Efficient estimation of variance components in nonparametric mixed-effects models with large samples

Nathaniel E. Helwig

Assistant Professor of Psychology and Statistics
University of Minnesota (Twin Cities)

Modern Modeling Methods Talk
May 25, 2016

Outline of Talk

1) Regression Background
- Parametric Regression
- Nonparametric Regression

2) Mixed-Effects Regression
- Linear Mixed-Effects Model
- NP Mixed-Effects Model

3) Simulation Study
- Design & Analyses
- Results

4) Educational Exam Example
- Data & Analyses
- Results

5) Procedural Learning Example
- Data & Analyses
- Results

6) Facial Expression Example
- Data & Analyses
- Results

Helwig, N.E. (2015). *Statistics and Computing* [Advance online publication].

`http://link.springer.com/article/10.1007%2Fs11222-015-9610-5`

Regression Background

Regression Reminder

A regression model relates the response variable Y to the independent variables $X_1, \ldots, X_p$ using some systematic rule.

For example, the general linear model (GLM) assumes that

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \tag{1}$$

for $i = 1, \ldots, n$ where

- y_i is the observed response variable for the i-th subject
- x_{ij} is the j-th observed predictor for the i-th subject
- $(\beta_0, \beta_1, \ldots, \beta_p)$ are the unknown regression coefficients
- $\epsilon_i \stackrel{\text{iid}}{\sim} (0, \sigma^2)$ is independent, identically distributed measurement error

Implies that $(y_i|\mathbf{x}_i) \stackrel{\text{ind}}{\sim} (\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}, \sigma^2)$

Multiple Linear Regression Model in Matrix Form

The multiple linear regression model has the form

$$\mathbf{y} = \mathbf{1}_n \beta_0 + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2)$$

where

- $\mathbf{y} = (y_1, \ldots, y_n)' \in \mathbb{R}^n$ is the $n \times 1$ response vector
- $\mathbf{1}_n = \{1\}_{n\times 1}$ is an $n \times 1$ vector of ones
- $\mathbf{X} = \{x_{ij}\}_{n\times p} \in \mathbb{R}^{n\times p}$ is the $n \times p$ design matrix
- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^p$ is the $p \times 1$ vector of slope coefficients
- $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)' \sim (\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$ is the $n \times 1$ error vector

Implies that $(\mathbf{y}|\mathbf{X}) \sim (\mathbf{1}_n \beta_0 + \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$

Ordinary Least Squares Problem

The ordinary least squares (OLS) function is

$$\begin{aligned} \mathrm{OLS}(\beta_0, \boldsymbol{\beta}) &= \sum_{i=1}^{n} \left(y_i - \beta_0 - \textstyle\sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \\ &= \|\mathbf{y} - \mathbf{1}_n \beta_0 - \mathbf{X}\boldsymbol{\beta}\|^2 \end{aligned} \tag{3}$$

where $\|\mathbf{w}\| = \{\sum_{i=1}^{n} w_i^2\}^{1/2}$ denotes the Euclidean norm.

The coefficients that minimize the OLS function have the form

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \textstyle\sum_{j=1}^{p} \hat{\beta}_j \bar{x}_j \\ \hat{\boldsymbol{\beta}} &= (\mathbf{X}_c' \mathbf{X}_c)^{-1} \mathbf{X}_c' \mathbf{y}_c \end{aligned} \tag{4}$$

where $\mathbf{X}_c = (\mathbf{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}_n')\mathbf{X}$ and $\mathbf{y}_c = (\mathbf{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}_n')\mathbf{y}$

Parametric versus Nonparametric Regression

The GLM is a form of parametric regression, where the relationship between Y and $X_1, \ldots, X_p$ has some predetermined form.

- Parameterizes relationship between Y and $X_1, \ldots, X_p$
- Then estimates the specified parameters
- Great if you know the form of the relationship (e.g., linear)

In contrast, nonparametric regression tries to estimate the form of the relationship between Y and $X_1, \ldots, X_p$.

