

# 15

## Fitting Generalized Linear Mixed Models with SAS

### 15.1 Introduction

Nowadays, many software packages allow for fitting of generalized linear mixed models, using one or several of the estimation procedures discussed in Chapter 14. Amongst the commercially available packages, SAS is the most flexible package, with most of the discussed methods included. In this chapter, we will show how the various methods can be implemented in the SAS package. The examples will be worked out using SAS version 9.1. It is by no means the intention to give a full detailed overview of all available options. Instead, emphasis will be on general guidelines with respect to the choice of the appropriate SAS procedures as well as with respect to how models are specified in the various available procedures. We refer to the online SAS manuals for a full description of the available procedures and their possible options.

As a guiding example, we reconsider the toenail data, with the same random-effects model as used in Section 14.8. More specifically, it will be assumed that, conditionally on subject-specific, random, intercepts  $b_i$ ,  $Y_{ij}$  is Bernoulli distributed with mean  $\pi_{ij}$ , modeled as

$$\text{logit}(\pi_{ij}) = \beta_0 + b_i + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}, \quad (15.1)$$

in which  $T_i$  is the treatment indicator for subject  $i$  (1 for group B, 0 for group A),  $t_{ij}$  is the time point at which the  $j$ th measurement is taken for the  $i$ th subject, and  $b_i$  is the random intercept assumed to be normally distributed with mean zero and variance  $\tau^2$ . Note that the marginal version

of this model was used in Chapter 10 to illustrate how marginal models can be fitted within the SAS environment.

As in Chapter 10, it will be assumed that the data have been stored in the SAS data file ‘test,’ which contains the variables ‘onyresp,’ ‘treatn,’ ‘time,’ and ‘idnum.’ The variable ‘response’ is the binary outcome variable defined as 1 for a severe toenail infection, and equal to 0 otherwise. Further, ‘treat’ is a binary treatment indicator to be 1 for group B and 0 for group A. The variable ‘time’ contains the time-point at which the outcome has been measured, and ‘idnum’ is the variable containing the subject’s identification label. Finally, it will be assumed that the data are organized such that each record corresponds to the information available for one specific subject, at one specific point in time, and it will be assumed that the data have been ordered according to the variable ‘idnum.’ For example, our toenail data set is set up in the following way:

Obs	time	treatn	idnum	onyresp
1	0	1	1	1
2	1	1	1	1
3	2	1	1	1
4	3	1	1	0
5	6	1	1	0
6	9	1	1	0
7	12	1	1	0
....	..	.	...	.
1903	0	1	383	1
1904	1	1	383	1
1905	2	1	383	1
1906	3	1	383	1
1907	6	1	383	0
1908	9	1	383	0

Note that subject #383 left the study prematurely after 9 months of follow-up, but before month 12.

## 15.2 The GLIMMIX Procedure for Quasi-Likelihood

The marginal and penalized quasi-likelihood methods have been implemented in the SAS procedure GLIMMIX, which is still experimental under SAS version 9.1. As an example, we will fit Model (15.1) using the PQL method. The procedure has many more statements and options than those

TABLE 15.1. SAS Procedure GLIMMIX. Available options for specification of the estimation method.

GLIMMIX option	Quasi-likelihood type	Inference pseudo-data
	PQL/MQL	ML/REML
'method=RSPL'	PQL	REML
'method=MSPL'	PQL	ML
'method=RMPL'	MQL	REML
'method=MMPL'	MQL	ML

presented here, but we restrict to the basic statements needed to fit a generalized linear mixed model.

### 15.2.1 The SAS Program

The following SAS code can be used to fit Model (15.1) using PQL based on REML estimation for the linear mixed models for the pseudo data:

```
proc glimmix data=test method=RSPL ;
class idnum;
model onyresp (event='1') = treatn time treatn*time
/ dist=binary solution;
random intercept / subject=idnum;
run;
```

Users of the SAS procedure MIXED for linear mixed models will recognize that the code here is very similar to that used in PROC MIXED. As explained in Section 14.4, this is because the estimation methods implemented in the GLIMMIX procedure iteratively fit linear mixed models to newly updated pseudo data.

