22

Random-effects Models with Serial Correlation

22.1 Introduction

In the previous chapter, we presented a random-effects based probit model and applied pseudo-likelihood ideas for parameter estimation. The model was generated from a multivariate normally distributed latent variable. This means that the latent variable follows a linear mixed model. An obvious extension is the inclusion of serial correlation, or autocorrelation, as can be done for the standard linear mixed-effects model. The extension proposed by Renard, Molenberghs, and Geys (2004) is the basis for this chapter. The model presented in Section 21.3 exhibits residual correlation between the surrogate and true endpoints on the same subject, in addition to the correlation induced by the random effects. The approach formulated in this chapter can be seen as a general version of this.

Barbosa and Goldstein (2000) propose to extend the standard multilevel model for binary outcomes, and hence the standard generalized linear mixed model, by allowing the residuals at the individual level to be correlated. These authors wrote the covariance between residuals for individual i at occasions j and k as

$$\sqrt{\pi_{ij}(1-\pi_{ij})\pi_{ik}(1-\pi_{ik})}f(|t_{ij}-t_{ik}|),$$

where the conditional mean, given random effects $\boldsymbol{b}_i$, $\pi_{ij} = E(y_{ij}|\boldsymbol{b}_i)$ is modeled as usual and $f(u)$ is a function of u, the time lag between measurement times t_{ij} and t_{ik}, i.e., $|t_{ij} - t_{ik}|$. For example, Barbosa and Goldstein

(2000) proposed the form:

$$f(u) = \alpha + \exp[-\kappa(u)], \tag{22.1}$$

for some function κ of the time lag, and they used the PQL algorithm to estimate parameters. In what follows, we will propose different parametric shapes for autocorrelation functions. A drawback of this approach, common with other PQL applications, especially with binary data, is the severe bias that can result. Also, the correction described above is *ad hoc* and falls outside the likelihood framework.

We first propose a full probabilistic model, starting from a general probit model, based on an underlying latent linear mixed model with serial correlation. The model is proposed in Section 22.2. Full likelihood estimation of this model is computationally demanding, however, and we therefore propose to use pairwise likelihood for estimation purposes in Section 22.3, building on the methodology presented in Chapter 21. In Section 22.4, a generalized linear mixed models augmented with autocorrelation is presented. The psychiatric study, analyzed before in Section 21.4, will be analyzed again in Section 22.5, using both autocorrelation methods. Whereas the analysis in Section 21.4 considered the specific context of surrogate marker evaluation, here we focus on the CGI outcome only. In Section 22.6, SAS code to fit the random-effects multivariate probit model, with or without serial correlation, as well as the generalized linear mixed model with serial correlation, is presented.

22.2 A Multilevel Probit Model with Autocorrelation

The model we propose for repeated binary data extends model (21.2), i.e., it extends the standard hierarchical or multilevel probit model. It is related to the model discussed in Heagerty and Lele (1998), which deals with binary spatial data. We will focus on a two-level hierarchy, or two-level model with, using multilevel terminology, subjects at the second level and measurements within subjects at the first level.

As in Section 21.2, we will introduce the model from a latent variable perspective. As usual, let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$ denote the vector of binary measurements on subject i ($i = 1, \ldots N$). We posit the existence of an unobserved continuous variable $\widetilde{Y}_{ij}$ and assume that the observed binary response is obtained by dichotomizing $\widetilde{Y}_{ij}$ based on a certain threshold or cut-off value. This threshold can be chosen to be 0 without loss of generality, provided an intercept term is included in the model. In other words, it is assumed that a positive response, $Y_{ij} = 1$, is recorded if $\widetilde{Y}_{ij} > 0$ and a negative response ($Y_{ij} = 0$) otherwise. On the latent variable scale the

model, generalizing (21.2), can be written as:

$$\widetilde{Y}_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i + \widetilde{\varepsilon}_{ij}. \tag{22.2}$$