- No predetermined form for relationship between Y and $X_1, \ldots, X_p$
- Great for discovering relationships and building prediction models

Need for NP Regression: Anscombe's (1973) Quartet

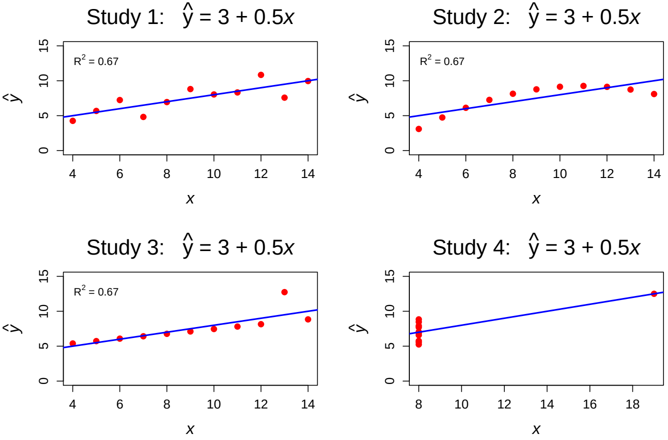

Figure 1: Estimated linear relationship with corresponding data.

Smoothing Spline ANOVA

A smoothing spline analysis of variance (SSANOVA) model is a statistical framework for estimating functional relationships.

- See Gu (2013), Gu and Wahba (1991), Helwig and Ma (2015), Kim and Gu (2004), Ma et al. (2015), Wahba (1990)

An SSANOVA model has the form

$$y_i = \eta(x_{i1}, \ldots, x_{ip}) + \epsilon_i \tag{5}$$

for $i = 1, \ldots, n$ where

- y_i is the observed response variable for the i-th subject
- x_{ij} is the j-th observed predictor for the i-th subject
- η is an unknown function relating Y and $X_1, \ldots, X_p$
- $\epsilon_i \overset{\text{iid}}{\sim} (0, \sigma^2)$ is independent, identically distributed measurement error

Penalized Least Squares Problem

To estimate η, we minimize the penalized least squares functional

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \eta(\mathbf{x}_i))^2 + \lambda J_{\boldsymbol{\theta}}(\eta) \tag{6}$$

where

- $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$ is the predictor vector
- $J_{\boldsymbol{\theta}}(\cdot) = \sum_{k=1}^{s} \theta_s J_s(\cdot)$ is a non-negative penalty functional
- λ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)'$ are non-negative smoothing parameters[1]

Generalized Cross-Validation criterion (Craven & Wahba, 1979) can be used to estimate the smoothing parameters.

[1] Global smoothing parameter λ controls overall influence of smoothness penalty, whereas local smoothing parameters in $\boldsymbol{\theta}$ control influence of marginal penalties.

SSANOVA Model Building

With $p > 1$ predictors, we need to specify the form of the function.

For example, with $p = 2$ predictors we have two possible models

$$\begin{aligned} \text{additive}: &\quad \eta(x_1, x_2) = \eta_0 + \eta_1(x_1) + \eta_2(x_2) \\ \text{interactive}: &\quad \eta(x_1, x_2) = \eta_0 + \eta_1(x_1) + \eta_2(x_2) + \eta_{12}(x_1, x_2) \end{aligned}$$

where

- η_0 is an analog of β_0 (unknown regression intercept)
- η_1 is the unknown main effect function for the first predictor
- η_2 is the unknown main effect function for the second predictor
- η_{12} is the unknown interaction effect function

Fitted Values and Smoothing Matrix

Given $\boldsymbol{\lambda} \equiv \{\lambda_k\}_{s\times 1}$ with $\lambda_k \equiv \lambda/\theta_k$, the fitted values are given by

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{K}\hat{\mathbf{d}} + \mathbf{J}_{\boldsymbol{\theta}}\hat{\mathbf{c}} \\ &= \mathbf{S}_{\boldsymbol{\lambda}}\mathbf{y}\end{aligned} \tag{7}$$

where

- $\mathbf{K}$ is $n \times m$ null space basis function matrix
- $\mathbf{J}_{\boldsymbol{\theta}}$ is $n \times t$ contrast space basis function matrix
- $\hat{\mathbf{d}}$ and $\hat{\mathbf{c}}$ are estimated basis function coefficients

Note that $\mathbf{S}_{\boldsymbol{\lambda}}$ is the smoothing matrix, which

- is an $n \times n$ symmetric matrix
- depends on the chosen smoothing parameters in $\boldsymbol{\lambda}$

Bayesian Interpretation of Smoothing Spline

Let $\eta = \eta_{\text{n}} + \eta_{\text{c}}$ denote the null and contrast space functions and assume the following prior distributions:

- η_{n} has a diffuse (vague) prior with mean zero
- η_{c} is a zero mean Gaussian process with covariance function proportional to ρ_{c} (reproducing kernel of contrast space)

Using the above prior assumptions and assuming $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$

- $\hat{\eta}$ can be interpreted as posterior mean of η given data $\mathbf{y}$
- we can derive posterior variance $Var(\eta|\mathbf{y})$

Bayesian Confidence Intervals

Using the Bayesian interpretation, we can form confidence intervals

$$\hat{\eta}(\mathbf{x}) \pm Z_{\alpha/2}\sqrt{Var(\eta|\mathbf{y})} \tag{8}$$

where $Z_{\alpha/2}$ is critical value from standard normal distribution.

