Three-Level Mixed-Effects Location Scale Model
With Modeling Random Scale Variance

BY

HONG LI
BS, Jiangnan University, Wuxi, China, 1998
MS, Chongqing University, Chongqing, China, 2004
MS, University of Illinois at Chicago, Chicago, 2009

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public Health Sciences
in the Graduate College of the
University of Illinois at Chicago, 2014

Chicago, Illinois

Defense Committee:

Li Liu, Chair
Donald Hedeker, Advisor, University of Chicago
Robin Mermelstein, Institute for Health Research and Policy
Sally Freels
Hui Xie
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>INTRODUCTION .................................................................</td>
</tr>
<tr>
<td>II.</td>
<td>LITERATURE REVIEW ..........................................................</td>
</tr>
<tr>
<td>A.</td>
<td>Two-level Mixed-Effects Regression Models ................................</td>
</tr>
<tr>
<td>B.</td>
<td>Two-Level Mixed-Effects Models with Multiple Random Effects ..........</td>
</tr>
<tr>
<td>C.</td>
<td>Heterogeneous Variance Modeling...........................................</td>
</tr>
<tr>
<td>D.</td>
<td>Three-Level Mixed-Effects Random Location Scale Models ..................</td>
</tr>
<tr>
<td>E.</td>
<td>Ecological Momentary Assessment.......................................</td>
</tr>
<tr>
<td>F.</td>
<td>The Proposed Model.........................................................</td>
</tr>
<tr>
<td>G.</td>
<td>Model Specification.......................................................</td>
</tr>
<tr>
<td>H.</td>
<td>Marginal Distribution of Outcome Variable................................</td>
</tr>
<tr>
<td>I.</td>
<td>Random Effects in Standardized Form ....................................</td>
</tr>
<tr>
<td>J.</td>
<td>Maximum Marginal Likelihood Estimation and Empirical Bayes Method ...</td>
</tr>
<tr>
<td>K.</td>
<td>First Derivative of Log Likelihood....................................</td>
</tr>
<tr>
<td>L.</td>
<td>Second Derivative of Log Likelihood..................................</td>
</tr>
<tr>
<td>M.</td>
<td>Numerical Integration.....................................................</td>
</tr>
<tr>
<td>III.</td>
<td>SIMULATION RESULTS ..........................................................</td>
</tr>
<tr>
<td>A.</td>
<td>Overview.....................................................................</td>
</tr>
<tr>
<td>B.</td>
<td>Evaluation Criteria.......................................................</td>
</tr>
<tr>
<td>C.</td>
<td>Simulation Results.........................................................</td>
</tr>
<tr>
<td>D.</td>
<td>Results of Three-level Random Intercept Model .........................</td>
</tr>
<tr>
<td>E.</td>
<td>Results of Three-level Random Location Scale Model ..................</td>
</tr>
<tr>
<td>F.</td>
<td>Results of Random Location Scale Model with Modeling Random Scale Variance</td>
</tr>
<tr>
<td>IV.</td>
<td>DATA ANALYSIS ..................................................................</td>
</tr>
<tr>
<td>A.</td>
<td>Research Interest of Adolescent Smoking Study ..........................</td>
</tr>
<tr>
<td>B.</td>
<td>Description of Study...........................................................</td>
</tr>
<tr>
<td>C.</td>
<td>Outcome Measures: Positive Affect and Negative Affect ...............</td>
</tr>
<tr>
<td>D.</td>
<td>Covariate Measure............................................................</td>
</tr>
<tr>
<td>E.</td>
<td>Model Setup..................................................................</td>
</tr>
<tr>
<td>F.</td>
<td>Comparison Results............................................................</td>
</tr>
<tr>
<td>G.</td>
<td>Analysis Results...............................................................</td>
</tr>
<tr>
<td>H.</td>
<td>Example SAS Codes............................................................</td>
</tr>
<tr>
<td>V.</td>
<td>CONCLUSIONS ..................................................................</td>
</tr>
<tr>
<td>CITED LITERATURE ..................................................................</td>
<td>97</td>
</tr>
<tr>
<td>VITA ..................................................................................</td>
<td>101</td>
</tr>
</tbody>
</table>
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. THE INITIALIZED VALUES IN SIMULATION PROCESS</td>
<td>53</td>
</tr>
<tr>
<td>II. 100 SIMULATION RESULTS UNDER 100 SIMULATED RL-RSS DATASETS</td>
<td>61</td>
</tr>
<tr>
<td>III. DESCRIPTIVE ANALYSIS OF OUTCOME MEASUREMENTS AND COVARIATES</td>
<td>86</td>
</tr>
<tr>
<td>IV. THREE-LEVEL RANDOM EFFECT LOCATION MODEL (RL MODEL) OF POSITIVE AFFECT AND NEGATIVE AFFECT</td>
<td>87</td>
</tr>
<tr>
<td>V. THREE-LEVEL RANDOM LOCATION RANDOM SCALE MODEL (RL-RS MODEL) OF POSITIVE AFFECT AND NEGATIVE AFFECT</td>
<td>88</td>
</tr>
<tr>
<td>VI. THREE-LEVEL RANDOM LOCATION RANDOM SCALE MODEL WITH MODELING RANDOM SCALE EFFECT (RL-RSS MODEL) OF POSITIVE AFFECT AND NEGATIVE AFFECT</td>
<td>89</td>
</tr>
<tr>
<td>VII. PARAMETERS AND VARIABLES IN EXAMPLE EMA DATA (RUN IN PROC NLMIXED SAS PROCEDURE)</td>
<td>93</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The box-and-whisker diagram of $\beta_0$</td>
<td>62</td>
</tr>
<tr>
<td>2. The box-and-whisker diagram of $\beta_1$</td>
<td>62</td>
</tr>
<tr>
<td>3. The box-and-whisker diagram of $\beta_2$</td>
<td>63</td>
</tr>
<tr>
<td>4. The box-and-whisker diagram of $\beta_3$</td>
<td>63</td>
</tr>
<tr>
<td>5. The box-and-whisker diagram of $\tau_0$</td>
<td>64</td>
</tr>
<tr>
<td>6. The box-and-whisker diagram of $\tau_1$</td>
<td>64</td>
</tr>
<tr>
<td>7. The box-and-whisker diagram of $\tau_2$</td>
<td>65</td>
</tr>
<tr>
<td>8. The box-and-whisker diagram of $\alpha_0$</td>
<td>65</td>
</tr>
<tr>
<td>9. The box-and-whisker diagram of $\alpha_1$</td>
<td>66</td>
</tr>
<tr>
<td>10. The box-and-whisker diagram of $\delta_0^2$</td>
<td>66</td>
</tr>
<tr>
<td>11. The box-and-whisker diagram of $\delta_1^2$</td>
<td>67</td>
</tr>
<tr>
<td>12. The box-and-whisker diagram of $\delta_{vw}$</td>
<td>67</td>
</tr>
<tr>
<td>13. Illustration of random intercept model</td>
<td>79</td>
</tr>
<tr>
<td>14. Illustration of random location scale model</td>
<td>80</td>
</tr>
<tr>
<td>15. Illustration of random location scale model with modeling random scale variance</td>
<td>81</td>
</tr>
</tbody>
</table>
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AW</td>
<td>Average Width</td>
</tr>
<tr>
<td>BS</td>
<td>Between-Subject</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>EB</td>
<td>Empirical Bayes</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation and Maximization</td>
</tr>
<tr>
<td>EMA</td>
<td>Ecological Momentary Assessment</td>
</tr>
<tr>
<td>ICC</td>
<td>Intraclass Correlation Coefficient</td>
</tr>
<tr>
<td>IG</td>
<td>Inverse Gamma</td>
</tr>
<tr>
<td>IID</td>
<td>Independent and Identically Distributed</td>
</tr>
<tr>
<td>LN</td>
<td>Log-Normal</td>
</tr>
<tr>
<td>MCAR</td>
<td>Missing Completely At Random</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td>MLE</td>
<td>Maximum Likelihood Estimate</td>
</tr>
<tr>
<td>MML</td>
<td>Maximum Marginal Likelihood</td>
</tr>
<tr>
<td>MRM</td>
<td>Mixed Random Effects Model</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean squared error</td>
</tr>
<tr>
<td>NA</td>
<td>Negative Affect</td>
</tr>
<tr>
<td>ΔNA</td>
<td>The Change of Negative Affect</td>
</tr>
<tr>
<td>NDSS</td>
<td>Nicotine Dependence Syndrome Scale</td>
</tr>
<tr>
<td>PA</td>
<td>Positive Affect</td>
</tr>
<tr>
<td>ΔPA</td>
<td>The Change of Positive Affect</td>
</tr>
<tr>
<td>RL</td>
<td>Random Intercept</td>
</tr>
<tr>
<td>RL-FS</td>
<td>Random Location Fixed-Scale</td>
</tr>
<tr>
<td>RL-RS</td>
<td>Random Location Random Scale</td>
</tr>
<tr>
<td>RL-RSS</td>
<td>Random Location Scale With Modeling Random Scale Variance</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>SE</td>
<td>Standard Error</td>
</tr>
<tr>
<td>SB</td>
<td>Standardized Bias</td>
</tr>
<tr>
<td>WLS</td>
<td>Weighted Least Squares</td>
</tr>
<tr>
<td>WS</td>
<td>Within-Subject</td>
</tr>
<tr>
<td>WS-BW</td>
<td>Within-Subject Between-Wave</td>
</tr>
<tr>
<td>WS-WW</td>
<td>Within-Subject Within-Wave</td>
</tr>
</tbody>
</table>
SUMMARY

In this dissertation, we propose a three-level mixed-effects random location scale model with modeling random scale variance (RL-RSS model). This model allows covariates to influence both error variance and random scale variance through a log-linear representation. The error variance varies across subjects through a subject-level normally distributed random scale effect, above and beyond the contribution of covariates on error variance. The subject-level random scale effect and random location effect are allowed to correlate with each other. Parameter estimation was based on the combination of maximum marginal likelihood (MML) method and Empirical Bayes (EB) method. An iterative Newton-Raphson solution was used to maximize the log likelihood, and multi-dimensional Gauss-Hermite quadrature is used to numerically approximate integral values. An SAS program via PROC NLMIXED using adaptive quadrature was developed to fit the proposed model.

The data from Ecological Momentary Assessment (EMA) Adolescent Smoking Study are used to illustrate the application of the proposed model. In this study, a three-level clustering data structure, level-1 smoking events/occasions nested within level-2 waves nested within level-3 subjects, was used in the data analysis. The proposed RL-RSS model was fit to the data.

Simulation process was carried out to validate the accuracy and reliability of the proposed three-level RL-RSS model. The simulation results show that RL-RSS resolves the intercept over-estimation of random scale variance occurring in the simple mixed-effect random location scale model.
I. INTRODUCTION

In the study of repeated measurements collected on the same units or clusters (longitudinal study), mixed random effects models (MRMs) containing both fixed effects and random effects are widely used to account for within-unit or within-cluster correlation. The independent and identically distributed (IID) error term is the basic model assumption across all observations. However, it is not necessarily true in the case of intensive longitudinal data when a large number of observations or measurements are collected within a cluster. Modern data collection procedures, such as EMA (Smyth, 2003; Stone, 1994), experience sampling (Feldman, 2001), and diary methods (Bolger, 2003) yield a relatively large number of subjects and observations per subject. The heterogeneous within-cluster variation may occur, and thus the IID assumption in the regular mixed-effect models may be no longer held. In some particular studies, the within-cluster and between-cluster variability have been addressed as the focus of research interest along with the examination of population mean (Smoking study in Robin Mermelstein’s group). Thus MRMs need to be developed to satisfy the needs of research. Hedeker (2008) described a mixed model that included the determinants of between-subject variance, and additionally allowed the determinants of the within-subjects (WS) variance plus a random subject scale effect. This model is referred to as a mixed-effect location scale model because subjects have both random location effects and random scale effects. The essence of mixed-effect location scale model is further modeling the variance and/or accounting for heterogeneous variance across individuals or clusters. It is worthy to note that the random scale effect allows WS variance to vary across individuals, above and beyond the contribution of covariates in the population mean and variance. Li (2012) extended random location random scale model (RL-RS) to three-level clustered data, referred as
three-level RL-RS model. However, the variance of random scale effect is assumed to be a constant in both Hedeker’s and Xue’s model. The WS variation is forced to be homogeneous across subjects in the analysis, and moreover, there is a lack of opportunity to explore the effects of covariates on random scale variance.

