25

High-dimensional Joint Models

25.1 Introduction

In Chapter 24, it has been discussed how multiple sequences of repeated measurements can be jointly analyzed. The examples given there all considered joint modeling of two (longitudinal) outcomes only. Here, we will extend this to (much) higher dimensions. The motivation for joint modeling will remain the same. In some cases, joint modeling is required because the association structure between the outcomes is of interest. For example, one may be interested in studying how the association between outcomes evolves over time or how outcome-specific evolutions are related to each other (Fieuws and Verbeke 2004). In other cases, joint modeling is needed in order to be able to draw joint inferences about the different outcomes. As examples, consider testing whether a set of outcomes shows the same average evolution, or testing for the effect of covariates on all outcomes simultaneously.

An example where joint modeling of many longitudinal outcomes has proven useful can be found in Fieuws and Verbeke (2005a), where longitudinally measured hearing thresholds were jointly analyzed, for the left ear and for the right ear, and for 11 different frequencies. This yielded a total of 22 longitudinal sequences per subject.

The possibly high dimension raises at least two additional problems, in addition to the issues discussed in Chapter 24. First, some of the models often used for the joint analysis of two longitudinal sequences are less applicable for higher dimensions. For example, when using conditional models

(Section 24.1), only two possibilities for the conditioning are possible in the case of two outcomes only: The first outcome can be modeled conditionally on the second, or vice versa. With (much) higher dimensions, (many) more possible conditioning strategies are possible, all yielding different models, of which parameters have different interpretations. Moreover, several of the research questions that require joint modeling are phrased in terms of the parameters in each of the univariate longitudinal models (i.e., longitudinal models for each repeated outcome separately), as was the case in the examples given earlier. Also, the models that are available for two outcomes often exploit the specific nature of those two outcomes, making extensions to higher dimensions far from straightforward. For example, the multivariate vector of responses may consist of outcomes of (many) different types, all requiring different models such as linear mixed models (Chapter 4), generalized linear mixed models (Chapter 14), as well as non-linear mixed models (Chapter 20). Second, even if a plausible joint model can be formulated, fitting of these high-dimensional models can become very cumbersome, unless under unrealistically strong assumptions.

In this chapter, we will focus on the random-effects approach, which can be viewed as an extension of the models discussed in Section 24.3. The model will be introduced in Section 25.2. Many applications of this type of joint models can be found in the statistical literature. For example, the approach has been used in a non-longitudinal setting to validate surrogate endpoints in meta-analyses (Buyse *et al* 2000, Burzykowski *et al* 2001) or to model multivariate clustered data (Thum 1997). Gueorguieva (2001) used the approach for the joint modeling of a continuous and a binary outcome measure in a developmental toxicity study on mice. Also in a longitudinal setting, Chakraborty *et al* (2003) obtained estimates of the correlation between blood and semen HIV-1 RNA by using a joint random-effects model. Other examples with longitudinal studies can be found in MacCallum *et al* (1997), Thiébaut *et al* (2002ab) and Shah *et al* (1997). All these examples refer to situations where the number of different outcomes is (very) low. Although the model formulation can be done irrespective of the number of outcomes to be modeled jointly, standard fitting procedures, such as maximum likelihood estimation, will only be feasible when the dimension in sufficiently low (typically dimension 2 or 3, at most). Therefore, Section 25.3 presents a model-fitting procedure which is applicable, irrespective of the dimensionality of the problem, and explains how inferences can be obtained for all parameters in the joint model. Finally, Section 25.4 applies the methodology for the joint analysis of 7 sets of questionnaires, each consisting of a number of binary outcomes. Other examples, simulation results, and more details on the models as well as on estimation and inference, can be found in Fieuws and Verbeke (2005ab).

In the remainder of this chapter, models for a single longitudinal outcome are called ‘univariate’ models, although they are, strictly speaking, multivariate models since they model a vector of repeated measurements, but all of the same outcome. Similarly, we will use the terminology 'bivariate' and 'multivariate' models to indicate joint longitudinal models for two or more outcomes, respectively.