Bayesian CIs have approximate "across-the-function coverage" when the smoothing parameters are selected according to GCV.

- On average contain $100(1-\alpha)\%$ of true function realizations
- See Wahba (1983), Nychka (1988), and Gu & Wahba (1993)

Mixed-Effects Regression

Introduction to Mixed-Effects Regression

A typical linear mixed-effects (LME) model can be written as

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon} \tag{9}$$

where

- $\mathbf{y} = (y_1, \ldots, y_n)' \in \mathbb{R}^n$ is the response vector
- $\mathbf{X} = \{x_{ij}\}_{n \times p} \in \mathcal{X}$ is the known fixed-effects design matrix
- $\mathbf{a} \in \mathbb{R}^p$ are the unknown fixed-effects coefficients
- $\mathbf{Z} = \{z_{ij}\}_{n \times q} \in \mathcal{Z}$ is the known random-effects design matrix
- $\mathbf{b} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma})$ are the unknown Gaussian random effects
- $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Psi})$ is unknown Gaussian error

Variance Components Model

A variance components (VC) model (Hartley & Rao, 1967; Harville, 1977) assumes that $\mathbf{b}_k \overset{\text{ind}}{\sim} \mathrm{N}(\mathbf{0}, \sigma^2 \tau_k \mathbf{I}_{q_k})$ with $\sum_{k=1}^{r} q_k = q$.

- $\mathbf{b} = (\mathbf{b}_1', \ldots, \mathbf{b}_r')'$ is the partitioned random effects vector
- $\boldsymbol{\Sigma} = \mathrm{bdiag}(\tau_k \mathbf{I}_{q_k})$ is a block diagonal matrix

For VC models the covariance matrix of $\mathbf{y}$ has the form

$$V(\mathbf{y}|\mathbf{a}) = \sigma^2 (\textstyle\sum_{k=1}^{r} \tau_k \mathbf{Z}_k \mathbf{Z}_k' + \boldsymbol{\Psi}) \tag{10}$$

where $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_r)$ can be partitioned according to the random effects.

Nonparametric Extensions of the VC Model

In this talk, we consider a nonparametric extension of the VC model

$$\mathbf{y} = \boldsymbol{\eta} + \mathbf{Zb} + \boldsymbol{\epsilon} \tag{11}$$

where $\boldsymbol{\eta} = \{\eta(\mathbf{x}_i)\}_{n \times 1}$ is an unknown smooth function of $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$.

- $\boldsymbol{\eta} = \mathbf{Kd} + \mathbf{J}_{\boldsymbol{\theta}}\mathbf{c}$
- $\mathbf{K} = \{\phi_v(\mathbf{x}_i)\}_{n \times m}$ where $\{\phi_v\}_{v=1}^m$ span the null space
- $\mathbf{J}_{\boldsymbol{\theta}} = \sum_{k=1}^s \theta_k \mathbf{J}_k$ where $\mathbf{J}_k = \{\rho_k^*(\mathbf{x}_i, \breve{\mathbf{x}}_h)\}_{n \times t}$ with ρ_k^* denoting a reproducing kernel function and $\{\breve{\mathbf{x}}_h\}_{h=1}^t \subset \{\mathbf{x}_i\}_{i=1}^n$ denoting the knots
- $\mathbf{d} = \{d_v\}_{m \times 1}$ and $\mathbf{c} = \{c_h\}_{t \times 1}$ are the unknown function coefficients

The goal is to estimate the unknown function η, as well as the unknown variance components $(\sigma^2, \boldsymbol{\tau})$ where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_r)$.

