5

Analyzing Longitudinal Data

Longitudinal data, or repeated measures data, can be viewed as multilevel data, with repeated measurements nested within individuals. In a simple application, this leads to a two-level model, with the series of repeated measures at the lowest level, and the individual persons at the highest level. Often, longitudinal data are collected to analyze individual change over time. This requires that the constructs that are under study be measured on a comparable scale at each occasion (cf. Plewis, 1985, 1996). When the time span is relatively short, this does not pose complicated problems. For instance, Tate and Hokanson (1993) report on a longitudinal study where the scores of students on the Beck Depression scale were collected at three occasions during the academic year. In such an application, we may assume that the research instrument remains constant, and that there is sufficient time between the measurements that memory effects are not a problem. In other applications, this may not be the case. For instance, in educational research pupils may be followed for many years. If the dependent variable is a construct like mathematics achievement, it is clear that the researchers cannot use the identical mathematics test at every occasion. Since the mathematical ability of the pupils increases, a test that discriminates well between individual pupils at the beginning of the study, will be much too easy for almost all the pupils at the end of the study. In this case, complex psychometric procedures are needed to ensure that the different tests are calibrated to indicate achievement on comparable scales. Finally, if data are collected that are closely spaced in time, we may expect considerable correlation between measurements collected at occasions that are close together. This should be reflected in the model. Multilevel models for longitudinal data are discussed by, amongst others, Bryk and Raudenbush (1987, 1992) and

Goldstein (1987, 1995), for an introduction see Snijders (1996) and Van der Leeden (1998).

5.1 FIXED AND VARYING OCCASIONS

It is useful to distinguish between repeated measures that are collected at fixed or varying occasions. If the measurements are taken at fixed occasions, all individuals provide measurements at the same set of occasions, usually regularly spaced, such as every year. When occasions are varying, we have a different set of measures taken at different points in time for different individuals. Such data occur, for instance, in growth studies, where physical or psychological characteristics are studied for a set of individuals at different moments in their development. The data collection could be at fixed moments in the year, but the individuals would have different ages at that moment. Or the original design is a fixed occasion design, but due to planning problems the data collection does not take place at the intended moments. For a multilevel analysis of the resulting data, the difference between fixed and varying occasions is not very important. For fixed occasion designs, especially when the occasions are regularly spaced and there are no missing data, repeated measures analysis of variance is a viable alternative for multilevel analysis. A comparison of the analysis of variance approach and multilevel analysis is given in the section 5.2. Another possibility in such designs is latent curve analysis, also known as growth curve analysis. This is a structural equation model (cf. Willett & Sayer, 1994; Duncan & Duncan, 1995) that models a repeated measures polynomial analysis of variance. This will be treated in chapter XX.

5.2 Example with Fixed Occasions

The example data are a longitudinal data set, with longitudinal data from 200 college students. The students' Grade Point Average (GPA) has been recorded for 6 successive semesters. At the same time, it was recorded whether the student held a job in that semester, and for how many hours. This is recorded in a variable 'job' (with categories 0=no job, 1=1-2 hours, 2=3-5 hours, 3=5-10 hours, 4=more than 10 hours), which for the purpose of this example is treated as an interval level variable. In this example, we also use the student variables high school GPA and sex (1=male, 2=female).

In a statistical package such as SPSS or SAS, these data are typically stored with the students defining the cases, and the repeated measurements as a series of variables, such as GPA1, GPA2,…, GPA6, and JOB1, JOB2,…, JOB6. For example, in SPSS the data structure would be as shown in Figure 5.1.

1:student 1

	student	sex	highgpa	gpa1	gpa2	gpa3	gpa4	gpa5	gpa6	job1	job2	job3	job4	job5	job6
1	1	2	2.8	2.3	2.1	3.0	3.0	3.0	3.3	2	2	2	2	2	2
2	2	1	2.5	2.2	2.5	2.6	2.6	3.0	2.8	2	3	2	2	2	2
3	3	2	2.5	2.4	2.9	3.0	2.8	3.3	3.4	2	2	2	3	2	2
4	4	1	3.8	2.5	2.7	2.4	2.7	2.9	2.7	3	2	2	2	2	2
5	5	1	3.1	2.8	2.8	2.8	3.0	2.9	3.1	2	2	2	2	2	2
6	6	2	2.9	2.5	2.4	2.4	2.3	2.7	2.8	2	3	3	2	3	3
7	7	1	2.3	2.4	2.4	2.8	2.6	3.0	3.0	3	2	3	2	2	2
8	8	2	3.9	2.8	2.8	3.1	3.3	3.3	3.4	2	2	2	2	2	2
9	9	1	2.0	2.8	2.7	2.7	3.1	3.1	3.5	2	2	3	2	2	2
10	10	1	2.8	2.8	2.8	3.0	2.7	3.0	3.0	2	2	2	3	2	2
11	11	2	3.9	2.6	2.9	3.2	3.6	3.6	3.8	2	3	2	2	2	2
12	12	2	2.9	2.6	3.0	2.3	2.9	3.1	3.3	3	2	2	2	2	2
13	13	1	3.7	2.8	3.1	3.5	3.6	3.9	3.9	2	2	2	2	2	2
14	14	2	3.5	2.4	3.0	2.9	3.0	3.3	3.4	2	2	2	2	2	2

Figure 5.1. Repeated measures data structure in SPSS.

The data structure for a multilevel analysis of these data is generally different, depending on the specific program that is used. Most multilevel software requires that the data is structured with the measurement occasions defining the cases, and student level variables repeated over the cases. For instance, in for the programs MLwiN or MixReg the data structure should be as shown in Figure 5.2.

1:student 1

	student	occasion	gpa	job	sex	highgpa
1	1	1	2.3	2	2	2.8
2	1	2	2.1	2	2	2.8
3	1	3	3.0	2	2	2.8
4	1	4	3.0	2	2	2.8
5	1	5	3.0	2	2	2.8
6	1	6	3.3	2	2	2.8
7	2	1	2.2	2	1	2.5
8	2	2	2.5	3	1	2.5
9	2	3	2.6	2	1	2.5
10	2	4	2.6	2	1	2.5
11	2	5	3.0	2	1	2.5
12	2	6	2.8	2	1	2.5
13	3	1	2.4	2	2	2.5
14	3	2	2.9	2	2	2.5

Figure 5.2. Repeated measures data structure for multilevel analysis.