A very important option is 'method=' in the GLIMMIX statement. Here, the type of quasi-likelihood is specified. In our example, the model is fitted using PQL, based on REML for the linear mixed models. This corresponds to the option 'method=RSPL.' An overview of the other available options is given in Table 15.1.

The CLASS statement specifies which variables should be considered as factors. Such classification variables can be either character or numeric. Internally, each of these factors will correspond to a set of dummy variables in the manner described in the SAS manual on linear models (1991, Section 5.5).

The MODEL statement names the response variable and all covariates corresponding to the fixed effects. By default, an intercept is added. In case no intercept is needed, the option 'noint' can be inserted. The option '(event='1')' has been added here in order to specify that the probability

to be modeled is  $P(Y_{ij} = 1)$  (the probability of a severe infection), rather than  $P(Y_{ij} = 0)$ . The ‘solution’ option is used to request printing of the estimates of all the fixed effects in the model, together with standard errors,  $t$ -statistics, corresponding  $p$ -values and confidence intervals. The ‘dist=’ is used to specify the conditional distribution of the data, given the random effects. Various distributions are available, including the normal, Bernoulli, binomial, and Poisson distribution. In our example, the Bernoulli distribution is specified as ‘dist=binary.’ The link function is then by default the natural link. In our example, this is the logit link. Others, such as probit, log-log, log, or identity, can be requested by adding an appropriate ‘link=’ option.

The RANDOM statement defines the vectors  $z_{ij}$  corresponding to the random effects in the model. Note that, when random intercepts are required (as in our example), this should be specified explicitly, which is in contrast to the MODEL statement where an intercept is included by default. The ‘subject=’ option is used to identify the subjects in our dataset. Here, ‘subject=idnum’ means that all records with the same value for ‘idnum’ are assumed to be from the same subject, whereas records with different values for ‘idnum’ are assumed to contain independent data. The variable ‘idnum’ is permitted to be continuous as well as categorical (specified in the CLASS statement). However, when ‘idnum’ is continuous, PROC GLIMMIX considers a record to be from a new subject whenever the value of ‘idnum’ is different from the previous record.

Suppose that random slopes for the time trend were to be included as well. This could be obtained by replacing the RANDOM statement in the above program by

```
random intercept time / subject=idnum type=un;
```

in which the option ‘type=un’ now specifies that the random-effects covariance matrix  $D$  is a general unstructured  $2 \times 2$  matrix. Special structures are available, such as models that assume equal variance for the intercepts and slopes, or models that assume independent intercepts and slopes.

### 15.2.2 The SAS Output

We now discuss some of the output produced by the original program presented in Section 15.2.1.

First, a table is given with some information about the fitted model and the estimation procedure. The ‘Residual PL’ estimation technique refers to PQL with REML (restricted or residual maximum likelihood) for the fitting of the linear models for the pseudo data:

#### Model Information

Data Set

WORK.TEST

```

Response Variable           onyresp
Response Distribution       Binary
Link Function               Logit
Variance Function          Default
Variance Matrix Blocked By idnum
Estimation Technique        Residual PL
Degrees of Freedom Method   Containment

Number of Observations Read      1908
Number of Observations Used      1908
    
```

The table labeled ‘Response Profile’ summarizes the number of severe and non-severe infections in the dataset, and reports that the probability that will be modeled is  $P(Y_{ij} = 1)$ , the probability of a severe infection.

Response Profile

Ordered Value	onyresp	Total Frequency
1	0	1500
2	1	408

The GLIMMIX procedure is modeling the probability that `onyresp='1'`.