25.2 Joint Mixed Model

A flexible joint model that meets the requirements discussed in Section 25.1 can be obtained by modeling each outcome separately using a mixed model (linear, generalized linear, or non-linear), by assuming that, conditionally on these random effects, the different outcomes are independent, and by imposing a joint multivariate distribution on the vector of all random effects. This approach has many advantages and is applicable in a wide variety of situations. First, the data can be highly unbalanced. For example, it is not necessary that all outcomes are measured at the same time points. Moreover, the approach is applicable for combining linear mixed models, non-linear mixed models, or generalized linear mixed models. The procedure also allows the combination of different types of mixed models, such as a generalized linear mixed model for a discrete outcome and a non-linear mixed model for a continuous outcome.

Let m be the dimension of the problem, i.e., the number of outcomes that need to be modeled jointly. Further, let Y_{rij} denote the jth measurement taken on the ith subject, for the rth outcome, $i = 1, \ldots, N$, $r = 1, \ldots, m$, and $j = 1, \ldots, n_{ri}$. Note that we do not assume that the same number of measurements is available for all subjects, nor for all outcomes. Let $\boldsymbol{Y_{ri}}$ be the vector of n_{ri} measurements taken on subject i, for outcome r. Our model assumes that each $\boldsymbol{Y_{ri}}$ satisfies a mixed model. Following our earlier notation of the Sections 13.2 and 20.5, let $f_{ri}(\boldsymbol{y_{ri}}|\boldsymbol{b_{ri}}, \boldsymbol{\theta_r})$ be the density of $\boldsymbol{Y_{ri}}$, conditional on a q_r-dimensional vector $\boldsymbol{b_{ri}}$ of random effects for the rth outcome on subject i. The vector $\boldsymbol{\theta_r}$ contains all fixed effects and possibly also a scale parameter needed in the model for the rth outcome. Note that we do not assume the same type of model for all outcomes: A combination of linear, generalized linear, and non-linear mixed models is possible. It is also not assumed that the same number q_r of random effects is used for all m outcomes.

In most applications, it will be assumed that, conditionally on the random effects $\boldsymbol{b_{1i}}, \boldsymbol{b_{2i}}, \ldots, \boldsymbol{b_{mi}}$, the m outcomes $\boldsymbol{Y_{1i}}, \boldsymbol{Y_{2i}}, \ldots, \boldsymbol{Y_{mi}}$ are independent. Extensions of this assumption can be found in Section 24.3 in the context of surrogate markers, or in Fieuws and Verbeke (2005a) in the analysis of the 22 longitudinal sequences of hearing thresholds. Finally, the model is completed by assuming that the vector $\boldsymbol{b}_i$ of all random effects for subject i is multivariate normal with mean zero and covariance D, i.e.,

$$\boldsymbol{b}_i = \begin{pmatrix} \boldsymbol{b_{1i}} \\ \boldsymbol{b_{2i}} \\ \vdots \\ \boldsymbol{b_{mi}} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} D_{11} & D_{12} & \cdots & D_{1m} \\ D_{21} & D_{22} & \cdots & D_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ D_{m1} & D_{m2} & \cdots & D_{mm} \end{pmatrix} \right].$$

The matrices D_{rs} represent the covariances between $\boldsymbol{b_{ri}}$ and $\boldsymbol{b_{si}}$, $r, s = 1, \ldots, m$. Finally, D is the matrix with blocks D_{rs} as entries.

A special case of the above model is the so-called shared-parameter model, which assumes the same set of random effects for all outcomes. An example of this is (24.19), where, in the context of surrogate marker evaluation, a random intercept b_i was used simultaneously in the model for the surrogate outcome as well as in the model for the true outcome. This clearly can be obtained as a special case of the above model by assuming perfect correlation between some of the random effects. The advantage of such shared-parameter models is the relatively low dimension of the random-effects distribution, when compared to the above model. The dimension of the random effects in shared parameter models does not increase with the number of outcomes to be modeled. In the above model, each new outcome added to the model introduces new random effects, thereby increasing the dimension of $\boldsymbol{b}_i$. Although the shared-parameter models can reasonably easy be fitted using standard software (Section 24.5), this is no longer the case for the model considered here. Estimation and inference under the above model will require specific procedures, which will be discussed in Section 25.3. A disadvantage of the shared-parameter model is that it is based on much stronger assumptions about the association between the outcomes, which may not be valid, especially in high-dimensional settings as considered in this chapter.