The multilevel regression model for longitudinal data is a straightforward application of the multilevel regression model described in chapter two. It can also be written as a sequence of models for each level. At the lowest, the repeated measures level, we have:

$$Y_{tj} = \pi_{0i} + \pi_{1i}T_{ti} + \pi_{2i}X_{ti} + e_{ti}. \quad (5.1)$$

In repeated measures applications, the coefficients at the lowest level are often indicated by the Greek letter π. This has the advantage that the person level coefficients, which are in repeated measures modeling at the second level, can be represented by the usual Greek letter β, and so on. In equation (5.1), Y_{ti} is the dependent variable of individual i measured at time point t, T is the time variable that indicates the time point, and X_{ti} is a *time varying covariate*. For example, Y_{ti} could be the GPA of a student at at time point t, T_{ti} indicates the occasion at which the GPA is measured, and X_{ti} the job status of the student at time t. Student characteristics, such as gender, are *time invariant covariates*, which enter the equation at the second level:

$$\pi_{0i} = \beta_{00} + \beta_{01}Z_i + u_{0i}, \tag{5.2}$$

$$\pi_{1j} = \beta_{10} + \beta_{11}Z_i + u_{1i}, \tag{5.3}$$

$$\pi_{2i} = \beta_{20} + \beta_{21}Z_i + u_{2i}. \tag{5.4}$$

By substitution, we get the single equation model:

$$Y_{tj} = \beta_{00} + \beta_{10}T_{ti} + \beta_{20}X_{ti} + \beta_{01}Z_i + \beta_{11}Z_iT_{ti} + \beta_{21}Z_iX_{ti} + u_{1i}T_{ti} + u_{ij}X_{ti} + u_{0i} + e_{ti} \tag{5.5}$$

Using variable labels instead of letters, the equation becomes

$$GPA_{tj} = \beta_{00} + \beta_{10}Time_{ti} + \beta_{20}Job_{ti} + \beta_{01}Sex_i + \beta_{11}Sex_iTime_{ti} + \beta_{21}Sex_iJob_{ti} + u_{1i}Time_{ti} + u_{ij}Job_{ti} + u_{0i} + e_{ti} \tag{5.6}$$

In longitudinal research, we sometimes have repeated measurements of individuals, who are all measured together on a small number of fixed occasions. This is typically the case with experimental designs involving repeated measures and panel research. If we simply want to test the null hypothesis that the means are equal for all occasions, we can use repeated measures Analysis of Variance. If we use repeated measures univariate analysis of variance (Stevens, 1996, p455), we must assume *sphericity*. Sphericity means that there are complex restrictions on the variances and covariances between the repeated measures, for details see Stevens (1996, chapter 13). A specific form of sphericity, which is easily understood, is *compound symmetry*, sometimes referred to as *uniformity*. Sphericity requires that all population variances of the repeated measures are equal, and that all population covariances of the repeated measures are also equal. If sphericity is not met, the F-ratio used in Analysis of Variance is positively biased, and we reject the null hypothesis too often. An alternative approach is to specify the repeated measures as observations on a multivariate response vector and use Multivariate Analysis of Variance (MANOVA). This does not require sphericity, and is generally considered the approach of choice if analysis of variance is to be used (cf. O'Brien & Kaiser, 1985; Stevens, 1996). However, the multivariate test is more complicated, because it is based on a transformation of the repeated measures, and what is tested are actually contrasts among the repeated measures.

A MANOVA analysis of the example data using the General Linear Model in SPSS (SPSS Inc., 1997) cannot easily incorporate a time-varying covariate such as job status. MANOVA can be used to test the trend over time of the repeated GPA measures by specifying polynomial contrasts, and to test the effects of sex and high school GPA. Sex is a dichotomous variable, that is entered as a factor, and high school GPA is a continuous variable that is entered as a covariate. Table 5.1 presents the results of the significance tests.

Table 5.1. MANOVA significance tests on GPA example data.

Effect tested:	F	df	p
GPA (repeated)	4.53	5/193	.001
GPA (linear trend)	12.77	1/197	.000
GPA*HighGPA	0.87	5/193	.505
GPA*Sex	1.42	5/193	.220
HighGPA	9.16	1/197	.003
Sex	7.23	1/197	.000

The MANOVA results indicate that there is a significant linear trend for the GPA measures. The higher polynomial trends, which are not in the Table, are not significant. Both Sex and High school GPA have significant effects, and there are no interactions. Table 5.2 presents the GPA means, rounded to one decimal, for the six occasions, for male and female students.

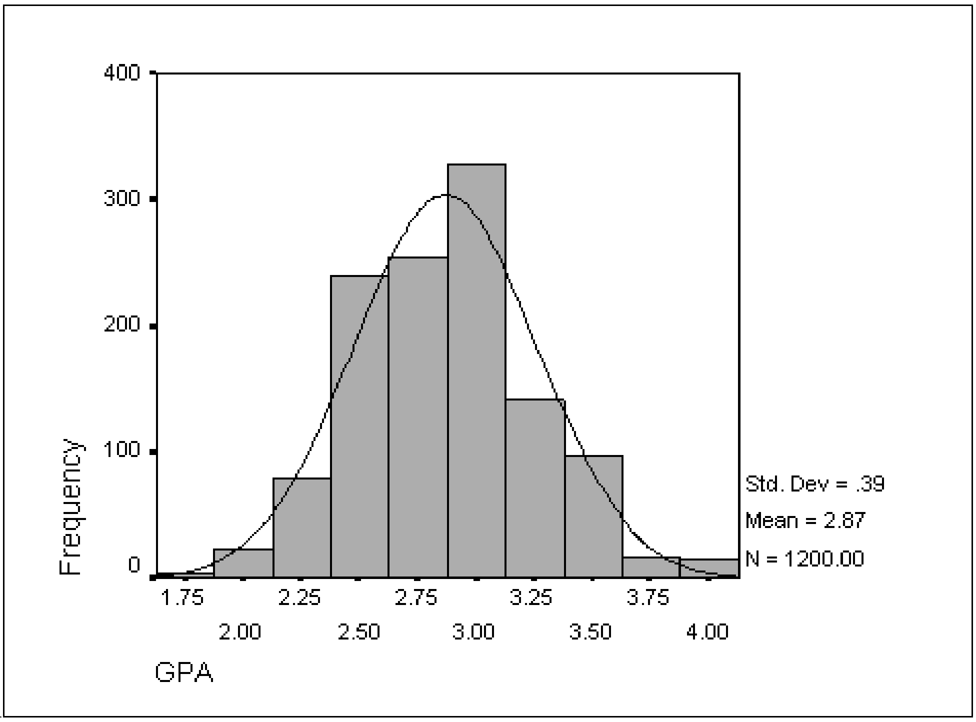
Table 5.2. GPA means at six occasions, for male and female students

Occasion:	1	2	3	4	5	6	Total
Male	2.6	2.7	2.7	2.8	2.9	3.0	2.8
Female	2.6	2.8	2.9	3.0	3.1	3.2	2.9
All students	2.6	2.7	2.8	2.9	3.0	3.1	2.9

Table 5.2 makes clear that there is a linear upward trend of about 0.1 for each successive GPA measurement. Female students have a GPA that is consistently about 0.1 higher than the male students. Finally, the SPSS output also contains the regression coefficient for the High school GPA at the six occasions; this coefficient varies but is generally positive, which indicates that students, who have a high GPA in High school, generally have a relatively high GPA in college.

In the multilevel regression model, the development over time is often modeled by a linear regression equation, with possibly different regression coefficients for different individuals. Thus, each individual can have their own regression curve, specified by individual regression coefficients that may depend on individual attributes. Quadratic and higher functions can be used to model nonlinear dependencies on time, and both time varying and person level covariates can be added to the model. It is useful to code the time points *T* as *t*=0,1,2,3,4,5. As a result, the intercept in the null-model can be interpreted as the expected value at the first occasion. Using time points t=1,2,3,4,5,6 would be completely equivalent, but slightly more difficult to interpret.