In this dissertation, we extend the mixed-effect location scale model to examine the determinants of WS variance heterogeneity across clusters by modeling random scale variance in three-level cluster data structure. The proposed model is referred to as mixed-effects RL-RSS model. The Adolescent Smoking Study using EMA was provided as an illustration example for three-level RL-RSS model. The research interest focuses on whether certain covariates could explain error variance (WS variance) heterogeneity across subjects, above and beyond the contribution of covariates in mean change and error variance. The smoking event observations (level-1) are nested within time period waves (level-2), and waves nested within subjects (level-3). For the sake of simplicity, random location intercepts are specified at subject level and wave level. The inclusion of random scale effects allows error variance (WS variance) to vary across subjects, over the consideration of covariate effects on WS variance through a log-normal function. The random scale effect is allowed to correlate with random location effect at the corresponding cluster level. Furthermore, the covariates are allowed to account for random scale variance on the basis of log-normal regression structure.

Maximum Marginal Likelihood (Bock, 1989) is proposed for parameter estimation. The Newton-Raphson optimization algorithm using Gauss-Hermite quadrature is conducted to maximize the log likelihood. The SAS PROC NLMIXED can be used to obtain the maximum likelihood estimate (MLE) for this model. The RL-RSS model can be easily generalized to a
variety of EMA studies in smoking and cancer-relevant research areas. The method can also be applied to other types of studies with three-level clustering data.

A simulation study is described to evaluate the accuracy and reliability of three-level RL-RSS model. Under the structure of RL-RSS model, 100 data sets, each with 800×2×50 observations (50 observations nested within each of the two waves nested within each of the 800 subjects), are generated. The proposed three-level RL-RSS model has two subject-level random effects (one random location effect and one random scale effect) and one wave-level random location effect. At most three covariates (either continuous or dichotomous), are considered to affect the population mean or variance. For the comparison purpose, three models (three-level RL-RSS model, three-level RL-RS and three-level random intercept model) are fit for each simulated dataset. The evaluation criteria of model fitting performance are assessed including bias, root mean square error (RMSE), 95% confidence interval (CI) coverage probability, and average length of 95% CIs. The simulation is also carried out using SAS PROC NLMIXED procedure.
II. LITERATURE REVIEW

A. Two-level Mixed-Effects Regression Models

In reality, the sample collection is often processed from a large population over a certain time period. The repeated measurements on the same variable occur in longitudinal study or the study with hierarchical clusters. Longitudinal data track the information on the same individual at multiple time points. The observations within the same cluster (e.g., repeated observations from the same subject) are more similar to each other than to those at other clusters (e.g., the observations from different subjects). Thus the independency assumption within the repeated observations is not appropriate. The ignorance of the dependency among correlated measurements would lead to invalid sample distribution assumption and further affect the accuracy of hypothesis testing, statistical inference, and parameter estimation.

In order to account for the correlation among repeated observations, random effect is introduced into MRMs. The random effects describe the unobserved variability due to subject or cluster, and thus explain the correlation of longitudinal repeated data. The variance of random effect indicates the degree of between-subject or between-clusters variation, or subject heterogeneity existing in the population. The MRM models (Laird, 1982) are capable to handle missing data such as subjects with incomplete data across time. The MRM models have the ability to analyze missing data and provide valid inferences in the case of MCAR (Rubin, 1976) and MAR (Rubin, 1976) by using MLE. The application of MRMs has been steadily increasing as a primary method for longitudinal normal distributed outcome or clustered data in various research fields.
B. Two-Level Mixed-Effects Models with Multiple Random Effects

To simplify the illustration of a general MRM model, we take the two-level mixed-effects model as an example. A longitudinal structure (repeated observations nested within subjects) is introduced in this section. Let \( i \) represent subjects \((i = 1 \ldots n)\), and \( j \) represent observations \( j = 1 \ldots n_i \) clustered within subject \( i \). A general mixed random effects model with multiple random effects is described as

\[
y_{ij} = x_{ij}'\beta + z_{ij}'v_i + \varepsilon_{ij} \tag{EQ. II-1}
\]

where \( y_{ij} \) is the value of continuous outcome for subject \( i \) at the observation \( j \), \( x_{ij} \) is a \( p \times 1 \) covariate vector, \( \beta \) is a \( p \times 1 \) vector of unknown fixed-effect regression coefficients, \( z_{ij} \) is a \( r \times 1 \) design vector for random effects, \( v_i \) is a \( r \times 1 \) vector of unknown random effects, and \( \varepsilon_{ij} \) is the error term. The error term \( \varepsilon_{ij} \), is assumed to have an IID normal distribution with mean 0 and variance \( \delta^2_{\varepsilon} \) across all subjects and all observations. The random effect, \( v_i \), is assumed to have a \( r \)-dimensional multivariate normal distribution with mean 0 and variance-covariance matrix \( \Sigma_v \).

The MRM in EQ. II-1 can be expressed in a more compact representation of matrix formulation, the MRMs for subject \( i \) is expressed as below:

\[
y_i = X_i\beta + Z_i v_i + \varepsilon_i \quad \text{with } i = 1, \ldots, n. \tag{EQ. II-2}
\]

Here \( y_i \) is a \( n_i \times 1 \) vector of continuous outcome for subject \( i \), \( X_i \) is a \( n_i \times p \) covariate matrix for subject \( i \), \( \beta \) is a \( p \times 1 \) vector of fixed regression parameters, \( Z_i \) is a \( n_i \times r \) design matrix for random effects, \( v_i \) is a \( r \times 1 \) vector of random effects for subject \( i \), and \( \varepsilon_i \) is a \( n_i \times 1 \) error vector. The first column of \( X_i \) and \( Z_i \) usually has the value one. The distribution assumptions for the
random effects and errors term are, respectively, $\mathbf{v}_i \sim N(\mathbf{0}, \Sigma_v)$ and $\mathbf{\varepsilon}_i \sim N(\mathbf{0}, \delta^2 \mathbf{I}_{n_i})$. As a result, $\mathbf{y}_i$ and $\mathbf{v}_i$ have a joint multivariate normal distribution:

\[
\begin{bmatrix} \mathbf{y}_i \\ \mathbf{v}_i \end{bmatrix} \sim N \left( \begin{bmatrix} X_i \mathbf{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} Z_i \Sigma_u Z_i' + \delta^2 \mathbf{I}_{n_i} & Z_i \Sigma_v \\ \Sigma_u Z_i' & \Sigma_v \end{bmatrix} \right)
\]

**EQ. II-3**

Model estimation generally utilizes a combination of two complementary methods (Laird 1982; Bock 1989), Maximum Likelihood (ML) method and EB approach. The random individual effects $\mathbf{v}_i$ and corresponding variance $V(\hat{\mathbf{v}}_i)$ are estimated using EB method, while ML is used for fixed-effects regression parameters $\mathbf{\beta}$, variance parameters $\delta^2$ and $\Sigma_v$.

The EB estimates of the random effects are derived from the posterior distribution of $\mathbf{v}_i$ given $\mathbf{y}_i$. The condition distribution of $\mathbf{v}_i$ given $\mathbf{y}_i$ can be easily derived from the assumed multivariate normal distribution of $\mathbf{v}_i$ and $\mathbf{y}_i$ in **EQ. II-3**. The EB estimate $\hat{\mathbf{v}}_i$ equals to the conditional mean of $\mathbf{v}_i$ given $\mathbf{y}_i$, and the variance estimate $\hat{V}(\hat{\mathbf{v}}_i)$ is the conditional variance of $\mathbf{v}_i$ given $\mathbf{y}_i$. Hedeker and Gibbons (2006) described the details of EB estimate derivation.

\[
\hat{\mathbf{v}}_i = E(\mathbf{v}_i|\mathbf{y}_i) = \left[ Z_i' Z_i + \delta^2 \Sigma_v^{-1} \right]^{-1} Z_i' (\mathbf{y}_i - X_i \mathbf{\beta})
\]

**EQ. II-4**

\[
\hat{V}(\hat{\mathbf{v}}_i) = \left[ Z_i' (\delta^2 \mathbf{I}_{n_i})^{-1} Z_i + \Sigma_v^{-1} \right]^{-1}
\]

Maximum likelihood (ML) method is used to estimate $\mathbf{\beta}$, $\delta^2$ and $\Sigma_v$. Several types of practical numerical iterations like Expectation and Maximization (EM) algorithm, Newton-Raphson or Fisher’s scoring methods are widely applied for maximum likelihood estimation. The EM algorithm alternates expectation step (EB estimates $\hat{\mathbf{v}}_i$ and $\hat{\Sigma}_{v_i}$ takes expectation of conditional distribution of $\mathbf{v}_i$ given $\mathbf{y}_i$) and maximization step (maximize the likelihood to get the MLEs of $\mathbf{\beta}$, $\delta^2$ and $\Sigma_v$ given the intermediate EB random effect parameter estimates) iteratively until convergence. The EB estimates
are viewed as unobserved latent values in EM. Finding a maximum likelihood solution requires taking the first derivatives of likelihood function with respect to $\beta$, $\delta^2_\varepsilon$ and $\Sigma_v$. On the negative side, EM algorithm has a relatively slow convergence rate. As an alternative approach, Newton-Raphson or Fisher’s scoring algorithm helps accelerate the convergence rate by maximizing the marginal log-likelihood directly with respect to $\beta$, $\delta^2_\varepsilon$ and $\Sigma_v$, with no need of the estimates of $\beta$, $\delta^2_\varepsilon$ and $\Sigma_v$ expressed in terms of EB estimates. Newton-Raphson or Fisher’s score algorithms require the second derivatives or expectation of the second derivatives. The EB estimates are calculated from the conditional distribution as illustrated in EQ. II-4 with the final MLEs $\beta$, $\delta^2_\varepsilon$ and $\Sigma_v$.

Expectation-maximization algorithm: Fixed-effect parameters $\beta$, $\delta^2_\varepsilon$ and $\Sigma_v$ can be estimated by the EM algorithm in Hedeker and Gibbons (2006). The distribution of $y_i$ given random effects $u_i$ has a multivariate normal distribution with mean vector $X_i\beta + Z_iu_i$ and covariance matrix $\delta^2_\varepsilon I_{n_i}$, noted as $y_i|u_i \sim N(X_i\beta + Z_iu_i, \delta^2_\varepsilon I_{n_i})$. The log density function of $y_i|u_i$ is

$$
\log f(y_i|u_i) = -\frac{n_i}{2} \log(2\pi) - \frac{1}{2} \log |\delta^2_\varepsilon I_{n_i}| - \frac{1}{2\delta^2_\varepsilon} (y_i - X_i\beta - Z_iu_i)'(y_i - X_i\beta - Z_iu_i).
$$

EQ. II-5

The marginal log-likelihood of $y_i$ is represented as below after integrating out random effects $u$.