The ‘Iteration History’ table gives a summary of the different steps in the iterative optimization procedure. Depending on the numerical optimization algorithm chosen, this table will contain different entries. The most important ones are:

Iteration History

Iteration	Objective Function	Change	Max Gradient
0	8517.0833042	0.97407150	0.000272
1	9474.2004261	1.19238147	0.000682
2	10389.283759	2.00000000	1.148E-6
..	.....	.....	.....
11	11147.900904	0.00001765	7.376E-8
12	11147.902006	0.00000000	3.708E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

At each intermediate step, minus the log-likelihood evaluated in the current parameter values is reported, together with how much this value differs from the value in the previous step. Further, the column labeled ‘Max Gradient’ reports the largest absolute value of the components in the gradient. At the optimum, this value equals zero.

#### Fit Statistics

-2 Res Log Pseudo-Likelihood	11147.90
Pseudo-AIC (smaller is better)	11149.90
Pseudo-AICC (smaller is better)	11149.90
Pseudo-BIC (smaller is better)	11153.59
Pseudo-CAIC (smaller is better)	11154.59
Pseudo-HQIC (smaller is better)	11151.38
Pearson Chi-Square	1455.03
Pearson Chi-Square / DF	0.76

The table termed ‘Fit Statistics’ gives minus twice the residual log-pseudo-likelihood value evaluated in the final solution, together with a number of information criteria, including the Akaike information criterion (AIC) and the Schwarz (BIC) information criterion. When REML estimation is used for the fitting of the linear mixed models for the pseudo-data, an objective function is maximized, which is called residual log-likelihood function, while, strictly speaking, the function is not a log-likelihood, and should not be used as a log-likelihood. We refer to Verbeke and Molenberghs (2000, Chapters 5 and 6) for a more detailed discussion with examples. Further, information criteria are statistics that are sometimes used to compare non-nested models that cannot be compared based on a formal testing procedure. The main idea behind information criteria is to compare models based on their maximized (residual) log-likelihood value (or equivalently minimized minus twice the log-likelihood value), but to penalize for the use of too many parameters. They should by no means be interpreted as formal statistical tests of significance. In specific examples, different information criteria can even lead to different model selections. An example of this is given in Section 6.4 of Verbeke and Molenberghs (2000) in the context of linear mixed models. More details about the use of information criteria can be found in Akaike (1974), Schwarz (1978), and Burnham and Anderson (1998). Finally, the ‘Pearson Chi-Square’ value and derived ratio over the degrees of freedom are based on the marginal distribution of the pseudo-data as well. It should be emphasized that, as all statistics in the above output table are based on the underlying model for the pseudo data, rather than on the model for the actually observed outcomes, they should be interpreted with extreme caution.

#### Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Intercept	idnum	4.7116	0.6031

In the table called ‘Covariance Parameter Estimates,’ estimates and associated standard errors are given voor de variance components in the model, i.e., for the elements in the random-effects covariance matrix  $D$ . In our example, this is the random-intercepts variance  $\tau^2$ .

Finally, two tables are reported containing estimates and inferences for the fixed effects in the model. As discussed in Section 14.6, the reported inferences immediately result from the linear mixed model fitted to the pseudo-data in the last step of the iterative estimation procedure.

#### Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	-0.7239	0.2370	292	-3.05	0.0025
treatn	0.000918	0.3363	1612	0.00	0.9978
time	-0.2883	0.03349	1612	-8.61	<.0001
treatn*time	-0.1106	0.05366	1612	-2.06	0.0395

#### Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treatn	1	1612	0.00	0.9978
time	1	1612	74.10	<.0001
treatn*time	1	1612	4.25	0.0395

## 15.3 The GLIMMIX Macro for Quasi-Likelihood

The GLIMMIX procedure can be viewed as a formal procedure, although still experimental in SAS version 9.1, which has grown out of the SAS macro GLIMMIX, applied earlier in Section 10.5 for fitting generalized estimating equations (GEE) based on linearization (Section 8.8). In GEE, the association between repeated measures is modeled through a marginal working correlation matrix. In our context, this correlation is modeled via the inclusion of random effects, conditionally on which repeated measures are assumed independent. This similarity implies that the same macro can

be used for fitting generalized linear mixed models as well. Without going into much detail, we present here the SAS code needed to repeat the analysis from Section 15.2 with the GLIMMIX macro. Afterwards, some selected output is shown.

