SAS® GLOBAL FORUM 2020 #SASGF

Paper 4181-2020

Modeling of Between- and Within-Subject Variances using Mixed Effects Location Scale (MELS) Models

Qianheng Ma, The University of Chicago; Donald Hedeker, The University of Chicago

ABSTRACT

Intensive longitudinal data are increasingly obtained in health studies to examine subjective experiences within changing environmental contexts. Such studies often utilize ecological momentary assessment (EMA) and/or experience sampling methods to obtain up to 30 or 40 observations for each subject within a period of a week or so. In this paper, we focus on data from an adolescent smoking study using ecological momentary assessment in which there was interest in examining mood variation associated with smoking. We describe the mixed-effects location scale (MELS) model which allows covariates to influence both the within-subject (WS) and between-subject (BS) variances, in addition to their influence on the mean. The model also includes a subject-level random effect to the within-subject variance specification. This permits subjects to have influence on the mean, or location, and variability, or (square of the) scale, of how soon the first smoking episode happened after wake-up. Additionally, we allow the location and scale random effects to be correlated. These mixed-effects location scale models have useful applications in many research areas where interest centers on the joint modeling of the mean and variance structure. In this presentation, we describe how SAS® PROC NLMIXED and SAS PROC MCMC can be used to estimate the parameters of the MELS model.

INTRODUCTION

Linear mixed effect models are frequently used for handling heterogeneity across clusters in multilevel longitudinal data. For a two-level mixed model for repeated measures (level-1) nested within-subjects (level-2), the random subject effect summarizes the latent subject-specific features apart from the observed covariate effects of interest. In traditional data analysis on longitudinal data using mixed effect models, the subject-level random effect is often specified in the mean function of the repeated outcome measurements. By including the random subject effects in the mean function, subjects are allowed to have different mean levels for their outcome trajectories.

The development of novel real-time data collection approaches such as ecological momentary assessments (EMA) even provides researchers intensive amount of longitudinal data which contains sufficient information to infer not only the subject mean but also the within-subject variability of the repeated outcome measurements. Motivated by these multilevel intensive longitudinal data, a mixed-effect location-scale (MELS) model was proposed as an extension of the linear mixed model (Hedeker, D. et al., 2008, 2013). In the MELS model, the random subject effect in the mean function of the outcome is called the random location effect; while the random subject effect in the within-subject variance function is called the random scale effect. Both are assumed to follow a bivariate normal distribution with mean zero and with a general variance-covariance matrix. The intuition of including the random scale effect is that, as is observed in many studies, some participants' outcome, such as mood, may display a more consistent pattern but others' mood can vary erratically across time. By including the scale effect, subjects are allowed to display different levels of variability in terms of their outcome trajectories.

Although the outcome variable in MELS is assumed to follow a normal distribution, MELS belongs to the class of generalized linear mixed effect model (GLMM) as the relationship between the outcome and the random scale effect is not linear. Flexible estimation approaches should be used to handle this complex distribution assumption and specification. In SAS/STAT, the NLMIXED modelling procedure enables us to specify a flexible formulation for within-subject variability as well as a user-defined likelihood function. In NLMIXED, maximum likelihood estimation is used for estimation. A critical step of this estimation method is to integrate out the random location/scale effects from the full likelihood function to obtain a marginal likelihood containing only the fixed effects. In other words, this integrated marginal likelihood function is the objective function for later optimization. For GLMM, usually there is no closed-form solution for this integral evaluation so that more general integration techniques such as adaptive Gaussian quadrature can be used. However, a shortcoming of this algorithm is that the complexity grows exponentially with the dimension of the random effects and so it is often time-consuming and computationally demanding.

With similar syntax as PROC NLMIXED, PROC MCMC also has flexibility for fitting unconventional generalized mixed effect models. Instead of maximizing the likelihood function, PROC MCMC utilizes a Bayesian approach to draw samples from the posterior distributions of the model parameters using the Metropolis-Hasting Algorithm. This procedure also produces the credible intervals or highest posterior density intervals for inference. According to our practice, the computation time of Markov Chain Monte Carlo methods doesn't grow significantly along with the dimension of random effects.

The paper is organized as follows. First, the MELS section introduces the notation of the MELS model, including transformation of the random effects. The next section "ADOLESCENT SMOKING SUDY" describes the real dataset that is used in the examples. Sections "ESTIMATE MELS USING PROC NLMIXED" and "ESTIMATE MELS USING PROC MCMC" provides syntax examples of PROC NLMIXED and PROC MCMC for estimating the parameters of the MELS model using the adolescent smoking data.