The standard multilevel probit model is obtained by assuming that the random effects $\boldsymbol{b}_i$ and residual error terms $\widetilde{\varepsilon}_{ij}$ are normally distributed. An additional assumption is that of conditional independence among responses, that is, conditionally on $\boldsymbol{b}_i$, the $\widetilde{Y}_{ij}$'s are independent. This implies that $\boldsymbol{b}_i \sim N(0, D)$ and, for reasons of identification, that $\widetilde{Y}_{ij} \sim N(0, 1)$. Whereas this assumption was made in Chapter 21, we relax it here by assuming instead that the $\widetilde{Y}_{ij}$'s are realizations from a stationary unit-variance Gaussian process $\widetilde{\varepsilon}(t)$ with autocorrelation function

$$\text{corr}[\widetilde{\varepsilon}(t), \widetilde{\varepsilon}(t')] = \rho(|t' - t|), \tag{22.3}$$

which is similar in spirit to (22.1). Indeed, following Goldstein, Healy, and Rasbash (1994), we assume that $\rho(u) = \exp[-\kappa(u)]$, where $\kappa(u)$ is a positive increasing function, not necessarily linear. Obvious choices include

$$\kappa(u) = \alpha u,$$

the exponential decay model,

$$\kappa(u) = \alpha u^2,$$

the Gaussian decay model, or, more generally,

$$\kappa(u) = \sum_{k=1}^{K} \alpha_k u^k$$

for any (fractional) polynomial constrained to take on positive values on $[0, +\infty[$. As pointed out by Goldstein, Healy, and Rasbash (1994), a difficulty when $\kappa(u)$ is a polynomial is that successive powers tend to be highly correlated and this may cause estimation difficulties. Another possible choice is then to add an inverse polynomial term such as in

$$\kappa(u) = \alpha_1 u + \alpha_2 u^{-1},$$

which avoids the high correlations associated with ordinary polynomial functions. One could even consider fractional polynomials within the κ function. Verbeke and Molenberghs (2000, Section 10.3) provide examples of serial correlation functions with fractional polynomials. Another useful extension is to make the parameters α_k explicitly dependent on explanatory variables. As to the choice of the κ function, Goldstein, Healy, and Rasbash (1994) state that it should "contain as few parameters as necessary to be flexible enough to describe real data. (...) There seems to be little substantive guidance on choice, and it is likely that different functional

forms will be appropriate for different kinds of data." Especially when covariates are allowed in the autocorrelation function, options for $\kappa(u)$ and hence for $\rho(u)$ in (22.3) are virtually unlimited.

As a result, $\widetilde{\boldsymbol{Y}}_i$ is a normally distributed vector with variance-covariance matrix $\Sigma_i = R(\boldsymbol{t}_i) = R_i$, since the variances are kept equal to unity, and for reasons of model identification. The matrix $R(\boldsymbol{t}_i)$ has its (j,k)th element equal to $\rho(|t_{ij} - t_{ik}|)$, where t_{ij} is the time at which the jth measurement on subject i is made.

22.3 Parameter Estimation for the Multilevel Probit Model

The log-likelihood for the observed (binary) data can be written

$$\begin{aligned}\ell &= \sum_{i=1}^{N} \sum_{a_{i1},\ldots,a_{in_i}=0}^{1} \delta_{a_{i1},\ldots,a_{in_i}} \\ &\quad \times \ln \int P(Y_{i1} = a_{i1}, \ldots, Y_{in_i} = a_{in_i} | \boldsymbol{b}_i) \phi(\boldsymbol{b}_i) d\boldsymbol{b}_i, \end{aligned} \tag{22.4}$$

with

$$\delta_{a_{i1},\ldots,a_{in_i}} = \begin{cases} 1 & \text{if } Y_{i1} = a_{i1}, \ldots, Y_{in_i} = a_{in_i}, \\ 0 & \text{otherwise.} \end{cases}$$

