , and compensation for this by applying the inverse of  $\mathbf{T}_0$  to  $\mathbf{B}_0, \dots, \mathbf{B}_U$  is the *only* possible transformation without changing the estimates of  $\mathbf{X}_i$ ,  $i=1, \dots, I$ , in the LSCA-P.

The estimates for the LSCA-P can be written as

$$\hat{\mathbf{X}}_i = \mathbf{F}_i^{\mathbf{S}} \mathbf{B}^{\mathbf{S}'}, \quad (7.14)$$

where  $\hat{\mathbf{X}}_i$  denotes the matrix of estimated scores of the  $i^{\text{th}}$  ( $i=1, \dots, I$ ) subject on  $J$  variables ( $j=1, \dots, J$ ) measured at  $K_i$  occasions ( $k=1, \dots, K_i$ );  $\mathbf{F}_i^{\mathbf{S}} = [\mathbf{S}_0^i \mathbf{F}_i | \dots | \mathbf{S}_U^i \mathbf{F}_i]$  is a supermatrix containing the component scores of lag zero to lag  $U$  of subject  $i$  positioned next to each other, and  $\mathbf{B}^{\mathbf{S}} = [\mathbf{B}_0 | \dots | \mathbf{B}_U]$  is a supermatrix containing the loading matrices of lag zero to lag  $U$  positioned next to each other. Remember that the matrix  $\mathbf{F}_i$  has  $K_i + U$  rows, denoted as row  $1-U$  to row  $K_i$ , and that row  $k_i$  of  $\mathbf{F}_i$  therefore contains the component scores of subject  $i$  at time point  $k_i$ . Hence, the supermatrix  $\mathbf{F}_i^{\mathbf{S}}$  can also be written as  $\mathbf{F}_i^{\mathbf{S}} = [\mathbf{F}_{i(1, \dots, K_i)} | \dots | \mathbf{F}_{i(1-U, \dots, K_i-U)}]$ , where  $\mathbf{F}_{i(1, \dots, K_i)}$  denotes the submatrix containing rows  $1, \dots, K_i$  of  $\mathbf{F}_i$ , and where the other submatrices are defined analogously.

Throughout Section 7.3.4, the following assumptions are made:

Assumption 1. There is at least one particular subject  $s$  for whom the columns of the matrices  $\mathbf{F}_{s(1-U, \dots, K_i-2U)}, \dots, \mathbf{F}_{s(U+1, \dots, K_i)}$  are linearly independent.

Assumption 2. The supermatrix  $\mathbf{B}^{\mathbf{S}}$  is of full column rank, namely of rank  $Q(U+1)$ .

Assumption 1, linear independence of the  $(2U+1)Q$  columns of the matrices  $\mathbf{F}_{s(1-U, \dots, K_i-2U)}, \dots, \mathbf{F}_{s(U+1, \dots, K_i)}$  implies that  $(2U+1)Q \leq K_i - U$ , because each such column has  $K_i - U$  elements. Thus, the number of occasions must be sufficiently large in comparison to the number of components and the maximal lag.

*Theorem 1.* Suppose that  $\hat{\mathbf{X}}_i = \mathbf{F}_i^S \mathbf{B}^{S'} = \tilde{\mathbf{F}}_i^S \tilde{\mathbf{B}}^{S'}$ , with  $\mathbf{F}_i^S \mathbf{B}^{S'} = [\mathbf{S}_0^i \mathbf{F}_i | \dots | \mathbf{S}_U^i \mathbf{F}_i]$  and  $\mathbf{B}^S = [\mathbf{B}_0 | \dots | \mathbf{B}_U]$ , and that  $\tilde{\mathbf{F}}_i^S$  and  $\tilde{\mathbf{B}}^S$  are defined analogously. Then, provided that Assumptions 1 and 2 are satisfied,  $\tilde{\mathbf{F}}_i^S = \mathbf{F}_i^S \mathbf{T}_0$ ,  $i=1, \dots, I$ , and  $\tilde{\mathbf{B}}_u^S = \mathbf{B}_u (\mathbf{T}_0^{-1})'$ ,  $u=0, \dots, U$ , for any non-singular matrix  $\mathbf{T}_0$ .

*Proof.* It follows from Assumption 1, that the supermatrix  $\mathbf{F}_s^S$  is of full column rank, namely of rank  $Q(U+1)$ . From

$$\hat{\mathbf{X}}_s = \mathbf{F}_s^S \mathbf{B}^{S'} = \tilde{\mathbf{F}}_s^S \tilde{\mathbf{B}}^{S'} \quad (7.15)$$

and the full column rank of  $\mathbf{F}_s^S$  and  $\mathbf{B}^S$ , it follows that the rank of  $\tilde{\mathbf{F}}_s^S \tilde{\mathbf{B}}^{S'}$  is  $Q(U+1)$ , and hence the column rank of  $\tilde{\mathbf{F}}_s^S$  as well as  $\tilde{\mathbf{B}}^S$  is full ( $Q(U+1)$ ). From (7.15), and from the full column rank of  $\mathbf{B}^S$ , it follows that  $\mathbf{F}_s^S = \tilde{\mathbf{F}}_s^S \tilde{\mathbf{B}}^{S'} \mathbf{B}^S (\mathbf{B}^{S'} \mathbf{B}^S)^{-1}$ . From the full column rank of  $\mathbf{F}_s^S$ , it follows that  $\tilde{\mathbf{B}}^{S'} \mathbf{B}^S (\mathbf{B}^{S'} \mathbf{B}^S)^{-1}$  is non-singular. Upon denoting  $\tilde{\mathbf{B}}^{S'} \mathbf{B}^S (\mathbf{B}^{S'} \mathbf{B}^S)^{-1} \equiv \mathbf{T}^{-1}$ , we have  $\tilde{\mathbf{F}}_s^S = \mathbf{F}_s^S \mathbf{T}$ , where  $\mathbf{T}$  is called a transformation matrix. Equation (7.15) boils down to  $\mathbf{F}_s^S \mathbf{B}^{S'} = \tilde{\mathbf{F}}_s^S \tilde{\mathbf{B}}^{S'} = \mathbf{F}_s^S \mathbf{T} \tilde{\mathbf{B}}^{S'}$ . Because  $\mathbf{F}_s^S$  has full column rank,  $\mathbf{F}_s^{S'} \mathbf{F}_s^S$  is non-singular. Premultiplication of both sides of  $\mathbf{F}_s^S \mathbf{B}^{S'} = \mathbf{F}_s^S \mathbf{T} \tilde{\mathbf{B}}^{S'}$  by  $(\mathbf{F}_s^{S'} \mathbf{F}_s^S)^{-1} \mathbf{F}_s^{S'}$  gives  $\tilde{\mathbf{B}}^S = \mathbf{B}^S (\mathbf{T}^{-1})'$ . The transformation matrix  $\mathbf{T}$  that transforms  $\mathbf{B}^S$  into  $\tilde{\mathbf{B}}^S$  by  $\tilde{\mathbf{B}}^S = \mathbf{B}^S (\mathbf{T}^{-1})'$  must be equal for all subjects  $i$ ,  $i=1, \dots, I$ . Thus,  $\mathbf{F}_i^S$  is transformed into  $\tilde{\mathbf{F}}_i^S$  by  $\tilde{\mathbf{F}}_i^S = \mathbf{F}_i^S \mathbf{T}$ , where  $\mathbf{T}$  is the transformation matrix, that is equal for all subjects  $i$ ,  $i=1, \dots, I$ .

From now on, the subscript  $i$  will be omitted for convenience, as the presented transformation results hold for all subjects  $i=1, \dots, I$ . The transformation matrix  $\mathbf{T}$  can be partitioned into  $(U+1)^2$  equally sized submatrices ( $Q \times Q$ ), which will be denoted as  $\mathbf{T}_{00}, \dots, \mathbf{T}_{UU}$ . Now,  $\tilde{\mathbf{F}}^S = \mathbf{F}^S \mathbf{T}$  can be expanded as

$$\begin{aligned} \tilde{\mathbf{F}}^S = \mathbf{F}^S \mathbf{T} = \\ [(\mathbf{S}_0 \mathbf{F} \mathbf{T}_{00} + \dots + \mathbf{S}_U \mathbf{F} \mathbf{T}_{U0}) | (\mathbf{S}_0 \mathbf{F} \mathbf{T}_{01} + \dots + \mathbf{S}_U \mathbf{F} \mathbf{T}_{U1}) | \dots | (\mathbf{S}_0 \mathbf{F} \mathbf{T}_{0U} + \dots + \mathbf{S}_U \mathbf{F} \mathbf{T}_{UU})] \end{aligned} \quad (7.16)$$

By noting that  $\tilde{\mathbf{F}}^S = [\mathbf{S}_0 \tilde{\mathbf{F}} | \mathbf{S}_1 \tilde{\mathbf{F}} | \dots | \mathbf{S}_U \tilde{\mathbf{F}}]$ , it follows that

$$\begin{aligned} \mathbf{S}_0 \tilde{\mathbf{F}} &= (\mathbf{S}_0 \mathbf{F} \mathbf{T}_{00} + \dots + \mathbf{S}_U \mathbf{F} \mathbf{T}_{U0}) \\ \mathbf{S}_1 \tilde{\mathbf{F}} &= (\mathbf{S}_0 \mathbf{F} \mathbf{T}_{01} + \dots + \mathbf{S}_U \mathbf{F} \mathbf{T}_{U1}) \\ &\dots \end{aligned} \quad (7.17)$$

$$\mathbf{S}_U \tilde{\mathbf{F}} = (\mathbf{S}_0 \mathbf{F} \mathbf{T}_{0U} + \dots + \mathbf{S}_U \mathbf{F} \mathbf{T}_{UU}).$$

The supermatrix  $\tilde{\mathbf{F}}^S = [\mathbf{S}_0 \tilde{\mathbf{F}} \mid \mathbf{S}_1 \tilde{\mathbf{F}} \mid \dots \mid \mathbf{S}_U \tilde{\mathbf{F}}]$  is composed of rows of  $\tilde{\mathbf{F}}$ . Row  $k$  of  $\tilde{\mathbf{F}}$ , which corresponds to the component scores at time point  $k$ , is denoted by  $\tilde{\mathbf{f}}_k'$  in the sequel; note that row  $k$  is not the  $k^{\text{th}}$  row in  $\tilde{\mathbf{F}}$ , since the numbering of rows of  $\mathbf{F}$  starts from  $1-U$ . The supermatrix  $\tilde{\mathbf{F}}^S$  is composed as

$$\tilde{\mathbf{F}}^S = \begin{bmatrix} \tilde{\mathbf{f}}_1' & \tilde{\mathbf{f}}_0' & & & \tilde{\mathbf{f}}_{1-U}' \\ \cdot & \tilde{\mathbf{f}}_1' & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \tilde{\mathbf{f}}_1' \\ \tilde{\mathbf{f}}_{K-U}' & \cdot & & & \\ \cdot & \tilde{\mathbf{f}}_{K-U}' & \dots & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \tilde{\mathbf{f}}_K' & \tilde{\mathbf{f}}_{K-1}' & & & \tilde{\mathbf{f}}_{K-U}' \end{bmatrix}.$$