MIXED-EFFECT LOCATION-SCALE (MELS) MODEL

The model structure of the MELS model is similar to the linear mixed effect model. Assume the data are two-level and the outcome variable y_{ij} of subject i $(i = 1, 2, \dots, N\ subjects)$ at the j^{th} assessment occasion within each subject $(j = 1, 2, \dots, n_i\ occations)$ is expressed as Equation (1). $\boldsymbol{\beta}$ is the $(p+1)$ -dimension vector of fixed effects and ν_i is the i^{th} subject's random location effect in the mean function of the outcome, which represents the additional deviation of the covariate effects/intercept of subject i from the population

$$y_{ij} = \boldsymbol{X_{ij}^T}\boldsymbol{\beta} + \nu_i + \epsilon_{ij} \tag{1}$$

The variance of ν_i, $\sigma^2_{\upsilon_{ij}}$ represents the between-subject (BS) variance. Instead of assuming it to be constant across subjects, $\sigma^2_{\upsilon_{ij}}$ can be further expressed by an exponential form in the log-linear equation (2) to allow covariates $\boldsymbol{u_{ij}}$ to affect the BS variance. $\boldsymbol{\alpha}$ is the fixed effect vector for the BS variance associated with the $\boldsymbol{u_{ij}}$ covariates. The use of the exponential function ensures a positive value for the resulting variance.

$$\sigma^2_{\upsilon_{ij}} = exp(\boldsymbol{u}_{ij}^{\boldsymbol{T}}\,\boldsymbol{\alpha}) \tag{2}$$

ϵ_{ij} in Equation (1) is the within-subject (WS) measurement error and is assumed to follow a normal distribution $\boldsymbol{N}\left(0,\ \sigma^2_{\epsilon_{ij}}\right)$. Similar to Equation (2), the WS variance can be represented by a log-linear function (Equation (3)) as follows:

$$\sigma^2_{\epsilon_{ij}} = exp(\boldsymbol{w}_{ij}^{\boldsymbol{T}}\boldsymbol{\tau} + \omega_i) \quad (3)$$

Here, $\boldsymbol{\tau}$ is the fixed effect vector affecting WS variance corresponding to covariates $\boldsymbol{w_{ij}}$. ω_i is the i^{th} subject's random scale effect in the WS variance. By including ω_i, even subjects with the same covariate values can have different levels of variation in their longitudinal outcome trajectories.

ν_i and ω_i are not necessarily independent. Therefore, the distribution of $(\nu_i,\ \omega_i)$ is

$$\begin{pmatrix}\nu_i\\ \omega_i\end{pmatrix} \sim N\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}\sigma^2_{\upsilon_{ij}} & \sigma_{\nu\omega}\\ \sigma_{\nu\omega} & \sigma^2_{\omega}\end{pmatrix}\right)$$

where $\sigma_{\nu_{ij}}$ and σ_ω are standard deviations of $(\nu_i,\ \omega_i)$, and $\sigma_{\nu\omega}$ is the covariance between the random location and scale effects.

ADOLESCENT SMOKING STUDY

We will use real data collected from participants in an Adolescent Smoking Study, which included 461 participants that were either 9th or 10th graders. In this study, participants carried a PDA for a week at baseline, and responded to random prompts, occurring approximately 4 to 5 times per day, and self-initiated smoking events. These subjects were similarly measured at the 6-month, 15-month, 2nd year, 5th and 6th year follow-up waves. For this part of the study, we focused on a subset of 242 subjects ($N = 242$) that had data at years 5 and/or 6 and contributed at least two waves with ≥ 2 smoking events at each wave. Therefore, subjects can have different numbers of smoking days and the number of smoking days n_i ranged from 2 to 18 with mean 8.37 for each subject (median=8). The longitudinal outcome of interest in this example is the time of the first report of smoking in the day (after 5am), which is a biomarker for smoking dependence. It is thought that subjects who smoke their first cigarette earlier in the day are more dependent smokers. From descriptive analysis, the outcome variable ranged from 0.02 to 23.13 hours (past 5am) with a mean=9.55 hours.

Conventionally, time-to-event data are analyzed by survival models. As we observed that the outcome was distributed close to a normal distribution, we assumed it to follow a normal distribution as specified in the MELS model of the last section. The outcome variable, time to the first cigarette of the day, is expressed in units of hours after 5am in the morning:

```
TimeHour = timeT - 5;
```

Before fitting models, we first preprocess some independent variables in the dataset. From the weekday variable (1 to 7), seven binary dummies were created. Six of these seven dummies will be used in the analysis.

```
mon=0;tue=0;wed=0;thu=0;fri=0;sat=0;sun=0;
if wkday5am=1 then mon=1;
if wkday5am=2 then tue=1;
if wkday5am=3 then wed=1;
if wkday5am=4 then thu=1;
if wkday5am=5 then fri=1;
```

```
if wkday5am=6 then sat=1;
if wkday5am=7 then sun=1;
```

MELS MODEL ESTIMATION USING PROC NLMIXED

As PROC NLMIXED can be sensitive to the specification of initial values, we first fit a heterogeneity mixed model using PROC MIXED and use the parameter estimates as starting values for many of the parameters of our MELS model (that will be fit with PROC NLMIXED).

```
PROC MIXED METHOD=ML COVTEST;
  CLASS id;
  MODEL TimeHour=tue wed thu fri sat sun wave6 Smk7rate/s;
  RANDOM INT/SUBJECT=id;
  REPEATED/LOCAL=exp(tue wed thu fri sat sun wave6 Smk7rate);
RUN;
```

The PROC MIXED procedure is fitting the model as below.