Exactly as in Chapter 21, this expression entails the evaluation of multivariate normal probabilities. For instance, we have

$$\begin{aligned} &P(Y_{i1} = 1, \ldots, Y_{in_i} = 1 | \boldsymbol{b}_i) \\ &= P(\widetilde{Y}_{i1} > 0, \ldots, \widetilde{Y}_{in_i} > 0 | \boldsymbol{b}_i) \\ &= \int_{-\infty}^{\xi_{i1}} \cdots \int_{-\infty}^{\xi_{in_i}} \phi\big[x_1, \ldots, x_{n_i}; R(\boldsymbol{t}_i)\big] dx_1 \ldots dx_{n_i}, \end{aligned} \tag{22.5}$$

where we define

$$\xi_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + z'_{ij}\boldsymbol{b}_i,$$

$\phi(\boldsymbol{x}; R)$ denotes the standardized multivariate normal density function, in the sense of having unit variances, with correlation matrix R.

As in Section 21.2, we propose the use of maximum pairwise likelihood (PL) to overcome the computational burden of full likelihood. In this case, we assemble all possible pairwise probabilities $P(Y_{ij} = \ell, Y_{ik} = m)$ $(\ell, k = 0, 1)$ within the ith unit. For the present model, these marginal bivariate probabilities can all be expressed in terms of univariate and bivariate probits that are computationally inexpensive to evaluate. For instance,

we have

$$P(Y_{ij} = 1, Y_{ik} = 1) = \Phi_2\left(\xi_{ij}, \xi_{ik}; \rho_{ijk}\right), \tag{22.6}$$

with

$$\xi_{ij} = \frac{\boldsymbol{x}'_{ij}\boldsymbol{\beta}}{\sqrt{\text{var}(\widetilde{Y}_{ij})}} \tag{22.7}$$

and overall correlations, induced in part by the random-effects structure and in part by the autocorrelation,

$$\widetilde{\rho}_{ijk} = \frac{\boldsymbol{z}'_{ij} D \boldsymbol{z}_{ik} + \rho(|t_{ij} - t_{ik}|)}{\sqrt{1 + \boldsymbol{z}'_{ij} D \boldsymbol{z}_{ij}}\sqrt{1 + \boldsymbol{z}'_{ik} D \boldsymbol{z}_{ik}}}, \tag{22.8}$$

where D denotes the variance-covariance matrix of $\boldsymbol{b}_i$, the function Φ_2 denotes the standard bivariate Gaussian distribution function, and $\text{var}(\widetilde{Y}_{ij})$, $\text{var}(\widetilde{Y}_{i'j})$ and $\rho_{ii'j}$ are obtained by selecting the appropriate 2×2 submatrix of the (marginal) covariance matrix of $\widetilde{\boldsymbol{Y}}_i$,

$$V_i = Z_i D Z'_i + R(\boldsymbol{t}_i).$$

Parameter estimation and inference follows from the methodology described in Section 21.2, built upon estimation and inferential tools laid out in Sections 9.2 and 9.3. In particular, the sandwich estimation ought to be used for precision estimation, and hypothesis testing can proceed using the test statistics laid out in Section 9.3.

A SAS macro was written to implement the methodology in the case of a model with random intercept and autocorrelation function $\rho(u) = \exp(-\alpha u^k)$. The algorithm was implemented in SAS IML (SAS Institute Inc. 1995) and maximization of the log PL performed using the NLPDD (Double-Dogleg) optimization routine (SAS Institute Inc. 1995). This optimization procedure requires only function and gradient calls that are less expensive to evaluate than second-order derivatives. To avoid constrained optimization, a Cholesky decomposition for D was used and the parameter α was log transformed. To estimate the covariance matrix of the PL estimator by way of the sandwich estimator, it should be observed that (9.6) requires only gradient calls, whereas (9.5) can be computed using numerical second-order derivatives (e.g., by forward difference approximation).