The matrices  $\mathbf{S}_0 \tilde{\mathbf{F}}$ ,  $\mathbf{S}_1 \tilde{\mathbf{F}}$ , ...,  $\mathbf{S}_U \tilde{\mathbf{F}}$  all cover the component scores at time points 1 to  $K-U$ , albeit in different positions. The submatrix with rows  $1, \dots, K-U$  of  $\tilde{\mathbf{F}}$  will be denoted as  $\tilde{\mathbf{F}}_{(1, \dots, K-U)}$  in the sequel, where the subscripts indicate the rows concerned. Using (7.17), the rows of  $\mathbf{S}_0 \tilde{\mathbf{F}}$ ,  $\mathbf{S}_1 \tilde{\mathbf{F}}$ , ...,  $\mathbf{S}_U \tilde{\mathbf{F}}$  that cover measurements at common time points, collected in  $\tilde{\mathbf{F}}_{(1, \dots, K-U)}$ , can be written in terms of submatrices of  $\mathbf{F}$  as

$$\begin{aligned} \tilde{\mathbf{F}}_{(1, \dots, K-U)} &= \mathbf{F}_{(1, \dots, K-U)} \mathbf{T}_{00} + \mathbf{F}_{(0, \dots, K-U-1)} \mathbf{T}_{10} + \dots + \mathbf{F}_{(1-U, \dots, K-2U)} \mathbf{T}_{U0} \\ \tilde{\mathbf{F}}_{(1, \dots, K-U)} &= \\ &\quad \mathbf{F}_{(2, \dots, K-U+1)} \mathbf{T}_{01} + \mathbf{F}_{(1, \dots, K-U)} \mathbf{T}_{11} + \mathbf{F}_{(0, \dots, K-U-1)} \mathbf{T}_{21} + \dots + \mathbf{F}_{(2-U, \dots, K-2U+1)} \mathbf{T}_{U1} \\ &\quad \dots, \\ \tilde{\mathbf{F}}_{(1, \dots, K-U)} &= \mathbf{F}_{(U+1, \dots, K)} \mathbf{T}_{0U} + \dots + \mathbf{F}_{(2, \dots, K-U+1)} \mathbf{T}_{(U-1)U} + \mathbf{F}_{(1, \dots, K-U)} \mathbf{T}_{UU}, \end{aligned} \quad (7.18)$$

respectively. From the equality of the first and the last rows of (7.18), it follows that

$$\begin{aligned} &\mathbf{F}_{(1-U, \dots, K-2U)} \mathbf{T}_{U0} + \dots + \mathbf{F}_{(0, \dots, K-U-1)} \mathbf{T}_{10} + \mathbf{F}_{(1, \dots, K-U)} (\mathbf{T}_{00} - \mathbf{T}_{UU}) - \\ &\mathbf{F}_{(2, \dots, K-U+1)} \mathbf{T}_{(U-1)U} - \dots - \mathbf{F}_{(U+1, \dots, K)} \mathbf{T}_{0U} = \mathbf{0}_{(K-U) \times Q}, \end{aligned} \quad (7.19)$$

with  $\mathbf{0}_{(K-U) \times Q}$  a  $(K-U) \times Q$  matrix having each element equal to 0. Using Assumption 1, it follows that  $\mathbf{T}_{00} - \mathbf{T}_{UU} = \mathbf{0}_{Q \times Q}$ , and  $\mathbf{T}_{0U} = \dots = \mathbf{T}_{(U-1)U} = \mathbf{0}_{Q \times Q}$ , and  $\mathbf{T}_{10} = \dots = \mathbf{T}_{U0} = \mathbf{0}_{Q \times Q}$ . The equality of the right hand terms in the first and second terms in (7.18) leads to

$$\begin{aligned} & \mathbf{F}_{(1-U, \dots, K-2U)} \mathbf{T}_{U0} + \mathbf{F}_{(2-U, \dots, K-2U+1)} (\mathbf{T}_{(U-1)0} - \mathbf{T}_{U1}) + \dots \\ & + \mathbf{F}_{(0, \dots, K-U-1)} (\mathbf{T}_{10} - \mathbf{T}_{21}) + \mathbf{F}_{(1, \dots, K-U)} (\mathbf{T}_{00} - \mathbf{T}_{11}) - \mathbf{F}_{(2, \dots, K-U+1)} \mathbf{T}_{01} = \mathbf{0}_{(K-U) \times Q} \end{aligned} \quad (7.20)$$

Equation (7.20), Assumption 1, and the result that  $\mathbf{T}_{10} = \dots = \mathbf{T}_{U0} = \mathbf{0}_{Q \times Q}$ , lead to the conclusion that  $\mathbf{T}_{01} = \mathbf{T}_{21} = \dots = \mathbf{T}_{U1} = \mathbf{0}_{Q \times Q}$ , and that  $\mathbf{T}_{00} - \mathbf{T}_{11} = \mathbf{0}_{Q \times Q}$ . Analogously, the intermediate combinations can be written down, resulting in  $\mathbf{T}_{uu} = \mathbf{T}_{vv} \equiv \mathbf{T}_0$  for  $u, v = 0, \dots, U$ , and  $\mathbf{T}_{uv} = \mathbf{0}_{Q \times Q}$  for  $u \neq v$ . This implies that

$$\begin{aligned} \mathbf{S}_u \tilde{\mathbf{F}} &= \mathbf{S}_u \mathbf{F} \mathbf{T}_0, \\ \tilde{\mathbf{B}}_u &= \mathbf{B}_u (\mathbf{T}_0^{-1})', \end{aligned} \quad (7.21)$$

for  $u=0, \dots, U$ . Thus, postmultiplication of  $\mathbf{S}_u \mathbf{F}$ , for any  $u=0, \dots, U$ , by a non-singular matrix  $\mathbf{T}_0$ , and compensating for this by applying  $(\mathbf{T}_0^{-1})'$  to  $\mathbf{B}_u$  is the *only* possible transformation of  $\mathbf{S}_u \mathbf{F}$  without changing the estimates of  $\mathbf{X}$ . For all  $u$ ,  $u=0, \dots, U$ , the transformation of  $\mathbf{S}_u \mathbf{F}$  is the same, and hence it is the same for all rows of  $\mathbf{F}$ , as can readily be seen by  $\mathbf{S}_0 \tilde{\mathbf{F}} = \mathbf{S}_0 \mathbf{F} \mathbf{T}_0 = \tilde{\mathbf{F}}_{(1, \dots, K)} = \mathbf{F}_{(1, \dots, K)} \mathbf{T}_0$  and  $\mathbf{S}_U \mathbf{F} = \mathbf{S}_U \mathbf{F} \mathbf{T}_0 = \tilde{\mathbf{F}}_{(1-U, \dots, K-U)} = \mathbf{F}_{(1-U, \dots, K-U)} \mathbf{T}_0$ . The latter, combined with the notion that the presented transformation results hold for all subjects  $i$ , implies that

$$\tilde{\mathbf{F}}_i = \mathbf{F}_i \mathbf{T}_0, \quad (7.22)$$

for  $i=1, \dots, I$ . Thus, postmultiplication of  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , by any non-singular matrix  $\mathbf{T}_0$ , and compensating this by applying  $(\mathbf{T}_0^{-1})'$  to  $\mathbf{B}_u$ ,  $u=0, \dots, U$ , is the *only* possible transformation of the matrices  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , without changing the estimates of  $\mathbf{X}_i$  in the LSCA-P. ~

*Corollary 1.* Let  $\mathbf{F}_i$ ,  $i=1, \dots, I$ , and  $\mathbf{B}_u$ ,  $u=0, \dots, U$ , be a given LSCA-P solution. Suppose that alternative solutions exist, namely  $\tilde{\mathbf{F}}_i$  for  $\mathbf{F}_i$ , and  $\tilde{\mathbf{B}}_u$  for  $\mathbf{B}_u$ , that lead to the same estimate  $\hat{\mathbf{X}}_i$  for  $\mathbf{X}_i$  as do  $\mathbf{F}_i$  and  $\mathbf{B}_u$ , that is  $\hat{\mathbf{X}}_i = \mathbf{F}_i^S \mathbf{B}^S = \tilde{\mathbf{F}}_i^S \tilde{\mathbf{B}}^S$ . Then, provided that Assumptions 1 and 2 are satisfied,

$$\tilde{\mathbf{F}}_i \tilde{\mathbf{B}}_u' = \mathbf{F}_i \mathbf{B}_u', \quad (7.23)$$

for  $i=1, \dots, I$  and  $u=0, \dots, U$ .

*Proof.* Provided that Assumptions 1 and 2 are satisfied, it follows from Theorem 1 that  $\tilde{\mathbf{F}}_i = \mathbf{F}_i \mathbf{T}_0$ ,  $i=1, \dots, I$ , and  $\tilde{\mathbf{B}}_u = \mathbf{B}_u (\mathbf{T}_0^{-1})'$ ,  $u=0, \dots, U$ . Hence,  $\tilde{\mathbf{F}}_i \tilde{\mathbf{B}}_u' = \mathbf{F}_i \mathbf{T}_0 \mathbf{T}_0^{-1} \mathbf{B}_u' = \mathbf{F}_i \mathbf{B}_u'$ , for  $i=1, \dots, I$ , and  $u=0, \dots, U$ .  $\sim$

Corollary 1 allows us to examine the transformational freedom of constrained versions of the LSCA-P starting from (7.23) rather than starting from the more complex (7.14). Additionally, (7.23) equals the SCA model, with the constraint on  $\mathbf{F}_i$  for the model under consideration. Therefore, the results concerning transformational freedom in each of the four SCA models (see Section 6.2.7) hold for each of the corresponding LSCA models as well.

### 7.3.5. Testing the LSCA-P and LSCA-IND algorithms

To investigate some of the estimation properties of the LSCA-P algorithm, a small simulation study was performed. To limit the size of the study, we chose to examine only one of the four methods for LSCA, LSCA-P. It is noteworthy, however, that the LSCA-IND algorithm was included also in our preliminary analyses, and the results of those analyses were essentially the same as those of the LSCA-P algorithm.

To test the LSCA-P algorithm, we constructed 375 datasets according to the model

$$\mathbf{X}_i = \mathbf{S}_0^i \mathbf{F}_i \mathbf{B}_0' + \mathbf{S}_1^i \mathbf{F}_i \mathbf{B}_1' + \varepsilon \mathbf{E}_i, \quad (7.24)$$

based on fixed shift matrices  $\mathbf{S}_0^i$  and  $\mathbf{S}_1^i$ , and fixed loading matrices  $\mathbf{B}_0$  and  $\mathbf{B}_1$  (see below); the unconstrained component score matrices  $\mathbf{F}_i$  were generated as described below,  $\mathbf{E}_i$  is a random matrix with elements sampled from the standard normal distribution, and  $\varepsilon$  is a scalar.

In the present simulation study, three variables were varied: the expected degree of first order autocorrelation of the components of  $\mathbf{F}_i$ , the degree of multicollinearity of  $\mathbf{B}_0$  and  $\mathbf{B}_1$ , and the error level. It was expected that higher autocorrelation of the components, higher degree of multicollinearity and higher error level would lead to poorer performance of the LSCA-P algorithm. The simulated data were analyzed by a lag one LSCA-P, where the 'correct' number of components was estimated.