Before we start the analysis, we examine the distribution of the outcome variable GPA in the disaggregated data file with 200x6=1200 observations. The histogram with embedded best fitting normal curve is in Figure 5.3. The distribution is fairly normal, so we can proceed with the analysis.



11

Figure 5.3 Histogram of GPA values in disaggregated data file.

Table 5.3 presents the results of a multilevel analysis of these longitudinal data. Model (1) is a model that contains only an intercept term; this model serves as a null model. The intercept of 2.87 in this model is simply the overall mean for GPA. The intercept-only model estimates the repeated measures variance as 0.098, and the person level variance as 0.057 (because these numbers are so small, they are given in 3 decimals). This estimates the total GPA variance as 0.155. Using equation (2.9), we estimate the intraclass correlation or the proportion variance at the person level is estimated as $\rho = 0.057/0.155 = 0.37$. About one-third of the variance of the GPA measures is variance between individuals, and about two-third is variance within individuals across time.

In model (2), the Time variable is added as a predictor with the same coefficient for all persons. The model predicts a value of 2.60 at the first occasion, which increases by 0.11 on each succeeding occasion. Just as in the MANOVA analysis, adding higher order polynomial trends for Time to the model does not improve prediction. Model (3) adds the time varying covariate Job status to the model. The effect of Job status is clearly significant; the more hours are worked, the lower the GPA. Model (4) adds the person level (time invariant) predictors High school GPA and Sex. Both effects are significant; high school GPA correlates with average GPA in college, and female students perform better than male students.

If we compare the variance components of model (1) and model (2), we see that entering the time variable has decreased the occasion level variance, while increasing the person level variance by 11%. This is typical for multilevel analysis of repeated measures. It makes it difficult to apply the suggestion by Bryk and Raudenbush (1992) to use the residual error variance of the intercept-only model as a benchmark, and to examine how much this goes down when explanatory variables are added to the model.

Table 5.3. Results multilevel analysis of GPA, fixed effects only

Model:	M1: null model		M2: + Time		M3: + Job status		M4: + High sch. GPA & Sex	
Fixed part								
Predictor	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
intercept	2.87	.02	2.60	.02	2.97	.04	2.49	.11
time			0.11	.00	0.10	.00	0.10	.00
job status					-0.17	.02	-0.17	.02
high GPA							0.09	.03
sex							0.15	.03
Random part								
intercept$_1$	0.098	.004	0.058	.003	0.055	.002	0.055	.002
intercept$_2$	0.057	.007	0.063	.007	0.052	.006	0.045	.005
Deviance	913.5		393.7		308.4		282.8	
AIC	919.5		401.7		318.4		296.8	

In Table 5.3, this strategy leads to inconsistencies, because in model (2) the second level residual error variance for the intercept actually goes up when the time variable is added to the model! The reason is, as is explained in chapter 2, that in multilevel models with random coefficients the notion of 'amount of variance explained at a specific level' is not a simple concept. As Snijders and Bosker (1994) explain in detail, the problem arises because the statistical model behind multilevel models is a hierarchical sampling model: groups are sampled at the higher level, and at the lower level individuals are sampled within groups. This sampling process creates some variability in all variables between the groups, even if there are in fact no real group differences. In time series, the lowest level is a series of measurements, which in many cases are evenly spaced and (almost) the same for all individuals in the sample. Therefore, the variability between persons in the time series variable is usually *much* higher than the hierarchical sampling model assumes. Consequently, the intercept-only model overestimates the variance at the occasion level, and underestimates the variance at the person level. The time variable that is included in the second model, models the occasion level variance in the dependent variable GPA. Conditional upon this effect, the variance estimated at the person level is much more realistic.

Chapter 2 in this book describes procedures based on Snijders and Bosker (1994) to correct the problem. Snijders and Bosker (1999) describe the present the reasoning and the procedures themselves in more detail.

With complex models, the procedures proposed by Snijders and Bosker (1994, 1999) are complicated. A simple approximation is to use as a benchmark for the occasion level variance the variance component estimated in the intercept only model (model 1). For the person level variance we use the variance component of the model that includes the time variable (model 2). Then we observe that the error variance at the occasion level goes down from 0.098 to 0.058, which means that the time variable explains about 41% of the GPA variance between the occasions. To see how much variance the time varying covariate Job status explains, we compare the occasion level variance components from models (3) and (1), which shows that model (3) explains about 44% of the occasion level variance. Clearly, the effect of the linear rend over time is much stronger than the effect of Job status.

To assess the effects on the person level variables, we regard the person level intercept variance of 0.063 in model (2) as the error variance, and observe that in model (3) this variance decreases to 0.052. This means that adding Job status to the model explains about 18% of the variance between persons. Apparently, although Job status is a time-varying covariate, is typically shows more variation between persons than from one semester to the next. In model (4) the variance estimate it goes down to 0.045. Thus, at the person level, the covariates High school GPA and Sex explain an additional 10% of the variation between the persons.

The models presented in Table 5.3 all assume that the rate of change is the same for all individuals. In the models presented below in Table 5.4, the regression coefficient of the Time variable is assumed to vary across individuals.

Table 5.4 Results multilevel analysis of GPA, including random slopes

Model:	M5: + Time random		M6:+ cross level interaction		
Fixed part					
Predictor	coeff.	s.e.	coeff.	s.e.	stand.coeff.
intercept	2.44	.10	2.51	.11	
time	0.10	.01	0.06	.02	0.26
job status	-0.13	.02	-0.14	.02	-0.19
high GPA	0.09	.03	0.09	.03	0.14
sex	0.12	.03	0.08	.04	0.10
time*sex			0.03	.01	0.23
Random part					
intercept$_1$	0.042	.002	0.042	.002	
intercept$_2$	0.038	.006	0.038	.006	
time$_2$	0.004	.001	0.004	.001	
int*time$_2$	-0.002	.002	-0.002	.001	
$r_{int*time}$	-0.21		-0.19		
Deviance	170.1		163.0		
AIC	188.1		183.0		

In model (5), the slope of the time variable is allowed to vary across individuals. In this model, the variance components for the intercept and the regression slope for the time variable are both significant. The significant intercept variance means that individuals have different initial states, and the significant slope variance means that individuals also have different rates of change. In model (6), the interaction of the time variable with the person level predictor Sex is added to the model. The interaction is significant, but including it does not decrease the slope variance for the time variable.

The variance component of 0.38 for the slopes of the Time variable does not seem very large. However, model (5) and (6) both assume a normal distribution for these slopes, with a standard deviation of $\sqrt{0.038} = 0.195$. Compared to the value of 0.10 for the average time slope in model (5), this is not small. There is appreciable variation among the time slopes, which is not modeled very well using the available student variables.

In both model (5) and (6) there is a small negative correlation between the initial status and the growth rate; students who start with a relatively low value of their GPA, increase their GPA faster than the other students. Note that the correlation between the intercept and slope is different in model (5) and (6). In a model without other predictors except the time variable this correlation can be interpreted as an ordinary correlation, but in models (5) and (6) it is a partial correlation, conditional on the predictors in the model.