Renard, Molenberghs, and Geys (2004) assessed the proposed methodology by means of a simulation study. Their simulations indicate that the mean and dependence parameters are strongly biased with a small number of subjects ($N = 100$). Increasing the number of measurements somewhat reduces the extent of bias. With a medium number of subjects ($N = 500$), parameters are still largely biased when the number of measurement occasions is small ($n_i = 5$) but the bias falls within more acceptable limits with

an increased number of measurement occasions per subject. The autocorrelation parameter is noticeably biased, though. With a large number of subjects ($N = 1000$), the bias for the mean parameters and the random-effect variance parameter becomes small but for the autocorrelation parameter it is still sizeable with datasets containing as many as $N = 20,000$ observations. Regarding precision estimation, the estimated standard errors somewhat overestimate the sampling variability, especially for the random-effect variance. These authors also reported on various convergence problems. This is not surprising for complicated models of this nature. Already for the general linear random-effects model, involving fixed effects, random effects, and serial correlation, convergence can be very involved. Here, the model additionally has a non-linear link structure and further binary data carry way less information than continuous outcomes. Model fitting for models this complex should therefore proceed with caution.

22.4 A Generalized Linear Mixed Model with Autocorrelation

In Section 8.8, marginal models based on linearization were considered, based on the concept of data approximation which later was employed in Section 14.4. In the first case, dependence among repeated measures is introduced by means of a residual covariance matrix, Σ_i in (8.36). In the second case, random effects are introduced. In both cases, the SAS procedure GLIMMIX could be used for parameter estimation, using PQL or MQL approximation.

The basis for this model development is the decomposition, in line with (14.6):

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i, \tag{22.9}$$

where $\boldsymbol{\mu}_i$ is specified by means of a GLMM and $\boldsymbol{\varepsilon}_i$ is the residual error structure. In a standard GLMM, $\boldsymbol{\varepsilon}_i$ is assumed to be uncorrelated and hence does not lead to additional parameters, as the variances follow from the mean-variance link. In the linearization based method of Section 8.8, $\boldsymbol{\mu}_i$ does not contain random effects, but $\boldsymbol{\varepsilon}_i$ is assumed to be correlated. One can choose an autocorrelation model to determine the variance of $\boldsymbol{\varepsilon}_i$ in (22.9), i.e., the matrix Σ_i in (8.36). Obvious choices include spatial exponential or spatial Gaussian models, an AR(1) structure if measurements are equally spaced, or any autocorrelation structure described in Section 22.2.

Combining both ideas produces a generalized linear mixed model with autocorrelation, just as the model in Section 22.2. The main difference is that (22.2) specifies a linear mixed model with autocorrelation in terms of the latent outcome underlying the multivariate probit model, whereas here the random effects are introduced at the level of the linear predictor

TABLE 22.1. *Meta-analysis in Schizophrenia. Maximum pseudo-likelihood parameter estimates (standard errors) for the probit random-intercept model with and without autocorrelation. The exponential and Gaussian models were taken for the autocorrelation structure. Coding for 'Treat': 0 = standard, 1 = experimental.*

		Random intercept + autocorrelation	
Effect	Random intercept	Expon.	Gaussian
Intercept	-0.27 (0.16)	-0.18 (0.12)	-0.22 (0.14)
Week 1	-1.88 (0.18)	-1.34 (0.20)	-1.62 (0.17)
Week 2	-1.17 (0.17)	-0.88 (0.16)	-1.08 (0.15)
Week 4	-0.70 (0.16)	-0.52 (0.13)	-0.62 (0.15)
Week 6	-0.21 (0.14)	-0.16 (0.11)	-0.18 (0.13)
Treat×Week 1	0.29 (0.21)	0.19 (0.15)	0.23 (0.18)
Treat×Week 2	0.58 (0.21)	0.43 (0.16)	0.52 (0.18)
Treat×Week 4	0.54 (0.21)	0.39 (0.16)	0.47 (0.19)
Treat×Week 6	0.33 (0.22)	0.24 (0.16)	0.29 (0.19)
Treat×Week 8	0.20 (0.22)	0.14 (0.17)	0.17 (0.20)
R.I. s.d. τ	1.83 (0.11)	1.12 (0.23)	1.53 (0.12)
R.I. var. τ^2	3.53 (0.40)	1.25 (0.52)	2.34 (0.37)
Autocorr. par. $\ln \phi$		-1.34 (0.33)	-1.21 (0.17)
Autocorr. $\rho = \rho(u = 1)$		0.27 (0.03)	0.26 (0.02)
$\log PL$	-1727.0	-1722.2	-1726.3

