7

Multilevel and Related Models for Longitudinal Data

Anders Skrondal[1] and Sophia Rabe-Hesketh[2]

[1] Department of Statistics and Methodology Institute, London School of Economics & Division of Epidemiology, Norwegian Institute of Public Health.
[2] Graduate School of Education and Graduate Group in Biostatistics, University of California, Berkeley & Institute of Education, University of London.

7.1 Introduction

Longitudinal data, often called repeated measurements in medicine and panel data in the social sciences, arise when units provide responses on multiple occasions. Such data can be thought of as clustered or two-level data with occasions i at level 1 and units j at level 2.

One feature distinguishing longitudinal data from other types of clustered data is the chronological ordering of the responses, implying that level-1 units cannot be viewed as exchangeable. Another feature of longitudinal data is that they often consist of a large number of small clusters.

A typical aim in longitudinal analysis is to investigate the effects of covariates both on the overall level of the responses and on changes of the responses over time. An important merit of longitudinal designs is that they allow the separation of cross-sectional and longitudinal effects. They also allow the investigation of heterogeneity across units both in the overall level of the response and in the development over time. Heterogeneity not captured by observed covariates produces dependence among responses even after controlling for those covariates. This violates the typical assumptions of ordinary regression models and must be accommodated to avoid invalid inference.

It is useful to distinguish between longitudinal data with balanced and unbalanced occasions. The occasions are balanced if all units are measured at the same time points t_i, $i = 1, \ldots, n$, and unbalanced if units are measured at different time points, $t_{ij}, i = 1, \ldots, n_j$. In the case of balanced occasions, the data can also be viewed as single-level multivariate data where responses at different occasions are treated as different variables. One advantage of the univariate multilevel approach taken here is that unbalanced occasions and

J. de Leeuw, E. Meijer (eds.), *Handbook of Multilevel Analysis*,

missing data are accommodated without resorting to complete case analysis (sometimes called listwise deletion). We will use maximum likelihood estimation, which produces consistent estimates if responses are missing at random (MAR) as defined by Rubin [59]; see Chapter 10 [40] for other approaches in the case of MAR and Verbeke and Molenberghs [65] for approaches in the case of responses not missing at random (NMAR).

In this chapter we will consider both linear mixed models and generalized linear mixed models. A linear mixed model is written in Chapter 1, equation (1.4), as

$$\underline{\boldsymbol{y}}_j = \boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\underline{\boldsymbol{\delta}}_j + \underline{\boldsymbol{\epsilon}}_j, \tag{7.1}$$

where $\underline{\boldsymbol{y}}_j$ is the vector of continuous responses for unit j. In this book the covariate matrices $\boldsymbol{X}_j$ and $\boldsymbol{Z}_j$ are treated as fixed. Extra assumptions are required when these matrices are treated as random; see, for instance, Rabe-Hesketh and Skrondal [54].

A generalized linear mixed model also accommodates non-continuous responses and can be written as

$$g(E(\underline{\boldsymbol{y}}_j \mid \underline{\boldsymbol{\delta}}_j)) = \boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\underline{\boldsymbol{\delta}}_j \stackrel{\Delta}{=} \underline{\boldsymbol{\eta}}_j, \tag{7.2}$$

where $g(\cdot)$ is a link function and $\boldsymbol{\eta}_j$ is a vector of linear predictors. Conditional on the random effects $\underline{\boldsymbol{\delta}}_j$, the elements $\underline{y}_{ij}$ of $\underline{\boldsymbol{y}}_j$ have a distribution from the exponential family and are mutually independent. See Rabe-Hesketh and Skrondal [54] and Chapter 9 [58] for treatments of generalized linear mixed models.

For dichotomous and ordinal responses, generalized linear mixed models with logit and probit links can also be defined using a latent response formulation. A linear mixed model is in this case specified for an imagined continuous latent response $\underline{y}^*_{ij}$. The observed dichotomous or ordinal response $\underline{y}_{ij}$ with $S > 1$ categories results from partitioning $\underline{y}^*_{ij}$ into S segments using $S - 1$ cut-points or thresholds; see Chapter 6 [31] for details.

We will use an example dataset to illustrate some of the ideas discussed in this chapter. The dataset comes from an American panel survey of 545 young males taken from the National Longitudinal Survey (Youth Sample) for the period 1980–1987. The data were previously analyzed by Vella and Verbeek [64] and can be downloaded from the web pages of Wooldridge [70] and Rabe-Hesketh and Skrondal [53]. The response variable is the natural logarithm of the hourly wage in US dollars and the following covariates will be used:

- `educ`: Years of schooling (x_{1j})
- `black`: Dummy variable for being black (x_{2j})
- `hisp`: Dummy variable for being Hispanic (x_{3j})
- `labex`: Labor market experience (in 2-year periods) (x_{4ij})

- `labexsq`: Labor market experience squared (x_{5ij})
- `married`: Dummy variable for being married (x_{6ij})
- `union`: Dummy variable for being a member of a union (x_{7ij})

The first three covariates are time-constant, whereas the next four are time-varying.

7.2 Models with Unit-Specific Intercepts

In longitudinal data it is usually impossible to capture all between-unit variability using observed covariates. If the remaining "unobserved heterogeneity" is ignored, it induces longitudinal dependence among the responses for the same unit (after controlling for the included covariates). A simple way of representing "unobserved heterogeneity" is by including unit-specific intercepts, which could be either random or fixed.

7.2.1 Random Intercept Models

Consider the response $\underline{y}_{ij}$ of unit j on occasion i ($i = 1, \ldots, n_j$). In a linear random intercept model, sometimes referred to as a one-way error component model, it is assumed that the unit-specific effects are realizations of a random variable $\underline{\delta}_j$,

$$\underline{y}_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \underline{\delta}_j + \underline{\epsilon}_{ij},$$

where $\underline{\delta}_j$ and $\underline{\epsilon}_{ij}$ are independently distributed $\underline{\delta}_j \sim \mathcal{N}(0, \omega^2)$ and $\underline{\epsilon}_{ij} \sim \mathcal{N}(0, \sigma^2)$. The random intercept or "permanent component" $\underline{\delta}_j$ allows the level of the response to vary across units, whereas the "transitory component" $\underline{\epsilon}_{ij}$ varies over occasions within units. The model is a special case of a linear mixed model (7.1) with $\boldsymbol{Z}_j = \mathbf{1}_{n_j}$.

The variance-covariance matrix of the responses $\underline{\boldsymbol{y}}_j$, after controlling for $\boldsymbol{X}_j$, is given by

$$\text{Cov}(\underline{\boldsymbol{y}}_j) = \text{Cov}(\mathbf{1}_{n_j}\underline{\delta}_j + \underline{\boldsymbol{\epsilon}}_j) = \omega^2 \mathbf{1}_{n_j}\mathbf{1}'_{n_j} + \sigma^2 \boldsymbol{I}_{n_j},$$

with diagonal elements $\omega^2 + \sigma^2$ and off-diagonal elements ω^2. The residual intraclass correlation becomes

$$\text{Corr}(\underline{y}_{ij}, \underline{y}_{i'j}) = \frac{\omega^2}{\omega^2 + \sigma^2}. \tag{7.3}$$

This covariance structure is also shown in panel A of Table 7.1. It is sometimes referred to as exchangeable since the joint distribution of the residuals for a given unit remains unchanged if the residuals are exchanged across occasions. The covariance structure also satisfies the sphericity property that the conditional variances $\text{Var}(\underline{y}_{ij} - \underline{y}_{i'j})$ of all pairwise differences are equal. Note that

Table 7.1 Common dependence structures for longitudinal data ($\boldsymbol{\Psi}_j \triangleq \mathrm{Cov}(\underline{\boldsymbol{y}}_j)$).