The component scores matrix  $\mathbf{F}_i$  was constructed by first generating an AR(1) time series with an expected mean of zero by

$$\mathbf{f}_{ik} = a \mathbf{f}_{i(k-1)} + \mathbf{e}_{ik}, \quad (7.25)$$

where  $a$  is a constant for manipulating the first order autocorrelation,  $\mathbf{e}_{ik}$  and  $\mathbf{f}_{i0}$  are both  $Q \times 1$  vectors with elements drawn from a  $N(0,1)$  distribution, and the vectors  $\mathbf{f}_{i1}, \dots, \mathbf{f}_{i(K+1)}$  were computed using (7.25). The matrix  $\mathbf{F}_i$  was obtained by collecting the vectors  $\mathbf{f}_{ik}'$ ,  $k=1, \dots, K+1$ , in its rows. The values of  $a$  were chosen as 0, 0.2, 0.4, 0.6

and 0.8, leading to expected first order autocorrelations for  $\mathbf{F}_i$  of 0, 0.2, 0.4, 0.6 and 0.8, respectively (see e.g. Jones, 1993). The loading matrices  $\mathbf{B}_0$  and  $\mathbf{B}_1$  were chosen as

$$\mathbf{B}_0 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 2-c & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 2-c \end{bmatrix}, \text{ and } \mathbf{B}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ c & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & c \end{bmatrix},$$

where the scalar  $c$  was varied so as to control the degree of multicollinearity of  $\mathbf{B}^S$ , the supermatrix containing the loading matrices  $\mathbf{B}_0$  and  $\mathbf{B}_1$  positioned next to each other. The values of  $c$ , and the corresponding condition numbers of  $\mathbf{B}^S$  in parentheses are 0 (1.4), 0.33 (2.3), 0.67 (5.2) and 1 (infinite). Note that the condition  $c=1$  in fact leads to data matrices with underlying data matrices (i.e., simulated data matrices without error) with zero lag structure. The expected error level was varied so that the expected percentage of error sums of squares in the data matrices was 0%, 20%, 50% and 69%. The number of subjects ( $I$ ) was five, the number of variables ( $J$ ) was six, the number of occasions ( $K$ ) was 30, and the number of components ( $Q$ ) was two. The design was fully crossed, except for the 0% error level condition combined with the condition with an infinite condition number of  $\mathbf{B}^S$ , that was excluded from the analyses: The latter condition leads to data matrices  $\mathbf{X}_i$ ,  $i=1, \dots, I$ , of exactly rank two, and hence to estimation problems if the data matrices are analyzed by the LSCA-P algorithm with  $Q=2$  and  $U=1$ . It is supposed that the data matrices constructed in this way cover a reasonable range of empirical data sets, where some data sets are likely to be quite extreme (like the one constructed with an infinite condition number of  $\mathbf{B}^S$ ). The number of replications in each condition was five.

The 375 LSCA-P data sets were analyzed by the LSCA-P algorithm (programmed in MATLAB5 (1998)). One rational, and four random starts were used. The convergence criterion was set at  $10^{-6}$ .

### 7.3.5.1. Retrieval of the loading matrices

The first issue studied pertained to the quality of retrieval of underlying loading matrices. For each simulated data matrix, the estimated loading matrices  $\hat{\mathbf{B}}_0$  and  $\hat{\mathbf{B}}_1$  were compared to the underlying loading matrices  $\mathbf{B}_0$  and  $\mathbf{B}_1$ .

The estimated loading matrices of an LSCA-P solution may be transformed without loss of fit, provided that the transformation is equal for all loading matrices and that the transformation is compensated in the component scores matrices of all subjects. Therefore, the matrix  $\hat{\mathbf{B}}_1^0$ , in which  $\hat{\mathbf{B}}_0$  and  $\hat{\mathbf{B}}_1$  of an LSCA-P are positioned

below each other, was projected on the matrix  $\mathbf{B}_1^0$ , in which  $\mathbf{B}_0$  and  $\mathbf{B}_1$  are positioned below each other, using ordinary regression. Hence, an optimal, in the least squares sense, nonsingular transformation matrix  $\mathbf{T}$  to transform  $\hat{\mathbf{B}}_1^0$  towards  $\mathbf{B}_1^0$ , was found. Then, the two columns of  $\hat{\mathbf{B}}_1^0 \mathbf{T}$  were compared to  $\mathbf{B}_1^0$  by computing the Proportion of Agreement as

$$\text{PA} = 1 - \frac{\|\mathbf{B}_1^0 - \hat{\mathbf{B}}_1^0 \mathbf{T}\|^2}{\|\mathbf{B}_1^0\|^2}. \quad (7.26)$$

The average Proportion of Agreement (PA) is displayed per error level by condition number of the loading matrices and by expected first order autocorrelation (AC1), in Figure 7.2.

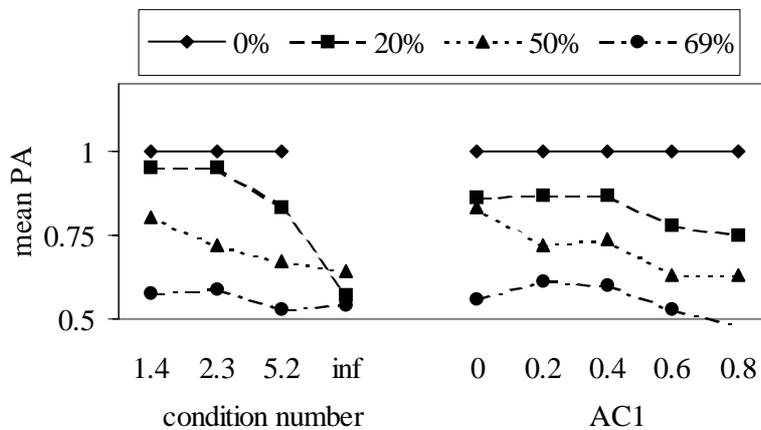


Figure 7.2. Average Proportion of Agreement (PA) of the LSCA-P simulation study per error level by condition number of the loading matrices (left) and by expected first order autocorrelation (AC1) (right); ‘inf’ denotes infinity.

As can be seen in Figure 7.2, an error level of 0% led to perfect recovery of the underlying loading matrices in all cases in the simulation study. Generally, a higher condition number of the underlying loading matrices, a higher error level, and a higher first order autocorrelation (AC1) level of the underlying component scores led to worse recovery of the underlying loading matrices, as is indicated by lower PA. The difference in recovery for the different error levels gets larger with decreasing condition numbers.

An ANOVA was performed to test whether the observed effects of the various manipulated factors could be distinguished from random fluctuations. The condition

with 0% error level was excluded from the analysis, because the PA-values were 1 for all observations in this condition. For the ANOVA, the PA-values were transformed to correct for the observed heterogeneity of variances for the groups by computing  $\tilde{P}\tilde{A} = \arcsin(\text{PA})^{1/2}$  (Stevens, 1992). The effects explicitly described in the previous paragraph were found to be significant at  $\alpha=0.05$  in the ANOVA.

In addition to the recovery measure, we inspected the fitting percentages of the LSCA-P model to the simulated data. The average fitting percentages were 100.0%, 87.8%, 70.1%, and 59.8% for the expected error levels of 0%, 20%, 50%, and 69%, respectively. This finding indicates a considerable amount of overfitting, implying that parts of the non-structural part of the data are fitted as well. Not surprisingly, the overfitting appears to be larger in the case of higher error levels.

### 7.3.5.2. Sensitivity to hitting suboptimal solutions

A disadvantage of using an alternating least squares algorithm is the possibility of ending up with sub-optimal solutions. A standard approach to this problem is to use multiple starts, and to choose the solution that produces the best fit of the model to the data. The second issue studied was the sensitivity of the LSCA-P algorithm to hitting suboptimal solutions. The LSCA-P algorithm was run five times, once started rationally and four times randomly. A solution with a fit value lower than 0.999 times the fit of the optimal solution (out of five) was considered to be a sub-optimal solution. The total number of sub-optimal solutions over the five runs appeared to be related to the error level and condition number of the loading matrices, and was only slightly related to autocorrelation of the component scores. On average, in the case of an error level of 0% (fit percentage 100%), 0.12 of the five starts ended up in a sub-optimal maximum, whereas in the case of higher error levels (20%, 50% or 69%) the number of sub-optimal solutions increased gradually (on average 1.52, 2.54 and 2.61, respectively). With increasing condition number of the loading matrices, the number of sub-optimal solutions also increased (on average 1.41, 1.44, 1.84 and 2.76 for condition numbers 1.4, 2.3, 5.2 and infinity, respectively). On average, rationally started runs ended up in sub-optimal solutions almost as often as randomly started runs. In ‘easier’ conditions (low autocorrelation between component scores ( $AC1=0$ ), low error level (0% and 20%) and low condition number (1.4, and 2.3)), the rationally started run ended less frequently in a sub-optimal solution. In ‘more difficult’ conditions, the average number of sub-optimal solutions of randomly started runs was smaller than for the rationally started run.

### 7.3.5.3. Discussion and conclusion of the simulation study

The results of this small simulation study revealed that recovery of the underlying loading matrices in LSCA-P is worse in the case of higher error level, higher condition number of  $\mathbf{B}^S = [\mathbf{B}_0 | \mathbf{B}_1]$ , and higher first order autocorrelation of the component scores. The recovery of the loading matrices still appears reasonable

( $PA > 0.80$ ) in the case of error levels lower than 20% combined with condition number and first order autocorrelations smaller than or equal to 5.2 and 0.6, respectively; the recovery also appears reasonable in the case of 50% error level combined with zero autocorrelation between component scores, or with condition number lower than 1.4.

A high first order autocorrelation of the component scores  $\mathbf{F}_i$  implies that  $\mathbf{S}_0\mathbf{F}_i$  is approximately proportional to  $\mathbf{S}_1\mathbf{F}_i$  (see (7.24)). As a result, the model in (7.24) almost reduces to

$$\mathbf{Y}_i = \mathbf{S}_0\mathbf{F}_i(\mathbf{B}_0 + \mathbf{B}_1)' + \varepsilon\mathbf{E}_i, \quad (7.27)$$

and this implies that the estimates of  $\mathbf{B}_0$  and  $\mathbf{B}_1$  becomes unstable if the error level is larger than zero.

A higher condition number of  $\mathbf{B}^S = [\mathbf{B}_0 | \mathbf{B}_1]$  led to poorer recovery of the loading matrices as well. We did not expect this finding because  $\mathbf{S}_0\mathbf{F}_i$  and  $\mathbf{S}_1\mathbf{F}_i$  are constrained to be lagged versions of each other and, as a result, a higher condition number of  $\mathbf{B}^S$  should not necessarily lead to unstable estimates.

When the error part was considerably larger than 20%, it appeared hard to disentangle the structural and the error parts. Besides the structural part, some part of the error is also fitted. The latter is illustrated in the simulation study by the phenomenon of overfitting: higher fit percentages were observed than could be expected on the basis of the construction of the data.

The number of sub-optimal solutions increased with increasing error level, condition number of  $\mathbf{B}^S$  and autocorrelation between the component scores. It appears wise to use multiple runs with different initial starting values to increase the chance of ending up in the global minimum.

The results of the simulation study suggest the following expectations for empirical LSCA-P. If empirical data have a lag one LSCA-P structure, and the data are not too noisy (say, error level lower than 20%), the correct loading matrices can be estimated to a reasonable level by performing an LSCA-P. If the autocorrelation between the component scores is zero, and/or the loading matrices lie in clearly different column spaces, reasonable recovery of the loading matrices can still be obtained in quite noisy data (error level 50%). Threshold values of the several factors that have been found to influence recovery, either for 'good' or for 'bad' recovery, cannot be given on the basis of the results of this small simulation study. Incidentally, those values are of limited value in practice, because one cannot infer the degree of either error level or autocorrelation in the components from empirical data. In social sciences, however, one generally expects a high error level, and a high autocorrelation in the component scores is also likely to appear. As a result, one may doubt whether good recovery of the loading matrices in LSCA-P can be obtained in social sciences. For example, a questionnaire with several questions concerning 'depression' and 'fear' filled out daily by several subjects, could be analyzed by one of the LSCA's. If two components, that could usefully be labeled 'depression' and 'fear', are extracted,

it is likely that the two series of component scores of the several subjects will show a high autocorrelation: a person depressed on one day is likely to be depressed on the next day, too. For the reasons mentioned above, it may be difficult to obtain an interpretable LSCA. Therefore, we expect the LSCA models to be of limited value in social sciences. On the contrary, the SCA models, which are special cases of the LSCA models, are promising, as shown by successful application to empirical data in Chapter 6.