Note also that, joining valid univariate mixed models does not necessarily lead to a correct joint model. Fieuws and Verbeke (2004) illustrate this in the context of linear mixed models for two continuous outcomes. It is shown how the joint model may imply association structures between the two sets of longitudinal profiles that may strongly depend on the actual parameterization of the individual models and that are not necessarily valid.

As before, estimation and inference will be based on the marginal model for the vector $\boldsymbol{Y}_i$ of all measurements for subject i. Assuming independence of the outcomes conditionally on the vector $\boldsymbol{b}_i$ of random effects, the log-likelihood contribution for subject i equals

$$\begin{aligned} & \ell_i(\boldsymbol{y_{1i}}, \boldsymbol{y_{2i}}, \ldots, \boldsymbol{y_{mi}} | \boldsymbol{\Psi^*}) \\ &= \ln \int \prod_{r=1}^{m} f_{ri}(\boldsymbol{y_{ri}} | \boldsymbol{b_{ri}}, \boldsymbol{\theta_r}) f(\boldsymbol{b}_i | D) d\boldsymbol{b}_i, \end{aligned} \quad (25.1)$$

in which all parameters present in the joint model (fixed effects parameters as well as covariance parameters) have been combined into the vector $\boldsymbol{\Psi}^*$.

Clearly, expression (25.1) shows that the joint model can be interpreted as one mixed-effects model, with conditional density

$$f_i(\boldsymbol{y}_i|\boldsymbol{b}_i) = \prod_{r=1}^{m} f_{ri}(\boldsymbol{y_{ri}}|\boldsymbol{b_{ri}}, \boldsymbol{\theta_r})$$

and with random effect $\boldsymbol{b}_i$. Hence, fitting of the model can, strictly speaking, be based on standard methods and standard software, available for fitting mixed models in general. However, computational problems will arise as the dimension of the random-effects vector $\boldsymbol{b}_i$ in the joint model increases. For example, re-consider the hearing thresholds mentioned earlier. If each of the 22 outcomes is modeled by way of a linear mixed model with random intercepts and random slopes for the time-evolution, then the resulting joint model contains $22 \times 2 = 44$ random effects, resulting in a 44-dimensional matrix D which contains 990 unknown parameters. Even in this case of linear models for continuous data, where the marginal likelihood can be calculated analytically, standard maximization algorithms are no longer sufficient to maximize this marginal likelihood with respect to this many parameters. Moreover, when approximation methods are needed in the calculation of the likelihood, as is the case for generalized or non-linear mixed models (Chapters 14 and 20), maximizing the joint likelihood becomes completely impossible using optimization techniques currently implemented for single outcomes. In Section 25.3, we will describe how estimates and inferences for all parameters can be obtained from pairwise fitting of the model, i.e., from separately fitting the implied joint model for each pair of outcomes.

25.3 Model Fitting and Inference

The general idea behind the pairwise fitting approach is straightforward. Instead of maximizing the likelihood of the full joint model presented in the previous section, all pairwise bivariate models will be fitted separately in a first step. Note the similarity between the pairwise approach used here and the pairwise pseudo-likelihood approach used in the Sections 9.4.1 and 21.3. In a second step, the parameters obtained by fitting the pairwise models will be combined to obtain one single estimate for each parameter in the full joint model.