A. Random intercept structure:

$$\boldsymbol{\Psi}_j = \omega^2 \mathbf{1}_{n_j}\mathbf{1}'_{n_j} + \sigma^2 \boldsymbol{I}_{n_j} = \begin{bmatrix} \omega^2+\sigma^2 & & & \\ \omega^2 & \omega^2+\sigma^2 & & \\ \vdots & \vdots & \ddots & \\ \omega^2 & \omega^2 & \cdots & \omega^2+\sigma^2 \end{bmatrix}$$

B. Random coefficient structure:

$$\boldsymbol{\Psi}_j = \boldsymbol{Z}_j \boldsymbol{\Omega} \boldsymbol{Z}'_j + \sigma^2 \boldsymbol{I}_{n_j}$$

C. Autoregressive residual structure AR(1):

$$\boldsymbol{\Psi}_j = \frac{\sigma_u^2}{1-\alpha^2} \begin{bmatrix} 1 & & & \\ \alpha & 1 & & \\ \vdots & \vdots & \ddots & \\ \alpha^{n_j-1} & \alpha^{n_j-2} & \cdots & 1 \end{bmatrix}$$

D. Moving average residual structure MA(1):

$$\boldsymbol{\Psi}_j = \sigma_u^2 \begin{bmatrix} 1+a^2 & & & & \\ a & 1+a^2 & & & \\ 0 & a & 1+a^2 & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \cdots & 1+a^2 \end{bmatrix}$$

E. Autoregressive response structure AR(1):

$$\boldsymbol{\Psi}_j = \frac{\sigma_\epsilon^2}{1-\gamma^2} \begin{bmatrix} 1 & & & \\ \gamma & 1 & & \\ \vdots & \vdots & \ddots & \\ \gamma^{n_j-1} & \gamma^{n_j-2} & \cdots & 1 \end{bmatrix}$$

the covariances ω^2 are restricted to be nonnegative in the random intercept model. If this restriction is relaxed, the above covariance structure is often called compound symmetric. In the case of balanced occasions, we could also allow the variance of $\underline{\epsilon}_{ij}$ to take on a different value $\boldsymbol{\Sigma}_{ii}$ for each occasion.

Typically, the random intercept model is estimated by either maximum likelihood or restricted maximum likelihood [42]. The likelihood has a closed form, but iterative methods such as the EM algorithm, Newton-Raphson, Fisher scoring, or iterated generalized least squares (IGLS) must be used to estimate the parameters; see Chapter 1 [15].

Maximum likelihood estimates of the random intercept model for the wage panel data, obtained using Stata's [63] `xtmixed` command, are given in the first column of Table 7.2. As might be expected, more years of schooling, more labor market experience, being married, and being a union member are all associated with higher hourly wages, whereas being black decreases the wage compared with being white, and Hispanics' wages are similar to those of whites (controlling for the other covariates). The residual intraclass correlation is estimated as 0.47; 47% of the variance not explained by the covariates is therefore between individuals and 53% within individuals.

For generalized linear mixed models, the dependence among observed responses is generally difficult to express because the model-implied correlations and variances depend on the covariates. However, for a generalized linear random intercept model, obtained by substituting $\boldsymbol{Z}_j = \mathbf{1}_{n_j}$ in (7.2), with dichotomous or ordinal responses, the intraclass correlation of the latent responses is constant and given by (7.3) with σ^2 replaced by $\pi^2/3$ for logit models and 1 for probit models. An important interpretational issue in generalized linear mixed models concerns the distinction between conditional and marginal effects, which correspond to unit-specific and population-averaged effects in the longitudinal setting. We return to this in Section 7.6.

Generally, the marginal likelihood does not have a closed form for generalized linear mixed models, making estimation more difficult. Common approaches include penalized quasilikelihood [23], maximum likelihood using adaptive quadrature [56] and Markov Chain Monte Carlo (MCMC) [10]; see also Chapter 9 [58]. For dichotomous responses and counts, closed-form likelihoods can be achieved by specifying a conjugate distribution for the random intercepts, giving the beta-binomial and negative-binomial models, respectively [38].

Simulation studies [5, 26, 48, 69] suggest that inference for the random intercept model and similar models is relatively robust to violation of the normality assumption for the random intercept. However, to safeguard against distributional misspecification, the random intercept distribution can be left unspecified by using nonparametric maximum likelihood estimation [30, 34, 37]. The nonparametric maximum likelihood estimator of the random intercept distribution is discrete with estimated locations and masses, their number being determined to reach the largest maximized likelihood.

For the wage panel data, `gllamm` [53, 55] in Stata was used to estimate models with a discrete random effects distribution. The directional derivative [37] was used to determine whether the nonparametric maximum likelihood estimator (NPMLE) was achieved as described in Rabe-Hesketh et al. [52]. In the example, the NPMLE appears to have eight mass points whose estimated locations and masses are shown in Fig. 7.1. This estimated discrete distribution is quite symmetric apart from a tiny mass at 1.77, which appears to accommodate one outlying individual whose log wage exceeded the 99th

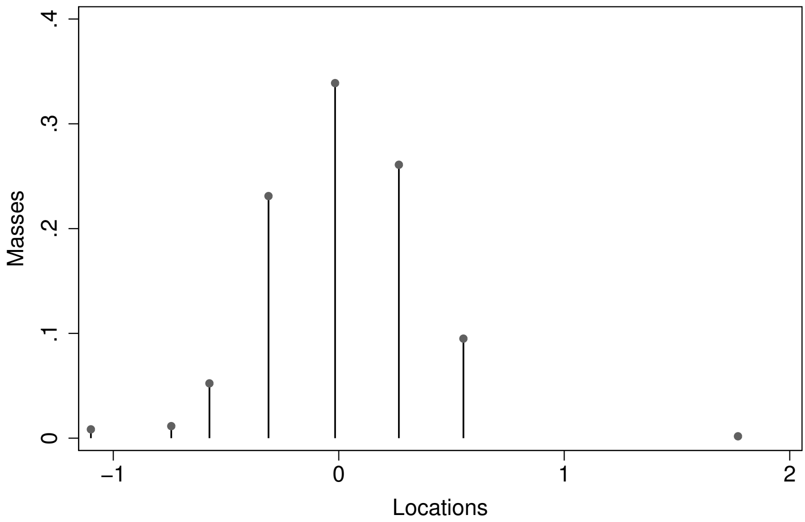

Fig. 7.1 Estimated discrete random intercept distribution from NPMLE.

percentile (across individuals and time) in 1981–1987. The standard deviation of the discrete distribution is very close to $\hat{\omega}_0$ for the conventional random intercept model, as are the estimates of the regression parameters $\boldsymbol{\beta}$ and variance parameter σ^2 given in the second column of Table 7.2.

As discussed for linear models in Chapter 3 [61], violation of the assumption that $\underline{\delta}_j$ has zero expectation can invalidate inference. Specifically, if $E(\underline{\delta}_j) = \boldsymbol{z}_j'\boldsymbol{\gamma}$, where $\mathbf{1}_{n_j}\boldsymbol{z}_j$ and $\boldsymbol{X}_j$ are nonorthogonal, the estimates of the regression coefficients $\boldsymbol{\beta}$ will be inconsistent. When the covariates $\boldsymbol{X}_j$ are treated as random variables $\underline{\boldsymbol{X}}_j$, this problem is referred to as endogeneity in econometrics because the covariates are correlated with the random effects.

The standard approach to handling endogeneity in econometrics is instrumental variables estimation [70]. In the present context, a simpler solution is to estimate the within-unit effects of $\boldsymbol{X}_j$, which can be achieved by also controlling for the cluster mean covariates $\bar{\boldsymbol{X}}_{\cdot j}$. An alternative for linear models is to use a fixed effects approach, which will be discussed next. Unfortunately, there are no easy fixes for violation of the assumption that $E(\underline{\epsilon}_{ij}) = 0$.

7.2.2 Fixed Intercept Models

A simple linear fixed intercept model or fixed effects model has the form

$$\underline{y}_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \delta_j + \underline{\epsilon}_{ij}, \tag{7.4}$$

where δ_j are unit-specific intercepts or “fixed effects” and $\underline{\epsilon}_{ij}$ are identically and independently normally distributed residuals with $E(\underline{\epsilon}_{ij}) = 0$. Due to the

Table 7.2 Estimates for wage panel data.

