

BIOS6643 Longitudinal

L16 Linearization

EJC

Department of Biostatistics & Informatics

Linearization

References

Summary

Linearization

References

Summary

- ▶ Generalized linear mixed models (GzLMM)
- ▶ GzLMM: the linearization approach

Related reading: Sections 5 in Non-normal notes

A GzLMM can be fit by 'linearizing' the data and employing LMM's; the transformed 'pseudo-data' P is fit with the linear mixed model $P = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$ to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{b}}$.

Using the fitted model, pseudo data are recomputed, and the process is repeated iteratively until estimates converge.

This is a doubly-iterated procedure, since we update estimates after each linear mixed model fitting, and within each linear mixed model fitting we use an iterative procedure as well.

The benefit of this approach is that we can take advantage of what standard LMM theory has to offer. For example, we can fit a model that has both random effects and higher level structures for \mathbf{R} , or that has multiple-level or complex (e.g. crossed) random effects.

For GEE, we could specify a working covariance structure for \mathbf{R} but not include random effects; for GzLMM methods that use techniques to approximate the likelihood, we can specify random effects but cannot have non-simple \mathbf{R} matrices, or random effects at multiple levels.

One drawback to the linearization method is estimator bias that has been reported (see the SAS Help Documentation: **Notes on Bias of Estimators page**). The situations considered are more extreme cases, though (e.g., binary data, matched pairs designs, i.e., clusters of size 2). The bias diminishes quickly as sample sizes increase.

A bit more history of the methods...

Breslow and Clayton (B&C, 1993, JASA) and Wolfinger and O'Connell (W&O, 1993, Journal of Statistical Computation and Simulation) published articles on pseudo-likelihood methods (i.e., linearization methods) for GzLMM's. (Also see earlier work referenced in W&O.)

Wolfinger and O'Connell (W&O) referred to their methods Psuedo-likelihood approaches. This article is relatively easy to read, and lays out the algorithm in a nice, organized fashion.

B&C refer to their methods as forms of quasi-likelihood estimation.

To better distinguish the quasi-likelihood methods used for GzLM/GEE's for those here, we'll stick with W&O's 'Pseudo-likelihood' terminology.

The algorithms for the 2 sets of authors are the same when the scale/dispersion parameter ϕ in W&O's article is set to 1. Note that this will be the case by default for binary/binomial and Poisson outcomes.

The linearization approach uses a Taylor series expansion.

If the expansion is about the current fixed-effect and random-effect estimates, it is a subject-specific (SS) type of expansion (or Penalized Quasi-likelihood [PQL] in B&C).

If the expansion is about around current fixed-effect estimates and where random effects are set to 0, it is a population averaged (PA) type of expansion (or Marginal Quasi-likelihood [MQL] in B&C).

These differences should be kept in mind when interpreting the estimates. See Fitzmaurice for more detail.

For PROC GLMMIX, you can specify one of 4 specific types of linearization, RSPL, MSPL, RMPL or MMPL using the METHOD= option.

The first letter indicates whether the residual likelihood (R) or maximum likelihood (M) is used, i.e., REML or ML-type estimation while employing the LMM.

The second letter indicates whether the SS or PA type of expansion is used. The last 2 letters stand for pseudo-likelihood estimation.

Approaches other than linearization can also be used with PROC GLIMMIX, but involve approximating the true likelihood (quadrature and Laplace methods). The default method for PROC GLIMMIX is RSPL.

SAS needs initial values of \mathbf{P} to start the iterative process. If no specification is made, the GLIMMIX output indicates what was used (e.g., 'Starting from: GLM estimates' or '...data'). [I believe the 'GLM' they are referring to are what we call GzLM.] The method used depends on what types of covariance parameters are specified in the model (\mathbf{R} -side or \mathbf{G} -side).