To facilitate interpretation, standardized regression coefficients have been calculated for the last model in Table 5.4 using equation (2.14). The standardized regression coefficients indicate that the change over time is the largest effect. Job status also seems important, and the interaction seems more important now. To investigate this further, we can construct the regression equation of the time variable separately for both male and female students. Substituting the mean for all other variables in the equation for model (6), we get for male students the regression equation: GPA = 2.47 + 0.09 * Time, and for female students the regression equation GPA = 2.55 + 0.12 * Time.

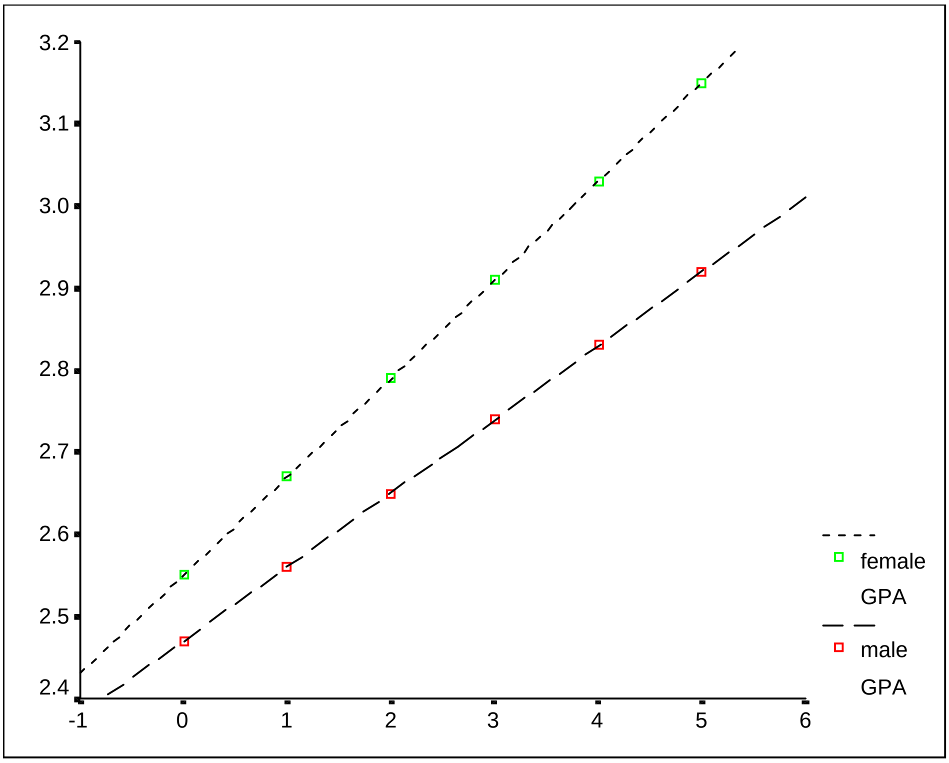


Figure 5.4 Regression lines for Time, separate for male and female students

Figure 5.4 presents a plot of these regression lines. The expected difference between male and female students, which is 0.08 in the first semester, increases to 0.23 in the second semester.

Since the time variable is coded in such a way that the first occasion is coded as zero, the negative correlation between the intercepts and slopes refers to the situation on the first measurement. As is explained in section 4.2 of chapter 4, the estimates of the variance components in the random part can change of the scale of the explanatory variables is changed. In many models this is not a real problem, because the interest is mostly in estimation and interpretation of the regression coefficients in the fixed part of the model. In repeated measures analysis, the correlation between the intercepts and the slopes of the time variable is often an interesting parameter, to be interpreted along with the regression coefficients. In this case, it is important to realize that the correlation is not invariant; it changes if the scale of the time variable is changed. Table 5.5 illustrates this point. In Table 5.5, we have the parameter estimates for model (5) in Table 5.4, for different scalings of the time variable. In model 5a, the time variable is scaled as in all our analyses so far, with the first occasion coded as zero. In model 5b, the time variable is coded with the last occasion coded as zero, and the earlier occasions with negative values –5,…,-1. In model 5c, the time variable is centered around its overall mean.

Table 5.5. Results model 5 for different time scalings

Model:	M5a 1^{st} occasion = 0		M5b last occasion = 0		M5c time centered	
Fixed part						
Predictor	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
intercept	2.44	.10	2.96	.11	2.70	.10
time	0.10	.01	0.10	.01	0.10	.01
job status	-0.13	.02	-0.13	.02	-0.13	.02
high GPA	0.09	.03	0.09	.03	0.09	.03
sex	0.12	.03	0.12	.03	0.12	.03
Random part						
$intercept_1$	0.042	.002	0.042	.002	0.042	.002
$intercept_2$	0.038	.006	0.109	.013	0.050	.006
$time_2$	0.004	.001	0.004	.001	0.004	.001

int*time2	-0.002	.002	0.017	.003	0.007	.001
$r_{int*time}$	-0.21		0.82		0.51	
Deviance	170.1		170.1		170.1	
AIC	188.1		188.1		188.1	

From the correlations in the three models we conclude in model 5b that students who end with a relatively high GPA, on average have a steeper GPA increase over time. In the centered model, 5c, this correlation is lower, but still quite clear. Note that the three models have exactly identical estimates for all parameters that do not involve Time, and also exactly the same deviance. The models are in fact equivalent. The different ways that time is coded lead to what statisticians call a *re-parameterization* of the model. The three models all describe the data equally well, and are equally valid. Nevertheless, they are not identical. The situation is comparable to viewing a landscape from different angles. The landscape does not change, but some views are more interesting than others are. The important lesson here is that in repeated measures analysis, careful thought must be given to the coding of the Time variable.

5.3 EXAMPLE WITH VARYING OCCASIONS

The data in the next example are a study of children's vocabulary development (Huttenlocher, Haight, Bryk & Seltzer, 1991), which is included with the HLM program (Bryk, Raudenbush & Congdon, 1996), and discussed at length by Bryk and Raudenbush (1992). The data set combines data from two studies. In the first study, children were observed on six or seven occasions between 12 to 26 months of age. In the second study, children were observed on four occasions between 12 and 24 months of age. For some cases, the 14-month occasion is missing. On each occasion, the vocabulary size was measured. In addition, we have the children's sex (coded 0 for male and 1 for female), and the amount of maternal speech measured at the 16-month data point (Bryk & Raudenbush, 1992, p 141).[1]

[1] The description of the vocabulary growth data here differs slightly from the description in

Although the measures are taken at regular ages, combining two studies and having missing data at one time point results in a quite unbalanced data file, as shown in Table 5.6. The total number of observations in these data is 126, on 22 children, or about 5.7 occasions per child. So, the sample size at both the child and the occasion level is rather small. In addition, for 11 children we have data on four occasions only. If we fit models with a quadratic trend, allowing varying coefficients, the fit for these children will almost be perfect.

Table 5.6. Data structure of childrenn's vocabulary data

Age-point (months)	12	14	16	18	20	22	24	26
Number of children	22	5	22	11	22	11	22	11

Most Manova software removes incomplete cases from the analysis, which for the data in Table 5.6 would result in removal of 17 out of 22 cases. In multilevel analysis, different numbers of observations at different time intervals pose no special problems, and all cases can be included in the analysis. Growth generally follows a nonlinear pattern, so the square of the child's age is added to the explanatory variables.