$$
h_i = h(y_i) = \int f(y_i|u_i)g(u) \, d(u).
$$

EQ. II-6

The normal distribution of $u$ is simply $\log g(u) = -\frac{r}{2 \ln |\Sigma_v|} - u'\Sigma_v^{-1}u/2$. Hence the overall log-likelihood is $\log L = \sum_{i=1}^{N} \log h_i = \sum_{i=1}^{N} \log \int_{u} f(y_i|u_i)g(u) \, d(u)$. Let $p_i = f(u_i|y_i) = $
\( f(y_i|v_i)g(v) \) denote the posterior distribution of \( v_i \) given \( y_i \). To estimate \( \beta \), take the first derivative of \( logL \) with respect to \( \beta \)

\[
\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} \frac{\log h_i}{h_i} \frac{\partial h_i}{\partial \beta} = \sum_{i=1}^{n} \frac{1}{h_i} \frac{\partial h_i}{\partial \beta} = \sum_{i=1}^{n} \int_{\mathcal{V}} \frac{f(y_i|v_i)g(v)}{h_i} \frac{\partial \log f(y_i|v)}{\partial \beta} d(v)
\]

\[
= \sum_{i=1}^{n} \int_{\mathcal{V}} p_i \frac{\partial \log f(y_i|v)}{\partial \beta} d(v)
\]

\[
= \sum_{i=1}^{n} \int_{\mathcal{V}} p_i X_i'(\delta_i^2 I_{n_i})^{-1}(y_i - X_i \beta - Z_i \tilde{v}_i) d(v)
\]

\[
= \delta_i^{-2} \sum_{i=1}^{n} X_i'(y_i - X_i \beta - Z_i \tilde{v}_i)
\]

And equate \( \frac{\partial \log L}{\partial \beta} \) to 0, the maximum likelihood estimator of \( \beta \) is

\[
\tilde{\beta} = \left[ \sum_{i=1}^{n} X_i' X_i \right]^{-1} \left[ \sum_{i=1}^{n} X_i'(y_i - Z_i \tilde{v}_i) \right]
\]

EQ. II-8

where \( \tilde{v}_i \) is the EB estimate of \( v \) as specified in EQ. II-4, which is the posterior mean of \( v \) given \( y_i \). To estimate the residual variance, take the first derivative of \( logL \) with respect to \( \delta_i^2 \)

\[
\frac{\partial \log L}{\partial \delta_i^2} = \sum_{i=1}^{n} \frac{\log h_i}{h_i} \frac{\partial h_i}{\partial \delta_i^2} = \sum_{i=1}^{n} \frac{1}{h_i} \frac{\partial h_i}{\partial \delta_i^2}
\]

\[
= \sum_{i=1}^{n} \int_{\mathcal{V}} \frac{f(y_i|v_i)g(v)}{h_i} \frac{\partial \log f(y_i|v)}{\partial \delta_i^2} d(v)
\]

EQ. II-9
\[=
\sum_{i=1}^{n} \int \left[ p_i \left( -\frac{n_i}{2\delta^2} + \frac{1}{2\delta^2} (y_i - X_i\beta - Z_i\nu_i)'(y_i - X_i\beta - Z_i\nu_i) \right) \right] d\nu \]

And equate \( \frac{\partial \log L}{\partial \delta^2} \) to 0, the ML estimator of \( \delta^2 \) is

\[ \delta^2 = \left( \sum_{i=1}^{n} \frac{1}{n_i} \right)^{-1} \left( \sum_{i=1}^{n} \hat{e}_i'\hat{e}_i + \text{tr}[Z_i\hat{\nu}_i(Z_i)' - 1] \right) \]

EQ. II-10

where \( \hat{\nu}_i \) is the posterior variance of \( \nu \) given \( y_i \) derived in EQ. II-4,

\[ \hat{\Sigma}_{i|y_i} = V(\nu_i|y_i) = \left[ Z_i'\hat{\Sigma}_e^{-1} Z_i + \Sigma_{\nu}^{-1} \right]^{-1} \].

The vector \( \hat{e}_i \) is defined as \( e_i = y_i - X_i\beta - Z_i\hat{\nu}_i \) with \( \hat{e}_i = y_i - X_i\hat{\beta} - Z_i\hat{\nu}_i \).

The similar derivation process is applied to estimate the unique elements in the covariance matrix \( \Sigma_{\nu} \) for random effects. The ML estimate of \( \Sigma_{\nu} \) is calculated by equating the first derivative of \( \log L \) with respect to vech \( \Sigma_{\nu} \) to 0 as

\[ \frac{\partial \log L}{\partial \text{vech} \Sigma_{\nu}} = \sum_{i=1}^{N} h_i \int \frac{\partial g(v)}{\partial \text{vech} \Sigma_{\nu}} f(y_i|v) dv = G' \Sigma_{\nu}^{-1} \text{vec} \left( \text{vech} \Sigma_{\nu}^{-1} \right) = 0. \]

EQ. II-11

The vech operator stacks the upper triangle of a square matrix including the diagonal elements to form one column vector. For an r-dimensional symmetric matrix \( \Sigma_{\nu} \), \( r(r+1)/2 \) unique elements are generated in \( \text{vech} \Sigma_{\nu} \). The vec operator of a square matrix is a column vector obtained by stacking the columns of the matrix on top of one another. Matrix \( G \) transforms \( \text{vech} \Sigma_{\nu} \) into \( \text{vec} \Sigma_{\nu} \) by \( \text{vec} \Sigma_{\nu} = G\text{vech} \Sigma_{\nu} \).
The EM algorithm alternates between performing an expectation (E-step) and maximization (M-step). The E-step starts with the initial values of \( \beta, \delta^2 \), and \( \Sigma_u \), and then computes the EB estimates of \( \hat{\theta_i}, \hat{V}(\hat{\theta_i}) \) as derived in EQ. II-4.

In M-step, the estimates of \( \beta, \delta^2 \), and \( \Sigma_u \) can be obtained by maximizing the log likelihood as derived in EQ. II-8, EQ. II-10, and EQ. II-11 with the EB estimates \( \hat{\theta_i}, \hat{V}(\hat{\theta_i}) \) in the current iteration. The iterations of E-step and M-step are repeated until convergence.

Newton-Raphson or Fisher’s scoring iteration: Comparing to EM algorithm, Newton-Raphson or Fisher’s scoring directly maximizes the marginal likelihood to estimate fix-effect parameters without integrating out random effects from conditional distribution of \( y_i | \nu \).

The general linear mixed model implies the marginal distribution

\[
y_i \sim \mathcal{N}(X_i \beta, Z_i \Sigma_u Z_i' + \delta^2 I_{n_i})
\]

**EQ. II-12**

The log-likelihood for the \( i \)th subject is

\[
\log f(y_i) = -\frac{n_i}{2} \ln(2\pi) - \frac{1}{2} \ln |Z_i \Sigma_u Z_i' + \delta^2 I_{n_i}| - \frac{1}{2} (y_i - X_i \beta)' (Z_i \Sigma_u Z_i' + \delta^2 I_{n_i})^{-1} (y_i - X_i \beta).
\]

**EQ. II-13**

and the overall log-likelihood is \( \log L = \sum_{i=1}^n \log f(y_i) \). The fixed-effect parameters \( \beta, \delta^2 \) and \( \Sigma_u \) are estimated by maximizing the above marginal likelihood directly without the EB estimates \( \hat{\theta_i} \) or \( \hat{V}(\hat{\theta_i}) \). The score equation for \( \beta \) is

\[
\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^n X_i' (Z_i \Sigma_u Z_i' + \delta^2 I_{n_i})^{-1} (y_i - X_i \beta)
\]

**EQ. II-14**

Equate it to 0, the ML estimator of \( \beta \) is
\[ \hat{\beta} = \left[ \sum_{i=1}^{n} X'_i (Z_i \Sigma_u Z'_i + \delta^2 \epsilon I_{n_i})^{-1} X_i \right]^{-1} \left[ \sum_{i=1}^{n} X'_i (Z_i \Sigma_u Z'_i + \delta^2 \epsilon I_{n_i})^{-1} y_i \right] \]

Equation II-15

The MLE \( \hat{\beta} \) is not expressed in terms of EB estimates \( \hat{\theta}_i \) and \( \hat{V}(\hat{\theta}_i) \) as it is in EM algorithm. Upon the similar strategy, ML estimators of \( \delta^2 \epsilon \) and \( \Sigma_u \) can also be achieved by maximizing marginal likelihood in a similar way. Newton-Raphson or Fisher’s scoring algorithm are applied for iterating solution of maximization. The EB estimates are calculated from the conditional distribution as illustrated in EQ. II-4 with the final MLEs \( \beta \), \( \delta^2 \epsilon \), and \( \Sigma_u \).

The MRM model fitting or estimation can be performed in a wide variety of popular statistical software packages such as SAS (MIXED procedure), R (LME function), STATA (XTMIXED function), SPSS (MIXED procedure), and S-plus (LME function).

C. Heterogeneous Variance Modeling

Mixed random effects model has acted as a primary analysis method for longitudinal data with normal outcome measures. Random subject effects are included in the MRM to account for the similarity among repeated observations within a subject. The random subject effects are often assumed to be IID normal-distributed random variables. The error terms in such models are also assumed to follow IID normal distribution across all subjects and all observations. The between-subject variation and WS variation denoted by random subject effects and the error variance respectively, are usually considered to be homogeneous across subjects. However, the homogeneous variance assumption of both within- and between-subject variation, can be violated in the study with a large number of subjects and observations. In addition, the random subject effects are not necessarily independent with the error terms in the real-world data. Lack-of-
homogeneous variance is often referred to as heteroscedasticity. Allowing for heteroscedasticity of within- and between-subject variations helps reduce the standard errors of the fixed-effects parameter estimates, and therefore improves the estimation precision by weighing down the large dispersions.

The study with a large number of subjects and observations per subject addresses the need to analyze the heteroscedasticity of both between- and WS variances. Modern data collection procedures, such as ecological momentary assessments (Stone, 1994; Smyth, 2003), experience sampling (Scollon, 2001), diary methods (Bolger, 2003), and real-time data captures provide such opportunities to explore the variance heteroscedasticity. Especially, EMA methods usually yield up to thirty observations per subject, and thus provide greater opportunities in heterogeneous variance modeling than conventional longitudinal studies. The data from such designs are also referred to as intensive longitudinal data (Walls, 2006), of which the research interest focuses on assessing individual variation and exploring the potential impact factors of variances.

Modeling the heteroscedasticity of error variance has been extensively investigated in simple linear regressions. The most common approaches to account for heterogeneous error variance are Box-Cox transformation (Box, 1964), and weighted least squares (WLS) estimation. The WLS applies extra weights on each observation when a simple transformation is not successful to adjust the heterogeneous variance. Additionally, jointly modeling mean and dispersion has been developed to account for heterogeneous variance. Harvey (1976) and Aitkin (1987) utilized a log-linear structure to allow error variance to be determined by a set of subject-level covariates, where maximum likelihood estimation was applied to estimate mean regression coefficient parameters and error variance parameters simultaneously. Aitkin (1987) developed the
approach of jointly modeling the mean and error variance parameters in the statistical computing package known as GLIM. The jointly modeling allowed the covariates to influence both the mean and error variance. Carroll (1982) proposed the nonparametric method to allow the variance to be determined by an unknown smooth function of design or mean response. Davidian (1987) stated that “most variance function estimation procedures can be looked upon as regressions with responses being transformations of absolute residuals from a preliminary fit or sample standard deviations.” Smyth (1989) included a linear predictor for the dispersion in generalized linear models, applying for normal, inverse Gaussian or gamma distributed data. “A log-linear dependence of the variances on suspected explanatory variables” was developed by Verbyla (1993). Detection of dependence, estimation, and tests of homogeneity were based on residual maximum likelihood estimation.