A bit more detail on transforming the data

Consider the GzLMM $\mathbf{g}(\mu) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$, where $\mathbf{g}(\cdot)$ is a differentiable monotone link function, $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$;

$$E(\mathbf{Y} | \mathbf{b}) = \mathbf{g}^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}) = \mathbf{g}^{-1}(\boldsymbol{\eta}) = \boldsymbol{\mu}; \text{Var}[\mathbf{Y} | \mathbf{b}] = \mathbf{A}^{1/2} \mathbf{R} \mathbf{A}^{1/2}$$

where \mathbf{A} is the variance function of $\boldsymbol{\mu}$ based on the mean-variance relationship of the particular distribution considered.

For example, for a Poisson distribution, \mathbf{A} is a diagonal matrix with elements of $\boldsymbol{\mu}$ down the diagonal, since for the Poisson the mean and variance are the same.

A first-order Taylor series expansion of $\boldsymbol{\mu}$ about current values $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{b}}$ is $\mathbf{g}^{-1}(\boldsymbol{\eta}) \approx \mathbf{g}^{-1}(\tilde{\boldsymbol{\eta}}) + \mathbf{g}'(\tilde{\boldsymbol{\mu}})\mathbf{X}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \mathbf{g}'(\tilde{\boldsymbol{\mu}})\mathbf{Z}(\mathbf{b} - \tilde{\mathbf{b}})$.

It is possible to add a dispersion parameter to the model. By default, log and logit link (i.e., for count and binary outcomes) will not have a dispersion parameter in SAS routines.

Rearranging this equation allows us to yield the mean of an LMM, given the random effects (on the right):

$$g'(\tilde{\boldsymbol{\mu}})(\boldsymbol{\mu} - g^{-1}(\tilde{\boldsymbol{\eta}})) + \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\mathbf{b}} \approx \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} .$$

The left side of the equation above is the expected value of $g'(\tilde{\boldsymbol{\mu}})(\mathbf{Y} - g^{-1}(\tilde{\boldsymbol{\eta}})) + \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\mathbf{b}}$, conditional on \mathbf{b} (treating the terms with tilde hats as fixed).

Thus, $\mathbf{P} = g'(\tilde{\boldsymbol{\mu}})(\mathbf{Y} - g^{-1}(\tilde{\boldsymbol{\eta}})) + \mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{Z}\tilde{\mathbf{b}}$ seems like a good candidate for the linearized data, since its expected value (given random effects) is $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$ via the approximation.

We can write \mathbf{P} more succinctly as $\mathbf{P} = g'(\tilde{\boldsymbol{\mu}})(\mathbf{Y} - \tilde{\boldsymbol{\mu}}) + g(\tilde{\boldsymbol{\mu}})$.

\mathbf{P} is assumed to be approximately normal with variances

$$\blacktriangleright \text{Var}[\mathbf{P} | \mathbf{b}] = g'(\tilde{\boldsymbol{\mu}}) \tilde{\mathbf{A}}^{1/2} \mathbf{R} \tilde{\mathbf{A}}^{1/2} g'(\tilde{\boldsymbol{\mu}})$$

$$\blacktriangleright \text{Var}[\mathbf{P}] = \mathbf{Z} \mathbf{G} \mathbf{Z}^\top + g'(\tilde{\boldsymbol{\mu}}) \tilde{\mathbf{A}}^{1/2} \mathbf{R} \tilde{\mathbf{A}}^{1/2} g'(\tilde{\boldsymbol{\mu}})$$

where all elements on the right are evaluated at the converged estimates; hence can really consider these as estimated variances, i.e., put ' $\hat{}$ ' on Var.

Recall for a standard LMM that $\text{Var}[\mathbf{Y} | \mathbf{b}] = \mathbf{R}$. For a mixed model, adding a weight statement allows us to fit $\text{Var}[\mathbf{P} | \mathbf{b}] = \mathbf{W}^{-1/2} \mathbf{R} \mathbf{W}^{-1/2}$, where \mathbf{W} is a diagonal weight matrix with weights w_{ij} down the diagonal.