25.3.1 Pairwise Fitting

The parameters in each univariate model can be estimated by fitting a model for that specific response only. Hence, the only parameters that cannot be estimated by fitting the univariate models are the parameters needed to model the association between the different outcomes. In the model introduced in Section 25.2, these are the parameters in the matrices D_{rs}, $r \neq s$. However, estimation of these parameters does not necessarily require fitting of the complete joint model for all outcomes, it is sufficient to fit all $m(m-1)/2$ bivariate models, i.e., all joint models for all possible pairs

$$(\boldsymbol{Y_1}, \boldsymbol{Y_2}), (\boldsymbol{Y_1}, \boldsymbol{Y_3}), \ldots, (\boldsymbol{Y_1}, \boldsymbol{Y_m}), (\boldsymbol{Y_2}, \boldsymbol{Y_3}), \ldots, (\boldsymbol{Y_2}, \boldsymbol{Y_m}), \ldots, (\boldsymbol{Y_{m-1}}, \boldsymbol{Y_m})$$

of the outcomes $\boldsymbol{Y_1}, \boldsymbol{Y_2}, \ldots, \boldsymbol{Y_m}$. Let the log-likelihood function corresponding to the pair (r, s) be denoted by $\ell(\boldsymbol{y_r}, \boldsymbol{y_s}|\boldsymbol{\Psi_{rs}})$. The vector $\boldsymbol{\Psi_{rs}}$ contains all parameters in the bivariate model for pair (r, s), i.e., the parameters in each of the univariate models, as well as the parameters in D_{rs}.

Let $\boldsymbol{\Psi}$ now be the stacked vector combining all $m(m-1)/2$ pair-specific parameter vectors $\boldsymbol{\Psi_{rs}}$. Estimates for the elements in $\boldsymbol{\Psi}$ are obtained by maximizing each of the $m(m-1)/2$ log-likelihoods $\ell(\boldsymbol{y_r}, \boldsymbol{y_s}|\boldsymbol{\Psi_{rs}})$ separately. It is important to realize that the parameter vectors $\boldsymbol{\Psi}$ and $\boldsymbol{\Psi^*}$ are not equivalent. Indeed, some parameters in $\boldsymbol{\Psi^*}$ will have a single counterpart in $\boldsymbol{\Psi}$, e.g., the parameters in D_{rs}, $r \neq s$, representing covariances between random effects from different outcomes. Other elements in $\boldsymbol{\Psi^*}$ will have multiple counterparts in $\boldsymbol{\Psi}$, e.g., the parameters in D_{rr}, representing variances and covariances of random effects from the same outcome. In the latter case, a single estimate for the corresponding parameter in $\boldsymbol{\Psi^*}$ is obtained by averaging all corresponding pair-specific estimates in $\widehat{\boldsymbol{\Psi}}$. Standard errors of the so-obtained estimates clearly cannot be obtained from averaging standard errors or variances. Indeed, the variability amongst the pair-specific estimates needs to be taken into account. Furthermore, two pair-specific estimates corresponding to two pairwise models with a common outcome are based on overlapping information and hence correlated. This correlation should also be accounted for in the sampling variability of the combined estimates in $\widehat{\boldsymbol{\Psi}}^*$. In the remainder of this section, we will use pseudo-likelihood ideas to obtain standard errors for the estimates, first in $\widehat{\boldsymbol{\Psi}}$, afterwards in $\widehat{\boldsymbol{\Psi}}^*$.

25.3.2 Inference for $\boldsymbol{\Psi}$

Fitting all bivariate models is equivalent to maximizing the function

$$\begin{aligned} p\ell(\boldsymbol{\Psi}) &\equiv p\ell(\boldsymbol{y_{1i}}, \boldsymbol{y_{2i}}, \ldots, \boldsymbol{y_{mi}}|\boldsymbol{\Psi}) \\ &= \sum_{r<s} \ell(\boldsymbol{Y_r}, \boldsymbol{Y_s}|\boldsymbol{\Psi_{rs}}), \end{aligned} \tag{25.2}$$

ignoring the fact that some of the vectors $\boldsymbol{\Psi_{rs}}$ have common elements, i.e., assuming that all vectors $\boldsymbol{\Psi_{rs}}$ are completely distinct. Obviously, (25.2), is of the form (9.3) and hence our pairwise fitting procedure fits within the general framework of pseudo-likelihood (Chapters 9 and 21). Our application of pseudo-likelihood methodology is different from most other applications in the sense that the same parameter vector is usually present in the different parts of the pseudo-likelihood function. Here, the set of parameters in $\boldsymbol{\Psi}_{\boldsymbol{rs}}$ is treated pair-specific, which allows separate maximization of each term in the pseudo log-likelihood function (25.2). In Section 25.3.3, we will account for the fact that $\boldsymbol{\Psi}_{\boldsymbol{rs}}$ and $\boldsymbol{\Psi}_{\boldsymbol{rs'}}$, $s \neq s'$, are not completely distinct, as they share the parameters referring to the rth outcome.