#### 7.4. Empirical example: Mood in individuals with Parkinson's disease

In Section 6.4.1, three SCA models of the data gathered in the study of mood in individuals with Parkinson's disease (Shifren, Hooker, Wood & Nesselroade, 1997) were discussed. In the previous chapter, the SCA-IND model with two components and the SCA-P models with two and four components appeared stable and easy to interpret. Note that the SCA models are special cases of the LSCA model: They are lag zero LSCA models. As we questioned whether lagged effects could also be discovered, we subjected the data to a series of LSCA. The algorithms were run five times, using one rational and four random starts. The convergence criterion was set at  $10^{-6}$ . We focus first on the fit and interpretability of the models. The fit percentages of the fitted models are presented in Table 7.1.

*Table 7.1. Fit percentages of the four LSCAs with one and two components and with zero and one lags for the mood-data. The models selected in Section 6.4.1 are indicated bold. The lag zero LSCA models (indicated with '\*') are equivalent to the associated SCA models; the LSCA-IND, LSCA-PF2 and LSCA-P models with  $Q=1$  (indicated with '^') are equivalent.*

Model	$Q=1; U=0$	$Q=1; U=1$	$Q=2; U=0$	$Q=2; U=1$
LSCA-ECP	24.0*	30.8	31.8*	43.52
LSCA-IND	30.8*^	34.0^	<b>42.8*</b>	45.11
LSCA-PF2	30.8*^	34.0^	43.1*	45.12
LSCA-P	30.8*^	34.0^	<b>43.5*</b>	47.83

The increase in fit gained from using a lag one model instead of a lag zero model is rather small, except for the LSCA-ECP case. On the basis of fit, the lag one LSCA-ECP appears interesting to inspect. However, the lag one LSCA-ECP models with one and two components were less easy to interpret than the zero lag LSCA-IND with two components (discussed in Chapter 6).

The one component lag one LSCA-ECP revealed a special structure: the first lag loading matrix revealed the same contrasts as the zero lag loading matrix, whereas the lag one loadings in absolute sense were somewhat lower than the lag zero loadings. Both the lag zero and the lag one loadings can be interpreted as denoting a positive/negative affect dimension. The component scores of eleven out of twelve

subjects showed high negative first order autocorrelations, ranging from  $-0.59$  to  $-0.97$ . (The deviant subject was Subject 5, who showed a first order autocorrelation of  $0.08$ . The pattern of the component scores over time of the two SCA-IND components for this subject were discussed briefly in Chapter 6.) Thus, the component scores showed large contrasts on successive occasions, making the solutions very difficult to interpret. A similar pattern was also found in the two components from the lag one LSCA-ECP with two components.

The behavior of the algorithm in fitting the lag one LSCA-ECP model, and the pattern of the model estimates is reminiscent of the occurrence of degenerate solutions in the PARAFAC model (see for example, Kruskal, Harshman, & Lundy, 1989). That is, the fitted data remain finite but some (or all) of the component scores approach plus and minus infinity. The fitted models are often unstable and unreliable. Typically, with degenerate solutions, the iteration process does not seem to converge: the function value decreases less and less from one iteration to the next, whereas the (absolute) component scores continuously increase. The LSCA-ECP algorithm also appeared to end in a degeneracy. Degeneracy problems in PARAFAC can be circumvented by extracting fewer components, or using restrictions in the model, for example orthogonality constraints (Harshman & Lundy, 1984b).

## 7.5. Discussion and conclusion

Four variants of lagged simultaneous component models, and their properties were discussed in the preceding sections. The results of the simulation study revealed that fitting an LSCA model with lag one to simulated data led to reasonably recovered loading matrices in the 'easy' conditions, for example low error level, and low autocorrelation between component scores. In more difficult conditions, the recovery of the loading matrices appeared quite poor. In the empirical example, the lag one LSCA-P appeared difficult to interpret due to large contrasts in the estimated component scores at successive time points, combined with loading matrices that showed equal patterns. This, and the behavior of the algorithm is reminiscent of degeneracy problems sometimes encountered in fitting the PARAFAC model to data. Further research is needed to examine the problems in fitting the LSCA models to empirical data. It would be interesting to investigate other possible analogies between the degeneracy problem in PARAFAC fitting and the problems in LSCA's. If the problems will appear to resemble each other, solutions may be found in the same direction.

Generally, the usefulness of LSCAs with lag larger than zero seems limited to data with a special data structure, for example, combinations of physiological and psychological variables, where the levels of the physiological variables influence the levels of the psychological variables at later time points. In Section 7.2, an example of a dynamic factor model was discussed. The observed variables pertained to measures for anxiety, namely 'physiological' and 'psychological' anxiety, where the

physiological anxiety at time point  $k$  was highly correlated with psychological anxiety at time point  $k+1$ . This kind of data could be analyzed well by LSCA.

When comparing an LSCA model with an SCA model for the (hypothetical) 'anxiety data' it is important to remember that a lag one LSCA with one component would, apart from error fitting and boundary effects, be equal to a lag zero LSCA model with two components. The component scores of the latter model would show a special structure, namely one series of component scores would be a first lag version of the other series of component scores. It is unlikely that one would recognize this special structure of the component scores in an SCA model, whereas in the LSCA model this structure is clearly revealed. When SCA models are fitted to data with an LSCA structure with two (or more) components, an additional problem arises in recognizing the lagged structure. Then, the components of the SCA models are unlikely to be oriented so that the lagged structure can be recognized.

One could also question whether empirical examples of fitted dynamic factor models with lagged loadings (see, for example, Molenaar, 1985; Hershberger, Corneal & Molenaar, 1994; Shifren, Hooker, Wood & Nesselroade, 1997), are useful representations of the structure underlying the observed data. A problem with the dynamic factor models is that, in the model, the latent factor series are specified as white noise (i.e., they are independent within and across time). Any possible time dependent mechanisms are covered via the lagged loading matrices only. This makes interpretation of the models extremely difficult, as one should take into account that the factor scores are white noise. It is not natural to think of a latent psychological time series as being a white noise series, and it may even conflict with the intuitive notion of psychological time series. Recently, Molenaar and Nesselroade (2001) proposed a method for rotating each univariate factor series (that is represented as white noise) to a univariate moving average. The latter is likely to be easier to interpret. However, if the univariate factor series are well described by moving averages, it may be easier to fit a dynamic factor model in which the factor scores are constrained to follow a moving average process. Such models have already been proposed by Engle and Watson (1981) and Immink (1986).

# 8. Conclusion

## 8.1. Summary

This thesis dealt with component models for multisubject multivariate longitudinal data. A distinction was between two types of multisubject multivariate longitudinal data on the basis of the comparability of measurement occasions across subjects. Component models were discussed for both types.

The CANDECOMP/PARAFAC (CP), Tucker3, Tucker2 and Tucker1 models were treated as possible models for longitudinal three-way data. It is often useful to use constrained variants of those models, mainly in order to reduce the degree of error fitting. We discussed two types of constraints that can be used in the case of longitudinal data, namely smoothness and latent curve constraints. Usually, requiring smoothness is a weaker and a more flexible type of constraint than imposing latent curves on the solution. Imposing latent curve constraints typically requires extensive knowledge about the data which is being modeled, but latent curve constraints can be very attractive, especially if the parameters of the function have a physical interpretation.

We discussed four variants of Simultaneous Component Analysis (SCA) and their four lagged counterpart models for modeling multisubject multivariate time series. The four SCA models differ in the degree of interindividual variability that is covered in the model. Hence, one can explicitly choose the most parsimonious model possible for a particular data set, without ignoring important aspects of the data. In this way, the interpretation of the model is facilitated, and the chance of mistakenly fitting a part of the error term in the model is reduced.

The lagged counterpart models of the four SCA models allow not only for simultaneous, but also for lagged relationships between component scores and observed scores. In the simulation experiment, the lagged SCA models appeared to be estimated reasonably. The attempts at fitting a lagged SCA model to empirical data were disappointing, as the estimated models were difficult to interpret. The algorithm seemed to end in a degeneracy, a problem sometimes encountered in fitting the CP model. The problems might occur for a particular type of multivariate time series, but it is also quite possible, that the LSCA models are too weakly constrained to be useful in practice. Further investigation is required to test this.

## 8.2. Discussion and future work

The SCA models are useful for modeling multivariate multisubject time series. This was illustrated by the two empirical examples in Chapter 6. Constrained versions of the SCA models could be developed, for example by applying smoothness constraints to the component scores. To achieve smoothness in an SCA model, one could apply a procedure analogous to the procedure used in the CP and Tucker3 models. However, one should be cautious when doing this. Frequently, time series are collected to investigate the degree of intraindividual variability of processes that are subject to fluctuations. The application of smoothness constraints implies a reduction of the intraindividual variability in the model, compared to the unconstrained counterpart model. It is important to realize that this reduction does not have necessarily an equal effect for all subjects. Suppose there are two subjects with equal intraindividual variability in the unconstrained model, then, in the smoothness constrained model, the subject with a high level of short-term fluctuation will show less intraindividual variability than the subject with a high level of long-term fluctuation.

In estimating the parameters of a smoothness or latent curve constrained model, the whole series of observed scores is taken into account. This implies that, in the models, the estimated score at a certain time point is influenced by past as well as future scores. If one is interested in the predictive value of scores, one should apply models that use only past information to model scores at a certain time point, for example, autoregressive moving average models (ARMA). The linear dynamic system model (Bijleveld, 1989) for multivariate time series is interesting in this respect, as the component scores at the successive time points are constrained to follow a Markov model (which is a special case of an ARMA model). The extension of the linear dynamic system model to model multisubject multivariate time series (Bijleveld & Bijleveld, 1997) can be written as a version of an SCA-P model, with the component scores constrained to follow a Markov model. It might be interesting to formulate other SCA models with component scores constrained to follow an ARMA model as well.

In some cases, additional information on the subjects is available but is ignored in the models for longitudinal three-way data or the multisubject multivariate time series. The additional information, which is denoted by the term independent variables in the sequel, can refer to measured variables, but it can also follow from the design, for example 'treatment group versus control group'. There are various possibilities for actually relating the independent variables to the model at hand. These will be discussed later. First, we focus on the distinction between time-varying independent variables and variables that are constant across time, and the consequences for approaches of taking this information into account in both the models for longitudinal three-way data and multisubject multivariate time series.

In the longitudinal three-way models, independent variables that are constant across time could be related to the subject component scores. Obviously, the type of

preprocessing that has been performed on the raw data should be taken into account. For example, centering the scores across measurement occasions per subject and per variable eliminates the interindividual differences in general level from the data. As a result, after this type of preprocessing, these differences cannot be related to the independent variables. Time-varying independent variables could be related to the weights for the variable component scores at each of the measurement occasions (for example, in the case of the Tucker3 model, the columns of the matrix  $(\mathbf{A} \otimes \mathbf{C})\mathbf{G}'_b$ ).