In our case, we have $\text{Var}[\mathbf{P} | \mathbf{b}] = g'(\tilde{\boldsymbol{\mu}}) \tilde{\mathbf{A}}^{1/2} \mathbf{R} \tilde{\mathbf{A}}^{1/2} g'(\tilde{\boldsymbol{\mu}})$. If we let $\mathbf{W}^{(-0.5)} = g'(\boldsymbol{\mu}) \mathbf{A}^{0.5}$, it follows that $\tilde{\mathbf{W}} = [g'(\tilde{\boldsymbol{\mu}})]^{-2} \tilde{\mathbf{A}}^{-1}$ since both $g'(\boldsymbol{\mu})$ and \mathbf{A} are diagonal matrices with elements of interest down the diagonal. Thus we can fit a weighted LMM to get the right variance.

For the Poisson, $g'(\boldsymbol{\mu})$ has elements $1/\mu_{ij}$ down the diagonal; \mathbf{A} has μ_{ij} down the diagonal. Thus $\mathbf{W} = [g'(\boldsymbol{\mu})]^{-2} \mathbf{A}^{-1} = \boldsymbol{\mu}$. For the binomial, $\mathbf{W} = \boldsymbol{\mu}(1 - \boldsymbol{\mu})$. Can you derive this?

As the LMM for \mathbf{P} has parameters and random effects that are of interest in the GzLMM as well, we can essentially conduct inference for $\boldsymbol{\beta}$ and \mathbf{b} using standard LMM methods, using converged values in the model for \mathbf{P} .

For example, we can derive $Var[\hat{\boldsymbol{\beta}}] = (\mathbf{X}^\top [Var[\mathbf{P}]]^{-1} \mathbf{X})^{-1}$. However, the variability in ‘converged values’ of \mathbf{G} and \mathbf{R} (i.e., the fact that they are estimated) are not taken into account. SAS Documentation mentions previous research that incorporated inflation factors, but that for most data sets that are fairly well balanced, they do not make much difference.

More generally, inference for fixed and random effect estimates in the model follow in much the same way as for standard linear mixed models, but for the pseudo data model. For more detail, see W&O, or Fitzmaurice’s longitudinal text; you can also find some detail in the SAS Help Documentation but it is not as complete as the aforementioned references.

An example of the algorithm for count data (log link)

To demonstrate the algorithm for PL, we would like to see how rescue inhaler count data for children at the Morgridge Academy at NJH relate to different types of fine particulate matter ($PM_{2.5}$) they are exposed to. Specifically, we include cigarette smoke, ambient $PM_{2.5}$, their interaction, plus covariates of Friday indicator and temperature in the model. (The Friday indicator helps to account for pretreats in the outcome, since they did not have physical exercise on that day compared to the others.)

Step 1: fit a weighted LMM with raw data as the outcome and as weights.

Step 2: obtain the predicted values $\tilde{\mu} = X\tilde{\beta} + Z\tilde{b}$ from Step 1 and calculate the pseudo-data as $P = (Y - \tilde{\mu})/\tilde{\mu} + \ln(\tilde{\mu})$, based on the \ln link for the Poisson,

Step 3: Fit an LMM using latest versions of P (outcome) and $+\tilde{\mu}$ (weights).

Step 4: obtain predicted values $\tilde{\mu} = \exp(X\tilde{\beta} + Z\tilde{b})$ from Step 3 and recompute P .

Continue Steps 3 and 4 until convergence

A few important notes on this algorithm:

Since there are likely to be some $y = 0$ counts in the data, some adjustments need to be made since we don't want any records with '0 weight', which would be dropped from analysis). One approach is to add 1 to all the counts.

In *Step 4* we exponentiate to get predicted values since the LMM in *Step 3* essentially involves an outcome on the log scale, while in *Step 1* it's on the regular scale).