In linear mixed-effects models, the assumption of homogeneous variance is untenable, and the corrective action is also required to ensure an accurate and efficient analysis. Cleveland (2000, 2002) developed a general class of mixed-effects location scale model to attempt handling variance heterogeneity. The RL-RS is akin to include the random effects adjusting the subject variability for both the population mean (location) and variance (scale). The error term has a multiplication representation of a random scale effect and a normal distributed error element, of which the error variance is allowed to vary with the covariates. For random scale variable, Cleveland considered four positive distributions (log-normal, gamma, inverse gamma, and Weibull). The approach consists of a sequence of steps for model estimation and model diagnostic check. As a special case, the model with no random scale effect is referred to as random location fixed-scale (RL-FS) model.
Hedeker et al. (2007, 2008, 2009) extended the standard mixed model by adding random effects to the WS variance specification using EMA data. The proposed random location scale model permits the covariates to have influence on the population mean (location) and population variability (scale). The between-subject variance and WS variance are allowed to vary across covariates via a log-linear representation. Additionally, the approach allows the location and scale random effects to be correlated. The model estimation can be conducted via SAS procedure PROC NLMIXED using adaptive quadrature for integration of the random effects and quasi-newton optimization algorithm. Hedeker (2006) employed the same structure of RL-RS model for multilevel ordinal outcome data. Again, the RL-RS model without random scale effect is referred to as the RL-FS model. Cleveland’s multiplicative RL-RS model is closely connected Hedeker’s RL-RS model with an additive log scale random scale effect. Cleveland presented a general class of RL-RS model for heterogeneous variance modeling denoted by

$$y_{ij} = x_{ij}' \beta + z_{ij}' \gamma + \epsilon_{ij}$$

EQ. II-16

with the conditions:

$$v_i \sim N(0, \Sigma_v)$$

$$\gamma_i \sim F_\gamma(\mu_\gamma, \delta_\gamma^2)$$ and $$E(\gamma_i^2) = 1$$, hence $$\mu_\gamma = 1 - \delta_\gamma^2$$

$$\epsilon_{ij} \sim N \left(0, \delta_{\epsilon_{ij}(x)}^2 \right)$$

$$v_i, \gamma_i, \epsilon_{ij}$$ are mutually independent.

where $$y_{ij}$$ is the outcome measure for $$i$$th subject $$j$$th observation, $$x_{ij}$$ is a $$p \times 1$$ column vector of fixed-effects covariate, $$\beta$$ is a $$p \times 1$$ column vector of fixed-effects regression coefficient parameters, $$z_{ij}$$ is a $$q \times 1$$ design vector of random location effects, and $$\nu_i$$ is a $$q \times 1$$ vector of random location effects with an IID multivariate normal distribution $$N(0, \Sigma_v)$$, the random scale effect $$\gamma_i$$, is an IID random scalar with a distribution $$F_\gamma$$ and restrictions $$E(\gamma_i^2) = 1$$. The possible
choices for $\gamma_l$ distribution include gamma, inverse gamma, Weibull, and log-normal. The error term for $\epsilon_{ij}$ follows an independent normal distribution: $N\left(0, \delta_{ij}^2(\omega)\right)$, where $\delta_{ij}^2(\omega)$ depends on a categorical explanatory variable $z_{ij}$. Hedeker’s additive log scale RL-RS model can be simply expressed as

$$y_{ij} = x'_{ij}\beta + z'_{ij}u_i + \epsilon_{ij}$$  \textit{EQ. II-17}

with the conditions:

$$v_i \sim N\left(0, \delta_{v_i}^2\right)$$

$$\epsilon_{ij}^*|\omega_i \sim N\left(0, \delta_{\epsilon_{ij}}^2\right) \text{ and } \log \delta_{\epsilon_{ij}}^2 = u'_{ij}\tau + \omega_i, \omega_i \sim N\left(0, \delta_{\omega}^2\right)$$

$$\begin{bmatrix} v_i \\ \omega_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \delta_{v_i}^2 & \delta_{v\omega} \\ \delta_{v\omega} & \delta_{\omega}^2 \end{bmatrix}\right)$$

$v_i$, and $\epsilon_{ij}^*$ are dependent.

The above notations are the same as EQ. II-16 in terms of $y_{ij}$, $x'_{ij}\beta$, and $z'_{ij}u_i$. The error $\epsilon_{ij}^*$ is assumed to be normal with zero mean and variance $\delta_{\epsilon_{ij}}^2$. The error variance $\delta_{\epsilon_{ij}}^2$ is further modeled through a log-linear function $\log \delta_{\epsilon_{ij}}^2 = u'_{ij}\tau + \omega_i$, where $u_{ij}$ is a column vector of explanatory variable, and $\tau$ is denoted for regression coefficients of WS variance. The random scale effect $\omega_i$ is distributed as $N(0, \delta_{\omega}^2)$. With some algebra transformation, Hedeker’s model in EQ. II-17 can be rewritten as

$$y_{ij} = x'_{ij}\beta + z'_{ij}u_i + \delta^*_{\epsilon_{ij}}\epsilon_{ij}$$

$$= x'_{ij}\beta + z'_{ij}u_i + e^{\frac{(u'_{ij}\tau + \omega_i)}{2}}\epsilon_{ij} = x'_{ij}\beta + z'_{ij}u_i + e^{\frac{\omega_i}{2}}\delta_{\omega}^2 e^{\frac{u'_{ij}\tau}{2}} + e^{\frac{\delta_{\omega}^2}{4}}e_{ij}$$

$$= x'_{ij}\beta + z'_{ij}u_i + \gamma_i\delta_{\epsilon_{ij}}\epsilon_{ij} = x'_{ij}\beta + z'_{ij}u_i + \gamma_i\epsilon_{ij}$$
with \( e_{ij} \sim N(0,1) \). From the above derivation, Hedeker’s RL-RS model can be viewed as a special case of Cleveland’s model with the random scale effect \( \gamma_i \sim \text{Lognormal}\left( -\frac{\delta_\omega^2}{2}, \delta_\omega^2 \right) \) and fixed-scale effect.

\[
\varepsilon_{ij} \sim N\left(0, \sigma_{\varepsilon ij}^2\right) \quad \text{and} \quad \log \delta_{\varepsilon ij}^2 = \omega_i t + \frac{\delta_\omega^2}{2}.
\]

Extended from Cleveland’s general random location scale model, Hedeker’s method relaxes the independence assumption between random location effect and random scale effect. It also allows them to correlate with each other.

D. Three-Level Mixed-Effects Random Location Scale Models

The three-level clustering data is characterized with (level 1) observations nested within (level 2) clusters, which are in turn nested within (level 3) clusters. For example, the repeated observations are nested within time points, and time points further nested within subjects, or subjects nested in classes, classes nested in schools. The previously described two-level mixed-effect random location random scale model (RL-RS) was extended to analyze three-level clustering data in (Li, 2012). Li et al. developed an extended three-level mixed-effects model based on Hedeker’s two-level RL-RS model. Li’s three-level mixed random location scale is defined as

\[
y_{ijk} = x'_{ijk} \beta + v_{0ij} + v_{0i} + \varepsilon_{ijk}
\]

with the conditions:

\[
v_{0ij} \sim N\left(0, \sigma_{0ij}^2\right), \quad v_{0i} \sim N\left(0, \sigma_i^2\right).
\]
\[
\log(\delta_{v_{ij}}^2) = z_i' \alpha, \log(\delta_i^2) = \pi_i' \lambda,
\]

RL-RS model: \( \epsilon_{ijk} | \omega_i \sim N \left( 0, \delta_{\epsilon_{ijk}}^2 \right), \quad \omega_i \sim N(0, \delta_{\omega}^2) \)

\[
\log(\delta_{\epsilon_{ijk}}^2) = u_{ijk}' \tau + \omega_i
\]

\[
\begin{bmatrix}
  v_{0i} \\
  v_{0ij} \\
  \omega_i
\end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \delta_0^2 & 0 & \delta_{v\omega} \\ 0 & \delta_0^2 & 0 \\ \delta_{v\omega} & 0 & \delta_{\omega}^2 \end{bmatrix} \right)
\]

\[\text{EQ. II-18}\]

RL-FS model: \( \epsilon_{ijk} \sim N \left( 0, \delta_{\epsilon_{ijk}}^2 \right) \)

\[
\log(\delta_{\epsilon_{ijk}}^2) = u_{ijk}' \tau
\]

\[
\begin{bmatrix}
  v_{0i} \\
  v_{0ij}
\end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \delta_0^2 \delta_{v0} \\ \delta_{v0} \delta_{\omega}^2 \end{bmatrix} \right)
\]

Let \( N = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ijk}} n_{ijk} \) be the total number of observations for all subjects at all waves. There are total subjects (subject \( i = 1 \ldots n \)) at the waves (wave \( j = 1 \ldots \ldots n_i \)), and the number of observations (\( k = 1 \ldots n_{ijk} \)) at wave \( j \) for subject \( i \). Here, \( y_{ijk} \) is the \( k \)th observed continuous outcome measures for subject \( i \) at wave \( j \), \( x_{ijk} \) is the covariate vector/design vector, \( \beta \) is the vector of fixed-effect regression coefficient parameters. The random effect \( v_{0ij} \) indicates the variation of subject \( i \) at wave \( j \), \( v_{0i} \) indicates the random effect for \( i \)th subject. The distribution of random effects \( v_{0i} \), \( v_{0ij} \) are assumed to be normal distribution \( N(0, \delta_0^2) \) and \( N \left( 0, \delta_{v_{ij}}^2 \right) \) respectively, where \( \delta_0^2 \) represents between-subject (BS) variation, and \( \delta_{v_{ij}}^2 \) represents between-subject (BS) heterogeneity within wave \( j \). The variance \( \delta_i^2 \) is further modeled by \( \delta_i^2 = e^{\pi_i' \lambda} \), and the variance \( \delta_{v_{ij}}^2 \) is modeled by \( \delta_{v_{ij}}^2 = e^{z_{ij}' \alpha} \). The \( \epsilon_{ijk} \) is the error term for the \( k \)th observation of subject \( i \) at wave \( j \). The errors \( \epsilon_{ijk} \) are assumed to be normal distributed with zero mean and variance \( \delta_{\epsilon_{ijk}}^2 \). The error variance \( \delta_{\epsilon_{ijk}}^2 \) represents the WS (WS) variation. In RL-RS model, \( \delta_{\epsilon_{ijk}}^2 \) is further modeled through a log-linear mixed-effects structure, namely, \( \log(\delta_{\epsilon_{ijk}}^2) = u_{ijk}' \tau + \omega_i \) to
allow the WS variance to vary across individuals, above and beyond the effect of covariates on WS variance, where $u_{ijk}$ is a column vector of explanatory variables, and the coefficient vector $\tau$ indicates the degree of covariates’ influence on within-cluster variance. The random scale effect $\omega_i$ are distributed as normal $N(0, \delta^2_{\omega})$. The $v_{0ij}, v_{0i}$ are the random effects that influence the population mean (location) of outcome at different cluster levels, whereas $\omega_i$ is a random effect that influence the variance (scale). The random effects at the same cluster level such as $v_{0ij}$ and $\omega_{ij}$ are allowed to be correlated with the covariance $\delta_{v\omega}$. This covariance indicates the degree to which random location effect is correlated with random scale effects. The three-level mixed-effect RL-RS models can be estimated in SAS procedure PROC NL MIXED procedure (SAS Institute Inc., Cary, North Carolina, USA) by specifying the likelihood in general MODEL statement.

E. **Ecological Momentary Assessment**

Ecological Momentary Assessment is motivated by the limitations and inaccuracy of the traditional retrospective reports. The interests of both methods are on the real-world behaviors of the subjects. The retrospective reports are more static and summary—they are collected at clinical sites by asking questions like “how often,” “how many,” “how intensive,” for a given period of time. This method, however, is limited by recall bias and memory inaccuracy, and fails to capture the dynamic behaviors over the period of time.

An EMA, on the other hand, tries to capture subjects’ dynamic behaviors by running the assessments periodically or randomly in real time and in real-world contexts (Smyth, 2003; Stone, 1994). Advance technologies such as cell phones and e-diaries (Facebook and Twitter) can be used to improve the efficiency of information collection with minimum disturbance to the subjects’ life routines. Due to periodical assessments, EMA also enables the study of micro-processes that
could affect subjects’ behaviors. By trying to capture the dynamics of behavior, EMA has been shown to be a promising new method. Data collection methods for EMA usually yield thirty or forty observations per subject. Such a large number of repeated observations within subjects offer a great opportunity in heterogeneous variance modeling than conventional longitudinal studies.