describing $\boldsymbol{\mu}_i$ after application of the link function, whereas the autocorrelation structure is introduced at the level of $\boldsymbol{\varepsilon}_i$. In other words, whereas the random effects and autocorrelation structures sit 'side by side' in (22.2), this is not the case here. To illustrate this, consider a logit-based model with autocorrelation function:

$$\boldsymbol{Y}_i = \frac{e^{X_i\boldsymbol{\beta}+Z_i\boldsymbol{b}_i}}{1+e^{X_i\boldsymbol{\beta}+Z_i\boldsymbol{b}_i}} + \boldsymbol{\varepsilon}_i \tag{22.10}$$

where $\boldsymbol{\varepsilon}_i$ is assumed to exhibit residual correlation, entering the covariance expression as in (20.48). Both structures enter the pseudo data as in (14.7) and it may appear that then the random effects and the residual error are side by side. However, the residual error of the pseudo data (14.7) is now a transformed version of the original error $\boldsymbol{\varepsilon}_i$.

TABLE 22.2. *Meta-analysis in Schizophrenia. PQL parameter estimates (model-based standard errors) for a linearization-based marginal model with autoregressive autocorrelation structure, random-intercept model, and random-intercept model with autoregressive autocorrelation structure. Logit link. Estimates obtained using the SAS procedure GLIMMIX. Coding for 'Treat': 0 = standard, 1 = experimental.*

Effect	Auto-correlation	Random intercept	R.I. + Autocorr.
Intercept	-0.22 (0.13)	-0.15 (0.20)	-0.13 (0.19)
Week 1	-1.58 (0.18)	-2.33 (0.24)	-2.89 (0.21)
Week 2	-0.95 (0.16)	-1.42 (0.23)	-1.80 (0.20)
Week 4	-0.54 (0.15)	-0.86 (0.22)	-1.12 (0.19)
Week 6	-0.15 (0.13)	-0.28 (0.22)	-0.37 (0.18)
Treat×Week 1	0.30 (0.20)	0.28 (0.27)	0.22 (0.27)
Treat×Week 2	0.51 (0.16)	0.63 (0.25)	0.76 (0.25)
Treat×Week 4	0.44 (0.16)	0.59 (0.25)	0.76 (0.25)
Treat×Week 6	0.26 (0.17)	0.37 (0.26)	0.48 (0.26)
Treat×Week 8	0.17 (0.18)	0.21 (0.28)	0.28 (0.28)
R.I. var. τ^2		3.54 (0.30)	5.92 (0.49)
Autocorr. par. θ	3.00 (0.14)		0.77 (0.10)
Autocorr. $\rho = \rho(u = 1)$	0.72 (0.01)		0.27 (0.04)
Autocorr. var. σ^2	1.02 (0.03)		0.55 (0.02)

22.5 A Meta-analysis of Trials in Schizophrenic Subjects

We consider the same meta-analysis based on five trials as in Section 21.4, and focus on the CGI ('Clinician's Global Impression') outcome. This is somewhat different from Section 21.4, where PANSS and CGI were analyzed jointly, in the context of surrogate marker evaluation. More specifically, the CGI overall change versus baseline is considered. Dichotomization was obtained by defining a success ($Y_{ij} = 1$) as clinical improvement since baseline (i.e., CGI grade equal to 1 or 2) and a failure otherwise.

We will first consider the multilevel probit models of Section 22.2 and then turn to generalized linear mixed models with serial correlation in Section 22.4.

The parameterization includes a saturated treatment by time model for the mean structure and include a random intercept in the model. For the autocorrelation structure, we assumed that $\kappa(u) = \alpha u^\gamma$ and tried several values of $\gamma = -1, 0.5, 1, 2$. The exponential model ($\gamma = 1$) provided the best fit in terms of (pseudo-)likelihood value at maximum. Both $\gamma = 1$ and

$\gamma = 2$ are reported in Table 22.1. We also fitted a model with

$$\kappa(u) = \alpha_1 u + \alpha_2 u^{-1},$$

but a boundary solution was obtained.