In W&O, they mention that convergence continues until changes in \mathbf{G} and \mathbf{R} are sufficiently small. Another (indirect) approach is to continue until changes in the Beta estimates are sufficiently small.

It probably won't take to many LMM fits for convergence. For example, it took 6 or 7 repeated LMM fits for the count data to converge. (This is the outer shell. Each LMM fit also involves iterations.)

```
*Start with data. Use 'data+1' in first step for weights, since data has 0's;
data simp; set rc3; nu=dnew_spuff; mu0=dnew_spuff+1; run;

*Initial fit;
proc mixed data=simp;
class subj_num_new year;
model mu0=friday temperature year pred_w1 pred_w2 pred_w1*pred_w2 / solution outp=preddy;
random intercept / subject=subj_num_new g solution;
repeated / subject=subj_num_new type=sp(pow)(date);
weight mu0;
ods output solutionf=beta0; run;

*Get predicted values from fit. In this case, don't need to exponentiate predicted values
since started with original data rather than logged data (since logging would yield negative
values);
data next; set preddy;
nu=log(pred)+(dnew_spuff-pred)/pred;
keep subj_num_new date friday year temperature pred_w1 pred_w2 dnew_spuff pred nu;
rename pred=lastpred; run;

*Subsequent iterations;
%macro byhand(num);
*Get new model estimates (and subsequent predicted values);
proc mixed data=next;
class subj_num_new year;
model nu=friday temperature year pred_w1 pred_w2 pred_w1*pred_w2 / solution outp=preddy;
random intercept / subject=subj_num_new g solution;
repeated / subject=subj_num_new type=sp(pow)(date);
weight lastpred;
ods output solutionf=beta; run;
data betanum; set beta; iteration=snum;
data predout&num; set preddy; iteration=snum; keep subj_num_new date year nu pred iteration;

*Get new pseudo data values;
data next; set preddy;
pred=exp(pred);
keep subj_num_new date friday year temperature pred_w1 pred_w2 dnew_spuff pred nu;
nu=log(pred)+(dnew_spuff-pred)/pred;
rename pred=lastpred; run;
%mend; %byhand(1); %byhand(2); %byhand(3); %byhand(4); %byhand(5); %byhand(6); %byhand(7);
```


The same code in GLIMMIX form is here.

```
proc glimmix data=merged pconv=0.000001 method=mspl;  
  where year<8;  
  class subj_num new year;  
  model dnew spuff = friday temperature year pred_w1 pred_w2 pred_w1*pred_w2  
    / covb dist=poisson solution;  
  random intercept / subject=subj_num_new;  
  random _residual_ / subject=subj_num_new type=sp(exp)(date);  
  output out=outer pred=p resid=r;  
  ods output covb=covarb; run;
```

Note that instead of using the REPEATED statement here (which doesn't exist in PROC GLIMMIX), we add another RANDOM statement with the key word "*residual*". You can really think of this as a REPEATED statement, since it specifies the \mathbf{R} matrix.

A few notes on approaches to get convergence: I used sp(exp) instead of sp(pow) for convergence but they are intrinsically the same structure, just different parameterizations; I had to soften the convergence criteria just a bit using the pconv option.

Here I chose MSPL, mainly for more apples-to-apples comparisons with quadrature later. Remember the default is RSPL.

Walking through the output:

Walking through the output:

The GLIMMIX Procedure		Fit Statistics	
Model Information			
Response Variable	dnew_spuff	-2 Log Pseudo-Likelihood	4509.37
Response Distribution	Poisson	Generalized Chi-Square	1112.52
Link Function	Log	Gener. Chi-Square / DF	0.59
Estimation Technique	PL		
Number of Observations Used	1900		

Covariance Parameter Estimates				
Cov Parm	Subject	Estimate	Standard Error	
Intercept	subj num new	0.2150	0.04802	
SP(EXP)	subj num new	0.6544	0.06479	
Residual		0.5855	0.02093	

This might seem troubling, but scale is adjusted for in the analysis since we have R-side parameters in the model. This estimate is just the residual variance estimate.