	Intercept						Intercept and Slope						Autoregressive (1)			
	Random		NPML		Fixed		Both random		Random slope		Both fixed		AR Residual		AR Response	
	Est	(SE)	Est	(SE)	Est	(SE)	Est	(SE)	Est	(SE)	Est	(SE)	Est	(SE)	Est	(SE)
Fixed part																
β_0 [cons]	−0.11	(0.11)	−0.17	(0.10)			−0.16	(0.12)					−0.11	(0.11)	0.11	(0.06)
β_1 [educ]	0.10	(0.01)	0.11	(0.01)			0.11	(0.01)					0.10	(0.01)	0.05	(0.00)
β_2 [black]	−0.14	(0.05)	−0.14	(0.05)			−0.15	(0.04)					−0.14	(0.05)	−0.08	(0.02)
β_3 [hisp]	0.02	(0.04)	0.01	(0.03)			0.01	(0.04)					0.02	(0.04)	0.01	(0.02)
β_4 [labex]	0.22	(0.02)	0.22	(0.02)	0.23	(0.02)	0.21	(0.02)	0.22	(0.02)			0.23	(0.02)	0.03	(0.02)
β_5 [labexsq]	−0.01	(0.00)	−0.02	(0.00)	−0.02	(0.00)	−0.01	(0.00)	−0.01	(0.01)	−0.01	(0.01)	−0.02	(0.00)	0.00	(0.00)
β_6 [married]	0.06	(0.02)	0.07	(0.02)	0.05	(0.02)	0.07	(0.02)	0.05	(0.02)	0.04	(0.03)	0.06	(0.02)	0.05	(0.01)
β_7 [union]	0.11	(0.02)	0.10	(0.02)	0.08	(0.02)	0.11	(0.02)	0.08	(0.02)	0.04	(0.02)	0.09	(0.02)	0.07	(0.01)
γ [lag]															0.56	(0.01)
Random part																
ω_0	0.33	(0.01)	0.33	(–)			0.44	(0.02)					0.31	(0.01)		
ω_1							0.10	(0.01)	0.11	(0.01)						
ρ_{10}							−0.66	(0.04)								
σ	0.35	(0.00)	0.35	(0.00)	0.35	(0.00)	0.33	(0.00)	0.33	(0.00)			0.37	(0.00)	0.39	(0.00)
α													0.27	(0.01)		
Log-likelihood[a]	−2193.3		−2176.1				−2118.9						−2109.0			

[a] No log-likelihood given when estimates are based on transformed data or subset of data.

inclusion of fixed effects δ_j for each unit j, the mean structure of $\underline{\boldsymbol{y}}_j$ is saturated so that the regression coefficients $\boldsymbol{\beta}$ represent within-unit or longitudinal effects only. Unlike the random intercept model, the fixed intercept model no longer makes any assumptions regarding the cross-sectional component of the model, so that endogeneity bias can be avoided. The cost of this robustness is that regression parameters for time-constant covariates such as gender or treatment group cannot be estimated and all covariates must therefore be time-varying.

The fixed intercepts are rarely of interest in themselves and estimation can be involved when there are many units. An attractive alternative to estimating all parameters is to eliminate the fixed intercepts. This can be accomplished by transforming both the responses and covariates and then using ordinary least squares (OLS). In econometrics, two popular transformations are first-differencing: $\underline{y}_{ij} - \underline{y}_{i-1,j}$, $\boldsymbol{x}_{ij} - \boldsymbol{x}_{i-1,j}$, and cluster-mean centering: $\underline{y}_{ij} - \bar{\underline{y}}_{\cdot j}$, $\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{\cdot j}$. Both approaches yield consistent estimates of the remaining regression coefficients, but the latter, known as the fixed effects estimator, is more efficient if the residuals $\underline{\epsilon}_{ij}$ are mutually independent as assumed above [70]. Verbeke et al. [66] propose eliminating the intercepts by conditioning on the cluster mean responses and maximizing the resulting conditional likelihood. This can be implemented by premultiplying $\underline{\boldsymbol{y}}_j$ and $\boldsymbol{X}_j$ by a $n_j \times (n_j - 1)$ orthonormal contrast matrix. This approach yields identical estimates as the fixed effects estimator based on cluster mean centering, but has the advantage that the OLS standard error estimates need not be corrected for the loss of degrees of freedom.

Some insight can be gained [41] regarding the difference between fixed effects and random effects estimators of the regression coefficients by considering the generalized least squares (GLS) estimator for the latter. The GLS estimator is asymptotically equivalent to the maximum likelihood estimator but has a closed form. It can be shown that the GLS estimator is a matrix weighted average of the fixed effects (or within-unit) estimator and the between-unit estimator obtained by OLS estimation for the regression of the cluster-mean response on the cluster-mean covariates. If the random intercept model is correctly specified, the GLS estimator is more efficient since it uses cross-sectional information in addition to longitudinal information. However, if the cross-sectional component of the model is misspecified, the GLS estimator becomes inconsistent for the longitudinal effects in contrast to the fixed effects estimator. Thus, a difference between fixed effects and GLS estimates for $\boldsymbol{\beta}$ suggests that the random effects model is misspecified and is the basis for the popular Durbin-Wu-Hausman specification test [25] in this context.

Returning to the wage panel data, the fixed effects estimates of the coefficients of the time-varying covariates, obtained using Stata's `xtreg` command, are given in the third column of Table 7.2. The estimates are quite similar to the estimates for the random intercept model, suggesting that the cross-sectional component of the random intercept model is not severely misspecified.

In generalized linear models, except for linear Gaussian or log-linear Poisson models, inclusion of a fixed intercept for each unit leads to inconsistent estimates of the regression parameters $\boldsymbol{\beta}$, which is known as the incidental parameter problem [49]. For binary logistic models, the problem can be overcome by conditioning on the sum of the responses for each unit to eliminate the unit-specific intercepts, as mentioned above for linear models. In epidemiology, such a conditional maximum likelihood approach is used for matched case-control studies [7], in psychometrics for the Rasch measurement model [57], and in econometrics for panel data [8, 9]. In addition to the limitation of not permitting time-constant covariates, this approach also discards units with all responses equal to 0 or all equal to 1. Furthermore, conditional maximum likelihood estimation is impossible for some model types such as probit models.

7.3 Models with Unit-Specific Intercepts and Slopes

Sometimes units vary not just in the overall level of the response (controlling for covariates) but also in the effects of time-varying covariates on the response. A typical example is where the effect of time, i.e., the rate of change, varies between units. Such heterogeneity in the effects of covariates can be viewed as interactions between the included covariates and a categorical variable representing the units.

7.3.1 Continuous Random Coefficients

The random coefficient model [35] can be written as

$$\underline{y}_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\underline{\boldsymbol{\delta}}_j + \underline{\epsilon}_{ij},$$

where $\boldsymbol{x}_{ij}$ denotes both time-varying and time-constant covariates with fixed coefficients $\boldsymbol{\beta}$ and $\boldsymbol{z}_{ij}$ denotes time-varying covariates with random coefficients $\underline{\boldsymbol{\delta}}_j \sim \mathcal{N}(\boldsymbol{\emptyset}, \boldsymbol{\Omega})$. Since the random coefficients have zero means, $\boldsymbol{x}_{ij}$ will typically contain all elements in $\boldsymbol{z}_{ij}$, with the corresponding fixed effects interpretable as the mean effects. The first element of these vectors is invariably equal to 1, corresponding to a fixed and random intercept, respectively. The random intercept model is thus the special case where $\boldsymbol{z}_{ij} = 1$. The random coefficient covariance structure of the vector $\underline{\boldsymbol{y}}_j$ is presented in panel B of Table 7.1.

A useful version of the random coefficient model for longitudinal data is a growth curve model where individuals are assumed to differ not only in their intercepts but also in other aspects of their trajectory over time, for example in the linear growth (or decline) of the response. These models include random coefficients for (functions of) time. For example, a linear growth curve model can be written as

$$\underline{y}_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \underline{\delta}_{0j} + \underline{\delta}_{1j}t_{ij} + \underline{\epsilon}_{ij}, \tag{7.5}$$

where t_{ij}, the time at the i-th occasion for individual j, is one of the covariates in $\boldsymbol{x}_{ij}$. The random intercept $\underline{\delta}_{0j}$ and random slope $\underline{\delta}_{1j}$ represent unit-specific deviations from the mean intercept and slope, respectively. The random intercept and slope should not be specified as uncorrelated, because translation of the time scale t_{ij} changes the magnitude of the correlation [18, 39, 62].