The children's vocabulary data illustrate a number of points that are important in multilevel analysis of longitudinal data. Since the data are a combination of data collected in two separate studies, a dummy variable is added to the data to indicate the study, to model possible differences between the studies that may result from differences in the research procedures. This dummy variable is coded –0.5 for the first study, and +0.5 for the second study. This coding implies that the intercept and the variance components, which are estimated for the value zero on this explanatory variable, refer to the average study result, disregarding the different number of observations for each study. Since the difference between –0.5 and +0.5 equals 1.0, the regression coefficient for the study variable indicates the difference between the two studies. The study indicator is viewed as a control variable. Therefore, it is included in each and every model, including the intercept-only model, to statistically control for possible differences between the two studies.

Bryk & Raudenbush (1992), to reflect the actual composition of the data. Bryk & Raudenbush give the number of occasions for the second study as three, which should be four.

Thus, the intercept-only model, including the study indicator, is given by equation (5.7) as follows:

$$Y_{tj} = \pi_{0i} + \pi_{1i}T_{ti} + e_{ti} \quad (5.7)$$

or, using verbal labels for the variables:

$$Vocabulary_{tj} = \pi_{0i} + \pi_{1i}\,Study_{ti} + e_{ti}. \quad (5.8)$$

If we fit the intercept-only model of equation (5.7) (including the study-variable), the intercept is estimated as 132, which means that the average child in our data, at the average age in our data, knows about 132 words. If we add Age to the equation, the intercept is estimated as –420! In a regression equation, the intercept is the expected outcome when all explanatory variables have the value zero. Thus, the intercept of –420 indicates, that at age-point zero the children's vocabulary size is negative. This clearly illustrates the importance of using a zero-point for the time scale that is meaningful in some sense. For the moment, we decide to center the age variable on its overall mean of 18.9 months. The quadratic trend is now modeled by the square of the centered age-variable. Table 5.7 presents the results of a sequence of models for the centered age variables.

Table 5.7 Results vocabulary data, occasion level only, age centered on mean

Model:	M1		M2		M3		M4	
	Intercept only		+ age		+ age squared		+ age random	
Fixed part								
Predictor	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
intercept	132	20	138	18	85	19	88	17
study	-111	39	-70	35	-68	35	-4	5
age (cent.)			30	2	31	2	28	3
age squared					2.6	0.4	2.6	.2
Random part								
intercept$_1$	31073	4290	10379	1437	7721	1071	940	140
intercept$_2$	2403	2459	4878	2075	5169	1994	6344	1948
age$_2$							236	73
int*age$_2$							1246	376
$r_{int*age}$							1.02	
Deviance	1669		1551		1519		1290	

AIC	1677	1561	1531	1304

The sequence of models 1-3 shows again, that the interpretation of variance components in a multilevel model is not as straightforward as that of variances of raw data. In model (1), the intercept-only model, the child level residual variance is estimated as 4669. In model (2), which includes the centered age variable as a predictor variable, the child level residual variance *increases* to 5951. In model (3), which adds the squared trend for age, the child level variance increases again to 6223. We have observed the same effect in the previous example using the GPA data, although here the increase in the variance is much larger. The explanation of this anomalous result is, as is explained in chapter 2, that in multilevel models with random coefficients the notion of 'amount of variance explained at a specific level' is not a simple concept. The problem arises because the statistical model behind multilevel models is a hierarchical sampling model: groups are sampled at the higher level, and at the lower level individuals are sampled within groups. This sampling process creates some variability between the groups, even if there are in fact no real group differences. In time series, the lowest level is a series of measurements, which in many cases are (almost) the same for all individuals in the sample. As a consequence, the variability between persons in the time series variable is usually *much* higher than the hierarchical sampling model assumes. So, in the intercept-only model the occasion level variance is overestimated, and the child level variance is underestimated. Including the age variable, which models the occasion level variance in the vocabulary measures, results in much more realistic estimates for both the occasion level and the child level variances. Adding the squared trend for the age variable improves the estimates again. Adding higher order trends does not improve prediction significantly.

In section 5.2, which discusses fixed occasion data, the suggestion was made to use as a benchmark for the occasion level variance the variance component estimated in the intercept only model (model 1), and for the person level variance the variance component of the model that includes the time variable. This example illustrates the importance of having a well-specified model for the time variable; if that includes quadratic and higher trends, the benchmark for the person level variance is the residual variance in the model that includes all time effects.

Model (4) models the age effect as a random coefficient. The variance of the slope coefficient for age is significant. Allowing the slopes for the squared age variable to vary leads to a model that does not converge.[1] In model (4), the correlation between the slope and intercept residuals is estimated as 1.02, which is not an admissible value. It is close to the range of admissible values, so we can accept the estimated value, and interpret this result as an indication that the correlation in the population is very high, close to 1.00.

Before going on by adding child level explanatory variables, we will repeat the analysis with a slightly different parameterization. In their extensive analysis of this data set, Bryk and Raudenbush (1992, pp.141-147) argue that, since children generally begin to utter their first words around the age of 12 months, it makes sense to express the age variable as centered on the age of 12 months. As a result, the estimated values of the intercept and the variances in the intercept-only model refer to the situation at the age of 12 months, with vocabulary size almost zero and little variation between the children.

The same sequence of models as in Table 5.7, but now for age centered at 12 months, and age squared computed from the age variable centered at age 12 produces the results shown in Table 5.8. The results resemble those in Table 5.7. In the final model, the intercept is very small and insignificant, indicating that at age 12 months the average vocabulary size is zero. The variance around the intercept is also insignificant. The coefficient for age is now also insignificant, which indicates that at age 12 months the vocabulary growth is almost zero, which in a graph would give almost a horizontal line. The age-squared variable, which indicates the acceleration in growth, is significant at 2.2. So each month, the number of words learned increases by a factor of 2.2 times the squared difference. So at 14 months, the expected vocabulary growth is 8.8 or about 9 words, at 16 months it is 35, at 18 months it is 79 words, and so on.

[1] In MLwiN the model does not converge. The model with varying slopes also exhibits convergence problems, evidenced by slow convergence and numeric warnings. The scaling of the variables may cause convergence problems, because they lead to large numbers for the variance components. If this is the case, a transformation like dividing the outcome variable by 10 or 100 often helps. In this particular instance, rescaling the outcome variable does not solve the convergence problems.

Table 5.8 Results vocabulary data, occasion level only, age centered on 12 months

Model:	M1 Intercept only		M2 + age		M3 + age squared		M4 + age random	
Fixed part								
Predictor	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
intercept	132	20	-66	22	-5	23	-2	7
study	-111	39	-70	35	-68	35	-4	5
age (12)			30	2	4.6	5.9	-2.1	7.3
age squared					2.6	0.4	2.2	0.2
Random part								
intercept$_1$	31073	4290	10377	1438	7721	1071	940	140
intercept$_2$	2403	2459	4882	2067	5169	1994	399	274
age$_2$							240	75
int*age$_2$							-388	137
$r_{int*age}$							-1.25	
Deviance	1669		1551		1519		1290	
AIC	1677		1561		1531		1304	

As mentioned above, the vocabulary growth data are included as an example with the HLM program, and are discussed at length by Bryk and Raudenbush (1992). For reasons to be explained further on, it is instructive to compare the results in Table 5.8, which are estimated using MLwiN (vs. 1.03), to results estimated using HLM (vs. 4.04). The results are presented in Table 5.9.