PGML Estimation of NPME Model

Wang (1998a,b) and Zhang et al. (1998) use the Penalized Generalized Maximum Likelihood (PGML) criterion

$$(1/n)\|\boldsymbol{\Sigma}_*^{-1/2}(\mathbf{y} - \mathbf{K}\mathbf{d} - \mathbf{J}_{\boldsymbol{\theta}}\mathbf{c})\|^2 + \lambda \mathbf{c}'\mathbf{Q}_{\boldsymbol{\theta}}\mathbf{c} \tag{12}$$

where

- $\boldsymbol{\Sigma}_* = \boldsymbol{\Psi} + \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'$ is proportional to the marginal covariance matrix of $\mathbf{y}$
- $\mathbf{Q}_{\boldsymbol{\theta}} = \sum_{k=1}^{s} \theta_k \mathbf{Q}_k$ where $\mathbf{Q}_k = \{\rho_k^*(\breve{\mathbf{x}}_g, \breve{\mathbf{x}}_h)\}_{t \times t}$ is the penalty matrix
- $\lambda \geq 0$ is the smoothing parameter

Smoothing and variance parameters estimated from the marginal distribution of $\mathbf{z} = \mathbf{F}_2'\mathbf{y}$ where $\mathbf{K} = \mathbf{F}_1\mathbf{R}$ is the QR-decomposition of $\mathbf{K}$ (and $\mathbf{F}_1'\mathbf{F}_2 = \mathbf{0}$).

- Treats the smoothing parameters as variance parameters

GCV Estimation of NPME Model

Assuming that $\boldsymbol{\Psi} = \mathbf{I}_n$, Gu and Ma (2005) propose minimizing

$$(1/n)\|\mathbf{y} - \mathbf{K}\mathbf{d} - \mathbf{J}_{\boldsymbol{\theta}}\mathbf{c} - \mathbf{Z}\mathbf{b}\|^2 + (1/n)\mathbf{b}'\tilde{\boldsymbol{\Sigma}}\mathbf{b} + \lambda\mathbf{c}'\mathbf{Q}_{\boldsymbol{\theta}}\mathbf{c} \tag{13}$$

where $\tilde{\boldsymbol{\Sigma}}$ are tuning parameters corresponding to the random effects.

- Defining $\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{-1}$ the first two terms in Equation (13) are proportional to the negative of the joint log-likelihood of $(\mathbf{y}, \mathbf{b})$.

They treat variance components as "mean components" and use $\boldsymbol{\theta} = \{\theta_k\}_{k=1}^{s}$ and $\tilde{\boldsymbol{\Sigma}}$ as tuning parameters to estimate η and $\mathbf{b}$, respectively.

- Treats the variance parameters as smoothing parameters
- Generalized Cross-Validation (GCV) criterion (Craven & Wahba, 1979) provides asymptotically optimal tuning of parameters.

Limitations of Current NPME Estimation Methods

PGML (REML) treats the smoothing parameters as variance components.

GCV treats the variance components as smoothing parameters.

For multiple predictors and/or variance components, estimation becomes cumbersome because need to optimize criterion w.r.t. several parameters.

Two-stage estimation idea from Helwig (2015)

1. REML estimation to get various components
2. GCV estimation to get smoothing parameters

Stage 1: REML Estimation of Variance Components

Initialize $\theta_k = 1 \ \forall k$ and define the fixed-effects design matrix $\mathbf{X}_{\boldsymbol{\theta}} = (\mathbf{K}, \mathbf{J}_{\boldsymbol{\theta}})$.

- See Helwig (2015) for justification of initialization

REML estimation of the variance components via the error contrasts $\tilde{\mathbf{z}} = \tilde{\mathbf{F}}_2'\mathbf{y}$, where $\mathbf{X}_{\boldsymbol{\theta}} = \tilde{\mathbf{F}}_1\tilde{\mathbf{R}}$ is the QR-decomposition of $\mathbf{X}_{\boldsymbol{\theta}}$ (and $\tilde{\mathbf{F}}_1'\tilde{\mathbf{F}}_2 = \mathbf{0}$).

- Variance component estimates do not depend on unknown θ_k estimates

Use either Fisher Scoring or Expectation Maximization (EM) algorithm to maximize REML log-likelihood (see Helwig, 2015).

- Algorithms are scalable as long as $q \ll n$.

Stage 2: Penalized Generalized Least Squares

Given the estimated variance components $\hat{\sigma}^2$ and $\hat{\boldsymbol{\tau}}$, we propose estimating η by minimizing the penalized generalized least squares (PGLS) function

$$(1/n)\|\hat{\boldsymbol{\Sigma}}_*^{-1/2}(\mathbf{y} - \mathbf{K}\mathbf{d} - \mathbf{J}_{\boldsymbol{\theta}}\mathbf{c})\|^2 + \lambda \mathbf{c}'\mathbf{Q}_{\boldsymbol{\theta}}\mathbf{c} \tag{14}$$

where $\hat{\boldsymbol{\Sigma}}_* = \boldsymbol{\Psi} + \mathbf{Z}\hat{\boldsymbol{\Sigma}}\mathbf{Z}'$ is proportional to (estimated) covariance matrix of $\mathbf{y}$.