In a linear growth curve model, the variance of the responses (controlling for the covariates) varies over occasions t_{ij},

$$\mathrm{Var}(\underline{y}_{ij}) = \omega_0^2 + 2\omega_{10}t_{ij} + \omega_1^2 t_{ij}^2 + \sigma^2.$$

Note that the variance increases as a quadratic function of time if $t_{ij} \geq 0$ and $\omega_{10} \geq 0$. The covariance between two responses $\underline{y}_{ij}$ and $\underline{y}_{i'j}$ for a unit at different occasions i and i' becomes

$$\mathrm{Cov}(\underline{y}_{ij}, \underline{y}_{i'j}) = \omega_0^2 + \omega_{10}(t_{ij} + t_{i'j}) + \omega_1^2 t_{ij}t_{i'j},$$

which depends on the time associated with the occasions.

For the wage panel data, we would expect wages to increase more rapidly for some individuals as they gain more labor market experience than for others. We therefore estimated a model with a random slope for `labex` in addition to the random intercept. Maximum likelihood estimates using `xtmixed` are given in the fourth column of Table 7.2. The fixed part estimates remain practically the same as for the random intercept model. There is a negative estimated correlation between the random intercept and random slope. To visualize the model, the bottom panel of Fig. 7.2 shows the fitted trajectories (obtained by plugging in empirical Bayes predictions of the random intercepts and slopes and setting `married` and `union` to zero) for the first 40 individuals. For comparison, the corresponding trajectories for the random intercept model are given in the top panel of the figure. These trajectories are nonlinear due to the quadratic term `labexsq` in the fixed part of the model.

For balanced occasions with associated times $t_{ij} = t_i$, the linear growth curve model can also be formulated as a two-factor model with fixed factor loadings,

$$\underline{y}_{ij} = \lambda_{0i}\,\underline{\beta}_{0j} + \lambda_{1i}\,\underline{\beta}_{1j} + \underline{\epsilon}_{ij}, \qquad \lambda_{0i} = 1, \quad \lambda_{1i} = t_i\,,$$

where

$$\underline{\beta}_{0j} = \beta_0 + \underline{\delta}_{0j}, \quad \underline{\beta}_{1j} = \beta_1 + \underline{\delta}_{1j}.$$

Note that the means of the factors cannot be set to zero here as is usually done in factor models. A path diagram of this model is shown in Fig. 7.3, where

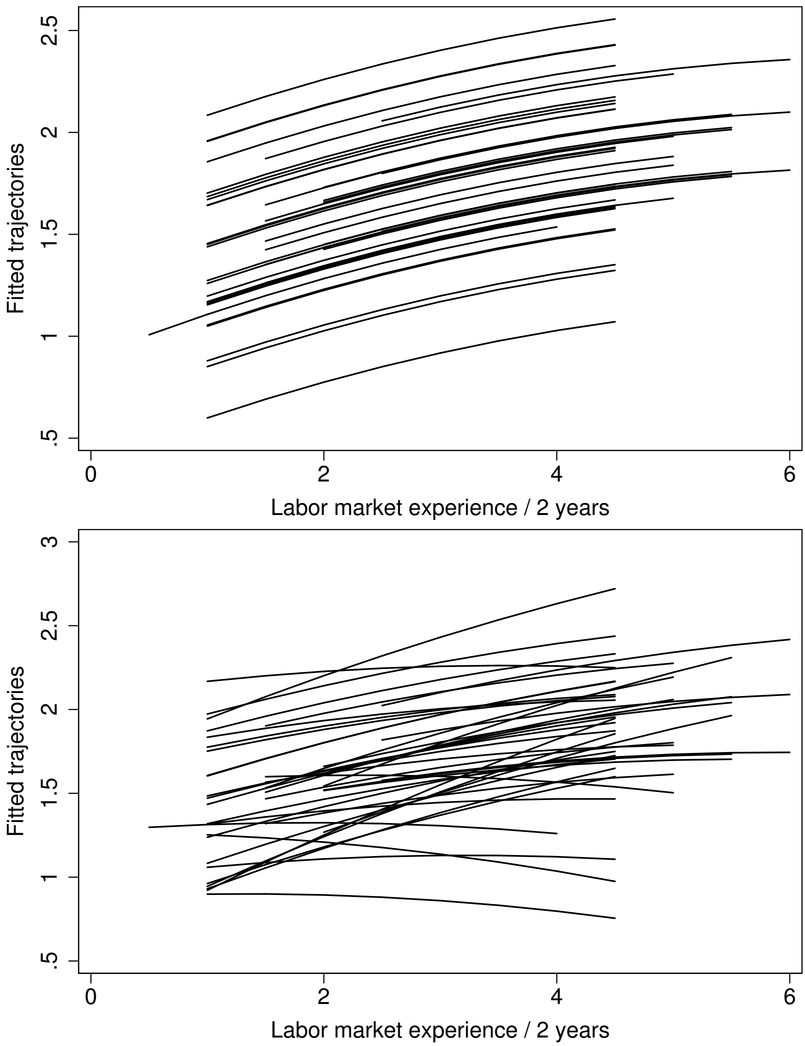

Fig. 7.2 Fitted trajectories for linear random intercept model (top) and random coefficient model (bottom). Empirical Bayes predictions are substituted for the random effects and `married` and `union` set to zero.

there are three occasions with times $t_1 = 0$, $t_2 = 1$, and $t_3 = 2$. Following the usual conventions, latent variables or random effects are represented by circles and observed variables by rectangles. Long arrows represent regressions and short arrows residual errors.

Meredith and Tisak [43] suggest using a two-factor model with free factor loadings λ_{1i} for $\underline{\beta}_{1j}$ (subject to identification restrictions, such as $\lambda_{11} = 0$ and

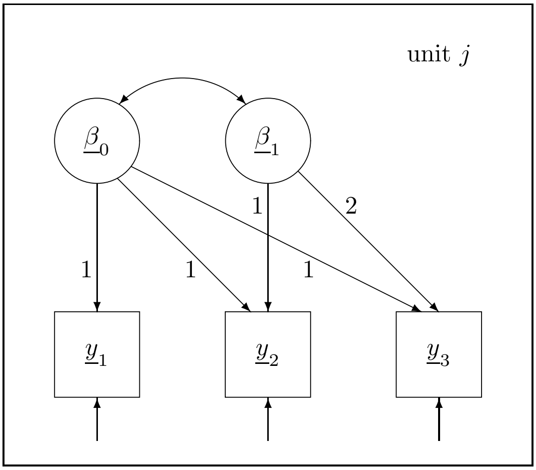

Fig. 7.3 Path diagram for growth curve model with balanced occasions.

$\lambda_{12}=1$) to model nonlinear growth. Estimation of this factor model requires balanced occasions, but modern software can handle missing data.

Generalized linear random coefficient models are defined analogously to the linear case. Maximum likelihood estimation using numerical integration becomes computationally more demanding as the number of random effects increases. Unfortunately, we can no longer exploit conjugacy to obtain closed-form likelihoods for counts and dichotomous responses.

7.3.2 Fixed Coefficients

Instead of considering the unit-specific intercepts and slopes as random, we can specify a model with fixed intercepts and slopes,

$$\underline{y}_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \delta_{0j} + \delta_{1j} z_{ij} + \underline{\epsilon}_{ij}.$$

If the data are balanced, $z_{ij} = z_i$, and the differences $z_i - z_{i-1}$ are constant, then the δ_{0j} and δ_{1j} can be eliminated by double-differencing [70]. Alternatively, first-differencing can be used to turn the model into a fixed-intercepts model, which can subsequently be estimated by any of the methods discussed in Section 7.2.2. This approach was used to obtain the estimates for the wage panel data given in the sixth column of Table 7.2. The estimated regression coefficients for `married` and `union` are considerably closer to zero than in the random coefficient model. Wooldridge [70] also describes an approach for eliminating the intercepts and slopes in more general models with unbalanced $\boldsymbol{z}_{ij}$.