The parameter α was rewritten as $\alpha = \exp(\phi)$. This implies that the overall autocorrelation function, for $\gamma = 1$, is

$$\rho(u) = \exp[-\kappa(u)] = \exp[-\exp(\phi)u], \tag{22.11}$$

and hence the correlation between, for example, two measurements one time unit apart is

$$\rho = \rho(1) = \exp[-\kappa(1)] = \exp[-\exp(\phi)]. \tag{22.12}$$

In Table 22.1, parameter estimates and standard errors are reported for the random-intercept model with and without exponential autocorrelation structure. Apart from the autocorrelation parameter ϕ, we also present ρ as in (22.12), for ease of reference and interpretation. As can be seen, parameter estimates for the model with exponential autocorrelation are all reduced in magnitude by an amount of roughly 30%. This is essentially due to the fact that the error terms in (22.2) are assumed to be autocorrelated; hence the autocorrelation explains a certain amount of variability that is otherwise captured in the residual variance. This residual variance itself depends on the regression parameters, which is why they are affected by such a change. The $\log PL$ value shows an improvement in the fit of the model. Formal testing needs to be done based on the method laid out in Section 9.3. As stated earlier, Gaussian autocorrelation fits the data less well than exponential autocorrelation. This also explains why the regression parameters in the Gaussian decay case change less.

Let us now switch to the generalized linear mixed models with autocorrelation. The autocorrelation function can be modeled using model (22.11). However, we will use a slightly reparameterized form, in agreement with the parameterization used by SAS, for convenience:

$$\rho(u) = \exp\left(-\frac{1}{\theta}u^{\gamma}\right)$$

and thus the correlation between two measurements one time unit apart is:

$$\rho = \rho(1) = \exp\left(-\frac{1}{\theta}\right). \tag{22.13}$$

Table 22.2 presents three models, with the same fixed-effects structure as in Table 22.1. Apart from the autocorrelation parameter θ, we also present the correlation ρ as in (22.13).

The first model exhibits an exponential autocorrelation structure only, and no random effects. The second model is the random-intercepts model,

and the third model combines both features. Observe that the correlation parameter ρ for the latter model is very similar to the one obtained in Table 22.1, which is not surprising. Fixed-effects parameter estimates are different, due to two causes. First, we use the logit link in Table 22.2 *versus* the probit link in Table 22.1. Second, PQL estimation in the GLMM case is known to lead to parameter attenuation, as reported in several instances (Tables 14.1 and 17.4).

To separate both issues, the same three models as in Table 22.2, but now with probit link, are presented in Table 22.3. Now, compare the second model in Table 22.3 to the first model in Table 22.1. Both are random-intercepts models, without serial correlation and with probit link. The attenuation in the PQL case is then clear, suggesting the use of integration based methods (Section 14.3) for pure random-effects GLMM, or of the pseudo-likelihood method when autocorrelation is additionally present. At least, this comparison issues caution regarding the use of PQL for generalized linear mixed models with autocorrelation, just as care is needed in the absence of autocorrelation.

The fixed effects in the first columns of Tables 22.2 and 22.3 are somewhat smaller than in the corresponding second and third columns. This is to be expected since these models are marginal, whereas the other two are random-effects based (Chapter 16). Recall the approximate relationship between a random-intercepts model and the corresponding marginal model, given by (16.3). In fact, the discrepancy is not as large as it could be, due to the attenuation of the PQL based methods.

Another comparison is between the fixed-effects parameter estimates in Table 22.3 and their counterparts in Table 22.2. This reveals, once more, the relationship between probit based parameters and their logit counterparts, the approximate conversion factor being $\pi/\sqrt{3}$, as explained in Section 3.4.

In both Tables 22.2 and 22.3, the autocorrelation in the first model is considerably larger than in the third model. This is to be expected, as in the third model a part of the autocorrelation is captured by the random intercept, whereas all correlation is accounted for by the autocorrelation process in the first model. In the first model in both tables, the autocorrelation variance σ^2 plays the role of an overdispersion parameter, indicating no evidence for overdispersion in this case. The same cannot be said for the third models, as the variance is captured by both the random-intercept variance and the serial variance, and the relationship between both is not straightforward because non-linear, as is clear from the position of the random effects *versus* the residual association in (22.10).