Table 5.9 Results vocabulary data, occasion level only, age centered on 12 months (HLM)

Model:	M1 Intercept only		M2 + age		M3 + age^2		M4 + age random		M5 + age, age^2 random	
Fixed part										
Predictor	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
intercept	132	20	-66	22	-5	6	-3	8	-4	6
study	-111	40	-70	37	-68	36	-5	8	-1	7
age (12)			30	2	2.6	0.4	-1.8	4	-0.4	2
age squared					2.6	0.4	2.2	0.2	2.0	0.2
Random part		χ^2_{20}		χ^2_{20}		χ^2_{20}		χ^2_{20}		χ^2_{19}

$intercept_1$	31209	-	10498	-	7883	-	888	-	678	-
$intercept_2$	2994	28.9	5471	87.9	5749	117.0	817	40.1	120	6.5
age_2							257	75	56	33.2
age^2_2									0.5	48.7
Deviance	1656		1536		1507		1297		1272	
AIC	1670		1540		1511		1305		1286	

One of the differences between Table 5.8 and Table 5.9 is the significance test on the variance components: MLwiN produces a standard error for the variances, while HLM carries out a chi-square test. More important differences are that HLM does not experience convergence problems, estimates somewhat different variance components, and is able to fit a model with both intercepts and slopes for age varying across children. These differences are not the result of using Restricted ML (in HLM), rather than Full ML (in MlwiN) estimation. If we use RIGLS estimation in MlwiN (equivalent to RML in HLM) we obtain estimates close to the earlier MlwiN estimates in Table 5.8. The most likely cause for the difference between the MlwiN and HLM results is the differece in the algorithm used. HLM uses the EM algorithm, which is sometimes slow, but has the advantage that it is very stable, and cannot produce estimates that are outside the acceptable parameter space. Indeed, the correlation between the slopes and intercepts, which MlwiN estimates as -1.25, is estimated by HLM as –0.99. This is very close to 1.00, but it is an acceptable value. In model (5) in Table 5.9, the correlations between the random coefficients range from 0.84 to 0.98, which indicates high collinearity for these parameters. Apparently, this poses a problem for the less robust estimation method used in MlwiN.

5.4 ADVANTAGES OF MULTILEVEL ANALYSIS FOR LONGITUDINAL DATA

Using multilevel models to analyze repeated measures data has several advantages. Bryk and Raudenbush (1992) mention five key points. First, by modeling varying regression coefficients at the occasion level, we have growth curves that are different for each person. This fits in with the way individual development is generally conceptualized (cf. Willett, 1988). Second, the number of repeated measures and their spacing may differ across persons. Other analysis method for longitudinal data cannot handle such data well. Third, the covariances between the repeated measures can be modeled as well, by specifying a specific structure for the variances and covariances at either level. Fourth, if we have balanced data and use RML estimation, the usual Analysis of Variance based F-tests and t-tests can be derived from the multilevel regression results (cf. Raudenbush, 1993 ; Maas & Snijders, 1999). This shows that Analysis of Variance on repeated measures is a specific case of the more general multilevel regression model. Fifth, in the multilevel model it is simple to add higher levels, to investigate the effect of family or social groups on individual development. A sixth advantage is that, in multilevel regression, it is straightforward to include time varying or time constant explanatory variables to the model, which allows us to model both the average group development and individual development.

5.5 SOME STATISTICAL ISSUES

5.5.1 Investigating and Analyzing Patterns of Change

In the previous sections, polynomial curves were used to model the pattern of change over time. Polynomial curves are often used for estimating developmental curves. They are convenient, because they can be estimated using standard linear modeling procedures, and they are very flexible. If there are *k* measurement occasions, they can be fitted exactly using a polynomial of degree *k-1*. In general, in the interest of parsimony, a polynomial of a lower degree would be preferred. Many inherently nonlinear functions can be approximated reasonably well by a polynomial function. Inherently nonlinear functions can be desirable, because they are considered to reflect some 'true' developmental process. For instance, Burchinal and Appelbaum (1991) consider the logistic growth curve and the exponential curve of special interest for developmental models. The logistic curve describes a developmental curve where the rate of development changes slowly in the beginning, accelerates in the middle, and slows again at the end. In the light of the example data used in section 5.3, it is interesting to note that they explicitly mention vocabulary growth in children as an example of logistic growth "…where children initially acquire new words slowly, beginning at about 1 year of age, then quickly increase the rate of acquisition until later in the preschool years when this rate begins to slow down again." (Burchinal & Appelbaum, 1991, pp 29-29). A logistic growth function is inherently nonlinear, because there is no transformation that makes it possible to model it as a linear model. It is harder to estimate than linear functions, because the solution must be found using iterative estimation methods. In multilevel modeling this becomes even more difficult, because these iterations must be carried out nested within the normal iterations of the multilevel estimation method. Estimating the nonlinear function is attractive from a theoretical point of view, because the estimated parameters have a direct interpretation in terms of the hypothesized growth process. An alternative is to use polynomial functions to approximate the true development function. Burchinal and Appelbaum (1991, p29) show that both logistic and exponential functions can be reasonably approximated using a cubic polynomial. However, the parameters of the polynomial model have no interpretation in terms of the growth process, and interpretation usually follows from plotting the average or some typical individual growth curves. Burchinal and Appelbaum (1991) provide a thoughtful discussion of these issues with the examples from the field of child development.

A general problem with polynomial functions is that they usually have very high correlations. This resulting collinearity problem may cause numerical problems in the estimation. If the occasions are evenly spaced and there are no missing data, transforming the polynomials to orthogonal polynomials offers a perfect solution. Tables for orthogonal polynomials are given in most handbooks on Analysis of Variance procedures (e.g., Hays, 1994). Even if the data are not nicely balanced, using orthogonal polynomials usually lessen the collinearity problem. If the occasions are unevenly spaced, or we want to use continuous time measurements, it often helps to center the time measures in such a way that the zero point is well within the range of observed data points.

Although polynomial curves are very flexible, other ways of specifying the change over time may be preferable. Snijders and Bosker (1999) discuss the use of piecewise linear functions and spline functions, which are both functions that break up the development curve in different adjacent pieces, each with it own development model.

If there are k fixed occasions, and there is no hypothesis involving specific trends over time, we can still model the differences between the occasions using a k-1 polynomial curve. However, in this case it is much more attractive to use simple dummy variables. The usual way to indicate *k* categories with dummy variables is to specify *k-1* dummy variables, with an arbitrary category as the reference category. In the case of fixed occasion data, it is preferable to remove the intercept term from the regression, and use *k* dummy variables to refer to the *k* occasions. This is taken up in more detail in section 5.5.3.