Note that Equation (14) is Equation (12) with $\hat{\boldsymbol{\Sigma}}_*$ in place of $\boldsymbol{\Sigma}_*$.

We propose estimating the smoothing parameters by minimizing the GCV score using the transformed data $\tilde{\mathbf{y}} = \hat{\boldsymbol{\Sigma}}_*^{-1/2}\mathbf{y}$.

Simulation Study

Design of Simulation B from Helwig (2015)

Assume that $x_{i1} \in [0,1]$, $x_{i2} \in \{1,2\}$, and

$$y_i = \eta(x_{i1}, x_{i2}) + b_{s_i x_{i2}} + \epsilon_i$$

where

- $\eta(x_1, x_2) = 2\sin(2\pi x_1) + I_{\{x_2=2\}} \sin(4\pi x_1)$
- $s_i \in \{1, \ldots, q\}$ is the subject indicator
- $\mathbf{b}_k = (b_{1k}, \ldots, b_{qk}) \sim \mathrm{N}(\mathbf{0}, \tau_k \mathbf{I}_q)$ for $k \in \{1,2\}$ and $\epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$

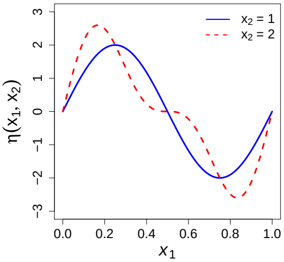

Figure 2: Simulation functions.

4 levels of number of subjects per group: $q \in \{10, 25, 50, 100\}$
4 levels of variance component: $\tau_1 \in \{0.09, 0.25, 0.5, 1\}$ (with $\tau_2 = 1$)
Assume 100 replications from each subject $\rightarrow n = 200q$

Analyses in Simulation B from Helwig (2015)

We compare three methods:

- REML (PGML) estimation using Wood's (2004) GAM approach via the `mgcv` R package (Wood, 2015)
- GCV estimation using Gu and Ma's (2005) approach via the `gss` R package (Gu, 2014)
- Two-stage (REML + PGLS) using Helwig's (2015) approach via the `bigsplines` R package (Helwig, 2016)

Used True Mean Squared Error (TMSE) with the random effects included

$$\text{TMSE} = \frac{1}{n} \sum_{i=1}^{n} [\eta(x_{i1}, x_{i2}) + b_{s_i x_{i2}} - \hat{\eta}(x_{i1}, x_{i2}) - \hat{b}_{s_i x_{i2}}]^2 \quad (15)$$

to assess the quality of each method.

Results of Simulation B from Helwig (2015): part 1

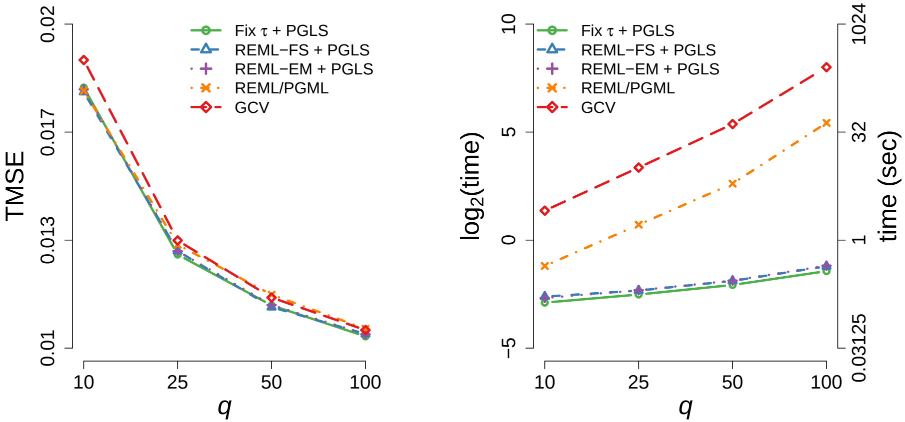

Figure 3: Median TMSE and runtime for each method with $\tau_1 = 0.09$ and $\tau_2 = 1$.