TABLE 22.3. *Meta-analysis in Schizophrenia. PQL parameter estimates (model-based standard errors) for a linearization-based marginal model with autoregressive autocorrelation structure, random-intercept model, and random-intercept model with autoregressive autocorrelation structure. Probit link. Estimates obtained using the SAS procedure GLIMMIX. Coding for 'Treat': 0 = standard, 1 = experimental.*

Effect	Auto-correlation	Random intercept	R.I. + Autocorr.
Intercept	-0.14 (0.08)	-0.11 (0.11)	-0.11 (0.11)
Week 1	-0.94 (0.11)	-1.38 (0.14)	-1.68 (0.12)
Week 2	-0.58 (0.10)	-0.84 (0.13)	-1.05 (0.11)
Week 4	-0.34 (0.09)	-0.51 (0.13)	-0.66 (0.10)
Week 6	-0.10 (0.08)	-0.16 (0.13)	-0.21 (0.10)
Treat×Week 1	0.17 (0.11)	0.17 (0.16)	0.15 (0.15)
Treat×Week 2	0.30 (0.10)	0.38 (0.14)	0.44 (0.14)
Treat×Week 4	0.27 (0.10)	0.37 (0.14)	0.45 (0.14)
Treat×Week 6	0.17 (0.10)	0.23 (0.15)	0.28 (0.15)
Treat×Week 8	0.11 (0.11)	0.13 (0.16)	0.17 (0.16)
R.I. var. τ^2		1.25 (0.10)	2.08 (0.15)
Autocorr. par. θ	3.00 (0.14)		0.75 (0.10)
Autocorr. $\rho = \rho(u = 1)$	0.72 (0.01)		0.26 (0.05)
Autocorr. var. σ^2	1.02 (0.03)		0.51 (0.02)

22.6 SAS Code for Random-effects Models with Autocorrelation

The method presented in Section 22.2 has been implemented, for the case of a random-intercept probit model with autocorrelation, by Didier Renard (Renard, Molenberghs, and Geys 2004) in a SAS macro, available from the authors upon request. A call to the macro to fit the random-intercept only model in Table 22.1 is:

```
%plrint_corr(dataset=cgi, y=cgi_bin, x=weekcls treat*weekcls,
             classvar=weekcls, id=id, varinit=, weight=1,
             info=0, scorrtim=, scorrinit=, scorrpow=);
```

Most of the arguments to the macro are self-evident and in agreement with standard SAS statements. These include 'y' and 'x' for the response and independent variables, respectively, 'classvar' for the independent variables that need to be treated as class variables, and 'id' to indicate the levels of independent replication. Pseudo-likelihood is requested by 'weight=1,' whereas 'weight=0' refers to full maximum likelihood. Convergence in gra-

dient terms is governed by 'conv,' with a default value of 10^{-4}, and 'maxiter' controls the maximum number of iterations, with a default value of 100. The user can control whether the information matrix is calculated using first-order derivatives ('info=0') or rather numerically calculatead second-order derivatives ('info=1'). An initial value for the random-intercept variance can be passed on by way of 'varinit.' The remaining options control the autocorrelation process. The time variable used in the autocorrelation process is passed on through 'scorrtim.' If this argument is left empty, then no autocorrelation is included and hence a standard random-intercept probit model is obtained. The power p of the exponential process $\exp[-\alpha(t_{ij} - t_{ik})^k]$ is specified via 'scorrpow,' with a (default) value of 1 for exponential decay and $k = 2$ for Gaussian decay. The parameter α can be initialized using 'scorrinit.'