Verbeke et al. [66] suggest a hybrid approach, treating the intercepts as fixed but the slope(s) as random,

$$\underline{y}_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \delta_{0j} + \underline{\delta}_{1j} z_{ij} + \underline{\epsilon}_{ij}.$$

The fixed intercepts are eliminated by forming contrasts using an orthonormal coefficient matrix as described in Section 7.2.2, corresponding to conditional maximum likelihood estimation. Estimates for the wage panel data using this approach are given in the fifth column of Table 7.2 and are quite similar to the estimates for the random coefficient model.

7.3.3 Discrete Random Coefficients

It is sometimes believed that the population consists of different subpopulations or classes of units characterized by different unknown patterns of development over time. Since class membership is not known, the parameters characterizing the development trajectory can be treated as discrete latent variables or random effects.

In a linear latent trajectory model or latent profile model [22] analogous to (7.5), the model for a unit in latent class c ($c = 1, \ldots, C$) is given by

$$\underline{y}_{ijc} = e_{0c} + e_{1c} t_{ij} + \underline{\epsilon}_{ijc}.$$

Each latent class is characterized by a pair of coefficients e_{0c} and e_{1c}, representing the intercept and slope of the latent trajectory. Other covariates can be included in the regression model above, so that the e_{0c} and e_{1c} describe the distinct patterns of deviations from the mean trajectory given the covariates. Alternatively, other covariates could be included in a multinomial logit model for the latent class membership probabilities, as is often done in conventional latent class models [14].

Interestingly, the number of classes cannot be increased indefinitely. If it is attempted to exceed the maximum possible number of classes, then estimated locations of some classes will either coincide or the probabilities of some classes tend to zero. The solution with the maximum number of classes then corresponds to the nonparametric maximum likelihood estimator [1]. An extension of the model would be to allow the variance of residuals $\underline{\epsilon}_{ijc}$ to differ between classes.

For balanced occasions, we do not have to assume that the latent trajectories are linear or have another particular shape but can, instead, specify an unstructured model with latent trajectory

$$\underline{y}_{ijc} = e_{ic} + \underline{\epsilon}_{ijc}, \qquad i = 1, \ldots, n,$$

for class c.

In the case of categorical responses, latent trajectory models are typically referred to as latent class growth models [47] and represent an application of mixture regression models [51, 68] to longitudinal data.

All these models assume that the responses on a unit are conditionally independent given latent class membership. Muthén and Shedden [46] relax this assumption for continuous responses in their growth mixture models by allowing the residuals $\underline{\epsilon}_{ijc}$ to be correlated conditional on latent class membership with covariance matrices differing between classes.

7.4 Models with Correlated Residual Errors

In the models considered so far, the residuals $\underline{\epsilon}_{ij}$ have been assumed to be mutually independent and the longitudinal dependence among the responses (given the covariates) has been accommodated by including either fixed or random unit-specific effects. In the case of random effects, the responses are conditionally independent given the random effects but marginally dependent with covariance structures for linear models given in Table 7.1.

These covariance structures may be overly restrictive, particularly for a random intercept model when there are a large number of occasions. For instance, the correlations between pairs of responses often decrease as the time lag increases, which is at odds with the constant correlations induced by the random intercept model. For such reasons, the residuals $\underline{\epsilon}_{ij}$ are sometimes allowed to be correlated. Caution should be exercised when combining a complex unit-level random part with a covariance structure for the residuals, as the resulting model may not be identified.

Allowing for dependence among the residuals can also be motivated as follows. Unit-specific intercepts and slopes accommodate the effects of only time-constant influences (not represented by the covariates). The independence assumption for the residuals then implies that time-varying random influences are immediate and do not persist over more than a single occasion. There is often no compelling reason to exclude a third type of random influence that is neither everlasting nor fleeting, but persists for an intermediate length of time, leading to serially correlated residual errors.

In the following subsections, we follow the treatment in Skrondal and Rabe-Hesketh [60]. We discuss the case of continuous responses, sometimes indicating how the models are modified for other response types. The models to be described can be generalized to dichotomous and ordinal responses using the latent response formulation.

7.4.1 Autoregressive Residuals

When occasions are equally spaced in time, a first-order autoregressive model AR(1) can be expressed as

$$\underline{\epsilon}_{ij} = \alpha \underline{\epsilon}_{i-1,j} + \underline{u}_{ij}, \tag{7.6}$$

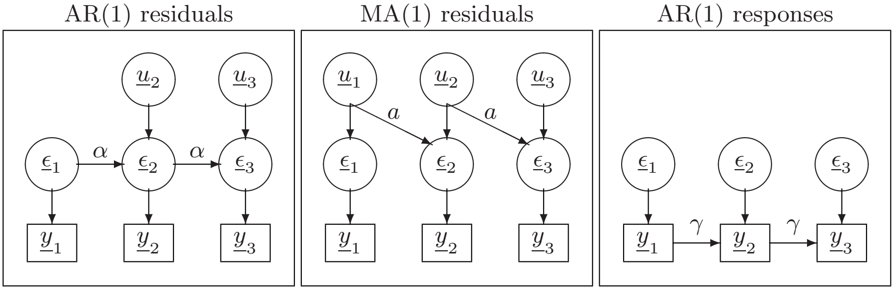

Fig. 7.4 Path diagrams for autoregressive responses and autoregressive and moving average residuals. Covariates and paths from covariates to responses omitted (Source: Skrondal and Rabe-Hesketh [60]).

where $\underline{\epsilon}_{i-1,j}$ is independently distributed from the "innovation errors" $\underline{u}_{ij}$, $\underline{u}_{ij} \sim \mathcal{N}(0, \sigma_u^2)$. This is illustrated in path diagram form in the first panel of Fig. 7.4. A "random walk" is obtained if the restriction $\alpha = 1$ is imposed.

Assuming that the process is weakly stationary, $|\alpha| < 1$, the covariance structure is as shown in panel C of Table 7.1. The correlations between responses at different occasions become

$$\mathrm{Corr}(\underline{\epsilon}_{ij}, \underline{\epsilon}_{i+k,j}) = \alpha^k.$$

For non-equally spaced occasions, the correlation structure is often specified as

$$\mathrm{Corr}(\underline{y}_{ij}, \underline{y}_{i+k,j}) = \alpha^{|t_{i+k} - t_i|},$$

where the correlation structure for unbalanced occasions is simply obtained by replacing t_i by t_{ij} [16]. In the case of balanced occasions, we can also specify a different parameter α_i for each occasion, giving an antedependence structure [21] for the residuals.

For the wage panel data, we estimated a random intercept model with AR(1) residuals by maximum likelihood using the `lme()` function in S-PLUS giving the estimates in column 7 of Table 7.2 (Stata's `xtregar` command can be used to estimate the model using the generalized least squares estimator proposed by Baltagi and Wu [4]). The autoregressive coefficient is estimated as $\hat{\alpha} = 0.27$ and the estimates of the regression parameters $\boldsymbol{\beta}$ are very similar to those given for the random intercept model in the first column. The random intercept model with AR(1) residuals has a considerably larger log-likelihood than the random intercept model with uncorrelated residuals. Introducing a random slope increases the log-likelihood to -2095.7 and reduces the estimated autoregressive coefficient to $\hat{\alpha} = 0.17$ (estimates not shown).

First-order autoregressive covariance structures are often as unrealistic as the random intercept structure since the correlations fall off too rapidly with increasing time lags. One possibility is to specify a higher-order autoregressive process of order k, AR(k),

$$\underline{\epsilon}_{ij} = \alpha_1 \underline{\epsilon}_{i-1,j} + \alpha_2 \underline{\epsilon}_{i-2,j} + \cdots + \alpha_k \underline{\epsilon}_{i-k,j} + \underline{u}_{ij}.$$

7.4.2 Moving Average Residuals

Random shocks disturb the response variable for some fixed number of periods before disappearing and can be modeled by moving averages [6]. A first-order moving average process MA(1) for the residuals can be specified as

$$\underline{\epsilon}_{ij} = \underline{u}_{ij} + a\, \underline{u}_{i-1,j}.$$

A path diagram for this model is given in the second panel of Fig. 7.4 and the covariance structure is presented in panel D of Table 7.1. The MA(1) process "forgets" what happened more than one period in the past, in contrast to the autoregressive processes.