In the SCA models for multivariate time series, the time-varying independent variables for each subject can easily be related to the component scores on the successive occasions for each subject. Independent variables that are constant across time are somewhat more difficult to relate to the SCA models: since the interpretation of the model parameters is easiest when the raw scores are centered across time per variable and *per subject* (see Section 6.2.1), the set of  $I \times J$  univariate series is analyzed in deviation from its own mean. Hence, differences in level between subjects are eliminated before analysis. As a result, if the goal is to relate the independent variable(s) to the general level, one must carry out separate analyses. However, if the goal is to model the effect of certain independent variable(s) on the degree of intraindividual variability, it is possible to relate those variables to the SCA-IND, SCA-PF2, and SCA-P models. For example, in the analysis of the ‘mood data’ (Section 6.4.1), one might be interested in relating the variability in ‘Emotional stability’ to a specific independent variable, like degree of social support.

In the sequel, we will refer to the component scores or parameters of the model that are to be related to the independent variables as the ‘dependent model parameters’. As mentioned before, various possibilities for relating the independent variables to the dependent model parameters exist. A first approach is to estimate a model for longitudinal three-way data or multisubject multivariate time series, and subsequently to relate quantitative independent variables to the dependent model parameters by computing correlations between the independent variable scores and the dependent model parameters; categorical independent variables can be related by computing category averages of the dependent model parameters (Kiers & Van Mechelen, 2001). A second approach is to incorporate the independent variables in the model. After the dependent model parameters have been estimated, the dependent model parameters could be regressed on the independent variable scores. This approach will be called the ‘subsequent regression approach’. Another method is to constrain the dependent model parameters so that they are a linear combination of the independent variables. This approach has been proposed in the CANDELINC context by Carroll, Pruzansky and Kruskal (1980), and can be extended to the present models as well. In the linear dynamic system model, Bijleveld (1989) used an approach to modeling independent variables which can be easily extended for use in the current context. Bijleveld’s method can be viewed as covering the two approaches for actually modeling independent variables. In the linear dynamic system model (Bijleveld, 1989; Bijleveld & Bijleveld, 1997), the dependent model parameters are required to be a linear combination of the independent variables *to a given extent*. The

extent is manipulated using a weight for the part of the least squares loss function that deals with the linear combination of the independent variables. The weight times the loss function is usually denoted by the term 'penalty' (see, for an example in a different context, Eilers & Marx, 1996). The respective two extremes of the continuum are reached by setting the penalty to infinity or zero, that is, to the Carroll, Pruzansky and Kruskal (1980) approach, and to the subsequent regression approach, respectively. This approach using a penalty has been elaborated for applying to the case for which both the independent and the dependent data are three-way arrays (Smilde & Kiers, 1999).

Each of the models that has been discussed in this thesis is fitted to data by ordinary least squares, hence by minimizing the unweighted least squares loss function. Under the single assumption that the errors have mean zero, ordinary least squares parameter estimates (OLS) are unbiased (Rice, 1995). The presence of autocorrelated errors with mean zero still provides unbiased OLS estimates, but the estimates may be inefficient (see e.g., Seber & Wild, 1989). Serial correlation in the errors may occur in models for longitudinal data. This can be understood as follows: An observed univariate sequence of scores collected at successive measurement occasions can be viewed as consisting of a true series of measurements plus random measurement error. In the case of longitudinal data, the true series can usually be viewed as evaluations of a certain, more or less intricate function. In modeling the observed series as evaluations of a smooth function that is *different* from the true function, autocorrelated residuals are induced. The more the fitted function deviates from the true function, the stronger the serial correlations between the errors. Having more sampling points in the same interval also leads to stronger serially correlated errors. To deal with this problem, one could try to approximate the true function as closely as possible, which usually boils down to fitting a rather complex function. This is not always desirable; for example, one may be interested in a simple description of the trend in the data. Alternatively, one could take the autocorrelated error structure into account in the model and the fitting procedure, for example by using a generalized least squares (GLS) instead of an OLS fitting procedure. However, specifying the error structure correctly may be a difficult task in practice, and the gain may be small.

In our applications, it is to be expected that a certain degree of autocorrelation in the errors is present. As a result, the fitting procedures that have been used throughout this thesis are probably not as efficient as could be. We chose to use OLS instead of GLS, because OLS fitting procedures are easier to handle than GLS, and because we expected the loss of efficiency to be within the acceptable range. Obviously, however, one can use different estimation procedures that do take the autocorrelation in the error structure into account.

The models for multivariate longitudinal data collected from more than one subject which have been discussed here, aim at providing an optimal summary of the data. In practice, these models can rarely be fitted to the data in a straightforward way. In rather complex types of data, such as the type discussed here, it is wise first to

---

model the data piece by piece before eventually turning it into a complex model that encompasses the full complexity of the data. After the usual checks of the data set, modeling the data is a process that consists of choosing particular models on the basis of the features a model should cover, fitting the model to data, interpreting the estimated model, and checking the stability of the model, where the different stages are usually passed through several times.



# References

- Alsberg, B.K. & Kvalheim, O.M. (1993). Compression of  $n$ th-order data arrays by B-splines. Part 1: Theory. *Journal of Chemometrics*, 7, 61-73.
- Anderson, T.W. (1963). The use of factor analysis in the statistical analysis of multiple time series. *Psychometrika*, 28 (1), 1-25.
- Andersson, C.A. & Bro, R. (1998). Improving the speed of multi-way algorithms: Part I. Tucker3. *Chemometrics and Intelligent Laboratory Systems*, 42, 93-103.
- Bentler, P.M. & Lee, S.-Y. (1978). Statistical aspects of a three-mode factor analysis model. *Psychometrika*, 43, 343-352.
- Bentler, P.M. & Lee, S.-Y. (1979). A statistical development of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology*, 32, 87-104.
- Bijleveld, C.C.J.H. (1989). *Exploratory linear dynamic systems analysis*. Leiden: DSWO Press.
- Bijleveld, C.C.J.H. & Bijleveld, F.D. (1997). A comparison of the applicability of two approaches to linear dynamic systems analyze for N subjects. *Kwantitatieve Methoden*, 55, 37-57.
- Bijleveld, C.C.J.H. & Van der Kamp, L.J.Th. (1998). Longitudinal data analysis: designs, models and methods. London: Sage.
- Bloxom, B. (1968). A note on invariance in three-mode factor analysis. *Psychometrika*, 33, 347-350.
- Borkenau, P. & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor Model to describe intraindividual variations over time? *Journal of Research in Personality*, 32 (2), 202-221.
- Brillinger, D.R. (1975). *Time series: data analysis and theory*. New York: Holt, Rinehart and Winston, Inc.
- Bro, R. (1998). *Multi-way analysis in the food industry. Models, algorithms and applications*. Unpublished doctoral thesis. Amsterdam: University of Amsterdam.
- Bro, R., Andersson, C.A. & Kiers, H.A.L. (1999). PARAFAC2-Part II. Modeling chromatographic data with retention time shifts. *Journal of Chemometrics*, 13, 295-309.
- Bro, R. & De Jong, S. (1997). A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics*, 11, 393-401.
- Bro, R. & Sidiropoulos, N.D. (1998). Least squares algorithms under unimodality and non-negativity constraints. *Journal of Chemometrics*, 12, 223-247.
- Bro, R. & Smilde, A.K. (in preparation). *Centering and scaling in component analysis*.
- Browne, M.W. (1993). Structured latent curve models. In: *Multivariate analysis: Future directions 2*. C.M. Cuadras & C.R. Rao (Eds.). Amsterdam: North-Holland.

- Browne, M.W. & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21 (2), 230-258.
- Browne, M.W. & Du Toit, S.H.C. (1991). Models for learning data. In: *Best methods for the analysis of change*. L.M. Collins & J.L. Horn (Eds.). Washington: American Psychological Association.
- Butcher, P.R., Kalverboer, A. F. & Geuze, R. H. (in press). Infants' shifts of gaze from a central to a peripheral stimulus: A longitudinal study of development between 6 and 26 weeks. *Infant Behavior and Development*, xx, xx-xx.
- Bus, A.G. & Kroonenberg, P.M. (1982). *Reading instruction and learning to read: A longitudinal study*. Internal report, SOL-Reeks, SOL/82-08. Groningen, The Netherlands: Department of Education, University of Groningen.
- Carroll, J.D. & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology*, 31, 607-649.
- Carroll, J.D. & Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of 'Eckart-Young' decomposition. *Psychometrika*, 35, 283-319.
- Carroll, J.D., Pruzansky, S. & Kruskal, J.B. (1980). CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, 1, 3-24.
- Cattell, R.B. (1944). 'Parallel proportional profiles' and other principles for determining the choice of factors by rotation. *Psychometrika*, 9, 276-283.
- Cattell, R.B. (1952). *Factor analysis*. New York: Harper.
- Cattell, R.B. (1963). The structuring of change by P-technique and incremental R-technique. In: C.W. Harris (Ed.). *Problems in measuring change*. Madison: The University of Wisconsin Press.
- Cattell, R.B. & Cattell, A.K.S. (1955). Factor rotation for proportional profiles: analytical solution and an example. *British Journal of Statistical Psychology*, 8, 83-92.
- Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, 31, 33-42.
- De Boor, C. (1978). *A practical guide to splines*. Berlin: Springer.
- Eilers, P.H.C. & Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89-121.
- Engle, R. & Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, 76, 774-781.
- Gabriel, K.R. & Zamir, S. (1979). Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics*, 21 (4), 489-498.
- Geweke, J.F. & Singleton, K.J. (1981). Maximum likelihood 'confirmatory' factor analysis of economic time series. *International Economic Review*, 22 (1), 37-54.
- Goldberg, L.R. (1990). An alternative 'description of personality': the Big-Five factor structure. *Journal of Personality and Social Psychology*, 59 (6), 1216-1229.

- Harshman, R.A. (1970). Foundations of the PARAFAC procedure: models and conditions for an 'exploratory' multi mode factor analysis. *UCLA Working Papers in Phonetics*, 16, 1-84.
- Harshman, R.A. (1972). Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA Working Papers in Phonetics*, 22, 111-117.
- Harshman, R.A. & Lundy, M.E. (1984a). The PARAFAC model for three-way factor analysis and multidimensional scaling. In: H.G. Law, C.W. Conrad jr., J.A. Hattie & R.P. McDonald (Eds). *Research methods for multimode data analysis*. New York: Praeger.
- Harshman, R.A. & Lundy, M.E. (1984b). Data preprocessing and the extended PARAFAC model. In: H.G. Law, C.W. Conrad jr., J.A. Hattie & R.P. McDonald (Eds). *Research methods for multimode data analysis*. New York: Praeger.
- Hastie, T.J. & Tibshirani, R.J. (1990). *Generalized additive models*. London: Chapman & Hall.
- Hershberger, S.L., Corneal, S.E. & Molenaar, P.C.M. (1994). Dynamic Factor Analysis: An application to emotional response patterns underlying daughter/father and stepdaughter/stepfather relationships. *Structural Equation Modeling*, 2 (1), 31-52.
- Immink, W. (1986). *Parameter estimation in Markov models and dynamic factor analysis*. Unpublished doctoral thesis, University of Utrecht.
- Jackson, J.E. (1991). *A user's guide to principal components*. New York: J. Wiley & Sons, Inc.
- Jansen, G.G.H. & Bus, A.G. (1982). *An application of a logistic bio-assay model to achievement test data. Technical report*. Groningen: Department of Education, University of Groningen.
- Jones, R.H. (1993). *Longitudinal data with serial correlation: a state-space approach*. London: Chapman and Hall.
- Jones, C.J. & Meredith, W. (1996). Patterns of personality change across life span. *Psychology and Aging*, 11 (1), 57-65.
- Jones, C.J. & Nesselroade, J.R. (1990). Multivariate, replicated, single-subject, repeated measures designs and P-technique factor analysis: a review of intraindividual change studies. *Experimental Aging Research*, 16 (4), 171-183.
- Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.
- Kiers, H.A.L. (1990). Majorization as a tool for optimizing a class of matrix functions. *Psychometrika*, 55, 417-428.
- Kiers, H.A.L. (1991). Hierarchical relations among three-way methods. *Psychometrika*, 56, 449-470.
- Kiers, H.A.L. (1997a). Three-mode orthomax rotation. *Psychometrika*, 62 (4), 579-598.
- Kiers, H.A.L. (1997b) Weighted least squares fitting using ordinary least squares algorithms. *Psychometrika*, 62, 251-266.