The use of these options implies that for the model with random intercept and exponential autocorrelation, the call changes to:

```
%plrint_corr(dataset=cgi, y=cgi_bin, x=weekcls treat*weekcls,
             classvar=weekcls, id=id, varinit=%str(1.117),
             weight=1, info=0, scorrtim=weekcls,
             scorrinit=%str(0.5), scorrpow=1);
```

Turning attention to the generalized linear mixed model with autocorrelation, the following code can be used:

```
proc glimmix data=m.cgi method=RSPL;
 class id weekcls;
 nloptions maxiter=50 technique=newrap absftol=1e-4;
 model cgi_bin (event='1') = weekcls treat*weekcls
                           / dist=binary link=probit solution;
 random intercept / subject=id type=un;
 random _residual_ / subject=id type=sp(exp)(timecls);
run;
```

The RANDOM statement with 'intercept' argument produces the random intercept model, whereas the serial process is invoked by means of the RANDOM statement with '_residual_' argument. The 'type=sp(exp)' requests exponential decay. Removing the first RANDOM statement produces a marginal model with autocorrelation process only. Removing the second one yields the classical random-intercept model. Removing the 'link=probit' option from the MODEL statement yields the logit link equivalents to these models. Since convergence can be challenging, it might be necessary to try several NLOPTIONS arguments to control updating, convergence criteria, etc. In our case, it has been necessary to switch the updating algorithm to Newton-Rahpson with the 'technique=newrap' option because quasi-Newton methods tend to get trapped in an infinite cycling between two or more values. Moreover, the number of iterations needs to be increased since for some analyses the default number of 20 was exceeded. Finally, the

convergence criterion was relaxed, either in terms of the function itself, using the 'absftol=' option, or in terms of the gradient, using the 'absgtol=' option.

22.7 Concluding Remarks

We have presented two approaches to deal with hierarchical generalized linear models, with both random effects and serial correlations. The first one is based on a probit model, overlaying a linear mixed model. The second one is based on the generalized linear mixed model framework, where the residual error terms are allowed to be correlated.

Both approaches have advantages and disadvantages. The hierarchical probit model is simple and appealing because the various effects enter the latent variable in a way very similar to the linear mixed model. On the other hand, the approach is restricted to a probit specification. Even though extensions could start from other fully specified marginal models, the properties and simplicity of an underlying multivariate normal are important factors rendering the probit specification unique. Although this seems to imply a restriction to binary data at the same time, the multilevel probit approach could in fact be applied to ordinal data, as in Section 7.6. Pseudo-likelihood provides a convenient estimation method. Renard, Molenberghs, and Geys (2004) reported good computational properties, but a loss in efficiency. A large sample size might be necessary for the asymptotic properties of the PL estimator to hold and the autocorrelation parameter may be subject to substantial bias in samples of small to moderate size. Nevertheless, in the analysis of our example, the autocorrelation parameter was estimated very similarly between the multilevel probit model and the GLMM-based approach.

Although the PL estimation procedure can, in principle, be applied to hierarchies with more than two levels, practical limitations on the number of levels will arise. For instance, in a three-level model all possible pairs within and between level 2 units pertaining to the same level 3 unit should be considered. This will become computationally prohibitive as the number of levels and the number of replicates per level increase.

The GLMM-based approach is very general and applies to all link functions. Nevertheless, because the random effects and the autocorrelation structure enter at different places into the model, irrespective of whether one consider the direct outcomes or the pseudo data derived from them, the model is somewhat less transparant and, for example, calculation of the overall variance or the overall correlation is far from straightforward. Although PQL is convenient, it suffers from potentially severe attenuation bias in the fixed efects parameter estimates, the estimates of the variance

components, as well as in all standard errors. This phenomenon has been reported before and switching from PQL to MQL would make things worse.

Finally, convergence difficulties should be anticipated to occur quite frequently in applications, regardless of which of the two routes were chosen. Even in linear mixed models, convergence failures are relatively common when modeling of the covariance structure involves joint specification of random effects, serial correlation, and measurement error, simply because these components of variability cannot easily be disentangled. An example in the context of the linear mixed model can be found in Verbeke and Molenberghs (2000, Section 9.4). Not surprisingly, this phenomenon amplifies with binary data, which contain less information than their continuous counterparts.