The moving average model of order k, MA(k), is given as

$$\underline{\epsilon}_{ij} = \underline{u}_{ij} + a_1 \underline{u}_{i-1,j} + a_2 \underline{u}_{i-2,j} + \cdots + a_k \underline{u}_{i-k,j},$$

with "memory" extending k periods in the past.

7.5 Models with Lagged Responses

Lags of the response $\underline{y}_{ij}$ can be included as covariates in addition to $\boldsymbol{x}_{ij}$. Such models are usually called autoregressive models but are sometimes also referred to as transition models [17], Markov models [17], or conditional models [11].

When occasions are equally spaced in time, a first-order autoregressive model for the responses $\underline{y}_{ij}$ can be written as

$$\underline{y}_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \gamma \underline{y}_{i-1,j} + \underline{\epsilon}_{ij}.$$

A path diagram for this model is shown under "AR(1) responses" in the third panel of Fig. 7.4. Assuming that the process is weakly stationary, $|\gamma| < 1$, the covariance structure is shown in panel E of Table 7.1. An extension of the autoregressive model is the antedependence model, which specifies a different parameter γ_i for each occasion.

A first-order autoregressive model for the responses was estimated for the wage panel data giving the estimates shown in the last column of Table 7.2. The regression coefficient of the lagged response is estimated as $\hat{\gamma} = 0.56$. As would be expected, many of the other regression coefficients change considerably due to controlling for the lagged response.

As for the residual autoregressive structure, the first-order autoregressive structure for responses is often deemed unrealistic, since the correlations fall off too rapidly with increasing time lags. Once again, this may be rectified by specifying a higher-order autoregressive process AR(k),

$$\underline{y}_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \gamma_1 \underline{y}_{i-1,j} + \gamma_2 \underline{y}_{i-2,j} + \cdots + \gamma_k \underline{y}_{i-k,j} + \underline{\epsilon}_{ij}.$$

Apart from being of interest in its own right, the lagged response model is useful for distinguishing between different longitudinal models. Consider two simple models; a model with a lagged response and lagged covariate but independent residuals $\underline{\epsilon}_{ij}$

$$\underline{y}_{ij} = \gamma\, \underline{y}_{i-1,j} + \beta_1 x_{ij} + \beta_2 x_{i-1,j} + \underline{\epsilon}_{ij}, \tag{7.7}$$

and an autocorrelation model without lagged response or lagged covariate

$$\underline{y}_{ij} = \beta\, x_{ij} + \underline{\epsilon}_{ij},$$

but residuals $\underline{\epsilon}_{ij}$ having an AR(1) structure. Substituting first for $\underline{\epsilon}_{ij} = \alpha \underline{\epsilon}_{i-1,j} + \underline{u}_{ij}$ from (7.6), then for $\underline{\epsilon}_{i-1,j} = \underline{y}_{i-1,j} - \beta\, x_{i-1,j}$, and reexpressing, the autocorrelation model can alternatively be written as

$$\underline{y}_{ij} = \alpha\, \underline{y}_{i-1,j} + \beta\, x_{ij} - \alpha\beta\, x_{i-1,j} + \underline{u}_{ij}.$$

This model is equivalent to model (7.7) with the restriction $\beta_2 = -\gamma\beta_1$. This means that we can use (7.7) to discriminate between autocorrelated residuals and lagged responses.

Use of lagged response models should be conducted with caution. First, lags should be avoided if the lagged effects do not have a “causal” interpretation since the interpretation of $\boldsymbol{\beta}$ changes when $\underline{y}_{i-1,j}$ is included as an additional covariate. Second, the models require balanced data in the sense that all units are measured on the same occasions. It is problematic if the response for a unit is missing at an occasion. In practice, the entire unit is often discarded in this case. Third, lagged response models reduce the sample size. This is because the $\underline{y}_{ij}$ on the first occasions can only serve as covariates and cannot be regressed on lagged responses (which are missing). Alternatively, if the lagged responses are treated as endogenous, the sample size is not reduced, but an initial condition problem arises for the common situation where the process is ongoing when we start observing it [28]. Finally, if random effects are also included in the model, even the initial response (at the start of the process) becomes endogenous [28].

An advantage of lagged response models as compared to models with autoregressive residuals is that they can easily be used for response types other than the continuous. Heckman [29] discusses a very general framework for longitudinal modeling of dichotomous responses, for instance combining lagged responses with random effects.

7.6 Marginal Approaches

As is clear from the general form of generalized linear mixed models (including linear mixed models) in (7.2), the model linking the expectation to the covariates is specified conditional on the unit-specific random effects $\underline{\boldsymbol{\delta}}_j$. The regression coefficients $\boldsymbol{\beta}$ therefore have a conditional or unit-specific interpretation.

The marginal or population averaged expectations of the responses can be obtained by integrating the inverse link function of the linear predictor over the random effects distribution

$$E(\underline{y}_{ij}) = \int g^{-1}(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\delta}_j)\ \phi(\boldsymbol{\delta}_j; \boldsymbol{\emptyset}, \boldsymbol{\Omega})\ \mathrm{d}\boldsymbol{\delta}_j\,, \tag{7.8}$$

where $\phi(\boldsymbol{\delta}_j; \boldsymbol{\emptyset}, \boldsymbol{\Omega})$ is the multivariate normal density of the random effects.

For linear mixed models, the link function $g(\cdot)$ is the identity and the population averaged expectation is simply the fixed part $\boldsymbol{x}'_{ij}\boldsymbol{\beta}$ of the model. Therefore, the regression coefficients $\boldsymbol{\beta}$ also have a population averaged interpretation in this case. In the linear case, it could therefore be argued that it does not matter whether the model is interpreted conditionally or marginally. However, in the marginal interpretation of the random part, only the covariance matrix $\boldsymbol{\Psi}_j$ of the total random part (as shown in Table 7.1) is interpreted, not the individual covariance matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}_j \overset{\Delta}{=} \mathrm{Cov}(\underline{\boldsymbol{\epsilon}}_j)$. Thus, Verbeke and Molenberghs [65] argue that the covariance matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}_j$ need not be positive semi-definite in this case as long as $\boldsymbol{\Psi}_j$ is positive semi-definite.

For link functions other than the identity, the expectation in (7.8) differs from the fixed part of the model. For generalized linear mixed models with probit links, we can derive a simple form for the population averaged expectation using the latent response formulation. The model can be specified as

$$\underline{y}^*_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\underline{\boldsymbol{\delta}}_j + \underline{\epsilon}_{ij}, \qquad \underline{\boldsymbol{\delta}}_j \sim \mathcal{N}(\boldsymbol{\emptyset}, \boldsymbol{\Omega}), \quad \underline{\epsilon}_{ij} \sim \mathcal{N}(0,1),$$

with $\underline{y}_{ij} = 1$ if $\underline{y}^*_{ij} > 0$ and $\underline{y}_{ij} = 0$ otherwise. The unit-specific model then becomes

$$E(\underline{y}_{ij} \mid \underline{\boldsymbol{\delta}}_j) = \Pr(\underline{y}_{ij} = 1 \mid \underline{\boldsymbol{\delta}}_j) = \Phi(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\underline{\boldsymbol{\delta}}_j),$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, the inverse probit link. The corresponding marginal model is given by