- Kiers, H.A.L. (1998a). A three-step algorithm for CANDECOMP/PARAFAC analysis of large data sets with multicollinearity. *Journal of Chemometrics*, 12, 155-171.
- Kiers, H.A.L. (1998b). Joint orthomax rotation of the core and component matrices resulting from three-mode principal component analysis. *Journal of Classification*, 15, 245-263
- Kiers, H.A.L. (2000). Towards a Standardized Notation and Terminology in Multiway Analysis. *Journal of Chemometrics*, 14, 105-122.
- Kiers, H.A.L. & Harshman, R.A. (1997). Relating two proposed methods for speedup of algorithms for fitting two- and three-way principal component and related multilinear models. *Chemometrics and Intelligent Laboratory Systems*, 36, 31-40.
- Kiers, H.A.L., Kroonenberg, P.M. & Ten Berge, J.M.F. (1992). An efficient algorithm for TUCKALS3 on data with large numbers of observations. *Psychometrika*, 57 (3), 415-422.
- Kiers, H.A.L. & Smilde, A.K. (1998). Constrained three-mode factor analysis as a tool for parameter estimation with second-order instrumental data. *Journal of Chemometrics*, 12, 125-147.
- Kiers, H.A.L. & Ten Berge, J.M.F. (1994). Hierarchical relations between methods for simultaneous component analysis and a technique for rotation to a simple simultaneous structure. *British Journal of Mathematical and Statistical Psychology*, 47, 109-126.
- Kiers, H.A.L., Ten Berge, J.M.F. & Bro, R. (1999). PARAFAC2-Part 1. A direct fitting algorithm for the PARAFAC2 model. *Journal of Chemometrics*, 13, 275-294.
- Kiers, H.A.L., Ten Berge, J.M.F. & Rocci, R. (1997). Uniqueness of three-mode factor models with sparse cores: the  $3 \times 3 \times 3$  case. *Psychometrika*, 62 (3), 349-374.
- Kiers, H.A.L. & Van Mechelen (2001). Three-way component analysis: principles and illustrative application. *Psychological Methods*, 6, 84-110.
- Kroonenberg, P.M. (1983). *Three-mode principal component analysis. Theory and applications*. Leiden: DSWO Press.
- Kroonenberg, P.M. & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45, 69-97.
- Kroonenberg, P.M. & Oort, F.J. (2001). *Three-mode analysis of multimode covariance matrices*. Submitted.
- Kruskal, J.B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Linear Algebra and Its Applications*, 18, 95-138.
- Kruskal, J.B., Harshman, R.A. & Lundy, M.E. (1989). How 3-MFA data can cause degenerate PARAFAC solutions, among other relationships. In: R. Coppi & S. Bolasco (Eds.). *Multiway data analysis*. Amsterdam: North Holland.
- Lawson, C.L. & Hanson, R.J. (1974). *Solving least squares problems*. Englewood Cliffs, N.J: Prentice-Hall.

- Levin, J. (1966). Simultaneous factor analysis of several gramian matrices. *Psychometrika*, 31 (3), 413-419.
- Linhart, H. & Zucchini, W. (1986). *Model selection*. New York: J. Wiley & Sons, Inc.
- Little, R.J.A. & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Louwse, D.J., Smilde, A.K. & Kiers, H.A.L. (1999). Cross-validation of multiway component models. *Journal of Chemometrics*, 13, 491-510.
- Magnus, J.R. & Neudecker, H. (1991). *Matrix differential calculus with applications in statistics and econometrics*. New York: Wiley.
- MATLAB5 (1998). *MATLAB. The language of technical computing*. Natick, MA: The Mathworks, Inc.
- Meredith, W. & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107-122.
- Mitchell, B.C. & Burdick, D.S. (1994). Slowly converging PARAFAC sequences: Swamps and two-factor degeneracies. *Journal of Chemometrics*, 8, 155-168.
- Molenaar, P.M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50, 181-202.
- Molenaar, P.C.M. & Nesselrode, J.R. (1998). A comparison of pseudo-maximum likelihood and asymptotically distribution-free dynamic factor analysis parameter estimation in fitting covariance-structure models to block-Toeplitz matrices representing single-subject multivariate time series. *Multivariate Behavioral Research*, 33 (3), 313-342.
- Molenaar, P.C.M. & Nesselrode, J.R. (2001). Rotation in the dynamic factor modeling of multivariate stationary time series. *Psychometrika*, 66 (1), 99-107.
- Murakami, T., Ten Berge, J.M.F. & Kiers, H.A.L. (1998). A case of extreme simplicity of the core matrix in three-mode principal components analysis. *Psychometrika*, 63 (3), 255-261.
- Nesselrode, J. R., & Molenaar, P. C. M. (1999). Pooling lagged covariance structures based on short, multivariate time series for dynamic factor analysis. In R. H. Hoyle (Ed.). *Statistical strategies for small sample research* (pp. 223-250). Thousand Oaks, CA: Sage Publications, Inc.
- Niesing, J. (1997). *Simultaneous component and factor analysis methods for two or more groups: a comparative study*. Leiden: DWO Press.
- Oort, F.J. (1999). Stochastic three-mode models. *British Journal of Mathematical and Statistical Psychology*, 52, 243-272.
- Oort, F.J. (2001). *Three-mode models for multivariate longitudinal data*. *British Journal of Mathematical and Statistical Psychology*, xx, xxx-xxx.
- Paatero, P. (1999). The multilinear engine -A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *Journal of Computational and Graphical Statistics*, 8 (4), 854-888.
- Paatero, P. & Andersson, C.A. (1999). *Further improvements of the speed of the Tucker3 three-way algorithm*. *Chemometrics and Intelligent Laboratory Systems*, 47, 17-20.

- Penrose, R. (1956). On the best approximate solutions of linear matrix equations. *Proc. Cambridge Phil. Soc.*, 52-17-19.
- Ramsay, J.O. (1988). Monotone regression splines in action. *Statistical Science*, 3 (4), 425-461.
- Ramsay, J.O. & Silverman, B.W. (1997). *Functional data analysis*. New York: Springer.
- Rao, C.R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, 14, 1-17.
- Rao, C.R. & Mitra, S.K. (1971). *Generalized inverse of matrices and its applications*. New York: Wiley.
- Rice, J.A. (1995). *Mathematical statistics and data analysis*. Belmont: Duxbury Press.
- Richards, F.J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany*, 10 (29), 290-300.
- Scholing, A. (1993). *Social phobia. Cognitive and behavioural treatment*. Unpublished doctoral thesis. Groningen: State University of Groningen.
- Seber, G.A.F. & Wild, C.J. (1989). *Nonlinear regression*. New York: Wiley.
- Shifren, K., Hooker, K., Wood, P. & Nesselroade, J.R. (1997). Structure and variation of mood in individuals with Parkinson's disease: a dynamic factor analysis. *Psychology and Aging*, 12 (2), 328-339.
- Smilde, A.K. & Kiers, H.A.L. (1999). Multiway covariates regression models. *Journal of Chemometrics*, 13, 31-48.
- Spikman, J.M., Timmerman, M.E., Van Zomeren, A.H. & Deelman, B.G. (1999). Recovery versus retest effects in attention after closed head injury. *Journal of Clinical and Experimental Neuropsychology*, 21 (5), 585-605.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Hillsdale: Lawrence Erlbaum Ass.
- Timmerman, M.E. & Kiers, H.A.L. (2000). Three Mode Principal Components Analysis: Choosing the numbers of components and sensitivity to local optima. *British Journal of Mathematical and Statistical Psychology*, 51, 1-16.
- Tucker, L.R. (1951). *A method for synthesis of factor analysis studies*. (Personnel Research Section Report No.984). Washington D.C.: Dept. of the Army.
- Tucker, L.R. (1958). Determination of parameters of a functional relationship by factor analysis. *Psychometrika*, 23, 19-23.
- Tucker, L.R. (1966a). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31, 279-311.
- Tucker, L.R. (1966b). Learning theory and multivariate experiment: illustration by determination of generalized learning curves. In: R.B. Cattell (Ed). *Handbook of multivariate experimental psychology*. Chicago: Rand McNally.
- Tupes, E.C. & Cristal, R.E. (1961). Recurrent personality factors based on trait ratings. *Technical Report, ASD-TR-61-97*, Lackland Air Force Base, TX, U.S. Air Force. Republished in: (1992) *Journal of Personality*, 60, 225-251.
- Van Buuren, S. (1990). *Optimal scaling of time series*. Leiden: DSWO Press.

- Van Rijckeversel, J.L.A. (1988). Fuzzy coding and B-splines. In: J.L.A. Van Rijckeversel & J. De Leeuw (Eds.). *Component and correspondence analysis. Dimension reduction by functional approximation*. Chichester: John Wiley & Sons.
- Velicer, W.F. & Jackson, D.N. (1990a). Component analysis versus common factor analysis: some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25 (1), 1-28.
- Velicer, W.F. & Jackson, D.N. (1990b). Component analysis versus common factor analysis: some further observations. *Multivariate Behavioral Research*, 25 (1), 97-114.
- Wahba, G. (1990). *Spline models for observational data*. Philadelphia, Pa: Society for Industrial and Applied Mathematics
- Watson, D. (1988). Intraindividual and interindividual analyses of positive and negative affect: their relation to health complaints, perceived stress, and daily activities. *Journal of Personality and Social Psychology*, 54 (6), 1020-1030.
- Watson, D., Clark, L.A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54 (6), 1063-1070.
- Weesie, J. & Van Houwelingen, H. (1983). *GEPCAM users' manual. Generalized principal components analysis with missing values*. Utrecht: Institute of Mathematical Statistics, State University of Utrecht, The Netherlands.
- Wold, S. (1978). Cross-validatory estimation of the number of components in factor and principal component models. *Technometrics*, 20, 397-405.



## Summary in Dutch (Samenvatting)

Het kenmerk van longitudinaal onderzoek is dat gegevens verzameld worden op verschillende tijdstippen bij hetzelfde individu of object. Met het verzamelen van longitudinale gegevens beoogt men inzicht te krijgen in het verloop van een bepaald proces, dat in het algemeen aan verandering onderhevig zal zijn. Om inzicht te krijgen in het verloop van een proces is het nodig om dit proces te meten. In de sociale wetenschappen is het vrijwel nooit doenlijk om een proces continu te meten, en zal een beperkt aantal meetmomenten gekozen worden. Belangrijke overwegingen bij de keuze van die meetmomenten zijn de mate van precisie waarmee een proces gevolgd moet worden om de onderzoeksvragen te kunnen beantwoorden, de kennis die bij voorbaat aanwezig is over de vorm van het te volgen proces, en de grootte van de meetfout.