Because the pairwise approach fits within the pseudo-likelihood framework, an asymptotic multivariate normal distribution for $\widehat{\boldsymbol{\Psi}}$ can be derived, using the general pseudo-likelihood theory presented in Section 9.2. More specifically, we have that $\widehat{\boldsymbol{\Psi}}$ asymptotically satisfies

$$\sqrt{N}(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) \quad \approx \quad N(\mathbf{0}, I_0^{-1} I_1 I_0^{-1})$$

in which $I_0^{-1} I_1 I_0^{-1}$ is a ‘sandwich-type’ robust variance estimator, and where I_0 and I_1 can be constructed using first- and second-order derivatives of the components in (25.2). Strictly speaking, I_0 and I_1 depend on the unknown parameters in $\boldsymbol{\Psi}$, but these are traditionally replaced by their estimates in $\widehat{\boldsymbol{\Psi}}$.

25.3.3 Combining Information: Inference for $\boldsymbol{\Psi}^*$

In a final step, estimates for the parameters in $\boldsymbol{\Psi}^*$ can be calculated, as suggested before, by taking averages of all the available estimates for that specific parameter. Obviously, this implies that $\widehat{\boldsymbol{\Psi}}^* = A'\widehat{\boldsymbol{\Psi}}$ for an appropriate weight matrix A. Hence, inference for the elements in $\widehat{\boldsymbol{\Psi}}^*$ will be based on

$$\begin{aligned} \sqrt{N}(\widehat{\boldsymbol{\Psi}}^* - \boldsymbol{\Psi}^*) &= \sqrt{N}(A'\widehat{\boldsymbol{\Psi}} - A'\boldsymbol{\Psi}) \\ &\approx N(\mathbf{0}, A' I_0^{-1} I_1 I_0^{-1} A). \end{aligned} \qquad (25.3)$$

As explained in Section 9.2, pseudo-likelihood methods often are less efficient than full maximum likelihood. However, simulation results of Fieuws and Verbeke (2005ab) suggest that, in the present context, this loss of efficiency is negligible, if any.

25.4 A Study in Psycho-Cognitive Functioning

To illustrate the pairwise approach for fitting high-dimensional multivariate repeated measurements, we analyze data from an experiment in which 105 Dutch-speaking elderly participants (54 females and 51 males) were randomly assigned to one of two physical activity oriented exercise programs. The first is a classical fitness program consisting of 3 weekly visits

TABLE 25.1. *Psycho-Cognitive Functioning. Parameter estimates (standard errors) for the fixed effects in model (25.4) obtained by fitting 7 separate univariate models, as well as obtained by fitting the joint model with the pairwise fitting approach.*

	7 Univariate models	
	$\widehat{\beta}_{r0}$ (s.e.)	$\widehat{\beta}_{r1}$ (s.e.)
Physical well-being	1.63 (0.26)	−0.13 (0.37)
Psychological well-being	1.56 (0.30)	1.22 (0.61)
Self-esteem	1.69 (0.30)	0.43 (0.42)
Physical self-perception	−0.55(0.14)	0.58 (0.24)
Degree of opposition	1.48 (0.17)	0.06 (0.24)
Self-efficacy	1.71 (0.25)	−0.24 (0.33)
Motivation	0.95 (0.11)	−0.35 (0.16)
	Joint model	
	$\widehat{\beta}_{r0}$ (s.e.)	$\widehat{\beta}_{r1}$ (s.e.)
Physical well-being	1.62 (0.25)	−0.12 (0.37)
Psychological well-being	1.71 (0.32)	1.00 (0.68)
Self-esteem	1.68 (0.32)	0.49 (0.39)
Physical self-perception	−0.52 (0.14)	0.52 (0.25)
Degree of opposition	1.47 (0.17)	0.07 (0.24)
Self-efficacy	1.70 (0.23)	−0.22 (0.33)
Motivation	0.94 (0.09)	−0.34 (0.16)