Results of Simulation B from Helwig (2015): part 2

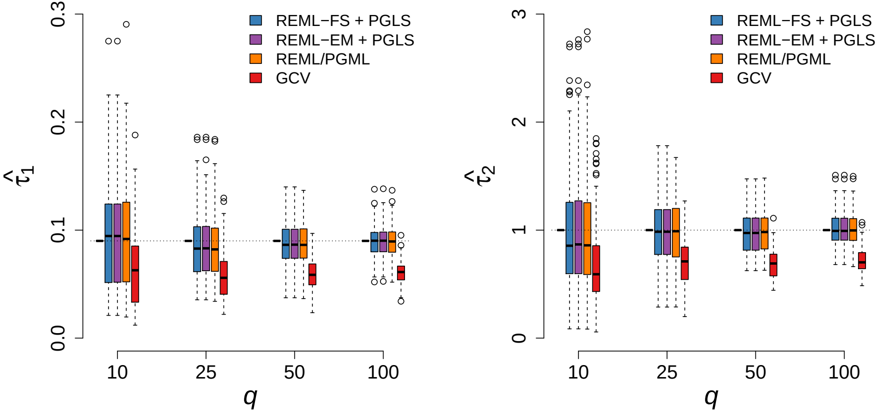

Figure 4: Estimated variance components for each method with $\tau_1 = 0.09$ and $\tau_2 = 1$.

Educational Achievement Example

Overview of Education Data from Paterson (1991)

Exam scores for $n = 3435$ students who attended 19 secondary schools that are cross classified with 148 primary schools (see Goldstein, 2011, Chap. 12).

We fit a NPME model of the form

$$y_i = \eta(x_i) + b_{p_i 1} + b_{s_i 2} + \epsilon_i$$

for $i \in \{1, \ldots, 3435\}$ where

- y_i is the secondary school leaving exam score
- x_i is a verbal reasoning exam score obtained in primary school
- $p_i \in \{1, \ldots, 148\}$ is the primary school indicator
- $s_i \in \{1, \ldots, 19\}$ is the secondary school indicator
- $b_{p1} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2 \tau_p)$, $b_{s2} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2 \tau_s)$, and $\epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$

Estimated Functional Relationship

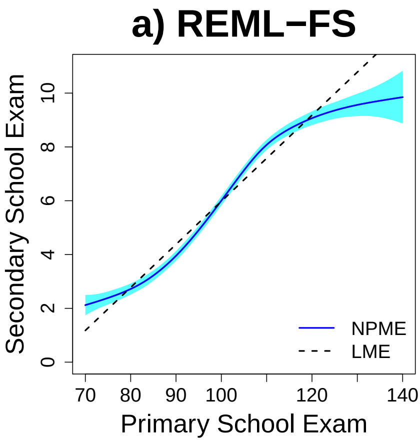

Figure 5: Predicted relationship between primary and secondary school exam.

Estimated Variance Components

Table 1: Variance parameter estimates for educational data.

Model	$\hat{\tau}_p$	$\hat{\tau}_s$	$\hat{\sigma}$
REML-FS + PGLS	0.0663	0.0054	2.0175
REML-EM + PGLS	0.0737	0.0098	2.0143
LMER	0.0646	0.0034	2.0627

Note. $\hat{\tau}_p$ and $\hat{\tau}_s$ are Primary and Secondary (respectively).

$\hat{\tau}_s < \hat{\tau}_p \Longrightarrow$ more variation due to primary school effect

$\hat{\tau}_s < \hat{\tau}_p \ll 1 \Longrightarrow$ more variation within the schools than between the schools

Procedural Learning Example

Response Time Data from Cudeck (1996)

Response times from $n = 214$ Air Force recruits who participated in a procedural learning experiment consisting of 12 trials.[2]

We fit a NPME model of the form

$$\ln(y_{ij}) = \eta(j) + b_{i0} + b_{i1}j + \epsilon_{ij}$$

for $i \in \{1, \ldots, 214\}$ and $j \in \{0, \ldots, 11\}$ where

- y_{ij} is the response time for the j-th trial of the i-th subject
- $b_{i0} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2\tau_0)$ are random intercepts
- $b_{i1} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2\tau_1)$ are random slopes
- $\epsilon_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$

[2] Thanks to Nidhi Kohli (UMN) and Jeffrey Harring (UMD) for providing the data.