### 15.3.1 The SAS Program

Before the GLIMMIX macro can be called, one has to specify where the code can be obtained from:

```
%inc 'path\glimm800.sas' / nosource;
run;
```

The following SAS code can now be used to repeat the analysis from Section 15.2 with the GLIMMIX macro:

```
%glimmix(
  data=test,
  stmts=%str(
    class idnum;
    model onyresp = treatn time treatn*time / solution;
    random intercept / subject=idnum;
    parms (4) (1) / hold=2;
  ),
  error=binomial
)
run;
```

The statements that appear in the STMTS statement are directly fed into the PROC MIXED calls needed for fitting the linear mixed models to the pseudo-data. Note that the GLIMMIX macro by default includes a residual overdispersion parameter. If the corresponding generalized linear mixed model does not contain such a parameter, it should explicitly be kept equal to one by the user. This is done using the 'hold=' option in the PARMS statement.

Because the MIXED procedure uses REML estimation by default, the above program requests PQL estimation, based on REML fitting for the pseudo-data. If ML fitting is required, this can be specified by adding the line

```
procopt=%str(method=ml),
```

into the above '%glimmix' call. In case MQL is required, rather than the default PQL, this can be specified by adding the line

```
options=MQL,
```



### 15.3.2 Selected SAS Output

Without discussing the output from the GLIMMIX macro in much detail, we here present some output tables, which are to be compared with the output from the GLIMMIX procedure, discussed in Section 15.2.2.

#### Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	idnum	4.7116
Residual		1.0000

#### Fit Statistics

-2 Res Log Likelihood	11147.9
AIC (smaller is better)	11149.9
AICC (smaller is better)	11149.9
BIC (smaller is better)	11153.6

#### Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	-0.7239	0.2370	292	-3.05	0.0025
treatn	0.000918	0.3363	1612	0.00	0.9978
time	-0.2883	0.03349	1612	-8.61	<.0001
treatn*time	-0.1106	0.05366	1612	-2.06	0.0395

#### Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treatn	1	1612	0.00	0.9978
time	1	1612	74.10	<.0001
treatn*time	1	1612	4.25	0.0395

Note that, indeed, the residual overdispersion parameter was kept equal to one, and the results are the same as obtained earlier from the GLIMMIX procedure.

## 15.4 The NLMIXED Procedure for Numerical Quadrature

Gaussian and adaptive Gaussian quadrature, as approximations to the integral in the marginal likelihood (Section 14.5) have been implemented in the SAS procedure NLMIXED. As an example, we will reproduce the results reported in Table 14.2 for (non-adaptive) Gaussian quadrature with 3 quadrature points. The procedure has many more statements and options than those presented here, but we restrict to the basic statements needed to fit a generalized linear mixed model.

### 15.4.1 The SAS Program

The following SAS code can be used to fit Model (15.1) using Gaussian quadrature with 3 quadrature points:

```
proc nlmixed data=test noad qpoints=3;
parms beta0=-1.6 beta1=0 beta2=-0.4 beta3=-0.5 tau=3.9;
teta = beta0 + b + beta1*treatn + beta2*time
      + beta3*time*treatn;
expteta = exp(teta);
p = expteta/(1+expteta);
model onyresp ~ binary(p);
random b ~ normal(0,tau**2) subject=idnum;
run;
```

Before presenting the results of this analysis, we briefly discuss the statements and options used in the above program. It is clear from the above code that the NLMIXED procedure requires completely different model-specifications than most other SAS procedures. The main advantage is that the user is given a very high degree of flexibility in the way the model is specified and parameterized. One of the consequences of this flexibility is that the user not only needs to specify the model but also has to specify names for all the parameters in the model. In this respect, it is important to know that SAS considers all symbols in the model specification that are not referring to variables in the input dataset as unknown parameters, to be estimated from the data.