F. **The Proposed Model**

As literature review revealed, Cleveland (2000, 2002) developed a general class of mixed-effects location scale model to account for heterogeneous variance in multiple-level data structure. Hedeker and Li extended regular mixed-effect model to mixed-effect random location scale model by modeling heterogeneous variance with a log linear structure. However, none of their methods have explored the determinants of random scale variance, which reflects the association between covariates and subject heteroscedasticity in terms of data dispersion. The WS variation is forced to be homogeneous across clusters in the existing methods. Less is known about whether random scale variance is associated with the potential covariates. In this dissertation, we extended the mixed-effect location scale model to examine determinants of subject heterogeneity of data dispersion by modeling random scale variance through a log-linear representation. On the basis of Li’s three-level RL-RS model, the proposed three-level mixed-effect random location scale model with modeling random scale variance (RL-RSS model) has been applied to the analysis for an EMA study of smoking as an illustration example. The model setup is based on a conventional three-level MRM with a random location intercept at each clustering level, but also allows covariates to influence the error variance (WS variance) and random scale variance by a log-linear representation. In other words, the log of error variance and random scale variance are modeled by
a regression structure with addressing the effects of covariates on them. The random scale effect is allowed to correlate with random location effects at the corresponding clustering level.

The three-level mixed-effects location scale models with modeling random scale variance (RL-RSS) was illustrated by applying to an Adolescent Smoking Study. The model analysis was performed in the commercial convenient software SAS PROC NLMIXED procedure.

G. Model Specification

The EMA data collected by the study was provided as an example of illustration for three-level mixed-effects location scale model with modeling random scale variance (RL-RSS model). In the longitudinal framework, the observations are repeatedly measured at each wave for each subject; in other words, multiple observations (level-1) are clustered or nested within the wave (level-2). Waves are further nested in subjects (level-3). Let \( N = \sum_{i=1}^{n_i} \sum_{j=1}^{n_{ij}} n_{ijk} \) be the total number of observations for all subjects at all waves. There are total subjects (subject \( i = 1 \ldots n \)) at the waves (wave \( j = 1 \ldots n_i \)), and the number of observations (\( k = 1 \ldots n_{ij} \)) at wave \( j \) for subject \( i \). To simplify the illustration, single random location effects at subject- and wave-level are considered. The RL-RSS model can be notated in the form of:

\[
y_{ijk} = x_{ijk}'\beta + v_{0ij} + v_{0i} + \epsilon_{ijk}
\]

with the conditions:

\[
\begin{align*}
v_{0ij} &\sim N(0, \delta_v^2), \\
v_{0i} &\sim N(0, \delta_0^2), \\
\epsilon_{ijk} | \omega_{ij} &\sim N\left(0, \delta_{\epsilon_{ijk}}^2\right), \\
\omega_{ij} &\sim N\left(0, \delta_{\omega_{ijk}}^2\right)
\end{align*}
\]

\[
\begin{bmatrix} v_{0i} \\ v_{0ij} \\ \omega_{ij} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \delta_0^2 \\ 0 \end{bmatrix}, \begin{bmatrix} \delta_v^2 & 0 & 0 \\ 0 & \delta_0^2 & \delta_{\omega_{ijk}}^2 \\ 0 & \delta_{\omega_{ijk}} & \delta_{\omega_{ijk}}^2 \end{bmatrix}\right)
\]
\[
\log \left( \delta_{ij}^2 \right) = u_{ij}^T \tau + \omega_{ij},\; \log \left( \delta_{\omega ij}^2 \right) = z_{ij}^T \alpha
\]

EQ. II-19

Where \( y_{ijk} \) is the \( k \)th observed continuous outcome measures for \( i \)th subject at \( j \)th wave, \( x_{ijk} \) is \( p \times 1 \) covariate vector/design vector (typically with the value “1” for the intercept as the first element), \( \beta \) is the corresponding \( p \times 1 \) vector of fixed-effect regression coefficient parameters. The random effect \( \nu_{0ij} \) indicates the random effect of subject \( i \) at wave \( j \), \( \nu_{0i} \) represents the random effect for \( i \)th subject. The distribution of random effects \( \nu_{0i}, \nu_{0ij} \) are assumed to be normal distribution \( N(0, \delta_0^2) \) and \( N(0, \delta_{0j}^2) \) respectively, where \( \delta_0^2 \) represents between-subject variation, and \( \delta_{0j}^2 \) represents between-subject variation at \( j \)th wave. The error term for the \( k \)th observation of subject \( i \) at wave \( j \) is \( \varepsilon_{ijk} \). The errors \( \varepsilon_{ijk} \) are assumed to be normal with zero mean and variance \( \delta_{\varepsilon ij}^2 \). The error variance \( \delta_{\varepsilon ij}^2 \) represents the WS within-wave (WS-WW) variation. To allow the WS-WW variance to vary across individuals and also be affected by covariates, above and beyond the contribution of covariates on the population mean, \( \delta_{\varepsilon ij}^2 \) is further modeled through a log-linear mixed-effects modeling structure, namely, \( \log \left( \delta_{\varepsilon ij}^2 \right) = u_{ij}^T \tau + \omega_{ij} \). Here, \( u_{ij} \) is \( q \times 1 \) column vector of explanatory variables, and the \( q \times 1 \) coefficient vector \( \tau \) indicates the degree of covariates’ influence on the error variance. The random scale effect \( \omega_{ij} \) is distributed as normal \( N \left( 0, \delta_{\omega ij}^2 \right) \). It is worthy to note that \( \nu_{0ij}, \nu_{0i} \) are the random effects over the population mean (we call it “location”) of outcome at different cluster levels, whereas \( \omega_{ij} \) is a random effect that influences data variance (the “scale” of data).

The skewed non-negative characteristic makes log-normal distribution to be a reasonable choice for variance modeling. It has been used in many diverse research areas for this purpose.
(Leonard, 1973; Shenk, 1998; Fowler, 1999). Log-normal distribution assumption allows $\omega_{ij}$ to be a random normal variable and leads to the joint distribution of $u_{0i}$, $u_{0ij}$, and $\omega_{ij}$ easily modeled as multivariate normal distribution. Other distributions like gamma, inverse-gamma, or Weibull do not have this nice property.

As an extension of Hedeker and Li’s mixed-effects location scale model, $\delta_{\omega_{ijk}}^2$, the variance of random scale effect $\omega_{ij}$, is further allowed to be influenced by observation-level, wave-level, subject-level covariates, and their possible interactions. The random scale variance model is denoted by $log\left(\delta_{\omega_{ijk}}^2\right) = z_{ijk}'\alpha$, where $z_{ijk}$ is $g \times 1$ column design vector of covariates, and $\alpha$ is the corresponding $g \times 1$ regression coefficients vector. The coefficient parameter $\alpha$ reflects how the covariates influence the subject heterogeneity of data dispersion.

The random effects at the same cluster level such as $u_{0i}$ and $\omega_{ij}$ are allowed to be correlated with the covariance $\delta_{u\omega}$. This covariance indicates the degree to which the random location effects and random scale effects are associated with each other. Random effects are assumed to be independent if they are at the different cluster levels, for instance, $u_{0i}$ is independent with $u_{0ij}$ and $\omega_{ij}$.

Based on the above assumptions EQ. II-19, $u_{0i}$, $u_{0ij}$, and $\varepsilon_{ijk}$ are pairwise independent with a multivariate normal conditional on $\omega_{ij}$:

$$\begin{bmatrix} u_{0i} \\ u_{0ij} \\ \varepsilon_{ijk} \end{bmatrix}_{\omega_{ij}} \sim N\left(\begin{bmatrix} 0 \\ \delta_{u\omega} \omega_{ij}/\delta_{\omega_{ijk}}^2 \\ 0 \end{bmatrix}, \begin{bmatrix} \delta_{\omega}^2 & 0 & 0 \\ 0 & \delta_{u\omega}^2/\delta_{\omega_{ijk}}^2 & 0 \\ 0 & 0 & \delta_{\varepsilon_{ijk}}^2 \end{bmatrix}\right)$$

EQ. II-20

H. **Marginal Distribution of Outcome Variable**

The conditional distribution of $y_{ijk}$ given random effects $u_{0i}$, $u_{0ij}$, and $\omega_{ij}$ is of the form:
\[ y_{ijk} | v_{0i}, v_{0ij}, \omega_{ij} \sim N \left( x'_{ijk} \beta + v_{0ij} + v_{0i}, \delta^2_{\epsilon_{ijk}} \right) \]

\[ \sim N \left( x'_{ijk} \beta + v_{0ij} + v_{0i}, e^{u'_{ijk} \tau + \omega_{ij}} \right) \]

As far as the marginal distribution of \( y_{ijk} \) is concerned, \( E(y_{ijk}) \), the marginal mean of \( y_{ijk} \) is clearly

\[ E(y_{ijk}) = E(x'_{ijk} \beta + v_{0ij} + v_{0i} + \epsilon_{ijk}) = E(x'_{ijk} \beta) + E(v_{0ij}) + E(v_{0i}) + E(\epsilon_{ijk}) = x'_{ijk} \beta. \]

The \( V(y_{ijk}) \), the variance of \( y_{ijk} \) is simply the summation of \( \delta^2_v \), \( \delta^2_0 \), and \( V(\epsilon_{ijk}) \) (the variance of \( \epsilon_{ijk} \)), because of pairwise independence assumption among \( v_{0ij}, v_{0i}, \) and \( \epsilon_{ijk} \) given \( \omega_{ij} \).

\[ V(y_{ijk}) = V(x'_{ijk} \beta + v_{0ij} + v_{0i} + \epsilon_{ijk}) = \delta^2_v + \delta^2_0 + V(\epsilon_{ijk}) \]

Based on the law of total variance, \( V(\epsilon_{ijk}) \), the variance of \( \epsilon_{ijk} \) can be derived by

\[ V(\epsilon_{ijk}) = E \left( V(\epsilon_{ijk} | \omega_{ij}) \right) + V \left( E(\epsilon_{ijk} | \omega_{ij}) \right) \]

\[ = E \left( \delta^2_{\epsilon_{ijk}} \right) = E \left( e^{u'_{ijk} \tau + \omega_{ij}} \right) = e^{u'_{ijk} \tau} E(e^{\omega_{ij}}) = e^{u'_{ijk} \tau + \delta_{\omega_{ijk}}^2 / 2} = e^{u'_{ijk} \tau + \epsilon_{ijk}^2 / 2} \]

Hence,

\[ V(y_{ijk}) = \delta^2_v + \delta^2_0 + e^{u'_{ijk} \tau + \epsilon_{ijk}^2 / 2} \]

However, the marginal distribution of \( y_{ijk} \) is not normal distributed any more even though \( v_{0i}, v_{0ij}, \omega_{ij}, \epsilon_{ijk} | \omega_{ij} \) are all assumed as normal distributions. The \( \epsilon_{ijk} \) can be expressed as the product of \( \delta_{\epsilon_{ijk}} \) and a standard normal distributed variable \( e_{ijk} \sim N(0, 1) \), where \( \delta_{\epsilon_{ijk}} \) is
independent with $e_{ijk}$ and follows log normal $LN\left(\frac{1}{2}u'_{ijk}\tau, \frac{1}{2}e^{-\frac{1}{2}e^2_{ijk}}\right)$. Obviously, $\varepsilon_{ijk}$ is no longer normal-distributed. Therefore, the marginal distribution of $y_{ijk}$ is actually a complex summation of normal distributions, and the product of a normal distribution and a log normal distribution, with the mean and variance, $x'_{ijk}\beta$ and $\delta_v^2 + e^{u'_{ijk}\tau + e^2_{ijk}}/2$ respectively, written as

$$y_{ijk}\sim\varphi\left(x'_{ijk}\beta, \delta_v^2 + e^{u'_{ijk}\tau + e^2_{ijk}}/2\right)$$

EQ. II-22

The marginal distribution of $y_{ijk}$ can be also derived from conditional distribution $y_{ijk}|v_{0i}, v_{0ij}, \omega_{ij}$ by integrating out random effects $v_{0i}, v_{0ij}, \omega_{ij}$. The conditional distribution of $y_{ijk}$, given $v_{0i}, v_{0ij}, \omega_{ij}$, was provided in EQ. II-21.

Let $\delta_i = (v_{0i}, v_{0ij}, \omega_{ij})'$, and $\Sigma = \begin{bmatrix} \delta_v^2 & 0 & 0 \\ 0 & \delta_{v\omega} & \delta_{\omega} \\ 0 & \delta_{\omega} & \delta_{\omega_{ijk}} \end{bmatrix}$.