Solutions for Fixed Effects						
Effect	year	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		-2.9893	2.1835	84	-1.37	0.1746
friday		-1.5012	0.07344	1807	-20.44	<.0001
temperature		0.004199	0.001941	1807	2.16	0.0307
year	4	-0.01433	0.09853	1807	-0.15	0.8844
year	5	0.006115	0.09637	1807	0.06	0.9494

This is equivalent to the estimated scale parameter. By including 'R-side' parameters, we can account for over or underdispersion for the Poisson (in this case, underdispersion).

Solutions for Fixed Effects

Effect	year	Estimate	Standard Error	DF	t Value	Pr > t
year	6	-0.1431	0.09245	1807	-1.55	0.1218
year	7	0
pred_w1		2.2923	1.4577	1807	1.57	0.1160
pred_w2		1.6328	1.1407	1807	1.43	0.1525
pred_w1*pred_w2		-1.1960	0.7607	1807	-1.57	0.1161

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
year	3	1807	2.31	0.0740

The `pred_w1` coefficient is the estimate of effect for SHS $PM_{2.5}$ exposure in the absence of ambient $PM_{2.5}$ exposure; `pred_w2` is vice versa. The ratio of the estimates is a measure of toxicity of SHS $PM_{2.5}$ relative to ambient $PM_{2.5}$.

Note that predicted values were obtained from another regression. SE's here are underestimated since they do not account for the variability of estimates in that first regression. Asymptotic or bootstrap methods can be used to properly estimate the SE's (see Measurement error chapters).

Comparing previous results with GEE and GLIMMIX quadrature

The table below shows different modeling approaches and impacts on estimates and SE's (in parentheses) of key predictors of interest.

The table below shows different modeling approaches and impacts on estimates and SE's (in parentheses) of key predictors of interest.

Predictor	GEE exch, empirical QIC=4927.8; $\hat{\phi}$ =0.69	GEE AR(1) QIC=4990.0; $\hat{\phi}$ =0.68	GzLMM, quad	GzLMM, PL [sp(exp)] and RI $\hat{\phi}$ =0.59*
Pred_w1	2.57 (emp=1.71; MB=1.42)	1.91 (emp=1.61; MB=1.48)	2.70 (1.91)	2.29 (1.46)
Pred_w2	1.89 (emp=1.32; MB=1.11)	1.24 (emp=1.22; MB=1.16)	1.92 (1.50)	1.63 (1.14)
Pred_w1*pred_w2	-1.35 (emp=0.88; MB=0.75)	-0.91 (emp=0.81; MB=0.77)	-1.37 (1.00)	-1.20 (0.76)

*If the random intercept is removed from this model, $\hat{\phi}$ =0.67, very close to that of GEE (for comparable model). GEE's reported 'scale' parameter estimate is $\sqrt{\hat{\phi}}$; the residual variance in GLIMMIX PL is $\hat{\phi}$.

Estimates are variable between modeling approaches, likely due to natural collinearity between main effects and interactions (Pred_w1 and Pred_w2 are positively correlated; the interaction is negatively correlated with each of Pred_w1 and Pred_w2). As a result, the magnitude of all estimates tend to increase or decrease collectively, from one analysis to the next. Functions of these estimates of main interest may differ less.

SE's are in strong agreement between the PL approach and model-based GEE approach, with empirical SE's somewhat higher. The GzLMM quadrature approach overestimates SE's since it does not account for the underdispersion.

An example with a binary outcome: exacerbation events for asthmatic children

PL used to model the fairly rare exacerbation events; left: PL using a random intercept (left), right: RI + AR(1) structure imposed on the errors.