to the gym. The second is a distance coaching program with an emphasis on incorporating physical activities in daily life. One of the aims of the study was to investigate whether the two programs have different impacts on the psycho-cognitive functioning of the participants. Different aspects of psycho-cognitive functioning referring to subjective well-being, self-esteem, self-perception and motivation were considered. A set of questionnaires has been used to measure these different aspects. More specifically, 7 sets of questions (items) were used, originating from different questionnaires and each set consisting of a different number of items: 10 items measuring physical well-being, 14 items for psychological well-being, 10 items for self-esteem, 30 items for physical self-perception, 21 items measuring the degree of opposition to physical activities, 5 items for perceived self-efficacy toward physical activity, and 16 items for motivation for the intervention program. All item scores were dichotomized, with a score equal to one expressing positive psycho-cognitive functioning. All subjects filled in at least one item for each of the seven sets. 64 subjects had no missing information for the 106 items. 20 subjects had one item missing. The missing item scores for the other subjects ranged from 2 to 22. The mean age equals

66.6 years (range 60–76 years) and the mean body mass index (BMI) is 27.0 kg/m^2 (range 20.7–38.0). Questionnaires considered in this analysis were completed by the participants, 6 months after the start of the study.

The aim of our analyses is to assess differences in efficacy between both exercise programs, as well as to study the strength of association between the 7 sets of questionnaires. Although not of a longitudinal nature, this data set clearly is an example of multivariate repeated measurements, of dimension 7, where a number of binary repeated measurements of psycho-cognitive functioning are available for each dimension. The random variable Y_{rij} now denotes the jth measurement (0 or 1), taken on the ith study participant, for the rth questionnaire, $i = 1, \ldots, 105$, $r = 1, \ldots, 7$, and $j = 1, \ldots, n_{ri}$. A score $Y_{rij} = 1$ reflects positive psycho-cognitive functioning, while $Y_{rij} = 0$ is an indication of negative psycho-cognitive functioning.

We will assume that each of the 7 questionnaires satisfy a random-intercepts logistic model, given by

$$\text{logit}[P(Y_{rij} = 1)] \quad = \quad \beta_{r0} + \beta_{r1}\text{DC}_i + b_{ri}, \tag{25.4}$$

in which DC_i is an indicator variable equal to 1 for the participants in the distance coaching program, and zero otherwise. Hence, $\exp(\beta_{r1})$ represents the multiplicative effect of this program on the odds for positive psycho-cognitive functioning measured by the items in questionnaire r, with $r = 1, \ldots, 7$ (1=physical well-being, 2=psychological well-being, 3=self-esteem, 4=physical self-perception, 5=degree of opposition, 6=self-efficacy, and 7=motivation). Note that this model allows for questionnaire-specific intercepts as well as intervention effects. More parsimonious models could be obtained by assuming, for example, the same regression parameters for all questionnaires, or by assuming some random effects to be common to a subset of the questionnaires (i.e., some of the b_{ri} are equal). Correlation between the items of the same set is modeled through the inclusion of the random effects b_{ri}. Correlation between the items of the different questionnaires is implied by the joint distribution for the 7 random intercepts, i.e.,

$$\left(\begin{array}{c} b_{1i}, b_{2i}, b_{3i}, b_{4i}, b_{5i}, b_{6i}, b_{7i} \end{array} \right)' \sim N(\mathbf{0}, \boldsymbol{D}),$$

where $\boldsymbol{D}$ is now the 7×7 unstructured covariance matrix of the random intercepts.

Table 25.1 shows the results from fitting the 7 univariate models separately, as well as from fitting the joint model using the pairwise fitting approach. Very similar estimates as well as inferences are obtained. Using approximate Wald-type tests (Z-tests), the separate analyses show significant differences between both groups on 3 of the 7 questionnaires. The DC-group scores better on physical self-perception and on psychological well-being, but worse on motivation.