Volgt men de processen van meerdere individuen, dan kan men niet alleen de individuele ontwikkeling van de afzonderlijke individuen onderzoeken, maar ook de interindividuele verschillen in ontwikkeling. Hierbij is het zinvol om twee typen onderzoeksdesigns te onderscheiden. Ten eerste, het design waarbij de meetmomenten van het gevolgde proces voor de individuen zinvol te vergelijken zijn. Hiervoor is het noodzakelijk dat een ijkpunt aan te wijzen is in het gevolgde proces. Een voorbeeld is onderzoek naar het effect van een therapie, waarbij de start van de therapie als ijkpunt geldt. Ten tweede, het design waarbij de meetmomenten niet vergelijkbaar zijn tussen de individuen, en geen zinvol ijkpunt aangewezen kan worden in het proces dat gevolgd wordt. Een voorbeeld is onderzoek naar stemmingswisselingen in de loop der tijd bij een aantal individuen. In dit type design is het niet zinvol om naar gemeenschappelijke ontwikkelingsprocessen voor de verschillende individuen op zoek te gaan.

In sociaal wetenschappelijk onderzoek worden vaak multivariate data verzameld, waarbij de variabelen indicatoren zijn voor één of meerdere constructen. Een veel gebruikte analysemethode om de scores van meerdere individuen op meerdere variabelen efficiënt samen te vatten is principale componentenanalyse. De samenvatters van de variabelen zijn hopelijk, gewoonlijk na rotatie, interpreteerbaar in termen van de constructen die onderzocht worden.

Worden multivariate gegevens van verscheidene individuen longitudinaal verzameld, dan kunnen twee typen data onderscheiden worden, op basis van de eerder genoemde vergelijkbaarheid van meetmomenten. Bij 'longitudinale drie-weg data' zijn de meetmomenten zinvol vergelijkbaar tussen proefpersonen, terwijl bij 'multisubject multivariate tijdreeksen' dit niet het geval is. Dit proefschrift behandelt een aantal technieken voor componentenanalyse van zowel longitudinale drie-weg data als van multisubject multivariate tijdreeksen.

Longitudinale drie-weg data kunnen geanalyseerd worden met standaard componentenanalyse-modellen voor drie-weg gegevens, zoals CANDECOMP/PARAFAC (CP), Tucker3, Tucker2 en Tucker1, welke hiërarchisch geordend kunnen worden naar mate van strengheid. Afhankelijk van de data-structuur en de mate waarin samenvatting van de gegevens gewenst is, kan gekozen worden voor een meer of minder streng model. Een analyse met één van de twee strengste modellen, het CP- of Tucker3-model, van longitudinale drie-weg data levert drie componentenmatrices op, namelijk één voor de subjecten, één voor de variabelen, en één voor de tijdstippen. In de tijdscomponentenmatrix zijn de componentscores verzameld op de gemeten tijdstippen. Deze scores vatten het scoreverloop in de tijd voor de proefpersonen en de variabelen samen. Per component kunnen de scores op de verschillende tijdstippen beschouwd worden als evaluaties van een onderliggende functie, waarvan de exacte vorm meestal niet bekend is. In veel gevallen is het gerechtvaardigd om aan te nemen dat de functie een gladde vorm heeft. Soms kan echter aangenomen worden dat de scores een meer specifiek gedetermineerde functie volgen. Deze aannames kunnen gebruikt worden bij de analyse, door gladheidsrestricties, of een bepaalde functionele vorm, op te leggen aan de componentscores. Het voornaamste voordeel van het juiste gebruik van restricties is een reductie van de mate waarin het niet-structurele deel in de modelschattingen wordt opgenomen.

In Hoofdstuk 4 wordt voorgesteld om gladheidsrestricties op te leggen in de CP en Tucker3-modellen met behulp van ‘B-splines’. B-splines kunnen op twee manieren aangewend worden om gladheid te bewerkstelligen in de CP- en Tucker3-modellen. Ten eerste kunnen de univariate reeksen met scores op opeenvolgende tijdstippen van de subjecten en de variabelen glad gemaakt worden, om daarna geanalyseerd te worden met een CP- of Tucker3-analyse. Ten tweede kunnen gladheidsrestricties opgelegd worden aan de opeenvolgende componentscores in de tijdscomponentenmatrix van het CP- of Tucker3-model. In een groot aantal gevallen hoeft de vraag welke van de twee benaderingen gebruikt dient te worden, niet gesteld te worden. Zoals in Hoofdstuk 4 bewezen wordt, leiden, onder bepaalde voorwaarden, deze twee methoden tot dezelfde modelschattingen. Additioneel aan gladheidsrestricties kunnen niet-negativiteits- en monotoniciteitsrestricties opgelegd worden met behulp van respectievelijk B-splines en de aan B-splines gerelateerde I-splines. In een simulatieonderzoek is de effectiviteit van het gebruik van gladheidsrestricties in de CP- en Tucker3-modellen onderzocht. Bij gesimuleerde drie-weg data met een gladde structuur bleek in het algemeen dat de parameters van de CP- en de Tucker3-modellen beter geschat worden als gladheidsrestricties worden opgelegd, dan wanneer een ongerestricteerd model gebruikt wordt. Een empirisch voorbeeld illustreert het gebruik van gladheidsrestricties in het Tucker3-model. Het Tucker3-model met gladheidsrestricties bleek eenvoudiger te interpreteren dan het ongerestricteerde Tucker3-model.

In Hoofdstuk 5 worden bepaalde functies aan de tijdscomponentscores in het Tucker3-model opgelegd. Het aldus verkregen model wordt het ‘Structured Latent

Curve (SLC) Tucker3-model' genoemd. Het SLC Tucker3-model is geïnspireerd op de 'Structured latent curve' factormodellen voor de analyse van univariate longitudinale gegevens van meerdere proefpersonen, zoals voorgesteld door Browne en Du Toit (1991) en Browne (1993). In het SLC Tucker3-model is iedere tijdscomponent een, vooraf gespecificeerde, functie van een beperkt aantal parameters. Analoot aan de methode van Browne en Du Toit wordt een eerste orde Taylor-polynoom gebruikt van een bepaalde doelfunctie om de functies van de 'latente curves' te beschrijven met behulp van een beperkt aantal basiscurven. Als doelfunctie is gekozen voor de Gompertzfunctie, welke vooral nuttig is om groeigegevens te modelleren. De parameters van de Gompertzfunctie en de basisfuncties zijn goed interpreteerbaar. De eigenschappen van het SLC Tucker3-model, en een algoritme om het model aan data te fitten zijn besproken. Voor de empirische gegevens waarvan in Hoofdstuk 4 een Tucker3-model met gladheidsrestricties is besproken, is een SLC Tucker3-model geschat. De interpretatie van de parameters vertoonde grote overeenkomsten. Echter, het SLC Tucker3-model is spaarzamer, en verdient daarom hier de voorkeur.

De analyse van multisubject multivariate tijdreeksen vereist een andere aanpak dan de analyse van longitudinale drie-weg data. Omdat de meetmomenten niet vergelijkbaar zijn voor de verschillende individuen is het niet zinvol om de tijdreeksen voor verschillende individuen samen te vatten, zoals in de besproken modellen voor longitudinale drie-weg data wordt gedaan. In Hoofdstuk 6 en 7 worden modellen voor de analyse van multisubject multivariate tijdreeksen behandeld, waarin zowel intra-individuele variabiliteit als interindividuele variabiliteit wordt gemodelleerd. In Hoofdstuk 6 worden vier varianten van Simultane ComponentenAnalyse (SCA) besproken. In elk van de vier modellen worden de multivariate tijdreeksen ontbonden in een beperkt aantal tijdreeksen van componentcores, en een variabelen componentenmatrix, waarop de interpretatie van de componenten is gebaseerd. De vier modellen kunnen hiërarchisch geordend worden van zwak naar streng gerestrictieerd. In het zwakste model wordt de meest variatie tussen individuen toegestaan, en in het strengste model de minste. Aandacht wordt besteed aan het schatten van de modelparameters, aan de keuze van het type model en het aantal componenten, en aan de rotatievrijheid van de vier modellen. Het gebruik van de SCA-modellen wordt geïllustreerd aan de hand van twee empirische voorbeelden.

In Hoofdstuk 7 worden de vier SCA-modellen uitgebreid, zodanig dat niet alleen gelijktijdige effecten, maar ook vertraagde effecten gemodelleerd kunnen worden. Deze vier 'Lagged SCA (LSCA)-modellen' zijn geïnspireerd op een variant van de klasse van dynamische factor-modellen voorgesteld door Molenaar (1986). Molenaar's dynamische factor-model wordt gewoonlijk gebruikt om multivariate tijdreeksen van één individu te modelleren, terwijl hier de multivariate tijdreeksen van meerdere individuen worden gemodelleerd. Algoritmen om de LSCA-modellen te fitten worden besproken, evenals de rotatievrijheid binnen de vier LSCA-modellen. Deze rotatievrijheid bleek, onder bepaalde assumpties, beperkt. In een klein

simulatieonderzoek zijn de schattingseigenschappen onderzocht van het algoritme voor het fitten van het minst gerestricteerde LSCA-model. De parameters van de gesimuleerde data bleken in het algemeen redelijk teruggeschat te worden, met uitzondering van de ‘moeilijke’ condities, bijvoorbeeld een hoog ruisniveau in de gesimuleerde gegevens. De analyse van een empirische data set leverde teleurstellende resultaten op, omdat de geschatte modelparameters nauwelijks interpreteerbaar bleken.

Hoofdstuk 8 sluit het proefschrift af met een samenvatting, een discussie van de gevonden resultaten, en mogelijke uitbreidingen van de gepresenteerde modellen. Ook wordt een mogelijke tekortkoming in de schattingsprocedures aangestipt.