The marginal density function of $y_{ijk}$ is

$$f(y_{ijk}) = \int_{-\infty}^{+\infty} f(y_{ijk}|\delta_i) f(\delta_i) d\delta_i = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\delta_i'\Sigma^{-1}\delta_i} \frac{1}{\sqrt{2\pi}\delta_{\varepsilon_{ijk}}} e^{-\frac{1}{2}(y_{ijk}-(x'_{ijk}\beta+v_{0ij}+v_{0i})^2)} \frac{1}{\sqrt{2\pi}\delta_{\varepsilon_{ijk}}} e^{-\frac{1}{2}u'_{ijk}\tau + e^2_{ijk}} d\delta_i$$

$$= \int_{-\infty}^{+\infty} e^{-\frac{1}{2}u'_{ijk}\tau + e^2_{ijk}} \frac{1}{\sqrt{2\pi}\delta_{\varepsilon_{ijk}}} e^{-\frac{1}{2}\delta_i'\Sigma^{-1}\delta_i} d\delta_i$$

The marginal distribution of $y_{ijk}$ is no longer normal density function after integrating out $v_{0i}, v_{0ij}, \omega_{ij}$. It is obviously no longer normal distribution. Instead it has a complex form of
summation of a normal variable $N \left( x'_{ijk}\beta, \delta_0^2 + \delta_{ij}^2 \right)$ and the product of a standard normal variable $e_{ijk}$ and a log-normal variable $LN \left( \frac{1}{2} u'_{ijk} \tau, \frac{1}{2} e^{z'_{ijk}a} \right)$ as discussed before. The covariance of outcome measurements is of the form:

$$\text{Cov}(y_{ijk}, y'_{ij'k'}) = \begin{cases} 0 & (i \neq i') \\ \delta_0^2 & (i = i', j \neq j') \\ \delta_0^2 + \delta_{ij}^2 + e^{u'_{ijk} \tau + e^{z'_{ijk}a}/2} & (i = i', j = j', k \neq k') \end{cases}$$

The intra-class correlation (ICC) is thus expressed as follows.

$$\text{ICC} = \text{Corr}(y_{ijk}, y'_{ij'k'})$$

$$= \begin{cases} 0 & (i \neq i') \\ \frac{\delta_0^2}{\sqrt{\delta_0^2 + \delta_{ij}^2 + e^{u'_{ijk} \tau + e^{z'_{ijk}a}/2}}} & (i = i', j \neq j') \\ \frac{\delta_0^2}{\sqrt{\delta_0^2 + \delta_{ij}^2 + e^{u'_{ijk} \tau + e^{z'_{ijk}a}/2}}} & (i = i', j = j', k \neq k') \\ 1 & (i = i', j = j', k = k') \end{cases}$$

I. **Random Effects in Standardized Form**

In order to enhance the numerical solution efficiency of likelihood estimation, Cholesky factorization is used to decompose the random effect vector $\delta_i = \begin{bmatrix} u_{0i} \\ u_{0ij} \\ \omega_{ij} \end{bmatrix}$ into a product of standardized multivariate normal vector $\theta_i$ and a lower triangular matrix $S_i$, 

\[ \begin{align*} 
\text{Cov}(y_{ijk}, y'_{ij'k'}) &= \begin{cases} 0 & (i \neq i') \\ \delta_0^2 & (i = i', j \neq j') \\ \delta_0^2 + \delta_{ij}^2 + e^{u'_{ijk} \tau + e^{z'_{ijk}a}/2} & (i = i', j = j', k \neq k') \end{cases} \\
\text{ICC} &= \text{Corr}(y_{ijk}, y'_{ij'k'}) \\
&= \begin{cases} 0 & (i \neq i') \\ \frac{\delta_0^2}{\sqrt{\delta_0^2 + \delta_{ij}^2 + e^{u'_{ijk} \tau + e^{z'_{ijk}a}/2}}} & (i = i', j \neq j') \\ \frac{\delta_0^2}{\sqrt{\delta_0^2 + \delta_{ij}^2 + e^{u'_{ijk} \tau + e^{z'_{ijk}a}/2}}} & (i = i', j = j', k \neq k') \\ 1 & (i = i', j = j', k = k') \end{cases} \\
\end{align*} \]
\[ \boldsymbol{\delta}_i = \begin{bmatrix} v_{0i} \\ v_{0ij} \\ \omega_{ij} \end{bmatrix} = \boldsymbol{S}_i \boldsymbol{\theta}_i = \begin{bmatrix} s_y \\ m_1 \\ s_1 \\ m_2 \\ s_2 \\ s_{ijk} \end{bmatrix} \begin{bmatrix} \theta_{0i} \\ \theta_{0ij} \\ \theta_{0ij} \end{bmatrix} \]

where

\[ \boldsymbol{\delta}_i = \begin{bmatrix} v_{0i} \\ v_{0ij} \\ \omega_{ij} \end{bmatrix}, \quad \boldsymbol{S}_i = \begin{bmatrix} s_y & 0 & 0 \\ m_1 & s_1 & 0 \\ m_2 & s_2 & s_{ijk} \end{bmatrix}, \quad \boldsymbol{\theta}_i = \begin{bmatrix} \theta_{0i} \\ \theta_{0ij} \\ \theta_{0ij} \end{bmatrix} \sim N(0, \mathbf{I}_3). \]

As specified in Eq. II-19, the variance of \( \boldsymbol{\delta}_i \),

\[ \text{Var}(\boldsymbol{\delta}_i) = \begin{bmatrix} \delta_0^2 & 0 & 0 \\ 0 & \delta_\nu^2 & \delta_{\nu\omega} \\ 0 & \delta_{\nu\omega} & \delta_{\nu\omega}^2 \end{bmatrix} = \begin{bmatrix} s_y & 0 & 0 \\ m_1 & s_1 & 0 \\ m_2 & s_2 & s_{ijk} \end{bmatrix} \]

The elements in lower triangular matrix \( \boldsymbol{S}_i \) can be solved:

\[ s_y = \delta_0, s_1 = \delta_\nu, s_2 = s_{ijk}, \quad s_{ijk} = \sqrt{\delta_{\nu\omega}^2 - \delta_{\nu\omega}^2 / \delta_\nu^2} = \sqrt{e^{z_{ijk}^2} - \delta_{\nu\omega}^2 / \delta_\nu^2}. \]

The Cholesky decomposition is then represented in the form of

\[ \boldsymbol{\delta}_i = \begin{bmatrix} v_{0i} \\ v_{0ij} \\ \omega_{ij} \end{bmatrix} = \begin{bmatrix} s_y & 0 & 0 \\ m_1 & s_1 & 0 \\ m_2 & s_2 & s_{ijk} \end{bmatrix} \begin{bmatrix} \theta_{0i} \\ \theta_{0ij} \\ \theta_{0ij} \end{bmatrix} \]

The model is now re-parameterized with the Cholesky decomposed components.

\[ y_{ijk} = x_{ijk}^t \beta + v_{0ij} + v_{0i} + e_{ijk} \]

\[ = x_{ijk}^t \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i} + e \left( \frac{1}{u_{ijk}^*} \frac{\delta_v \theta_{0ij}}{\delta_\nu} + \sqrt{\frac{e^{z_{ijk}^2}}{\delta_\nu^2} - \frac{\delta_{\nu\omega}^2}{\delta_\nu^2}} \right), \quad \text{Eq. II-23} \]

The conditional distribution of \( y_{ijk} \) given \( \boldsymbol{\theta}_i = (\theta_{0i}, \theta_{0ij}, \theta_{0ij})' \), is

\[ y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{0ij} \quad \text{Eq. II-24} \]
\begin{equation*}
\sim N \left( x'_{ijk} \beta + \delta_u \theta_{0ij} + \delta_0 \theta_{0i}, e^{u'_{ijk} + \frac{\delta_u \theta_{0ij}}{\delta_u}} + e^{u'_{ijk} + \frac{\delta_0 \theta_{0i}}{\delta_0}} \right)
\end{equation*}

The marginal distribution of $y_{ijk}$ has the following form with the mean $x'_{ijk} \beta$, and the variance $\delta_u^2 + \delta_0^2 + e^{u'_{ijk} + \frac{\delta_u \theta_{0ij}}{\delta_u} + \frac{\delta_0 \theta_{0i}}{\delta_0}}$.

$$y_{ijk} \sim \varphi \left( x'_{ijk} \beta, \delta_u^2 + \delta_0^2 + e^{u'_{ijk} + \frac{\delta_u \theta_{0ij}}{\delta_u} + \frac{\delta_0 \theta_{0i}}{\delta_0}} \right)$$

The expression of marginal distribution is the same as that derived in EQ. II-22. But it is not a normal distribution as discussed before.

**J. Maximum Marginal Likelihood Estimation and Empirical Bayes Method**

The fixed-effect parameters are estimated by the MML method. The detailed process of MML is described in this section. The marginal distribution is derived from the conditional distribution of $y_i$ given the standardized random effects by integrating out of random effects. The maximization is carried out by taking the first derivative and the second derivative with respect to log marginal likelihood as described in sections K and L. The integrals in the derivation process are numerically approximated by Gauss-Hermite Quadrature approach in 0. The random effect parameters are derived from the posterior distribution of random effects given $y_i$ by the EB approach.

Continue with the re-parameterized model in Cholesky decomposition form in I. Let $\theta^*_{oi} = (\theta_{0i}, \theta_{0i1}, \theta_{0i2}, \ldots \theta_{0in_i}, \theta_{0i1}, \theta_{0i2}, \ldots \theta_{0in_i})'$, denote all random effect variables for subject $i$, where $\theta_{0i}$ is the subject-level random location effect for subject $i$, $\theta^*_{oi} = (\theta_{0i1}, \theta_{0i2}, \ldots \theta_{0in_i})'$ is a vector including all wave-level random location effects for $ith$ subject.
and \( \boldsymbol{\theta}_{oi} = (\theta_{o1i}, \theta_{o2i}, \ldots, \theta_{oin_i})' \) is a vector including all wave-level random scale effects for \( i \) th subject. The \( \boldsymbol{\theta}_i^* \) follows a \( 2n_i + 1 \) dimensional standard multivariate normal distribution \( N(0, I_{2n_i+1}) \).