$$
\begin{aligned}
E(\underline{y}_{ij}) &= \Pr(\underline{y}_{ij} = 1)\\
&= \Pr(\underline{y}^*_{ij} > 0)\\
&= \Pr(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\underline{\boldsymbol{\delta}}_j + \underline{\epsilon}_{ij} > 0)\\
&= \Pr\big(-(\boldsymbol{z}'_{ij}\underline{\boldsymbol{\delta}}_j + \underline{\epsilon}_{ij}) \leq \boldsymbol{x}'_{ij}\boldsymbol{\beta}\big)\\
&= \Pr\left(\frac{\boldsymbol{z}'_{ij}\underline{\boldsymbol{\delta}}_j + \underline{\epsilon}_{ij}}{\sqrt{\boldsymbol{z}'_{ij}\boldsymbol{\Omega}\boldsymbol{z}_{ij} + 1}} \leq \frac{\boldsymbol{x}'_{ij}\boldsymbol{\beta}}{\sqrt{\boldsymbol{z}'_{ij}\boldsymbol{\Omega}\boldsymbol{z}_{ij} + 1}}\right)\\
&= \Phi\left(\frac{\boldsymbol{x}'_{ij}\boldsymbol{\beta}}{\sqrt{\boldsymbol{z}'_{ij}\boldsymbol{\Omega}\boldsymbol{z}_{ij} + 1}}\right),
\end{aligned}
\tag{7.9}
$$

where the denominator is greater than 1 if $\boldsymbol{\Omega} \neq \boldsymbol{\emptyset}$. For a random intercept probit model, the denominator is a constant $\sqrt{\omega^2 + 1}$ and the population averaged model has the same functional form as the unit-specific model but with attenuated regression coefficients $\boldsymbol{\beta}/\sqrt{\omega^2 + 1}$. This attenuation is shown graphically in Fig. 7.5, where the dashed curves represent unit-specific relationships for a random intercept probit model with a single covariate, whereas the flatter solid curve represents the population averaged relationship.

It can be seen from (7.9) that if any aspect of the random part of the model is altered, the regression coefficients must also be altered to obtain a

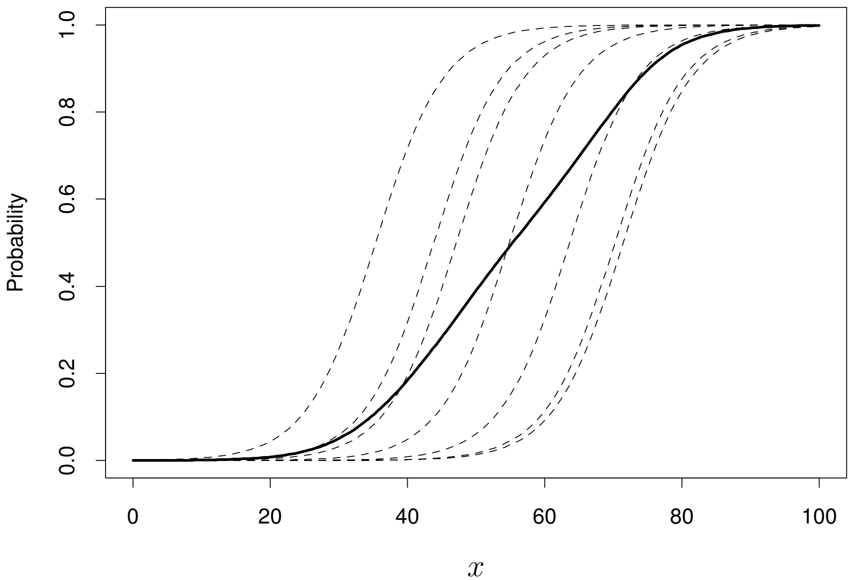

Fig. 7.5 Unit-specific versus population-averaged probit regression.

good fit to the empirical (marginal) relationship between the response and covariates. Therefore, estimates of the unit-specific regression parameters become inconsistent under misspecification of the random part except for linear mixed models.

Whether unit-specific or population averaged effects are of interest will depend on the context. For example, in public health, population averaged effects may be of interest, whereas unit-specific effects are important for the patient and clinician. An advantage of unit-specific effects is that they are more likely to be stable across populations, whereas marginal effects depend greatly on the between-unit heterogeneity, which will generally differ between populations.

If interest is focused on marginal effects and between-unit heterogeneity or longitudinal dependence are regarded as a nuisance, generalized estimating equations (GEE) [36, 71] can be used. The simplest version is to estimate the mean structure as if the responses were independent and then adjust standard errors for the dependence using the so-called sandwich estimator. The estimates of the population averaged regression parameters can be shown to be consistent, but if the responses are correlated, they are not efficient. To increase efficiency a "working correlation matrix" is therefore specified within a multivariate extension of the iteratively reweighted least squares algorithm for generalized linear models. Typically, one of the structures listed in Table 7.1 is used for the working correlation matrix of the residuals $\underline{y}_{ij} - g^{-1}(\boldsymbol{x}'_{ij}\boldsymbol{\beta})$, as well as unrestricted and independence correlation structures. The working correlation matrix is combined with the variance function of an appropriate generalized linear model, typically allowing for overdispersion if the responses are counts. It is important to note that, apart from continuous responses, the specified correlation structures generally cannot be derived from a statistical model. Thus, GEE is a multivariate quasi-likelihood approach with no proper likelihood.

There are also "proper" marginal statistical models with corresponding likelihoods. Examples include the Bahadur [2] and Dale [13] models, which parameterize dependence via marginal correlations and marginal bivariate odds ratios, respectively [19, 44]. See Molenberghs and Verbeke [45] for an overview of these models.

Heagerty and Zeger [27] introduce random effects models where the marginal mean is regressed on covariates as in GEE. In these models, the relationship between the conditional mean (given the random effects) and the covariates is found by solving the integral equation (7.8) linking the conditional and marginal means. Interestingly, the integral involved can be written as a unidimensional integral over the distribution of the sum of the terms in the random part of the model.

7.7 Concluding Remarks

It is straightforward to extend the longitudinal models discussed here to situations where units are clustered by including random effects varying at higher levels.

We have focused on linear and quadratic growth models, but nonlinear growth models can also be specified via linear mixed models using higher-order polynomials of time or splines [62]. Nonlinear mixed models [50] can be preferable if specific functional forms are suggested by substantive theory as in pharmacokinetics.

Useful books on modeling longitudinal data include Skrondal and Rabe-Hesketh [60], Hand and Crowder [24], Crowder and Hand [12], Vonesh and Chinchilli [67], Jones [33], Hsiao [32], Baltagi [3], Wooldridge [70], Lindsey [38], Verbeke and Molenberghs [65], Diggle et al. [17], and Fitzmaurice et al. [20].

References

1. M. Aitkin. A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55:117–128, 1999.
2. R. R. Bahadur. A representation of the joint distribution of responses to n dichotomous items. In H. Solomon, editor, *Studies in Item Analysis and Prediction*, volume 6, pages 158–168. Stanford University Press, Stanford, CA, 1961.
3. B. H. Baltagi. *Econometric Analysis of Panel Data*, 2nd edition. Wiley, Chichester, 2001.
4. B. H. Baltagi and P. X. Wu. Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory*, 15:814–823, 1999.
5. D. J. Bartholomew. The sensitivity of latent trait analysis to the choice of prior distribution. *British Journal of Mathematical and Statistical Psychology*, 41: 101–107, 1988.
6. G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*, 3rd edition. Prentice-Hall, Englewood Cliffs, NJ, 1994.
7. N. E. Breslow and N. Day. *Statistical Methods in Cancer Research. Vol. I — The Analysis of Case-Control Studies*. IARC, Lyon, 1980.
8. G. Chamberlain. Analysis of covariance with qualitative data. *Review of Economic Studies*, 47:225–238, 1980.
9. G. Chamberlain. Panel data. In Z. Griliches and M. D. Intriligator, editors, *Handbook of Econometrics*, volume 2, pages 1247–1318. North-Holland, Amsterdam, 1984.
10. D. G. Clayton. Generalized linear mixed models. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 275–301. Chapman & Hall, London, 1996.