The option 'noad' in the NLMIXED statement is needed to request non-adaptive quadrature as, by default, adaptive quadrature is used. The option 'qpoints=' specifies the number of quadrature points. If this option is omitted, the number of quadrature points is selected adaptively by evaluating the log-likelihood function at the starting values of the parameters until two successive evaluations show sufficiently small relative change. Remember that model fitting based on the Laplace approximation for the integrals

in the marginal likelihood (Section 14.3) can be specified by choosing adaptive Gaussian quadrature with one quadrature point.

The PARMs statement is used to specify starting values for all parameters in the model. Parameters not listed in the PARMs statement are given an initial value of 1. Here we are confronted with one of the major drawbacks of the current version of the NLMIXED procedure, i.e., the fact that the procedure does not automatically generate starting values, except for the default value of 1 for all the parameters that do not occur in the PARMs statement. In complex models however, convergence of the numerical optimization algorithms may highly depend on the specified starting values.

The MODEL statement is used to specify the conditional distribution of the data, given the random effects. Various distributions are available, including the normal, Bernoulli, binomial, and Poisson distributions. In our example, the Bernoulli distribution is specified as ‘binary( $p$ )’ in which  $p$  is the success probability that has been specified in the program lines prior to the MODEL statement. The user has full flexibility over the way the model is specified as well as the number of intermediate steps that are used to define the success probability. For example, the above program corresponds to the parameterization as given in (15.1). A different parameterization of the same model would be

$$\text{logit}(\pi_{ij}) = \begin{cases} \beta_0 + b_i + \beta_1 t_{ij}, & \text{Treatment A} \\ \beta_2 + b_i + \beta_3 t_{ij}, & \text{Treatment B} \end{cases}$$

This can be specified using the statements

```
teta = beta0*(1-treatn) + beta2*treatn + b
      + beta1*(1-treatn)*time + beta3*treatn*time;
expteta = exp(teta);
p = expteta/(1+expteta);
```

or, equivalently,

```
if treatn=0 then teta=beta0 + b + beta1*time;
if treatn=1 then teta=beta2 + b + beta3*time;
expteta = exp(teta);
p = expteta/(1+expteta);
```

In case models are needed that do not fit within any of the classical distributions, user-defined likelihoods can be specified through the option ‘model onyresp  $\sim$  general( $\ell\ell$ )’ in which  $\ell\ell$  is the user-defined log-likelihood.

The RANDOM statement defines the random effects in the model. In our example, if the RANDOM statement had been omitted, the parameter  $b$  would have been considered a fixed intercept, and this would have led to an over-parameterized model. Now,  $b$  is specified to be normally distributed

with mean 0 and standard deviation  $\tau$ . Again, the user has full flexibility here. For example, if one wishes to estimate the random-intercepts variance rather than the standard deviation, this can be achieved by specifying

```
random b ~ normal(0,tau2) subject=idnum;
```

Also, a mean model can be specified for the random effect  $b$ . For example, our original model can also be specified as

```
proc nlmixed data=test noad qpoints=3;
parms beta0=-1.6 beta1=0 beta2=-0.4 beta3=-0.5 tau=3.9;
teta = b + beta1*treatn + beta2*time + beta3*time*treatn;
expteta = exp(teta);
p = expteta/(1+expteta);
model onyresp ~ binary(p);
random b ~ normal(beta0,tau**2) subject=idnum;
run;
```