Let \( \mathbf{y}_i = (y_{11} \ldots y_{1n_i}, y_{i1} \ldots y_{in_i})' \) denote the vector of all outcome measures for subject \( i \). The marginal distribution of \( \mathbf{y}_i \) can be derived from the conditional distribution \( \mathbf{y}_i | \boldsymbol{\theta}_i^* \) by integrating out random effects \( \theta_{0i}, \theta_{0i}^*, \) and \( \theta_{oi}^* \).

\[
 f(\mathbf{y}_i) = \int_{\theta_{oi}} f(\mathbf{y}_i | \theta_{0i}) f(\theta_{0i}) d\theta_{0i}
\]

\[
 = \int_{\theta_{oi}} \left[ \int_{\theta_{oi}^*} f(\mathbf{y}_i | \theta_{0i}, \theta_{0i}^*, \theta_{oi}^*) f(\theta_{0i}) f(\theta_{0i}^*) d\theta_{0i} d\theta_{0i}^* \right] f(\theta_{0i}) d\theta_{0i}
\]

\[
 = \int_{\theta_{oi}} \int_{\theta_{oi1}, \ldots, \theta_{oin_i}} f(\mathbf{y}_i | \theta_{0i}, \theta_{0i1}, \ldots, \theta_{oin_i}, \theta_{oi1}, \ldots, \theta_{oin_i}) d(\theta_{0i1}, \theta_{0i2}, \ldots, \theta_{oin_i}, \theta_{oi1}, \theta_{oi2}, \ldots, \theta_{oin_i}) f(\theta_{0i}) d\theta_{0i}
\]

Given subject-level random effect \( \theta_{0i} \), the responses at different waves for \( i \)th subject are independent with each other. Therefore,

\[
 f(\mathbf{y}_i | \theta_{0i}) = \int_{\theta_{oi1}, \ldots, \theta_{oin_i}} f(\mathbf{y}_i | \theta_{0i}, \theta_{oi1}, \ldots, \theta_{oin_i}) f(\theta_{0i1}, \ldots, \theta_{oin_i}, \theta_{oi1}, \ldots, \theta_{oin_i}) d(\theta_{0i1}, \ldots, \theta_{oin_i}, \theta_{oi1}, \ldots, \theta_{oin_i})
\]

\[
 = \prod_{j=1}^{n_i} \int_{\theta_{oi1}, \theta_{oj1}} f(\mathbf{y}_{ij} | \theta_{0ij}, \theta_{oji}, \theta_{oij}) f(\theta_{0ij}, \theta_{oij}) d(\theta_{0ij}, \theta_{oij})
\]
Also, conditional on subject-level and wave-level random effects \( \theta_{0i}, \theta_{0ij}, \theta_{\omega ij} \), all \( n_{ij} \) responses for \( i \)th subject and \( j \)th are mutually independent. That is,

\[
f(y_{ij}|\theta_{0i}, \theta_{0ij}, \theta_{\omega ij}) = \prod_{k=1}^{n_{ij}} f(y_{ijk}|\theta_{0i}, \theta_{0ij}, \theta_{\omega ij})
\]

Therefore, the marginal distribution of \( y_i \) is written as

\[
f(y_i) = \int_{\theta_{0i}} \left[ \prod_{j=1}^{n_i} \int_{\theta_{0ij}\theta_{\omega ij}}^{n_{ij}} f(y_{ij}|\theta_{0i}, \theta_{0ij}, \theta_{\omega ij}) f(\theta_{0ij}, \theta_{\omega ij}) d(\theta_{0ij}, \theta_{\omega ij}) \right] f(\theta_{0i}) d\theta_{0i}
\]

\[
= \int_{\theta_{0i}} \left[ \prod_{j=1}^{n_i} \int_{\theta_{0ij}\theta_{\omega ij}}^{n_{ij}} f(y_{ijk}|\theta_{0i}, \theta_{0ij}, \theta_{\omega ij}) \prod_{k=1}^{n_{ij}} f(\theta_{0ij}, \theta_{\omega ij}) d(\theta_{0ij}, \theta_{\omega ij}) \right] f(\theta_{0i}) d\theta_{0i}
\]

The log likelihood for all subjects is simply the summation of log likelihood of each subject. \( \log L(\beta, \delta_{\omega}, \delta_{\omega}, \alpha, \tau) \) is short as \( \log L \) thereafter.

\[
\log L(\beta, \delta_{\omega}, \delta_{\omega}, \alpha, \tau) = \sum_{i=1}^{n} \log f(y_i).
\]

As stated in EQ. II-4, the random effect \( \theta_i \) and can be estimated using the EB approach.

\[
\tilde{\theta}_i = E(\theta_i|y_i)
\]

\[
= \int_{\theta_i} \theta_i f(\theta_i|y_i) d\theta_i = \int_{\theta_i} \theta_i \frac{f(y_i|\theta_i)f(\theta_i)}{f(y_i)} d\theta_i = \frac{1}{f(y_i)} E(\theta_i f(y_i|\theta_i))
\]

\[
= \frac{1}{f(y_i)} E(\theta_i \sum_{j} \sum_{k} f(y_{ijk}|\theta_i))
\]

EQ. II-25

and

\[
\hat{V}(\tilde{\theta}_i) = V(\theta_i|y_i)
\]

\[
= \int_{\theta_i} (\theta_i - \tilde{\theta}_i)(\theta_i - \tilde{\theta}_i)' f(\theta_i|y_i) d\theta_i = \int_{\theta_i} (\theta_i - \tilde{\theta}_i)(\theta_i - \tilde{\theta}_i)' \frac{f(y_i|\theta_i)f(\theta_i)}{f(y_i)} d\theta_i
\]

\[
= \frac{1}{f(y_i)} E(((\theta_i - \tilde{\theta}_i)(\theta_i - \tilde{\theta}_i)' f(y_i|\theta_i)) = \frac{1}{f(y_i)} E(((\theta_i - \tilde{\theta}_i)(\theta_i - \tilde{\theta}_i)' f(y_i|\theta_i))
\]

29
K. **First Derivative of Log Likelihood**

Let \( \lambda = (\beta, \delta_0, \delta_{01}, \alpha, \tau)' \) denote the unknown parameter vector to be estimated. The first derivative of log likelihood \( \log L \) with respect to \( \lambda \) is

\[
\frac{\partial \log L}{\partial \lambda} = \sum_{i=1}^{n} \frac{\partial \log(f(y_i))}{\partial \lambda} = \sum_{i=1}^{n} \frac{1}{f(y_i)} \frac{\partial f(y_i)}{\partial \lambda}
\]

where

\[
\frac{\partial f(y_i)}{\partial \lambda} = \int_{\theta_{oi}} f(y_i|\theta_{oi}) \frac{\partial \log(f(y_i|\theta_{oi}))}{\partial \lambda} f(\theta_{oi}) d\theta_{oi}
\]

\[
= \int_{\theta_{oi}} f(y_i|\theta_{oi}) \left[ \sum_{j=1}^{n_i} \frac{1}{f(y_{ij}|\theta_{oji})} \frac{\partial f(y_{ij}|\theta_{oi})}{\partial \lambda} \right] f(\theta_{oi}) d\theta_{oi}
\]

\[
= \int_{\theta_{oi}} f(y_i|\theta_{oi}) \left[ \sum_{j=1}^{n_i} \frac{1}{f(y_{ij}|\theta_{oi})} \int_{\theta_{oij}} \frac{\partial f(y_{ij}|\theta_{oi}, \theta_{oij}, \theta_{oij})}{\partial \lambda} f(\theta_{oi}), \theta_{oij}) d(\theta_{oij}, \theta_{oij}) \right] f(\theta_{oi}) d\theta_{oi}
\]

\[
= \int_{\theta_{oi}} f(y_i|\theta_{oi}) \left[ \sum_{j=1}^{n_i} \frac{1}{f(y_{ij}|\theta_{oi})} \int_{\theta_{oij}} f(y_{ij}|\theta_{oi}, \theta_{oij}, \theta_{oij}) L G f(\theta_{oi}), \theta_{oij}) d(\theta_{oij}, \theta_{oij}) \right] f(\theta_{oi}) d\theta_{oi}
\]

\[
= \int_{\theta_{oi}} f(y_i|\theta_{oi}) \left[ \sum_{j=1}^{n_i} \frac{1}{f(y_{ij}|\theta_{oi})} \int_{\theta_{oij}} f(y_{ij}|\theta_{oi}, \theta_{oij}, \theta_{oij}) \left( \sum_{k=1}^{n_{ij}} L G \right) f(\theta_{oi}), \theta_{oij}) d(\theta_{oij}, \theta_{oij}) \right] f(\theta_{oi}) d\theta_{oi}
\]

\[
= \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} \left[ \int_{\theta_{oij}} \int_{\theta_{oij}} \frac{f(y_{ij}|\theta_{oi}, \theta_{oij}, \theta_{oij})}{f(y_{ij}|\theta_{oi})} L G f(\theta_{oi}), \theta_{oij}) f(\theta_{oi}) d(\theta_{oij}, \theta_{oij}) d\theta_{oi} \right]
\]
The first derivative of log density \( \frac{\partial \log(f(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{0ij}))}{\partial \lambda} \) can be derived by the following steps. As shown in EQ. II-24, \( y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{0ij} \) follows normal distribution

\[
y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{0ij} \sim N \left( x'_{ijk} \beta + \delta_{i} \theta_{0ij} + \delta_{0} \theta_{0ij} e^{u'_{ijk} + \delta_{uo} \theta_{0ij} + \delta_{0} \theta_{0ij}} \right)
\]

The kernel of log density function is of the form:

\[
\log(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{0ij}) = \log \delta^2_{\epsilon_{ijk}} + \frac{\epsilon^2_{ijk}}{\delta^2_{\epsilon_{ijk}}}
\]

\[
\delta^2_{\epsilon_{ijk}} = e^{u'_{ijk} + \delta_{uo} \theta_{0ij} + \delta_{0} \theta_{0ij}}
\]

where

\[
\epsilon_{ijk} = y_{ijk} - (x'_{ijk} \beta + \delta_{i} \theta_{0ij} + \delta_{0} \theta_{0ij})
\]

The first derivative of log conditional density with respect to \( \lambda \) is

\[
\frac{\partial \log(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{0ij})}{\partial \lambda}
\]

\[
\frac{1}{\delta^2_{\epsilon_{ijk}}} \frac{\partial \delta^2_{\epsilon_{ijk}}}{\partial \lambda} + 2 \frac{\epsilon_{ijk} \delta_{\epsilon_{ijk}}}{\delta^2_{\epsilon_{ijk}}} \frac{\partial \delta_{\epsilon_{ijk}}}{\partial \lambda} - \frac{\epsilon^2_{ijk}}{\delta^2_{\epsilon_{ijk}}} \frac{\partial \delta^2_{\epsilon_{ijk}}}{\partial \lambda}
\]

\[
= \frac{1}{\delta^2_{\epsilon_{ijk}}} \left( 1 - \frac{\epsilon_{ijk}}{\delta^2_{\epsilon_{ijk}}} \right) \frac{\partial \delta^2_{\epsilon_{ijk}}}{\partial \lambda} + 2 \frac{\epsilon_{ijk} \delta_{\epsilon_{ijk}}}{\delta^2_{\epsilon_{ijk}}} \frac{\partial \delta_{\epsilon_{ijk}}}{\partial \lambda}
\]

with

\[
\frac{\partial \delta^2_{\epsilon_{ijk}}}{\partial \lambda} = \frac{u'_{ijk} + \delta_{uo} \theta_{0ij} + \delta_{0} \theta_{0ij}}{-\delta^2_{\epsilon_{ijk}}}
\]

31
\[
\begin{align*}
\delta_{e_{ijk}} u'_{ijk} &= \begin{cases} \\
0 \\
\frac{\delta^2_{e_{ijk}}}{\delta_{\epsilon}} \left( -\frac{z_{ijk}^t \theta_{oij} \delta_{o\omega} z_{ijk}^\alpha}{2 \sqrt{e^{z_{ijk}^\alpha} - \delta_{o\omega}^2 / \delta_{\epsilon}^2}} \right) \\
0 \\
\frac{\delta^2_{e_{ijk}}}{\delta_{\epsilon}} \left( -\frac{\theta_{oij}}{\delta_{o\omega}} + \frac{\theta_{oij}}{\delta_{o\omega}} \frac{\delta_{o\omega}}{\delta_{\epsilon}} \frac{\delta_{o\omega}^2}{\delta_{\epsilon}} \right) \\
\frac{\delta^2_{e_{ijk}}}{\delta_{\epsilon}} \left( -\frac{\delta_{o\omega}^2 \theta_{oij}}{\delta_{o\omega}} + \frac{\theta_{oij}}{\delta_{o\omega}} \frac{\delta_{o\omega}^2}{\delta_{\epsilon}} \frac{\delta_{o\omega}^2}{\delta_{\epsilon}} \right) \\
\delta_{\epsilon} \right) \\
\lambda = \tau \\
\lambda = \beta \\
\lambda = \alpha \\
\lambda = \delta_0 \\
\lambda = \delta_{v\omega} \\
\lambda = \delta_{v\omega} \\
\lambda = \delta_{v}
\end{cases}
\end{align*}
\]
\[
\frac{\partial \varepsilon_{ijk}}{\partial \lambda} = \frac{\partial}{\partial \lambda} \left( y_{ijk} - (x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i}) \right) = \begin{cases} 
0 & \lambda = \tau \\
-x_{ijk}' & \lambda = \beta \\
0 & \lambda = \alpha \\
-\theta_{0i} & \lambda = \delta_0 \\
0 & \lambda = \delta_{vo} \\
-\theta_{0ij} & \lambda = \delta_v 
\end{cases}
\]