# Subject Index

- alternating least squares, 14
- autocorrelated residuals, 120
- B-splines, 26–27
- CANDECOMP/PARAFAC, 11, 18–19
  - fitting, 13
  - transformational freedom, 15
  - with smoothness constraints, 25–48
- canonical class model, 69, 98
- centering, 15
- change
  - long-term, 1, 2
  - short-term, 1, 2
- coefficient of congruence. *See* Tucker's coefficient of congruence
- column-wise Kronecker product. *See* Khatri-Rao product
- constrained model, 15
- constraints, 21
  - in modeling longitudinal data, 21–23
  - latent curve, 51–67
  - smoothness, 25–48
- convergence, 14
- covariate. *See* independent variable modeling
- cross-validation, 27
  - expectation maximization cross-validation, 75
  - split-half analysis, 76–77
- degeneracy, 115
- degree of modeling error, 16
- dynamic factor model, 95–97
- empirical example
  - LSCA model, 114–15
  - SCA model, 79–91
  - SLC Tucker3 model, 40–46
  - Tucker3 model with smoothness constraints, 40–46
- error fitting, 17
- error of approximation, 16
- error of estimation, 16
- expectation maximization cross-validation, 16, 75
- factor analysis
  - common factor analysis, 5
  - component analysis, 5
- fit, 13
- fitting
  - CANDECOMP/PARAFAC, 13
  - LSCA-ECP, 104
  - LSCA-IND, 103–4
  - LSCA-P, 100–101
  - LSCA-PF2, 101–3
  - SCA-ECP, 79
  - SCA-IND, 78–79
  - SCA-P, 77–78
  - SCA-PF2, 78
  - SLC Tucker3 model, 59–62
  - Tucker1 model, 13
  - Tucker2 model, 13
  - Tucker3 model, 13
- functional data, 21
- functional form, 2–4
- generalized least squares, 120
- global minimum, 14
- Gompertz curve, 50, 53
- independent variable modeling, 118–20
- interval scale, 15, 71
- I-splines, 27
- Khatri-Rao product, 10
- Lagged Simultaneous Component Analysis (LSCA), 100–101
  - starting values, 104
- linear dynamic system model, 69, 118, 119
- local minimum, 14, 31, 34, 37, 112, 113

- longitudinal data, 1
  - longitudinal three-way data, 9
  - longitudinal two-way data, 4–5
  - multiple time series, 5
- longitudinal research, 1
- LSCA-ECP, 100
  - fitting, 104
  - transformational freedom, 105–9
- LSCA-IND, 100
  - fitting, 103–4
  - transformational freedom, 105–9
- LSCA-P, 98–99
  - fitting, 100–101
  - transformational freedom, 105–9
- LSCA-PF2, 99–100
  - fitting, 101–3
  - transformational freedom, 105–9
- measurement error, 2
- missing data, 23
- mode, 6, 10
- model selection
  - SCA model, 75–77
  - three-way component model, 15–17
- overall error, 16
- PARAFAC. *See*
  - CANDECOMP/PARAFAC
- preprocessing
  - multisubject multivariate time series, 70–71
  - three-way data, 15
- principal component analysis, 10
- P-technique, 69
- ratio scale, 15, 40
- sampling time points, 1–4
- SCA-ECP, 74
  - fitting, 79
  - transformational freedom, 74–75
- SCA-IND, 73–74
  - fitting, 78–79
  - transformational freedom, 74–75
- scaling, 15
- SCA-P, 71–72
  - fitting, 77–78
  - transformational freedom, 74–75
- SCA-PF2, 73
  - fitting, 78
  - transformational freedom, 74–75
- Simultaneous Component Analysis
  - empirical example, 79–91
  - model selection, 75–77
  - starting values, 79
  - transformational freedom, 74–75
  - with Equal average Cross-Products constraints (SCA-ECP), 74
  - with INDSCAL constraints (SCA-IND), 73–74
  - with invariant Pattern (SCA-P), 71–72
  - with PARAFAC2 constraints (SCA-PF2), 73
- SLC Tucker3 model, 58–59
  - fitting, 59–62
  - transformational freedom, 62
- SLC two-way model, equal time points, 50–52
- SLC two-way model, equal time points
  - fitting, 53–56
- SLC two-way model, unequal time points, 57
- smoothing parameter, 27
- split-half analysis, 16, 76
- starting values
  - LSCA models, 104
  - SCA models, 79
  - SLC Tucker3 model, 60
  - SLC two-way model, 55
- structured latent curve (SLC) component model, 51–67
- three-mode principal component analysis. *See* Tucker3 model
- three-way data, 6
- three-way data analysis
  - component analysis, 6
  - factor analysis, 6

- time series
  - component analysis, 6
  - factor analysis, 6
- transformational freedom
  - CANDECOMP/PARAFAC, 15
  - LSCA models, 105–9
  - SCA models, 74–75
  - SLC Tucker3 model, 62
  - Tucker1, Tucker2 and Tucker3 models, 14–15
- Tucker's coefficient of congruence, 34
- Tucker1 model, 20
  - fitting, 13
  - transformational freedom, 14–15
- Tucker2 model, 20
  - fitting, 13
  - transformational freedom, 14–15
- Tucker3 model, 12, 19–20
  - fitting, 13
  - transformational freedom, 14–15
  - with latent curve constraints, 58–59
  - with smoothness constraints, 25–48
- variability
  - interindividual, 5
  - intraindividual, 5, 69



# Author Index

- Alsberg, B.K., 26  
Anderson, T.W., 69  
Andersson, C.A., 14, 33, 91  
Arabie, P., 10  
Bentler, P.M., 6  
Bijleveld, C.C.J.H., 1, 6, 69, 70, 118, 119  
Bijleveld, F.D., 70, 118, 119  
Bloxom, B., 6  
Borkenau, P., 86, 87, 88, 90  
Brillinger, D.R., 98  
Bro, R., 14, 15, 21, 30, 33, 71, 73, 75, 78,  
79, 91, 101  
Browne, M.W., 16, 22, 23, 31, 49, 50, 53  
Burdick, D.S., 34  
Bus, A.G., 40, 62  
Butcher, P.R., 4  
Carroll, J.D., 6, 10, 11, 12, 14, 29, 33, 74,  
119  
Cattell, A.K.S., 11  
Cattell, R.B., 11, 69  
Chang, J., 6, 11, 12, 14, 33, 74  
Clark, L.A., 80  
Cliff, N., 79  
Corneal, S.E., 116  
Cristal, R.E., 86  
Cudeck, R., 16  
De Boor, C., 26  
De Jong, S., 30  
Deelman, B.G., 4  
Du Toit, S.H.C., 22, 23, 49, 50, 53  
Eilers, P.H.C., 120  
Engle, R., 69, 95, 116  
Geuze, R.H., 4  
Geweke, J.F., 95  
Goldberg, L.R., 86  
Hanson, R.J., 30  
Harshman, R.A., 6, 11, 14, 15, 16, 30, 33,  
71, 75, 76, 115  
Hastie, T.J., 21, 25, 26, 27  
Hershberger, S.L., 116  
Hooker, K., 5, 70, 80, 114, 116  
Immink, W., 6, 69, 95, 116  
Jackson, D.N., 6  
Jackson, J.E., 6  
Jansen, G.G.H., 40, 62  
Jones, C.J., 22, 69, 110  
Jones, R.H., 22, 69, 110  
Kaiser, H.F., 14, 22, 75, 76  
Kalverboer, A.F., 4  
Kiers, H.A.L., 10, 13, 14, 16, 17, 21, 30,  
34, 41, 43, 57, 71, 73, 75, 76,  
78, 79, 87, 91, 101, 102, 104,  
119  
Kroonenberg, P.M., 6, 11, 12, 13, 14, 19,  
40, 62  
Kruskal J.B., 119  
Kruskal, J.B., 15, 29, 115  
Kvalheim, O.M., 26  
Lawson, C.L., 30  
Lee, S.-Y., 6  
Linhart, H., 16  
Little, R.J.A., 23, 47, 58  
Louwerse, D.J., 16, 75  
Lundy, M.E., 6, 15, 16, 71, 76, 115  
Magnus, J.R., 10  
Marx, B.D., 120  
Meredith, W., 22  
Mitchell, B.C., 34  
Mitra, S.K., 10, 103  
Molenaar, P.C.M., 6, 69, 70, 95, 97, 116  
Murakami, T., 41  
Nesselroade, J.R., 5, 6, 69, 80, 97, 114,  
116  
Neudecker, H., 10  
Niesing, J., 91  
Oort, F.J., 6, 67  
Ostendorf, F., 86, 87, 88, 90  
Paatero, P., 14

- Penrose, R., 56, 60  
Pruzansky, S., 29, 119  
Ramsay, J.O., 5, 21, 22, 26, 27, 28  
Rao, C.R., 10, 22, 103  
Rice, J.A., 120  
Richards, F.J., 3, 50  
Rocci, R., 21  
Rubin, D.B., 23, 47, 58  
Scholing, A., 4  
Seber, G.A.F., 55, 56, 57, 120  
Shifren, K., 5, 70, 80, 114, 116  
Sidiropoulos, N.D., 21  
Silverman, B.W., 5, 21, 22, 26, 28  
Singleton, K.J., 95  
Smilde, A.K., 15, 16, 21, 71, 75, 120  
Spikman, J.M., 4  
Stevens, J., 39, 112  
Tellegen, A., 80  
Ten Berge, J.M.F., 14, 21, 41, 71, 73, 75,  
78, 79, 87, 91, 101  
Tibshirani, R.J., 21, 25, 26, 27  
Timmerman, M.E., 4, 17  
Tisak, J., 22  
Tucker, L.R., 6, 11, 14, 22, 33, 34, 44, 67,  
73, 86, 100  
Tupes, E.C., 86  
Van Buuren, S., 6, 69, 70, 98  
Van der Kamp, L.J.Th., 1, 6  
Van Houwelingen, H., 14, 63  
Van Mechelen, I., 16, 43, 76, 119  
Van Rijckevorsel, J.L.A., 27  
Van Zomeren, A.H., 4  
Velicer, W.F., 6  
Wahba, G., 21  
Watson, D., 80  
Watson, M., 69, 95, 116  
Weesie, J., 14, 63  
Wild, C.J., 55, 56, 57, 120  
Wold, S., 75  
Wood, P., 5, 70, 80, 114, 116  
Zamir, S., 57  
Zucchini, W., 16

# Acknowledgements

For some people, the term Ph.D. project is synonymous with at least four years of solitary confinement. Fortunately, this is not true. Research doesn't get far without the help of, and the discussions and interactions with other people. During my research project, a number of people were particularly important, and I am indebted to them for their help.

First of all, Henk Kiers, who was one of the initiators of the 'aandachtsgebied' 'Multi-way analysis of multivariate longitudinal data'. Henk was my supervisor during the whole project, although I experienced the contacts more as 'having collaborated' rather than 'being supervised'. I greatly appreciate his patience, not only when explaining difficult details, but also when dealing with my impatience. But what I value most, is the fact that Henk gave me great freedom in seeking my own way in multi-way world.

Jos ten Berge was my supervisor in the background. Jos read the several almost-final-versions of the manuscript carefully. His critical comments invariably improved the manuscript.

Madelaine van der Meulen was important in two respects. First, Madelaine was a nice roommate who desired the same ratio of socializing to working and who liked the same room climate as I do. Second, Madelaine provided the opportunity to share and to discuss early ideas, which was very helpful.

The project was embedded in an 'aandachtsgebied'. Our multi-way-meetings twice a year were valuable. The constructive criticism of Willem Heiser, Pieter Kroonenberg, Frans Oort, Mark de Rooij, and Peter Molenaar helped me learn to express and defend ideas, and to develop new ideas.

Michael Browne was kind enough to read and comment on an early version of the chapter on Structured latent curve component models, and on the final version of the manuscript.

Age Smilde made various useful comments on the final version of this manuscript.

Phillipa Butcher spent a lot of time in checking the whole manuscript for dutchified phrases. Her explanation of the rules of grammar will be effective for a long time!

Marijtje van Duijn kindly advised me on the statistical analyses.

Paul Eilers offered me some Matlab files to compute the spline bases. Hans Boelens put some Matlab code to perform the Levenberg-Marquardt algorithm at my disposal.

The empirical examples that are discussed in this thesis could not have been included without the kind permission of several people. Jeanet Bus put the ‘Learning-to-read data’ at my disposal. Kim Shifren kindly offered her dataset on ‘Mood in Individuals with Parkinson’s disease’, which was collected at Syracuse University, NY, USA. Kim also commented on an earlier version of the SCA chapter. Professor Borkenau and Professor Ostendorf offered the ‘Big-Five data’.

Finally, Cor van der Meulen prevented my focus from becoming too narrow. I am happy that he offered general support during the last four years.