11. D. R. Cox. The analysis of multivariate binary data. *Applied Statistics*, 21: 113–120, 1972.
12. M. J. Crowder and D. J. Hand. *Analysis of Repeated Measures*. Chapman & Hall, London, 1990.
13. J. R. Dale. Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, 42:909–917, 1986.
14. C. M. Dayton and G. B. MacReady. Concomitant variable latent class models. *Journal of the American Statistical Association*, 83:173–178, 1988.
15. J. de Leeuw and E. Meijer. Introduction to multilevel analysis. In J. de Leeuw and E. Meijer, editors, *Handbook of Multilevel Analysis*, Chapter 1. Springer, New York, 2008. (this volume)
16. P. J. Diggle. An approach to analysis of repeated measures. *Biometrics*, 44: 959–971, 1988.
17. P. J. Diggle, P. J. Heagerty, K.-Y. Liang, and S. L. Zeger. *Analysis of Longitudinal Data*, 2nd edition. Oxford University Press, Oxford, UK, 2002.
18. R. C. Elston. On estimating time-response curves. *Biometrics*, 20:643–647, 1964.
19. G. M. Fitzmaurice, N. M. Laird, and A. G. Rotnitzky. Regression models for discrete longitudinal responses. *Statistical Science*, 8:284–309, 1993.
20. G. M. Fitzmaurice, N. M. Laird, and J. H. Ware. *Applied Longitudinal Analysis*. Wiley, New York, 2004.
21. K. R. Gabriel. Ante-dependence analysis of an ordered set of variables. *Annals of Mathematical Statistics*, 33:201–212, 1962.
22. W. A. Gibson. Three multivariate models: factor analysis, latent structure analysis and latent profile analysis. *Psychometrika*, 24:229–252, 1959.
23. H. Goldstein. Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, 78:45–51, 1991.
24. D. J. Hand and M. J. Crowder. *Practical Longitudinal Data Analysis*. Chapman & Hall, London, 1996.
25. J. A. Hausman. Specification tests in econometrics. *Econometrica*, 46:1251–1271, 1978.
26. P. J. Heagerty and B. F. Kurland. Misspecified maximum likelihood estimates and generalized linear mixed models. *Biometrika*, 88:973–985, 2001.
27. P. J. Heagerty and S. L. Zeger. Marginalized multilevel models and likelihood inference. *Statistical Science*, 15:1–26, 2000. (with discussion)
28. J. J. Heckman. The incidental parameters problem and the problem of initial conditions in estimating a discrete stochastic process and some Monte Carlo evidence on their practical importance. In C. F. Manski and D. L. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 179–196. MIT Press, Cambridge, MA, 1981. (Downloadable from `http://elsa.berkeley.edu/~mcfadden/discrete.html`)
29. J. J. Heckman. Statistical models for discrete panel data. In C. F. Manski and D. L. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 114–178. MIT Press, Cambridge, MA, 1981. (Downloadable from `http://elsa.berkeley.edu/~mcfadden/discrete.html`)

30. J. J. Heckman and B. Singer. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52: 271–320, 1984.
31. D. Hedeker. Multilevel models for ordinal and nominal variables. In J. de Leeuw and E. Meijer, editors, *Handbook of Multilevel Analysis*, Chapter 6. Springer, New York, 2008. (this volume)
32. C. Hsiao. *Analysis of Panel Data*, 2nd edition. Cambridge University Press, Cambridge, UK, 2003.
33. R. H. Jones. *Longitudinal Data with Serial Correlation: A State-Space Approach.* Chapman & Hall, London, 1993.
34. N. M. Laird. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73:805–811, 1978.
35. N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38:963–974, 1982.
36. K.-Y. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22, 1986.
37. B. G. Lindsay and M. L. Lesperance. A review of semiparametric mixture models. *Journal of Statistical Planning and Inference*, 47:5–28, 1995.
38. J. K. Lindsey. *Models for Repeated Measurements*, 2nd edition. Oxford University Press, Oxford, UK, 1999.
39. N. T. Longford. *Random Coefficient Models.* Oxford University Press, Oxford, UK, 1993.
40. N. T. Longford. Missing data. In J. de Leeuw and E. Meijer, editors, *Handbook of Multilevel Analysis*, Chapter 10. Springer, New York, 2008. (this volume)
41. G. S. Maddala. The use of variance components models in pooling cross-section and time series data. *Econometrica*, 39:341–358, 1971.
42. C. E. McCulloch and S. R. Searle. *Generalized, Linear and Mixed Models.* Wiley, New York, 2001.
43. W. Meredith and J. Tisak. Latent curve analysis. *Psychometrika*, 55:107–122, 1990.
44. G. Molenberghs. Generalized estimating equations. In M. Aerts, H. Geys, G. Molenberghs, and L. M. Ryan, editors, *Topics in Modelling Clustered Data*, pages 47–75. Chapman & Hall/CRC, Boca Raton, FL, 2002.
45. G. Molenberghs and G. Verbeke. *Models for Discrete Longitudinal Data.* Springer, New York, 2005.
46. B. O. Muthén and K. Shedden. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55:463–469, 1999.
47. D. S. Nagin and K. C. Land. Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric mixed Poisson model. *Criminology*, 31:327–362, 1993.
48. J. M. Neuhaus, W. W. Hauck, and J. D. Kalbfleisch. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, 79:755–762, 1992.
49. J. Neyman and E. L. Scott. Consistent estimates based on partially consistent observations. *Econometrica*, 16:1–32, 1948.

50. J. C. Pinheiro and D. M. Bates. *Mixed-Effects Models in S and S-PLUS*. Springer, New York, 2000.
51. R. E. Quandt. A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67:306–310, 1972.
52. S. Rabe-Hesketh, A. Pickles, and A. Skrondal. Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation. *Statistical Modelling*, 3:215–232, 2003.
53. S. Rabe-Hesketh and A. Skrondal. *Multilevel and Longitudinal Modeling Using Stata*. Stata Press, College Station, TX, 2005.
54. S. Rabe-Hesketh and A. Skrondal. Generalized linear mixed effects models. In G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs, editors, *Longitudinal Data Analysis: A Handbook of Modern Statistical Methods*. Chapman & Hall/CRC, Boca Raton, FL, 2007.
55. S. Rabe-Hesketh, A. Skrondal, and A. Pickles. GLLAMM manual. Working Paper 160, U.C. Berkeley Division of Biostatistics, Berkeley, CA, 2004. (Downloadable from `http://www.bepress.com/ucbbiostat/paper160/`)
56. S. Rabe-Hesketh, A. Skrondal, and A. Pickles. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, 128:301–323, 2005.
57. G. Rasch. *Probabilistic Models for Some Intelligence and Attainment Tests*. Danmarks Pædagogiske Institut, Copenhagen, 1960.
58. G. Rodríguez. Multilevel generalized linear models. In J. de Leeuw and E. Meijer, editors, *Handbook of Multilevel Analysis*, Chapter 9. Springer, New York, 2008. (this volume)
59. D. B. Rubin. Inference and missing data. *Biometrika*, 63:581–592, 1976. (with discussion)
60. A. Skrondal and S. Rabe-Hesketh. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall/CRC, Boca Raton, FL, 2004.
61. T. A. B. Snijders and J. Berkhof. Diagnostic checks for multilevel models. In J. de Leeuw and E. Meijer, editors, *Handbook of Multilevel Analysis*, Chapter 3. Springer, New York, 2008. (this volume)
62. T. A. B. Snijders and R. J. Bosker. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage, London, 1999.
63. StataCorp. *Stata Statistical Software: Release 9.0*. Stata Corporation, College Station, TX, 2005.
64. F. Vella and M. Verbeek. Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men. *Journal of Applied Econometrics*, 13:163–183, 1998.
65. G. Verbeke and G. Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer, New York, 2000.
66. G. Verbeke, B. Spiessens, and E. Lesaffre. Conditional linear mixed models. *American Statistician*, 55:25–34, 2001.
67. E. F. Vonesh and V. M. Chinchilli. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, New York, 1997.

68. M. Wedel and W. A. Kamakura. *Market Segmentation: Conceptual and Methodological Foundations*, 2nd edition. Kluwer, Dordrecht, 2000.
69. M. Wedel and W. A. Kamakura. Factor analysis with (mixed) observed and latent variables. *Psychometrika*, 66:515–530, 2001.
70. J. M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, 2002.
71. S. L. Zeger and K.-Y. Liang. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42:121–130, 1986.