- The mean function $y_{ij} = (\beta_0 + \Sigma_{k=1}^{6} \beta_k D_{ijk} + \beta_7 wave_j + \beta_8 SmkRate_{ij}) + \nu_i + \epsilon_{ij}$
- For WS variance, $\sigma^2_{\epsilon_{ij}} = EXP(\tau_0 + \Sigma_{k=1}^{6} \tau_k D_{ijk} + \tau_7 wave_j + \tau_8 SmkRate_{ij})$; note that PROC MIXED does not allow a random scale effect.
- For BS variance, $\sigma^2_{\upsilon_{ij}} = EXP(\alpha_0)$ and contain only the intercept.

The output for the variance parameters from PROC MIXED is displayed below:

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Intercept	id	4.1784	0.6117	6.83	<.0001
EXP tue		-0.1332	0.1253	-1.06	0.2877
EXP wed		-0.2225	0.1276	-1.74	0.0812
EXP thu		-0.02743	0.1249	-0.22	0.8262
EXP fri		0.003630	0.1262	0.03	0.9771
EXP sat		0.04770	0.1313	0.36	0.7163
EXP sun		0.01055	0.1348	0.08	0.9377
EXP wave6		0.03296	0.07113	0.46	0.6431
EXP smk7rate		-0.1792	0.05263	-3.40	0.0007
Residual		17.6821	1.8989	9.31	<.0001

Display 1. PROC MIXED: BS And WS Variance

Estimates with initial of 'EXP' denotes the estimates for $\boldsymbol{\tau}_1$ to $\boldsymbol{\tau}_8$. The term "Residual" denotes the estimate of $EXP(\tau_0)$, and the term "Intercept" denotes the variance of the location random effect, the BS variance, namely $\sigma^2_{\upsilon_{ij}} = EXP(\alpha_0)$. Therefore, the estimate for τ_0 is $\ln(17.6821) = 2.8726$ and the estimate for α_0 is $\ln(4.1784) = 1.43$ which can be used as the initial values for the parameters in the BS variance as well as the WS variance in the subsequent PROC NLMIXED code.

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	11.4756	0.3382	241	33.93	<.0001
tue	-0.6326	0.3182	1775	-1.99	0.0469
wed	-0.7576	0.3133	1775	-2.42	0.0157
thu	-0.6899	0.3253	1775	-2.12	0.0341
fri	-0.5676	0.3297	1775	-1.72	0.0853
sat	-0.04530	0.3438	1775	-0.13	0.8952
sun	0.4450	0.3528	1775	1.26	0.2074
wave6	-0.2306	0.1831	1775	-1.26	0.2080
smk7rate	-1.3321	0.2280	1775	-5.84	<.0001

Display 2. PROC MIXED: Fixed Effects

Similarly, the estimated displayed in the output above provides starting values for the regression coefficients $\boldsymbol{\beta}$.

These PROC MIXED estimates are then used as starting values for $\boldsymbol{\beta}$ as well as $\boldsymbol{\tau}$, in other words the fixed effects in the mean function and the WS variance function. However, we only obtain the intercept in the BS variance function from PROC MIXED. Therefore, we plug in the estimates of $\boldsymbol{\beta}$ $\boldsymbol{\tau}$ and α_0 from PROC MIXED as starting values in the PROC NLMIXED PARMS statement, and assign values of zero for the few remaining parameters.

```
PARMS beta0=11.4756 beta_tue=-0.6326 beta_wed=-0.7576 beta_thu=-0.6899
beta_fri=-0.5676 beta_sat=-0.04530 beta_sun=0.4450 beta_wave6=-0.2306
beta_smk7rate=-1.3321

tau0=2.8726 tau_tue=-0.1332 tau_wed=-0.2225 tau_thu=-0.02743
tau_fri=0.003630 tau_sat=0.04770 tau_sun=0.01055 tau_wave6=0.03296
tau_smk7rate=-0.1792

alpha0=1.43 alpha_tue=0 alpha_wed=0 alpha_thu=0 alpha_fri=0 alpha_sat=0
alpha_sun=0 alpha_wave6=0 alpha_Smk7rate=0

corr=0 ln_varScale=0;
```

For estimation, PROC NLMIXED uses maximum likelihood estimation (MLE) method.

```
PROC NLMIXED GCONV=1e-12 TECH=QUANEW METHOD=GAUSS QPOINTS=11;
```

Prior to the optimization stage, full conditional likelihood will be integrated with respect to random effects and then the resulted marginal likelihood will be the objective function for optimization. The Adaptive Gaussian Quadrature method (METHOD=GAUSS) is a general approach for this integration step and is applicable for generalized linear models and 11 quadrature points are specified for this integral evaluation. For NLMIXED, the default optimization method is Quasi-Newton method (TECH=QUANEW) which uses the gradient function to approximate the Hessian matrix so that no extra steps are needed for computing the Hessian. As a result, the computation time for each iteration of the Quasi-Newton method is faster than the traditional Newton-Raphson algorithm, however more steps may be needed for convergence.

For the efficiency of estimation, the random effects (ν_i, ω_i) can be represented in terms of standardized random effects (θ_{1i}, θ_{2i}) with unit variance and $\rho_{\nu\omega}$ as the correlation. In other words, (ν_i, ω_i) = ($\sigma_{\nu_{ij}}\theta_{1i}$, $\sigma_\omega\theta_{2i}$) and (θ_{1i}, θ_{2i}) follow a bivariate normal distribution as follows:

$$\begin{pmatrix}\theta_{1i}\\ \theta_{2i}\end{pmatrix} \sim N\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}1 & \rho_{\nu\omega}\\ \rho_{\nu\omega} & 1\end{pmatrix}\right) \tag{4}$$

Using this standardization, the integral evaluation step is simplified so that the computation time taken for the quadrature method will be shorter and convergence will be easier to achieve. We could also use the Cholesky decomposition (which is what we will use in the PROC MCMC section below).