5.5.2 Handling Missing Data

An often-cited advantage of multilevel analysis of longitudinal data is the ability to handle missing data (Bryk & Raudenbush, 1992; Snijders, 1996; Maas & Snijders, 1999). More accurately, this refers to the ability to handle models with varying time points. In a fixed occasions model, observations may be missing because at some time points respondents were not measured. In MANOVA, the usual treatment of missing time points is to remove the case from the analysis, and analyze only the complete cases. Multilevel regression models do not assume equal numbers of observations, or even fixed time points, so respondents with missing observations pose no special problems here, and all cases can remain in the analysis. This is an advantage because larger samples increase the precision of the estimates and the power of the statistical tests. However, this advantage of multilevel modeling does not extend to missing observations on the explanatory variables. If explanatory variables are missing, the usual treatment is again to remove the case from the analysis.

The capacity to handle missing occasions is an important advantage. Little and Rubin (1987, 1989) distinguish between data that are missing completely at random (MCAR) and data that are missing at random (MAR). In both cases, the failure to observe a certain data point is assumed to be independent of the unobserved (missing) value. With MCAR data, the missingness is completely independent of all other variables as well. With with MAR data, the missingness may depend on other variables in the model. It is clear that MAR is a less restrictive assumption than MCAR. In longitudinal research, a major problem is the occurrence of panel attrition: individuals who after one or more measurement occasions drop out of the study altogether. Panel attrition is generally not totally random; some types of individuals are more prone to drop out than others. The complete cases method does assume that data are missing completely at random MCAR), an assumption which in panel attrition is generally violated.

Little (1995) shows that multilevel modeling of repeated measures with missing data assumes that the data are missing at random (MAR), provided ML estimation is used. In panel attrition, we typically have much information about the drop-outs from earlier measurement occasions. The assumption that, conditional on these variables (which includes the score on the dependent variable on earlier occasions), the missingness is random, appears reasonable. Thus, when the repeated measures data is MAR, but not MCAR, MANOVA leads to biased estimates, and multilevel analysis leads to unbiased estimates. Hox (2000) presents an analysis of repeated measures data where some data points were removed according to a simulated MAR attrition process. In this example, MANOVA estimates were quite different from the corresponding estimates obtained from the complete data set, while the multilevel results were very close to the corresponding complete estimates. Hedeker and Gibbons (1997) present a more intricate way to incorporate the missingness mechanism in the model. Using multilevel analysis for repeated measures, they divide the data into groups according to their missingness pattern, and then use include variables that indicate these groups as explanatory variables. This makes it possible to investigate the effect of missing data patterns on the outcome, and to estimate an overall outcome across the missingness patterns.

5.5.3 Complex Covariance Structures

If multilevel modeling as described above is used to model longitudinal data, the variances and covariances between different occasions have a very specific structure. The variance at any occasion has the value $\sigma_e^2 + \sigma_{00}^2$, and the covariance between any two occasions has the value σ_{00}^2. Thus, for the GPA example data, a simple linear trend model as specified by equation (5.9) is

$$GPA_{tj} = \beta_{00} + \beta_{10}Time_{ti} + u_{0i} + e_{ti} \qquad (5.9),$$

where the residual variance on the occasion level is given by σ_{ε}^2, and the residual error on the person level is given by σ_{00}^2. For this and similar models without random effects, the matrix of variances and covariances among the occasions is given by (Goldstein, 1995, Bryk & Raudebush, 1992, Snijders & Bosker, 1999):

$$\Sigma\,(\mathrm{Y}) = \begin{bmatrix} \sigma_e^2+\sigma_{00}^2 & \sigma_{00}^2 & \sigma_{00}^2 & \cdots & \sigma_{00}^2 \\ \sigma_{00}^2 & \sigma_e^2+\sigma_{00}^2 & \sigma_{00}^2 & \cdots & \sigma_{00}^2 \\ \sigma_{00}^2 & \sigma_{00}^2 & \sigma_e^2+\sigma_{00}^2 & \cdots & \sigma_{00}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{00}^2 & \sigma_{00}^2 & \sigma_{00}^2 & \cdots & \sigma_e^2+\sigma_{00}^2 \end{bmatrix} \qquad (5.10)$$

In the covariance matrix (5.9) all variances are equal, and all covariances are equal. This shows that using a standard multilevel model, assuming that the residual errors e_{ti} are independent and have constant variance over time, assumes compound symmetry, the same restrictive assumption that is made in univariate analysis of variance for repeated measures. According to Stevens (1996), if the assumption of compound symmetry is violated, the standard ANOVA significance tests are too lenient, and reject the null hypothesis more often than is warranted. Therefore, multivariate analysis of variance is preferred, which estimates all variances and covariances among occasions without restrictions.

Bryk and Raudenbush (1992, p132) argue that uncorrelated errors may be appropriate in short time series. However, the multilevel regression model can be extended to include an unconstrained covariance matrix at the lowest level (Goldstein, 1995; Maas & Snijders, 1999). We use a multivariate response model (Goldstein, 1995) with dummy variables indicating the different occasions. This, if we have *p* measurement occasions, we have *p* dummy variables, one for each occasion. The intercept term is removed from the model, so the lowest level is empty. The dummy variables are all allowed to have random slopes at the second level. Thus, for our grade point example with six occasions, we have six dummy variables O_1, O_2,…,O_6, and the equation for a model without further explanatory variables becomes:

$$Yti = \beta_{10}O_{1i} + \beta_{20}O_{2i} + \beta_{30}O_{3i} + \beta_{40}O_{4i} + \beta_{50}O_{5i} + \beta_{60}O_{6i} +$$

$$u_{10}O_{1i} + u_{20}O_{2i} + u_{30}O_{3i} + u_{40}O_{4i} + u_{50}O_{5i} + u_{60}O_{6i} \quad (5.11)$$

Having six random slopes at level two provides us with a 6×6 covariance matrix for the six occasions. The regression slopes β_{10} to β_{60} are simply the estimated means at the six occasions. Equation (5.11) defines a multilevel model that is equivalent to the MANOVA approach. Maas and Snijders (1999) discuss model (5.11) at length, and show how the familiar F-ratio's from the MANOVA approach can be obtained from the multilevel software output. An attractive property of the multilevel approach here is that it is not affected by missing data. Delucchi and Bostrom (1999) compare the MANOVA and the multilevel approach to longitudinal data using small samples with missing data. Using simulation, they conclude that the multilevel approach is more accurate than the MANOVA approach.

The model in equation (5.11) is equivalent to a MANOVA model. Since the covariances between the occasions are estimated without restrictions, it does not assume compound symmetry. However, the fixed part estimates the six means at the six time points. To model a linear trend over time, we must replace the fixed part of model (5.11) with the fixed part of the linear trend model (5.9). This gives us the following model:

$$GPA_{ti} = \beta_{00} + \beta_{10}T_{ti} + u_{10}O_{1i} + u_{20}O_{2i} + u_{30}O_{3i} + u_{40}O_{4i} + u_{50}O_{5i} + u_{60}O_{6i} \quad (5.12)$$

To specify model (5.11) in standard multilevel software we must specify an intercept term that has no variance component, and six dummy variables for the occasions that have no fixed coefficients.[1] The covariance matrix between the residual errors for the six occasions has no restrictions. If we impose the restriction that all variances are equal, and that all covariances are equal, we have again the compound symmetry model. This shows that the simple linear trend model in (5.9) is one way to impose the compound symmetry structure on the random part of the model. Consequently, we can use the overall chi-square test based on the deviance of the two models to test if the assumption of compound symmetry is tenable.