<pre>proc glimmix data=y5dat_red method=mspl; model exacerb(event='1') = pm25cen02 day weekend holiday / solution distribution=binary; random intercept / subject=id; run;</pre>	<pre>proc glimmix data=y5dat_red method=mspl; model exacerb(event='1') = pm25cen02 day weekend holiday / solution distribution=binary; random intercept / subject=id; random _residual_ / subject=id type=ar(1); run;</pre>																																																																								
<p>Number of Observations Used 6923 G-side Cov. Parameters 1 Columns in X 5 Columns in Z per Subject 1 Subjects (Blocks in V) 43 Max Obs per Subject 162 -2 Log Pseudo-Likelihood 47864.26 Gener. Chi-Square / DF 0.63</p>	<p>Number of Observations Used 6923 G-side Cov. Parameters 1 R-side Cov. Parameters 2 -2 Log Pseudo-Likelihood 38935.51 Gener. Chi-Square / DF 8.62</p>																																																																								
<table><tr><th>Cov Parm</th><th>Subject</th><th>Estimate</th><th>SE</th></tr><tr><td>Intercept</td><td>id</td><td>4.7531</td><td>1.2455</td></tr></table>	Cov Parm	Subject	Estimate	SE	Intercept	id	4.7531	1.2455	<table><tr><th>Cov Parm</th><th>Subject</th><th>Estimate</th><th>SE</th></tr><tr><td>Intercept</td><td>id</td><td>3.0102</td><td>0.8374</td></tr><tr><td>AR(1)</td><td>id</td><td>0.6911</td><td>0.008856</td></tr><tr><td>Residual</td><td></td><td>0.6214</td><td>0.01888</td></tr></table>	Cov Parm	Subject	Estimate	SE	Intercept	id	3.0102	0.8374	AR(1)	id	0.6911	0.008856	Residual		0.6214	0.01888																																																
Cov Parm	Subject	Estimate	SE																																																																						
Intercept	id	4.7531	1.2455																																																																						
Cov Parm	Subject	Estimate	SE																																																																						
Intercept	id	3.0102	0.8374																																																																						
AR(1)	id	0.6911	0.008856																																																																						
Residual		0.6214	0.01888																																																																						
<table><tr><th>Effect</th><th>Estimate</th><th>SE</th><th>DF</th><th>t Value</th><th>Pr> t </th></tr><tr><td>Intercept</td><td>-4.9398</td><td>0.4275</td><td>42</td><td>-11.55</td><td><.0001</td></tr><tr><td>pm25cen02</td><td>-0.00606</td><td>0.01032</td><td>6876</td><td>-0.59</td><td>0.5576</td></tr><tr><td>day</td><td>0.009547</td><td>0.001349</td><td>6876</td><td>7.08</td><td><.0001</td></tr><tr><td>weekend</td><td>-0.2313</td><td>0.1384</td><td>6876</td><td>-1.67</td><td>0.0948</td></tr><tr><td>holiday</td><td>-0.1285</td><td>0.1937</td><td>6876</td><td>-0.66</td><td>0.5072</td></tr></table>	Effect	Estimate	SE	DF	t Value	Pr> t	Intercept	-4.9398	0.4275	42	-11.55	<.0001	pm25cen02	-0.00606	0.01032	6876	-0.59	0.5576	day	0.009547	0.001349	6876	7.08	<.0001	weekend	-0.2313	0.1384	6876	-1.67	0.0948	holiday	-0.1285	0.1937	6876	-0.66	0.5072	<table><tr><th>Effect</th><th>Estimate</th><th>SE</th><th>DF</th><th>t Value</th><th>Pr> t </th></tr><tr><td>Intercept</td><td>-4.3896</td><td>0.4103</td><td>42</td><td>-10.70</td><td><.0001</td></tr><tr><td>pm25cen02</td><td>-0.00713</td><td>0.008261</td><td>6876</td><td>-0.86</td><td>0.3881</td></tr><tr><td>day</td><td>0.009147</td><td>0.002195</td><td>6876</td><td>4.17</td><td><.0001</td></tr><tr><td>weekend</td><td>-0.2085</td><td>0.07911</td><td>6876</td><td>-2.64</td><td>0.0084</td></tr><tr><td>holiday</td><td>-0.00787</td><td>0.1267</td><td>6876</td><td>-0.06</td><td>0.9505</td></tr></table>	Effect	Estimate	SE	DF	t Value	Pr> t	Intercept	-4.3896	0.4103	42	-10.70	<.0001	pm25cen02	-0.00713	0.008261	6876	-0.86	0.3881	day	0.009147	0.002195	6876	4.17	<.0001	weekend	-0.2085	0.07911	6876	-2.64	0.0084	holiday	-0.00787	0.1267	6876	-0.06	0.9505
Effect	Estimate	SE	DF	t Value	Pr> t																																																																				
Intercept	-4.9398	0.4275	42	-11.55	<.0001																																																																				
pm25cen02	-0.00606	0.01032	6876	-0.59	0.5576																																																																				
day	0.009547	0.001349	6876	7.08	<.0001																																																																				
weekend	-0.2313	0.1384	6876	-1.67	0.0948																																																																				
holiday	-0.1285	0.1937	6876	-0.66	0.5072																																																																				
Effect	Estimate	SE	DF	t Value	Pr> t																																																																				
Intercept	-4.3896	0.4103	42	-10.70	<.0001																																																																				
pm25cen02	-0.00713	0.008261	6876	-0.86	0.3881																																																																				
day	0.009147	0.002195	6876	4.17	<.0001																																																																				
weekend	-0.2085	0.07911	6876	-2.64	0.0084																																																																				
holiday	-0.00787	0.1267	6876	-0.06	0.9505																																																																				