The outcome y_{ij} is assumed to follow a normal distribution $N(\mu_{ij}, \sigma^2_{\epsilon_{ij}})$.

```
MODEL TimeHour ~ NORMAL(mu,varWS);
```

For the BS variance, $\sigma^2_{\upsilon_{ij}} = EXP(\alpha_0 + \Sigma_{k=1}^{6}\alpha_k D_{ijk} + \alpha_7 wave_j + \alpha_8 SmkRate_{ij})$,

```
varBS = EXP(alpha0 + alpha_tue*tue + alpha_wed*wed + alpha_thu*thu +
alpha_fri*fri + alpha_sat*sat + alpha_sun*sun + alpha_wave6*wave6 +
alpha_Smk7rate*Smk7rate);
```

For the mean function $\mu_{ij} = \beta_0 + \Sigma_{k=1}^{6}\beta_k D_{ijk} + \beta_7 wave_j + \beta_8 SmkRate_{ij} + \nu_i$ and $\nu_i = \sigma_{\nu_{ij}}\theta_{1i}$

```
nu = SQRT(varBS)*theta1;

mu = beta0 + beta_tue*tue + beta_wed*wed + beta_thu*thu + beta_fri*fri +
beta_sat*sat + beta_sun*sun + beta_wave6*wave6 + beta_Smk7rate*Smk7rate +
nu;
```

For the WS variance, $\sigma^2_{\epsilon_{ij}} = EXP(\tau_0 + \Sigma_{k=1}^{6}\tau_k D_{ijk} + \tau_7 wave_j + \tau_8 SmkRate_{ij} + \omega_i)$ and $\omega_i = \sigma_\omega\theta_{2i}$

```
StdDev_scale=SQRT(EXP(ln_varScale));
omega= StdDev_scale*theta2;

varWS = EXP(tau0 + tau_tue*tue + tau_wed*wed + tau_thu*thu + tau_fri*fri +
tau_sat*sat + tau_sun*sun + tau_wave6*wave6 + tau_Smk7rate*Smk7rate +
omega);
```

In the above, we represent σ^2_ω as an exponential function (which is always positive) so that this optimization is free of any constraints. The standard deviation of the scale random effect ω_i is the square-root of this exponential function.

Finally, the random effect distribution is assumed to follow s bivariate normal distribution (Equation (4)).

```
RANDOM theta1 theta2 ~ NORMAL([0,0], [1,corr,1]) SUBJECT=id;
```

Combining these statements together provides the PROC NLMIXED code for estimation of the parameters of the MELS model:

```
PROC NLMIXED GCONV=1e-12 TECH=QUANEW METHOD=GAUSS QPOINTS=11;

  PARMS beta0=11.4756 beta_tue=-0.6326 beta_wed=-0.7576 beta_thu=-0.6899
  beta_fri=-0.5676 beta_sat=-0.04530 beta_sun=0.4450 beta_wave6=-0.2306
  beta_smk7rate=-1.3321
```

```
  tau0=2.8726 tau_tue=-0.1332 tau_wed=-0.2225 tau_thu=-0.02743
  tau_fri=0.003630 tau_sat=0.04770 tau_sun=0.01055 tau_wave6=0.03296
  tau_smk7rate=-0.1792

  alpha0=1.43 alpha_tue=0 alpha_wed=0 alpha_thu=0 alpha_fri=0 alpha_sat=0
  alpha_sun=0 alpha_wave6=0 alpha_Smk7rate=0

  corr=0 ln_varScale=0;

  varBS = EXP(alpha0 + alpha_tue*tue + alpha_wed*wed + alpha_thu*thu +
  alpha_fri*fri + alpha_sat*sat + alpha_sun*sun + alpha_wave6*wave6 +
  alpha_Smk7rate*Smk7rate);

  StdDev_scale=sqrt(exp(ln_varScale));

  nu= SQRT(varBS)*theta1;
  omega= StdDev_scale*theta2;

  mu  = beta0 + beta_tue*tue + beta_wed*wed + beta_thu*thu + beta_fri*fri  +
  beta_sat*sat + beta_sun*sun + beta_wave6*wave6 + beta_Smk7rate*Smk7rate +
  nu;

  varWS = EXP(tau0 + tau_tue*tue + tau_wed*wed + tau_thu*thu + tau_fri*fri
  + tau_sat*sat + tau_sun*sun + tau_wave6*wave6 + tau_Smk7rate*Smk7rate +
  omega);

MODEL TimeHour ~ NORMAL(mu,varWS);
RANDOM theta1 theta2 ~ NORMAL([0,0], [1,corr,1]) SUBJECT=id;
RUN;
```

The parameter estimates from PROC NLMIXED are listed below. We conclude that, controlling for other covariate effects, higher smoking rate is significantly associated with earlier first cigarette (mean estimate = 1.46 hours). In terms of the WS variance, higher smoking rate is associated with greater consistency (i.e., lower WS variance). Exponentiating the estimate (exp(-0.2) = 0.82) yields a variance ratio estimate; with every unit increase in smoking rate the WS variance is reduced by 18%. Thus, subjects with higher smoking rates have their first cigarette earlier, on average, and exhibit greater consistency in the time for their first cigarette. Smoking rate does not have a significant effect on the BS variance. Another interesting observation is that the correlation between location and scale is positive, which indicates that subjects who smoke the first cigarette earlier/later in the morning tend are more consistent/erratic.