Estimated Functional Relationship

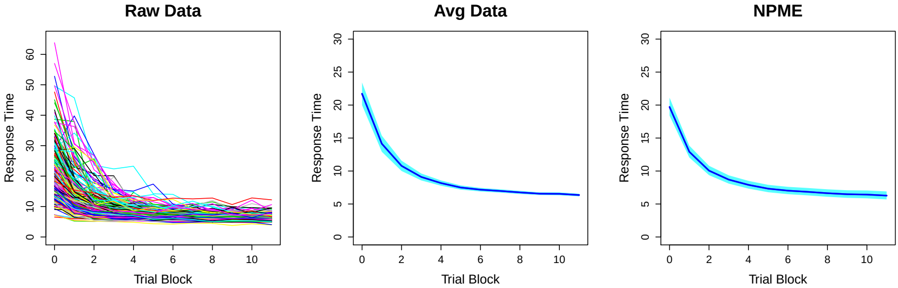

Figure 6: Data for all $n = 214$ subjects and $j = 12$ trials (left), average data across subjects for each trial (middle), and predicted mean for nonparametric mixed-effects model (right).

Subject Specific Predictions

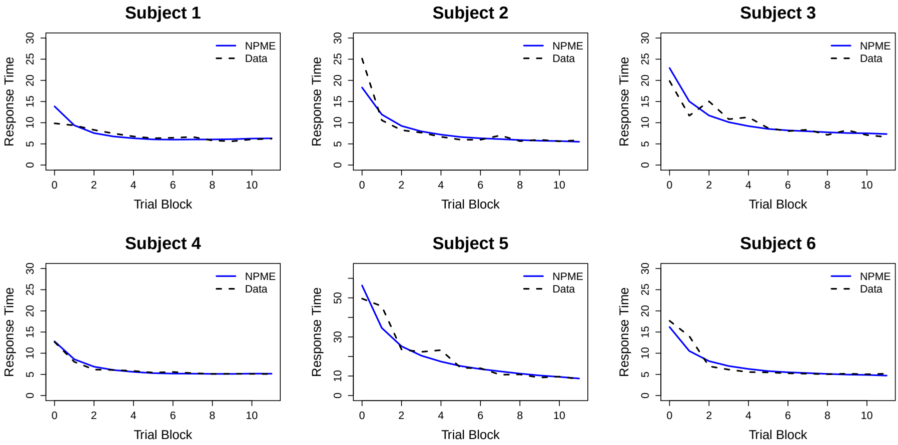

Figure 7: NPME model predictions for six subjects.

Facial Expression Example

Smile Effectiveness Data

Data from $n = 788$ participants collected at the 2015 Minnesota State Fair.[3]

We fit a NPME model of the form

$$y_{ij} = \eta(a_{ij}, e_{ij}, d_{ij}) + b_i + \epsilon_{ij}$$

for $i \in \{1, \ldots, 788\}$ and $j \in \{1, \ldots, 15\}$ where

- y_{ij} is the rating of j-th stimuli displayed to i-th subject
- a_{ij} is the angle of j-th stimuli displayed to i-th subject
- e_{ij} is the extent of j-th stimuli displayed to i-th subject
- d_{ij} is the dental show of j-th stimuli displayed to i-th subject
- $b_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2\tau)$ are random intercepts and $\epsilon_{ij} \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$

[3] Joint work with Sofía Lyford-Pike (MED UMN) and Stephen J. Guy (CSE UMN)

Visualization of Smile Animations

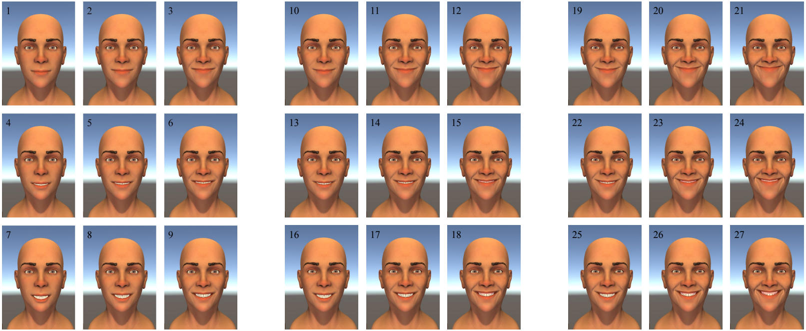

Figure 8: The 27 different smiling faces used in the last frame of the animation. Faces 1–9 are low angle, faces 10–18 are medium angle, and faces 19–27 are high angle. Within each block of nine, the columns represent extent (low, med, high) and the rows represent dental show (low, med, high). The numbers were included post hoc for labeling purposes.