in which the overall intercept  $\beta_0$  is now incorporated as average of the random effects. Inclusion of random slopes in Model (15.1) can be done with the following code:

```
proc nlmixed data=test qpoints=10 noad;
parms beta0=-1.6 beta1=0 beta2=-0.4 beta3=-0.5
      d11=16 d12=0 d22=0.1;
teta = beta0 + b1 + beta1*treatn + beta2*time
      + b2*time + beta3*time*treatn;
expteta = exp(teta);
p = expteta/(1+expteta);
model onyresp ~ binary(p);
random b1 b2 ~ normal([0, 0], [d11,d12,d22]) subject=idnum;
run;
```

with obvious parameterization for the means of all components in the random-effects vector, and with the random-effects covariance specified through its lower triangle. If for example, one wishes to incorporate independence of random intercepts and slopes, this is done by replacing the RANDOM statement in the above program by

```
random b1 b2 ~ normal([0, 0], [d11,0,d22]) subject=idnum;
```

When one wishes to directly estimate the correlation between random intercepts and slopes, rather than their covariance, the following PARMs and RANDOM statements can be used:

```
parms beta0=-1.6 beta1=0 beta2=-0.4 beta3=-0.5
      d11=16 rho=0 d22=0.1;
random b1 b2 ~ normal([0, 0],
      [d11,rho*sqrt(d11)*sqrt(d22),d22])
      subject=idnum;
```

The ‘subject=’ option determines when new realizations of the random effects occur. The procedure assumes the occurrence of a new realization whenever the value of the variable specified in the ‘subject=’ option changes from the previous observation. This is why the input dataset needs to be sorted according to this variable (Section 15.1). Further, the RANDOM statement allows inclusion of an output option of the form ‘out=dataset’ which requests an output dataset containing empirical Bayes estimates for the random effects, together with their approximate standard errors.

The current version of the NLMIXED procedure allows one RANDOM statement only, which poses some restrictions to flexibly specifying random-effects models with random effects at different levels. In the examples considered so far, we had two levels in the design: A first level representing the subjects, and a second level representing the measurements within the subjects. An example where more than two levels would be required would be the analysis of longitudinal profiles from children randomly sampled from randomly sampled schools. In order to correctly account for the different sources of sampling variability, random effects might be needed for schools as well as for children within the schools. Such multi-level models can, to some extent, be fitted within the NLMIXED procedure, but non-standard coding is required.

### 15.4.2 The SAS Output

We now discuss some of the output produced by the original program presented in Section 15.4.1. The parameter estimates and associated standard errors have already been reported in Table 14.2.

First, two tables are given, containing information about the specified model, the observations in the dataset, and the numerical optimization algorithms used in the model fitting process:

#### Specifications

Data Set	WORK.TEST
Dependent Variable	onyresp
Distribution for Dependent Variable	Binary
Random Effects	b
Distribution for Random Effects	Normal
Subject Variable	idnum
Optimization Technique	Dual Quasi-Newton
Integration Method	Gaussian Quadrature

#### Dimensions

Observations Used	1908
Observations Not Used	0
Total Observations	1908

Subjects	294
Max Obs Per Subject	7
Parameters	5
Quadrature Points	3

The table labeled ‘Parameters’ lists the parameters in the model, their starting values, and minus the log-likelihood evaluated in these initial parameter values:

Parameters					
beta0	beta1	beta2	beta3	tau	NegLogLike
-1.6	0	-0.4	-0.5	3.9	760.941002

The ‘Iteration History’ table gives a summary of the different steps in the iterative optimization procedure. Depending on the chosen numerical optimization algorithm, this table will contain different entries. The most important ones are:

Iteration History			
Iter	NegLogLike	Diff	MaxGrad
1	747.757703	13.1833	129.7817
2	728.271809	19.48589	133.3329
3	686.505096	41.76671	116.594
..	.....	.....	.....
10	672.074434	0.000686	0.012941
11	672.074433	8.32E-7	0.000319

NOTE: GCONV convergence criterion satisfied.

At each intermediate step, minus the log-likelihood evaluated in the current parameter values is reported, together with how much this value differs from the value in the previous step. Further, the column labeled ‘MaxGrad’ reports the largest absolute value of the components in the gradient. At the optimum, this value equals zero.