Parameter Estimates								
Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	95% Confidence Limits		Gradient
beta0	11.5724	0.3679	240	31.46	<.0001	10.8476	12.2971	-0.00001
beta_tue	-0.6221	0.3456	240	-1.80	0.0731	-1.3028	0.05859	6.201E-6
beta_wed	-0.7920	0.3366	240	-2.35	0.0194	-1.4551	-0.1290	-0.00003
beta_thu	-0.6630	0.3479	240	-1.91	0.0579	-1.3484	0.02230	-3E-6
beta_fri	-0.4988	0.3515	240	-1.42	0.1571	-1.1912	0.1935	0.000031
beta_sat	-0.09854	0.3677	240	-0.27	0.7890	-0.8230	0.6259	0.000011
beta_sun	0.4795	0.3824	240	1.25	0.2111	-0.2738	1.2327	-0.00004
beta_wave6	-0.3188	0.2029	240	-1.57	0.1174	-0.7185	0.08088	-0.00002
beta_smk7rate	-1.4570	0.2749	240	-5.30	<.0001	-1.9985	-0.9155	0.000034
alpha0	1.4618	0.3976	240	3.68	0.0003	0.6785	2.2451	3.147E-6
alpha_tue	-0.08216	0.4070	240	-0.20	0.8402	-0.8840	0.7197	-0.00004
alpha_wed	-0.1074	0.3921	240	-0.27	0.7844	-0.8798	0.6650	0.000010
alpha_thu	0.1520	0.3804	240	0.40	0.6898	-0.5974	0.9014	0.000025
alpha_fri	0.2355	0.3879	240	0.61	0.5443	-0.5287	0.9997	0.000053
alpha_sat	-0.2678	0.4686	240	-0.57	0.5681	-1.1909	0.6552	-1.78E-6
alpha_sun	0.08530	0.4227	240	0.20	0.8403	-0.7474	0.9180	-0.00003
alpha_wave6	-0.3327	0.2873	240	-1.16	0.2479	-0.8986	0.2332	0.000025
alpha_Smk7rate	0.04170	0.2652	240	0.16	0.8752	-0.4808	0.5642	8.909E-6
tau0	2.9031	0.1163	240	24.97	<.0001	2.6740	3.1321	0.000097
tau_tue	-0.1252	0.1314	240	-0.95	0.3414	-0.3840	0.1336	7.316E-6
tau_wed	-0.2581	0.1347	240	-1.92	0.0566	-0.5235	0.007355	0.000020
tau_thu	-0.07235	0.1312	240	-0.55	0.5820	-0.3309	0.1862	0.000036
tau_fri	-0.04127	0.1342	240	-0.31	0.7586	-0.3055	0.2230	0.000017
tau_sat	0.02432	0.1383	240	0.18	0.8606	-0.2482	0.2968	2.23E-6
tau_sun	0.02784	0.1414	240	0.20	0.8441	-0.2507	0.3064	0.000013
tau_wave6	0.07178	0.08076	240	0.89	0.3750	-0.08731	0.2309	0.000170
tau_smk7rate	-0.2007	0.06550	240	-3.06	0.0024	-0.3297	-0.07163	0.000055
corr	0.8456	0.1612	240	5.24	<.0001	0.5280	1.1632	-7.68E-6
ln_varScale	-2.6794	0.4229	240	-6.34	<.0001	-3.5124	-1.8463	4.783E-6

Display 3. PROC NLMIXED: Solutions

The computation time for this PROC NLMIXED analysis took almost 26 minutes for convergence. One reason is because a relatively large number (11) of quadrature points were specified. For faster computation, we can specify only 1 quadrature point in most of the cases which is equivalent to the Laplace Approximation (Pinheiro, J. C., Chao, E. C., 2016). Alternatively, when we did not specify the QPOINTS, then PROC NLMIXED determined that only 5 quadrature points were needed for this example and the computation time was about 6 minutes.

MELS MODEL ESTIMATION USING PROC MCMC

In PROC MCMC, which uses Bayesian methods, the parameters to be estimated are viewed as random variables and are assumed to follow certain probabilistic distributions. PROC MCMC uses Markov chain Monte Carlo sampling to generate sequential samples from their posterior distribution given the data. The MCMC method does not include integration steps so that the complexity of this method will not increase dramatically with the dimension of the random effects.