NPME Model Predictions

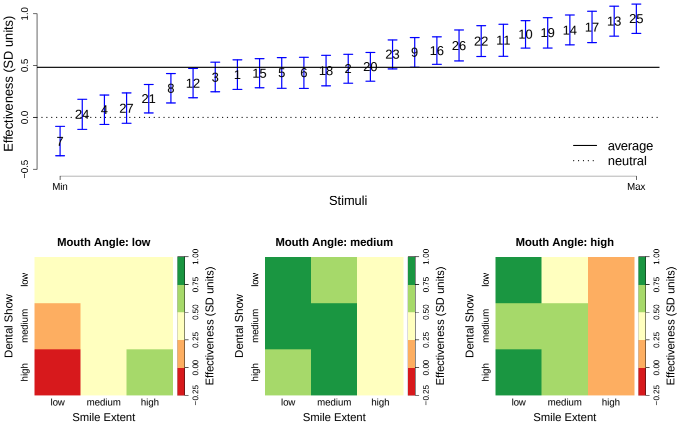

Figure 9: Smile effectiveness ratings for the 27 stimuli.

Conclusions and Future Directions

NPME models offer a flexible framework for data analysis and prediction.

- Applicable to any LMER situation
- Requires fewer assumptions than LMER

Now possible to efficiently fit VC models to large samples.

- No need for super-computer or lots of patience
- Algorithms applicable to LMER too

Future work will include extensions for…

- Correlated random effects
- Non-Gaussian response variables

References I

Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician 21*, 17–21.

Craven, P. and G. Wahba (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik 31*, 377–403.

Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate Behavioral Research 31*, 371–403.

Goldstein, H. (2011). *Multilevel Statistical Models* (Fourth ed.). West Sussex: Wiley.

Gu, C. (2013). *Smoothing Spline ANOVA Models* (Second ed.). New York: Springer-Verlag.

Gu, C. (2014). Smoothing spline anova models: R package gss. *Journal of Statistical Software 58*(5), 1–25.

Gu, C. and P. Ma (2005). Optimal smoothing in nonparametric mixed-effect models. *The Annals of Statistics 33*, 1357–1379.

Gu, C. and G. Wahba (1991). Minimizing GCV/GML scores with multiple smoothing parameters via the newton method. *SIAM Journal on Scientific and Statistical Computing 12*, 383–398.

Gu, C. and G. Wahba (1993). Smoothing spline ANOVA with component-wise Bayesian "confidence intervals". *Journal of Computational and Graphical Statistics 2*, 97–117.

Hartley, H. O. and J. N. K. Rao (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika 54*, 93–108.

Harville, D. A. (1977). Maximum-likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association 72*, 320–340.

Helwig, N. E. (2015). Efficient estimation of variance components in nonparametric mixed-effects models with large samples. *Statistics and Computing [Advance online publication]*, 1–18.

Helwig, N. E. (2016). *bigsplines: Smoothing Splines for Large Samples*. R package version 1.0-8.

Helwig, N. E. and P. Ma (2015). Fast and stable multiple smoothing parameter selection in smoothing spline analysis of variance models with large samples. *Journal of Computational and Graphical Statistics 24*, 715–732.

Kim, Y.-J. and C. Gu (2004). Smoothing spline gaussian regression: More scalable computation via efficient approximation. *Journal of the Royal Statistical Society, Series B 66*, 337–356.

References II

Ma, P., J. Huang, and N. Zhang (2015). Efficient computation of smoothing splines via adaptive basis sampling. *Biometrika 102*, 631–645.

Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association 83*, 1134–1143.

Paterson, L. (1991). Socio-economic status and educational attainment: A multidimensional and multilevel study. *Evaluation and Research in Education 5*, 97–121.

Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *Journal of the Royal Statistical Society, Series B 45*, 133–150.

Wahba, G. (1990). *Spline models for observational data*. Philadelphia: Society for Industrial and Applied Mathematics.

Wang, Y. (1998a). Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society, Series B 60*, 159–174.

Wang, Y. (1998b). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association 93*, 341–348.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association 99*, 673–686.

Wood, S. N. (2015). *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation and GAMMs by REML/PQL*. R package version 1.8-7.

Zhang, D., X. Lin, J. Raz, and M. Sowers (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association 93*, 710–719.