TABLE 25.2. *Psycho-Cognitive Functioning. Estimated correlation matrix for the random intercepts in Model (25.4).*

Physical well-being	1.00						
Psychological well-being	0.75	1.00					
Self-esteem	0.55	0.76	1.00				
Physical self-perception	0.66	0.46	0.53	1.00			
Degree of opposition	0.19	0.12	0.23	0.38	1.00		
Self-efficacy	0.29	0.24	0.25	0.36	0.23	1.00	
Motivation	0.42	0.31	0.28	0.40	0.47	0.30	1.00

Using the results from the joint model, an overall test can be constructed for the presence of any systematic difference between both exercise programs. Formally, this corresponds to testing the null-hypothesis

$$H_0 : \beta_{11} = \beta_{21} = \beta_{31} = \beta_{41} = \beta_{51} = \beta_{61} = \beta_{71} = 0$$

versus the alternative that at least one of these parameters differs from zero. Since this null hypothesis is of the general form $H_0 : L'\boldsymbol{\Psi}^* = \mathbf{0}$ for the appropriate matrix L, a Wald-type test (χ^2-test) can easily be derived from the asymptotic distribution (25.3) for $\widehat{\boldsymbol{\Psi}}^*$. This yields a test statistic value equal to 17.84, which is significant when compared to the χ^2_7 distribution ($p = 0.013$). Similarly, other hypotheses of interest can be tested as well.

An additional aim of our analyses was to study the strength of association between the 7 sets of questionnaires. Table 25.2 presents the correlations obtained from the fitted covariance matrix $\widehat{D}$. These correlations express the association between the different constructs underlying each of the seven scales. Performing a principal components analysis (PCA) on the 7×7 correlation matrix of the random effects reveals that the first principal component explains only 49% of the variability. One approach sometimes used to join multiple random-effects models in such a way that the joint model can still easily be fitted using standard software, assumes common random effects for all outcomes, leading to so-called shared-parameter models. An example in a slightly different context can be found in De Gruttola and Tu (1994). More specifically, it is then assumed that all $\boldsymbol{b_{ri}}$ equal $\boldsymbol{b}_i$. In our example, this would lead to univariate random intercepts common to all questionnaires. The advantage would be that this model can very easily be fitted because only one random effect is involved. However, the PCA results suggest that this would be a very unrealistic model for the data set at hand, which could result in biased inferences for the fixed effects of interest (Adams *et al* 1997, Folk and Green 1989).

Figure 25.1 plots the component loadings of the random intercepts for the seven questionnaires on the first two principal components, explaining 49% and 17.4% of the variation. In this reduced representation, we observe,

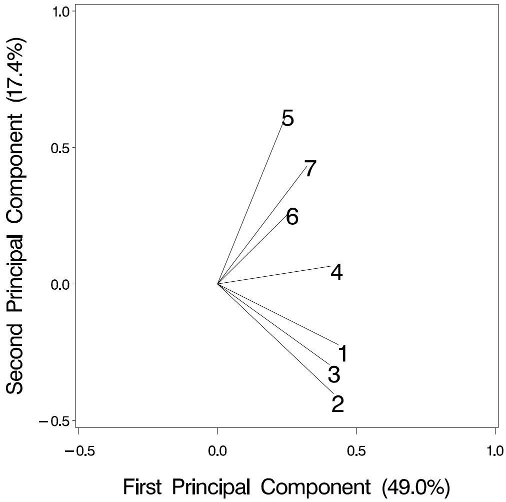

FIGURE 25.1. *Psycho-Cognitive Functioning. Component loadings for the seven questionnaires on the first two principal components for the* 7×7 *correlation matrix of the random intercepts in model (25.4).*
1: physical well-being; 2: psychological well-being; 3: self-esteem; 4: physical self-perception; 5: degree of opposition; 6: self-efficacy; 7: motivation.

not surprisingly, that the scales referring to well-being and self-esteem are strongly correlated with each other, as opposed to their relation with motivational oriented scales.

Part VI

Missing Data