The syntax of PROC MCMC is similar to PROC NLMIXED, with some additional specifications. Here, we specify 100,000 repetitions of MCMC sampling (NMC). The first few repetitions are called the burn-in samples which can be highly inter-dependent before the Markov Chain converges. In this example, these 500 burn-in samples (NBI=500) will be excluded from the calculation of posterior sample means. Thinning is used to select the posterior samples for the mean calculation. Here, we specify THIN=25 so that one sample is selected from every 25 repetitions for the mean calculation. In the Bayesian approach, the frequentist model selection criterion of AIC or BIC are not provided, instead the Deviance Information Criterion (DIC) will be provided.

```
PROC MCMC SEED=9879 NMC=100000 NBI=500 MAXTUNE=50 THIN=25 DIC;
```

Similar to PROC NLMIXED, we also specify the initial values using the estimates from PROC MIXED. It is convenient that PROC MCMC can do both the group-wise update and parameter-wise update. Multiple PARMS statement are allowed and each PARMS statement represents a block of parameters to be updated at the same time. Here we only use one PARMS statement which means one single block and all of the model parameters are updated at each step.

```
  PARMS beta0=11.4756 beta_tue=-0.6326 beta_wed=-0.7576 beta_thu=-0.6899
  beta_fri=-0.5676 beta_sat=-0.04530 beta_sun=0.4450 beta_wave6=-0.2306
  beta_smk7rate=-1.3321

  tau0=2.8726 tau_tue=-0.1332 tau_wed=-0.2225 tau_thu=-0.02743
  tau_fri=0.003630 tau_sat=0.04770 tau_sun=0.01055 tau_wave6=0.03296
  tau_smk7rate=-0.1792

  alpha0=1.43 alpha_tue=0 alpha_wed=0 alpha_thu=0 alpha_fri=0 alpha_sat=0
  alpha_sun=0 alpha_wave6=0 alpha_Smk7rate=0

  corr=0 ln_varScale=0;
```

In addition to the initial values, we also need to specify the prior distribution for the parameters. For the beta parameters, which are mutually independent and assumed to have the same univariate prior distribution, we can simply specify:

```
  PRIOR be: ~NORMAL (0, VAR=10000);
```

For a single parameter:

```
  PRIOR cov~NORMAL (0, VAR=10000);
```

PROC MCMC has flexible options for different specification of the priors. The priors we assigned to the parameters in this example are extremely flat priors, a normal distribution with extremely large variance, which is close to the (improper) uniform prior which contains almost no prior information for the parameters. As a result of this, estimation will not be seriously influenced by the prior.

The mean, BS variance, WS variance are actually specified the same as in PROC NLMIXED. However, PROC MCMC doesn't allow the random effects to be parameter-dependent. So before the specification of the mean/BS variance/WS variance, one extra step using the Cholesky decomposition is needed:

```
L11=SQRT(varBS);
L21=cov/L11;
L22=SQRT(varScale-L21**2);
```

By doing the Cholesky decomposition, we represent the location and scale random effects as two independent subject-level random effects (s_{1i}, s_{2i}) with unit variance, i.e.,

$$\begin{pmatrix} \sigma_{\nu_{ij}} & \sigma_{\nu\omega}/\sigma_{\nu_{ij}} \\ \sigma_{\nu\omega}/\sigma_{\nu_{ij}} & \sqrt{\sigma_{\omega}^2 - \sigma_{\nu\omega}^2/\sigma_{\nu_{ij}}^2} \end{pmatrix} \begin{pmatrix} s_{1i} \\ s_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\nu_{ij}}^2 & \sigma_{\nu\omega} \\ \sigma_{\nu\omega} & \sigma_{\omega}^2 \end{pmatrix} \right)$$

And,

$$\begin{pmatrix} s_{1i} \\ s_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

Therefore,

```
nu=    L11*sub_1;
omega= L21*sub_1+L22*sub_2;

mu  = beta0 + beta_tue*tue  + beta_wed*wed  + beta_thu*thu  + beta_fri*fri
+ beta_sat*sat  + beta_sun*sun  + beta_wave6*wave6  +
beta_Smk7rate*Smk7rate + nu;

varWS = EXP(tau0 + tau_tue*tue + tau_wed*wed + tau_thu*thu + tau_fri*fri +
tau_sat*sat + tau_sun*sun + tau_wave6*wave6 + tau_Smk7rate*Smk7rate +
omega);
```

Here, the random effects follow a multivariate (bivariate) normal distribution in which the mean vector and the covariance matrix are required to be arrays as follows. If the array contains all numeric values, parentheses and commas will be needed:

```
ARRAY location_scale_mean[2] (0,0);
ARRAY location_scale_cov[2,2] (1,0,0,1);
```

If the array consists of parameters or variables in the dataset then no parentheses are needed:

```
ARRAY location_scale_rf[2] sub_1 sub_2;
```

Then the random effect distribution is specified as:

```
  RANDOM location_scale_rf~MVN(location_scale_mean, location_scale_cov)
  SUBJECT=id;
```

Combining these statements provides the PROC MCMC code for estimation of the parameters of the MELS model:

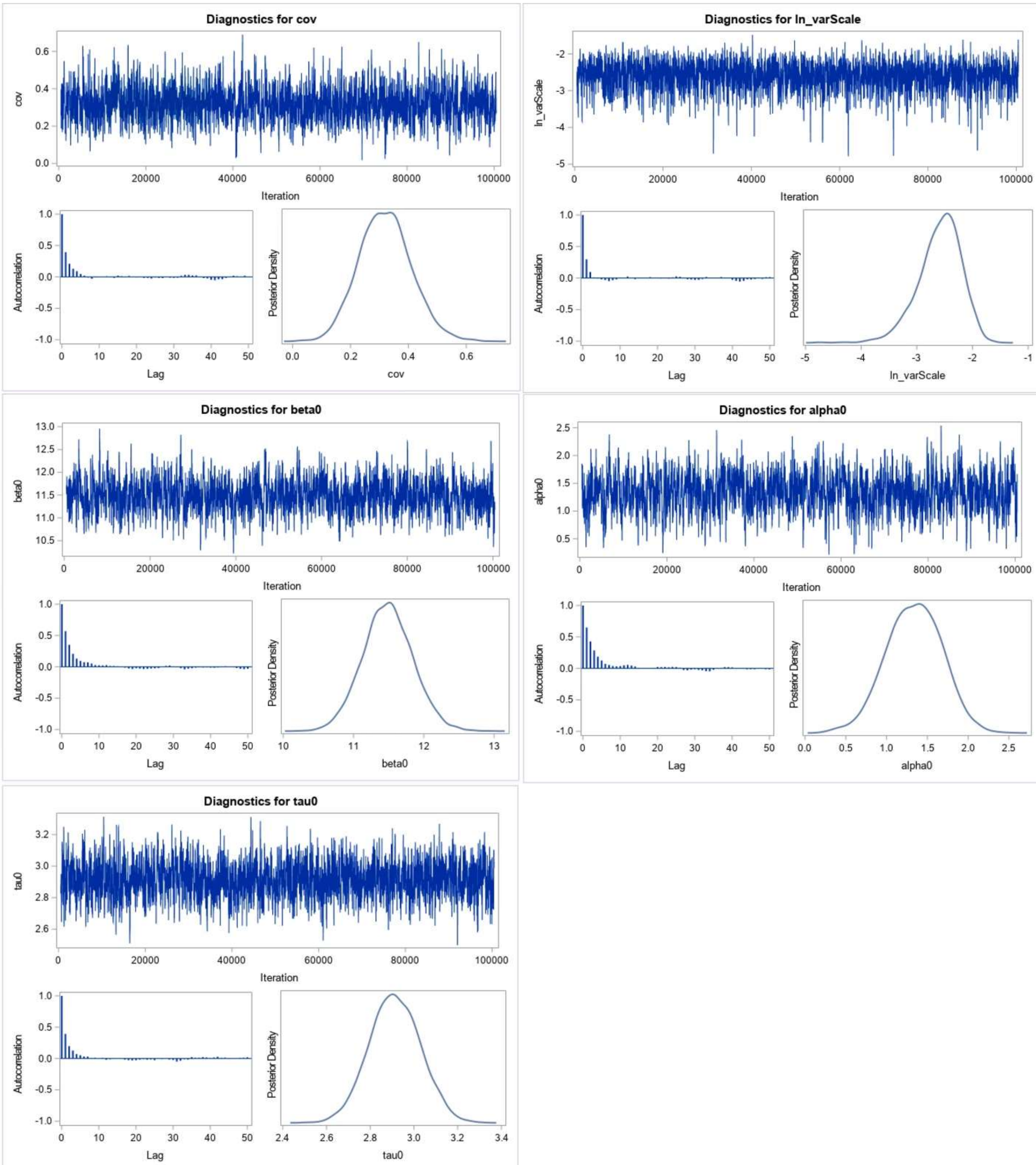
```
PROC MCMC SEED=9879 NMC=100000 NBI=500 MAXTUNE=50 THIN=25 DIC;
  PARMS beta0=11.4756 beta_tue=-0.6326 beta_wed=-0.7576 beta_thu=-0.6899
  beta_fri=-0.5676 beta_sat=-0.04530 beta_sun=0.4450 beta_wave6=-0.2306
  beta_smk7rate=-1.3321

  tau0=2.8726 tau_tue=-0.1332 tau_wed=-0.2225 tau_thu=-0.02743
  tau_fri=0.003630 tau_sat=0.04770 tau_sun=0.01055 tau_wave6=0.03296
  tau_smk7rate=-0.1792

  alpha0=1.43 alpha_tue=0 alpha_wed=0 alpha_thu=0 alpha_fri=0 alpha_sat=0
  alpha_sun=0 alpha_wave6=0 alpha_Smk7rate=0

  cov=0 ln_varScale=.005;

  PRIOR be: ~NORMAL (0, VAR=10000);
  PRIOR al: ~NORMAL (0, VAR=10000);
  PRIOR ta: ~NORMAL (0, VAR=10000);
  PRIOR cov: ~NORMAL (0, VAR=10000);
  PRIOR ln_varScale~NORMAL(0,VAR=10000);

  varBS = EXP(alpha0 + alpha_tue*tue + alpha_wed*wed + alpha_thu*thu +
  alpha_fri*fri + alpha_sat*sat + alpha_sun*sun + alpha_wave6*wave6 +
  alpha_Smk7rate*Smk7rate);

  varScale = EXP(ln_varScale);

  L11=SQRT(varBS);
  L21=cov/L11;
  L22=SQRT(varScale-L21**2);

  mu  = beta0 + beta_tue*tue + beta_wed*wed  + beta_thu*thu + beta_fri*fri  +
  beta_sat*sat + beta_sun*sun + beta_wave6*wave6  + beta_Smk7rate*Smk7rate +
  L11*sub_1;

  varWS = EXP(tau0 + tau_tue*tue + tau_wed*wed + tau_thu*thu + tau_fri*fri +
  tau_sat*sat + tau_sun*sun + tau_wave6*wave6 + tau_Smk7rate*Smk7rate +
  L21*sub_1+L22*sub_2);

  ARRAY location_scale_mean[2] (0,0);
  ARRAY location_scale_cov[2,2] (1,0,0,1);
  ARRAY location_scale_rf[2] sub_1 sub_2;

  MODEL TimeHour ~ NORMAL(mu,varWS);
  RANDOM location_scale_rf~MVN(location_scale_mean,location_scale_cov)
SUBJECT=id;

RUN;
```