Fit Statistics

-2 Log Likelihood	1344.1
AIC (smaller is better)	1354.1
AICC (smaller is better)	1354.2
BIC (smaller is better)	1372.6

The table termed ‘Fit Statistics’ gives minus twice the log-likelihood value evaluated in the final solution, together with the information criteria of Akaike (AIC) and Schwarz (BIC), as well as a finite-sample corrected version of AIC (AICC). They have the same interpretation as discussed earlier in Section 15.2.2. However, the information criteria are now defined in terms of the maximized likelihood (obtained from numerical integration) for the assumed model for the actually observed data, rather than on the likelihood for the underlying pseudo-data, as was the case in Section 15.2.2.

The final part of the output is the table labeled ‘Parameter Estimates,’ which contains estimates and associated inferences for all the parameters in the marginal likelihood:

Parameter Estimates							
Parameter	Estimate	Standard Error	DF	t Value	Pr >  t	Alpha	
beta0	-1.5221	0.3063	293	-4.97	<.0001	0.05	
beta1	-0.3932	0.3812	293	-1.03	0.3031	0.05	
beta2	-0.3198	0.03481	293	-9.19	<.0001	0.05	
beta3	-0.09098	0.05236	293	-1.74	0.0833	0.05	
tau	2.2555	0.1217	293	18.54	<.0001	0.05	
Parameter	Lower	Upper	Gradient				
beta0	-2.1250	-0.9192	-0.00007				
beta1	-1.1433	0.3570	-0.00002				
beta2	-0.3883	-0.2513	0.000058				
beta3	-0.1940	0.01207	0.000319				
tau	2.0161	2.4950	-0.00003				

As discussed in Section 14.6, the reported standard errors are obtained from the inverse Fisher information matrix. The ratio of the estimate over its standard error produces a *t*-value that is compared to a *t*-distribution in order to obtain a formal test of significance. One hereby uses an *ad hoc* number of degrees freedom equal to the number of subjects in the dataset, minus the number of random effects. In our example, this results in *t*-tests based on  $294 - 1 = 293$  degrees of freedom. In case one wishes classical Wald-type tests (*Z*-tests), these can be obtained by pre-specifying a large number of degrees of freedom. This is done through the ‘df=’ option in the NLMIXED statement. Based on the chosen *t*-approximation to the standardized parameter estimate, lower and upper confidence limits are reported based on the  $(1 - \text{Alpha})100\%$  confidence level. The default ‘Alpha’-value can be changed using the ‘alpha=’ option in the NLMIXED statement. Finally, the column labeled ‘Gradient’ contains the first-order derivative of the objective function with respect to each of the parame-

ters in the marginal likelihood. Note that the maximal gradient value of 0.000319 reported previously in the ‘Iteration History’ table is the gradient value for the parameter  $\beta_3$ , i.e., for the interaction between the time trend and the treatment indicator. Finally, it should be emphasized that, in general, the reported  $p$ -values for variance components should be interpreted with great care, due to possible occurrence of boundary problems, as explained in Section 14.6.

## 15.5 Alternative Software Tools

In this chapter, we have extensively illustrated the use of the SAS package for fitting generalized linear mixed models. Many other statistical software packages offer tools for fitting these models, including HLM (Raudenbush *et al* 2001), EGRET (Cytel Software Corporation 2000), *gllamm* in Stata (Rabe-Hesketh, Pickles, and Skrondal 2001), and MIXOR and MIXREG (Hedeker and Gibbons 1994, 1996).

As discussed in Chapter 14, there is a variety of methods available for fitting generalized linear mixed models. They differ in the type of approximation or in the order of the approximation. When using software, it is therefore very important to be aware of what precisely has been implemented. A full description of software tools can be found in Tuerlinckx *et al* (2004) and in Skrondal and Rabe-Hesketh (2004).