Output on left shows similarity to the previous results based on quadrature.

The model that uses the AR(1) structure (right) does have a substantially lower -2 log likelihood than the one on the left, but there are currently no commonly accepted goodness-of-fit statistics to compare models using PL, even nested ones.

Note that SE's are much bigger on the left (except for 'day'); this is probably due to underdispersion not being accounted for without R-side parameters in the model. Based on this, I would probably go with the model on the right.

In R, the **glmmPQL()** function in the MASS package will do the linearization method and Pseudo-likelihood estimation.

Up to this point I have experienced glmmPQL to be a bit more finicky than PROC GLIMMIX. For example, for the exacerbation data set above, a run with a random intercept yields estimates that are fairly close to that of SAS, but outrageously large SE's, indicating some problem with the fit. However, with the following medication use data (less rare events), the results are fairly close.

These data involve medication use (albuterol use, from inhaler). This is generally used as needed, so is modeled as an outcome, as a function of a pollutant variable averaged over 3 days (the last day aligning with the day where med use was measured); weekend, holiday and Friday are indicator variables to allow differences on those days. In particular, children were given pretreats on other school days (M-Th) when gym class, but not on Fridays or on weekends or holidays. Temperature, pressure and humidity are meteorological variables that are typically included as covariates in air pollution models.

Code and abbreviated output follow, for the model that includes a random intercept for subjects, and spatial power structure for error covariances.

R approach

```
library(MASS) library(nlme) gm1 <- glmmPQL(fixed = doser_bin ~  
pm25cen02 + day + weekend + holiday, random = ~1 | id, correlation  
= corCAR1(form = ~day|id), family = binomial, data = dat)
```

SAS approach

```
proc glimmix data=y5dat method=mspl;  
model doser_bin(event='1') = pm25cen02 day weekend holiday  
humidity temperature p/ solution distribution=binary;  
random intercept / subject=id;  
random _residual_ / type=sp(pow)(day) subject=id; run;
```

Notes: the nlme package is necessary for the correlation structure (i.e., that argument in the glmmPQL function), while the MASS package is for the glmmPQL function itself; Newton-Raphson with ridging for optimization in SAS; no direct mention in R. Output (n=43 with 6417 total records used for analysis).