Posterior Summaries and Intervals					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
beta0	4000	11.4879	0.3554	10.7689	12.1628
beta_tue	4000	-0.6005	0.3457	-1.2446	0.0959
beta_wed	4000	-0.7837	0.3339	-1.4107	-0.1145
beta_thu	4000	-0.6439	0.3498	-1.3396	0.0584
beta_fri	4000	-0.4852	0.3513	-1.1624	0.2338
beta_sat	4000	-0.0719	0.3665	-0.7503	0.6814
beta_sun	4000	0.5377	0.3753	-0.1543	1.3015
beta_wave6	4000	-0.2571	0.1943	-0.6238	0.1247
beta_smk7rate	4000	-1.4172	0.2515	-1.9407	-0.9432
alpha0	4000	1.3377	0.3395	0.6972	2.0023
alpha_tue	4000	-0.0304	0.3836	-0.8340	0.6785
alpha_wed	4000	-0.0832	0.3753	-0.8243	0.6334
alpha_thu	4000	0.1870	0.3698	-0.4975	0.9169
alpha_fri	4000	0.2655	0.3736	-0.4419	1.0461
alpha_sat	4000	-0.1863	0.4184	-1.0501	0.5614
alpha_sun	4000	0.2586	0.3881	-0.5555	0.9632
alpha_wave6	4000	-0.1998	0.2430	-0.6602	0.3197
alpha_Smk7rate	4000	0.0402	0.1655	-0.2786	0.3771
tau0	4000	2.9130	0.1155	2.6900	3.1353
tau_tue	4000	-0.1314	0.1321	-0.3743	0.1388
tau_wed	4000	-0.2562	0.1375	-0.5216	0.0121
tau_thu	4000	-0.0836	0.1331	-0.3454	0.1801
tau_fri	4000	-0.0498	0.1351	-0.3025	0.2153
tau_sat	4000	0.0264	0.1380	-0.2474	0.2932
tau_sun	4000	-0.0160	0.1483	-0.2994	0.2754
tau_wave6	4000	0.0461	0.0828	-0.1242	0.2041
tau_smk7rate	4000	-0.1910	0.0633	-0.3144	-0.0658
cov	4000	0.3204	0.0903	0.1495	0.5003
ln_varScale	4000	-2.5901	0.3898	-3.3767	-1.9142

Display 4. PROC MCMC: Posterior Means

In PROC MCMC, posterior means serve as the estimates for the parameters. PROC MCMC also provides the 95% highest posterior density (HPD) intervals or the 95% credible intervals for inference. If the interval doesn’t include 0, then we can conclude that the covariate has significant effect on the mean/BS variance/WS variance. Compared to the results from PROC NLMIXED, PROC MCMC produced similar estimates as well as similar conclusions to PROC NLMIXED.

In the trace plots below, we only show the convergence of the intercepts in mean/BS variance/WS variance as well as the scale variance and the covariance between location and scale. All of the parameters converge.



Display 5. PROC MCMC: Convergence

CONCLUSION

One can estimate the parameters of the MELS model using either frequentist (PROC NLMIXED) or Bayesian (PROC MCMC) methods. For both, PROC MIXED can be used to provide starting values for many, but not all, of the model parameters. For this example PROC MCMC was more efficient as it only took 8 minutes to complete the 100,000 iterations, compared to 26 minutes for PROC NLMIXED (QPOINTS=11).

REFERENCES

Hedeker, D., Mermelstein, R. J., Demirtas, H., 2008. "An Application of a Mixed-Effects Location Scale Model for Analysis of Ecological Momentary Assessment (EMA) Data." *Biometrics,* 64(2):627-634.

Hedeker, D., Nordgren, R., 2013. "MIXREGLS: A Program for Mixed-Effects Location Scale Analysis" *Journal of Statistical Software,* 52(12):1-38.

Pinheiro, J. C., Chao, E. C., 2006, "Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models," *Journal of Computational and Graphical Statistics,* 15(1):58-81.

ACKNOWLEDGMENTS

Data used in this publication are from the National Cancer Institute funded Adolescent Smoking Study (P01 CA098262, PI: Robin J. Mermelstein, University of Illinois at Chicago).

RECOMMENDED READING

- *SAS/STAT® 14.1 User's Guide the NLMIXED Procedure*
- *SAS/STAT® 13.1 User's Guide the MCMC Procedure*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Donald Hedeker
Department of Public Health Sciences
The University of Chicago
773-702-0566
hedeker@uchicago.edu