Models with a residual error structure over time as in (5.12) are very complex, because they assume for the error structure a saturated model. If there are *n* time points, the number of

[1] Some software, such as MixReg and Prelis 2.3 has built-in provisions for specifying specific patterns for the covariances between occasions.

elements in the covariance matrix for the occasions is *n(n+1)/2*. So, with six occasions, we have 21 elements to be estimated. If the assumption of compound symmetry is tenable, models based on (5.9) are preferable, because they are much smaller. Their random part requires only two elements to be estimated. The advantage is not only that smaller models are more parsimonious, but they are also easier to estimate. However, the compound symmetry model is very restrictive, because it assumes that there is no correlation between time points. This assumption is in many cases not very realistic, because the error term contains all omitted sources of variation (including measurement errors), which may be correlated over time. Different assumptions about the autocorrelation over time lead to different assumptions for the structure of the covariance matrix across the occasions. For instance, it is reasonable to assume that occasions that are close together in time display a higher correlation than occasions that are far apart. A more restricted version is to assume that the autocorrelation between the occasions follow the model

$$e_t = \rho\, e_{t-1} + \varepsilon \qquad (5.13)$$

where e_t is the error term at occasion t, ρ is the autocorrelation, and ε is a residual error with variance σ_ε^2. The error structure in equation (5.13) is a first order autoregressive process. This leads to a covariance matrix of the form

$$\Sigma\,(\mathrm{Y}) = \frac{\sigma_e^2}{(1-\rho^2)} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix} \qquad (5.14)$$

The first term $\sigma_\varepsilon^2/(1-\rho^2)$ is a constant, and the autocorrelation coefficient ρ is between –1 and +1, but typically positive. As a result, the elements in covariance matrix Σ become smaller the further away they are from the diagonal. Such a structure is called a *simplex*. It is possible to have second order autoregressive processes, and other models for the error structure over time. The first order autoregressive model that produces the simplex in (5.14) estimates two variances plus an autocorrelation. This is almost as parsimonious as the compound symmetry model, but does not assume constant variances and covariances.

It should be noted that allowing random slopes for the time trend variables, e.g. for the linear trend, also models a specific covariance matrix for the occasions. For instance, in a model with a randomly varying time variable, the variance of any specific occasion at time point t is given by

$$\text{var}(Y_t) = \sigma_0^2 + 2\,\sigma_{01}\,(t-t_0) + \sigma_1^2\,(t-t_0) + \sigma_e^2 \qquad (5.15),$$

and the covariance between any two specific occasions at time points *t* and *s* is given by

$$\text{cov}(Y_t, Y_s) = \sigma_0^2 + \sigma_{01}\,[(t-t_0) + (s-s_0)] + \sigma_1^2\,(t-t_0)(s-t_0) \qquad (5.16),$$

where t_0 is the value on which the time variable is centered (if the time variable is already centered, t_0 may be omitted from the equation). Such models usually do not produce the simple structure of a simplex or other autoregressive model, but their random part can be interpreted in terms of variations in developmental curves or growth trajectories. In contrast, complex random structures are usually interpreted in terms of underlying but unknown influences.

The important point is that, in longitudinal data, there are many interesting models between the extremes of the very restricted compound symmetry and the saturated MANOVA model. In general, if there are *n* time points, any model that estimates fewer than *n(n+1)/2* (co)variances for the occasions represents a restriction on the full MANOVA model. Thus, any such model can be tested against the MANOVA model using the chi-square deviance test. If the chi-square test is significant, there are correlations across time that are not modeled adequately.

If the *Time* variable, or one of its polynomials, has a random slope, it is not possible to have a saturated MANOVA model for the covariances across time. In fact, if we have *n* occasions, and use *n* polynomials with random slopes, we simply have an alternative way to specify the Manova model of equation (5.11). If there are random slopes for the time variable, an attractive and very general model for the covariances across time is to assume that each lag has its own autocorrelation. So, all occasions that are 1 time point apart, share a specific autocorrelation, all occasions that are 2 time points apart, share a different autocorrelation, and so on. This leads to a covariance matrix for the occasions which is called a Toeplitz matrix:

$$\Sigma\,(\mathrm{Y}) = \sigma_e^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix} \tag{5.17}$$

This model poses *n-1* unique autocorrelations. Typically, the autocorrelations with large lags will be small, so they can be removed from the model.

Table 5.10 presents three models using the GPA example data. The first model has a fixed slope for the Time variable. The second model has a random slope for the time variable, and the third model has no random effects for the intercept or the time variable, but models a saturated covariance matrix across the time points. For simplicity, the Table only shows the variances at the six occasions, and not the covariances. From a comparison of the deviances, it is clear that the saturated model fits better. However, the random Time model estimates 4 terms in the random part, and the saturated model estimates 21 terms. It would seem attractive to seek a more parsimonious model for the random part.

Table 5.5. Results model 5 for different time scalings

Model:	Time fixed comp symm.		Time random, comp. symm.		Time fixed, MANOVA	
Fixed part						
Predictor	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
intercept	2.49	.11	2.44	.10	2.39	.10
time	0.10	.00	0.10	.01	0.10	.01
job status	-0.17	.02	-0.13	.02	-0.10	.01
high GPA	0.09	.03	0.09	.03	0.08	.03
sex	0.15	.03	0.12	.03	0.12	.03
Random part						
intercept$_1$	0.055	.002	0.042	.002		
intercept$_2$	0.045	.005	0.038	.013		
time$_2$			0.004	.001		
int*time$_2$			-.002	.002		
O1					0.09	.01
O2					0.10	.01
O3					0.10	.01

O4			0.11	.01
O5			0.10	.01
O6			0.12	.01
Deviance	282.8	170.1	-10.7	
AIC	296.8	188.1	41.3	

The models described above can be modeled with multilevel software that allows restrictions on the random and fixed part. Some programs (such as MixReg, Hedeker & Gibbons, 1996b, or Prelis 2.3, Jöreskog, Sörbom, du Toit & du Toit, 1999) recognize the existence of longitudinal data, and allow direct specification of various types of autocorrelation structures. For a discussion of some of these structures in the context of multilevel longitudinal models, see Hedeker and Gibbons (1996b). If there are many different and differently spaced occasions, the Manova and related models become impractical. With varying occasions, it is still possible to specify an autocorrelation structure, but it is more difficult to interpret than with fixed occasions. The program MlwiN (Goldstein et al., 1998) can model very general autocorrelation structures using macro's available from the Multilevel Modelling Project at the University of London (Yang, Rasbash & Goldstein, 1998). A detailed discussion of multilevel models for longitudinal data, including suggestions for computing estimates for the explained variance, can be found in Snijders and Bosker (1999).