R approach						SAS approach					
<pre>> summary(gml) Linear mixed-effects model fit by maximum likelihood AIC BIC loglik NA NA NA Random effects: Formula: ~1 id (Intercept) Residual StdDev: 0.6038089 1.024816 Correlation Structure: Continuous AR(1) Formula: ~day id Parameter estimate(s): Phi 0.2728577 Fixed effects: doser_bin ~ pm25scn02 + day + weekend + holiday + humidity + temp + pressure</pre>						<pre>-2 Log Pseudo-Likelihood 28724.17 Generalized Chi-Square 6729.24 Gener. Chi-Square / DF 1.05 Response Variable doser_bin Response Distribution Binary Link Function Logit Estimation Technique PL Degrees of Freedom Method Containment</pre>					
<pre>Value SE DF t-value p-value (Interc.) 7.4092 4.8634 6366 -1.52 0.1277 pm25scn02 0.00093 0.00578 6366 0.16 0.8724 day -0.00228 0.000795 6366 -2.86 0.0042 Friday -1.6333 0.00267 6366 -19.76 <.0001 weekend -1.6514 0.07743 6366 -21.33 <.0000 holiday -1.3860 0.1096 6366 -12.65 0.0000 Humidity 0.00197 0.00265 6366 0.74 0.4569 Temp 0.00842 0.003905 6366 2.16 0.0311 Pressure 0.01234 0.007673 6366 1.61 0.1079</pre>						<pre>Cov Parm Subject Estimate SE Intercept id 0.3646 0.09993 AR(1) id 0.2729 0.01239 Residual 1.0487 0.02006 Solutions for Fixed Effects Effect Estimate SE DF t Value Pr> t Intercept -7.4092 4.8601 42 -1.52 0.1349 pm25scn02 0.00093 0.00577 6366 0.16 0.8723 day -0.00228 0.000795 6366 -2.87 0.0042 Friday -1.6333 0.00267 6366 -19.77 <.0001 weekend -1.6514 0.07737 6366 -21.34 <.0001 holiday -1.3860 0.1095 6366 -12.66 <.0001 humidity 0.00197 0.002652 6366 0.74 0.4566 temp. 0.00842 0.003902 6366 2.16 0.0310 p 0.01234 0.007668 6366 1.61 0.1076</pre>					

Variances are relatively close (remember to square these SD's to compare to right).

The estimates from SAS and R are the same; SE's are close.

Given it is PQL, it is a subject-specific locus expansion. Since R output states that the '...model fit by Maximum likelihood...', I believe that the default method used in R is equivalent to MSPL in PROC GLIMMIX (rather than RSPL, which is the default method in PROC GLIMMIX). This needs to be verified, though.

Summary of some key points in previous analyses and extra notes

For Poisson regression, the residual variance acts as the scale (or dispersion) parameter if you include 'R-side' parameters in the model.

If you compare 'apples-to-apples' PL versus GEE models, you may notice that the residual variance is about equal to the scale parameter, and if the SE's of the beta estimates are adjusted for the scale parameter in the GEE, they should be about the same as those from PL.

The 'Gener. Chi-Square / DF' fit statistic reflects over/underdispersion. From the Poisson models I've fit, it is equal to the dispersion parameter.

Fitting a negative binomial for the same underdispersed data yields a scale parameter of 0, meaning it is the same distribution as the Poisson, but in this case the statistic becomes 1. The fact that the scale parameter is 0 is not surprising since the NB is more suited for overdispersed data, but not sure why the fit statistics are different.

Breslow NE and Clayton DG. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88, 9-25.

Wolfinger, R.D. and O'Connell, M. (1993) Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* 48, 233-243.

Summary

BIOS6643
Longitudinal

EJC

Linearization

References

Summary