Hence,

\[
\frac{\partial \log(y_{ijk}|\theta_{0i}, \theta_{0ij}, \theta_{wij})}{\partial \lambda} = \begin{cases} 
(1 - \frac{y_{ijk} - (x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i})}{e^{u'_{ijk} + \frac{\delta_{vo} \theta_{0ij} + \sqrt{z'_{ijk}^2 - \delta_{vo}^2 \theta_{wij}}}{\delta_v}}} \cdot u'_{ijk}' \right) & \text{if } \lambda = \tau \\
-2x_{ijk}' \frac{y_{ijk} - (x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i})}{e^{u'_{ijk} + \frac{\delta_{vo} \theta_{0ij} + \sqrt{z'_{ijk}^2 - \delta_{vo}^2 \theta_{wij}}}{\delta_v}}} & \text{if } \lambda = \beta \\
(1 - \frac{y_{ijk} - (x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i})}{e^{u'_{ijk} + \frac{\delta_{vo} \theta_{0ij} + \sqrt{z'_{ijk}^2 - \delta_{vo}^2 \theta_{wij}}}{\delta_v}}} \cdot -\frac{z_{ijk}^2 \theta_{wij} e^{z_{ijk}'}}{2^{\frac{1}{2} \sqrt{z_{ijk}'^2 - \delta_{vo}^2}}} & \text{if } \lambda = \alpha \\
-2\theta_{0i}' \frac{y_{ijk} - (x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i})}{e^{u'_{ijk} + \frac{\delta_{vo} \theta_{0ij} + \sqrt{z'_{ijk}^2 - \delta_{vo}^2 \theta_{wij}}}{\delta_v}}} & \text{if } \lambda = \delta_0 
\end{cases}
\]

\text{EQ. II-26}
\[
\begin{align*}
&\left(1 - y_{ijk} - \left(x'_{ijk}\beta + \delta_v\theta_{0ij} + \delta_0\theta_{0i}\right)\right) \left(\frac{\theta_{0ij} - \theta_{wij}}{\delta_v} - \frac{\theta_{wij}}{\delta_v} \delta_{w\omega} \delta^2_v\right), \quad \text{if } \lambda = \delta_{w\omega} \\
&\left(-\frac{\delta_{w\omega}\theta_{0ij}}{\delta^2_v} + \frac{\theta_{wij}\delta^2_{w\omega}}{\delta^3_v e^{z'_{ijk}a} - \delta^2_{w\omega}/\delta^2_v} - \left(-\frac{\delta_{w\omega}\theta_{0ij}}{\delta^2_v} + \frac{\theta_{wij}\delta^2_{w\omega}}{\delta^3_v e^{z'_{ijk}a} - \delta^2_{w\omega}/\delta^2_v} - 2\theta_{0ij}\right) y_{ijk} - \left(x'_{ijk}\beta + \delta_v\theta_{0ij} + \delta_0\theta_{0i}\right)\right) e^{u'_{ijk}\tau + \delta_{w\omega}\theta_{0ij} + \sqrt{e^{z'_{ijk}a} - \delta^2_{w\omega}/\delta^2_v}}, \\
&\text{if } \lambda = \delta_v
\end{align*}
\]

The overall first derivative of \( \log L \) with respect to \( \lambda \) is

\[
\frac{\partial \log L}{\partial \lambda} = \sum_{i=1}^{n} \frac{1}{f(y_i)} \frac{\partial f(y_i)}{\partial \lambda} = \sum_{i=1}^{n} \frac{1}{f(y_i)} \int_{\theta_{0i}} f(y_i|\theta_0) \left[ \sum_{j=1}^{n_i} \frac{1}{f(y_{ij}|\theta_{0ij})} \int_{\theta_{wij}} f(y_{ij}|\theta_{0i},\theta_{0ij},\theta_{wij}) \left( \sum_{k=1}^{n_{ij}} LG \right) f(\theta_{0ij},\theta_{wij}) d\theta_{0ij} \right] f(\theta_{0i}) d\theta_{0i} \quad \text{EQ. II-27}
\]

The part \( LG = \frac{\partial \log(f(y_{ijk}|\theta_{0i},\theta_{0ij},\theta_{wij}))}{\partial \lambda} \) was derived in EQ. II-27.

L. **Second Derivative of Log Likelihood**

\[
\frac{\partial^2 \log L}{\partial \lambda \partial \eta} = \sum_{i=1}^{n} \frac{\partial^2 \log(f(y_i))}{\partial \lambda \partial \eta} = \sum_{i=1}^{n} \frac{1}{f^2(y_i)} \frac{\partial f(y_i)}{\partial \lambda \partial \eta} + \frac{1}{f(y_i)} \frac{\partial^2 f(y_i)}{\partial \lambda \partial \eta}
\]
As derived in EQ. II-27, the first derive is of the form:

\[
\frac{\partial f(y_i)}{\partial \lambda} = \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} \left[ \int_{\theta_{oi}}^{\theta_{oij}} \int_{\theta_{oij}}^{\theta_{oij}} \frac{f(y_i|\theta_{oi}) f(y_{ij}|\theta_{oi}, \theta_{oij}, \theta_{oij})}{f(y_{ij}|\theta_{oi})} L\Gamma f(\theta_{oij}, \theta_{oij}) f(\theta_{oi}) d\theta_{oij} d\theta_{oi} \right]
\]

The second derive of \( f(y_i) \) with respect to \( \lambda \) and \( \eta \) is

\[
\frac{\partial^2 f(y_i)}{\partial \lambda \partial \eta} = \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} \left[ \int_{\theta_{oi}}^{\theta_{oij}} \int_{\theta_{oij}}^{\theta_{oij}} \frac{\frac{\partial}{\partial \eta} \left( \frac{f(y_i|\theta_{oi}) f(y_{ij}|\theta_{oi}, \theta_{oij}, \theta_{oij})}{f(y_{ij}|\theta_{oi})} \right) \log \left( f(y_{ijk}|\theta_{oi}, \theta_{oij}, \theta_{oij}) \right) \right] \frac{\partial \log \left( f(y_{ijk}|\theta_{oi}, \theta_{oij}, \theta_{oij}) \right)}{\partial \lambda} d\theta_{oij} d\theta_{oi}
\]

\[
= \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} \left[ \int_{\theta_{oi}}^{\theta_{oij}} \int_{\theta_{oij}}^{\theta_{oij}} \frac{\partial f(y_i|\theta_{oi})}{\partial \eta} \frac{\partial f(y_{ij}|\theta_{oi}, \theta_{oij}, \theta_{oij})}{\partial \eta} \log \left( f(y_{ijk}|\theta_{oi}, \theta_{oij}, \theta_{oij}) \right) \right] \frac{\partial \log \left( f(y_{ijk}|\theta_{oi}, \theta_{oij}, \theta_{oij}) \right)}{\partial \lambda} d\theta_{oij} d\theta_{oi}
\]

\[
+ \frac{f(y_i|\theta_{oi})}{f(y_{ij}|\theta_{oi})} \frac{\partial f(y_{ij}|\theta_{oi}, \theta_{oij}, \theta_{oij})}{\partial \eta} \log \left( f(y_{ijk}|\theta_{oi}, \theta_{oij}, \theta_{oij}) \right) \frac{\partial^2 \log \left( f(y_{ijk}|\theta_{oi}, \theta_{oij}, \theta_{oij}) \right)}{\partial \lambda^2} d\theta_{oij} d\theta_{oi}
\]

\[
- \frac{f(y_i|\theta_{oi}) f(y_{ij}|\theta_{oi}, \theta_{oij}, \theta_{oij})}{f^2(y_{ij}|\theta_{oi})} \frac{\partial f(y_{ij}|\theta_{oi})}{\partial \eta} \frac{\partial \log \left( f(y_{ijk}|\theta_{oi}, \theta_{oij}, \theta_{oij}) \right)}{\partial \lambda} d\theta_{oij} d\theta_{oi}
\]

where
\[
\frac{\partial f(y_{ij}|\theta_{oi})}{\partial \eta} = \frac{\sum_{k=1}^{n_{ij}} \frac{\partial f(y_{ijk}|\theta_{oi})}{\partial \eta}}{f(y_{ijk}|\theta_{oi})} = \frac{\sum_{i=1}^{n_{oi}} \sum_{j=1}^{n_{ij}} \frac{1}{f(y_{ijk}|\theta_{oi},\theta_{oij},\theta_{wij})} \partial \log \left( f(y_{ijk}|\theta_{oi},\theta_{oij},\theta_{wij}) \right) / \partial \eta}{f(\theta_{oij},\theta_{wij}) d\theta_{oij} d\theta_{wij}}
\]

\[
\frac{\partial f(y_i|\theta_{oi})}{\partial \eta} = \frac{\partial \sum_{j=1}^{n_{oi}} \log(f(y_{ij}|\theta_{oi}))}{\partial \eta} = \frac{1}{f(y_i|\theta_{oi})} \frac{\sum_{j=1}^{n_{oi}} f(y_{ij}|\theta_{oi})}{\partial \eta}
\]

\[
= \frac{1}{f(y_i|\theta_{oi})} \sum_{j=1}^{n_{oi}} \frac{1}{f(y_{ij}|\theta_{oi})} \frac{\sum_{k=1}^{n_{ijk}} \frac{1}{f(y_{ijk}|\theta_{oi},\theta_{oij},\theta_{wij})} \partial \log \left( f(y_{ijk}|\theta_{oi},\theta_{oij},\theta_{wij}) \right) / \partial \eta}{f(\theta_{oij},\theta_{wij}) d\theta_{oij} d\theta_{wij}}
\]

\[
\frac{\partial f(y_{ij}|\theta_{oi},\theta_{oij},\theta_{wij})}{\partial \eta} = \sum_{k=1}^{n_{ijk}} \frac{\partial f(y_{ijk}|\theta_{oi},\theta_{oij},\theta_{wij})}{\partial \eta} = \sum_{k=1}^{n_{ijk}} \frac{f(y_{ijk}|\theta_{oi},\theta_{oij},\theta_{wij})}{\partial \eta} \partial \log \left( f(y_{ijk}|\theta_{oi},\theta_{oij},\theta_{wij}) \right)
\]

\[
\frac{\partial^2 \log \left( f(y_{ijk}|\theta_{oi},\theta_{oij},\theta_{wij}) \right)}{\partial \lambda \partial \eta} = \frac{\partial^2 \left( \log \delta_{\epsilon_{ijk}}^2 + \delta_{\epsilon_{ijk}}^2 \right)}{\partial \lambda \partial \eta}
\]

\[
= \frac{1}{2\delta_{\epsilon_{ijk}}^4} \left( 1 - \frac{2\delta_{\epsilon_{ijk}}^2}{\delta_{\epsilon_{ijk}}^2} \right) \frac{\partial^2 \delta_{\epsilon_{ijk}}^2}{\partial \lambda \partial \eta} + \frac{1}{2\delta_{\epsilon_{ijk}}^2 \delta_{\epsilon_{ijk}}^2} \left( \delta_{\epsilon_{ijk}}^2 - 1 \right) \frac{\partial^2 \delta_{\epsilon_{ijk}}^2}{\partial \lambda \partial \eta} - \frac{1}{\delta_{\epsilon_{ijk}}^2 \delta_{\epsilon_{ijk}}^2} \frac{\partial^2 \delta_{\epsilon_{ijk}}^2}{\partial \lambda \partial \eta}
\]

with
The second derivation of \( \log(y_{ijk}|\theta_{o_i}, \theta_{o_ij}, \theta_{o_ij}) \) with respect to \( \lambda \) and \( \eta \) is:

\[
\frac{\partial^2 \log(y_{ijk}|\theta_{o_i}, \theta_{o_ij}, \theta_{o_ij})}{\partial \tau \partial \eta}
\]
\[
\begin{align*}
\left\{ \begin{array}{l}
y_{ijk} - \left( x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i} \right) \\
e^{u_{ijk} + \delta_{\omega_0} \theta_{0ij} + \sqrt{\sum_{i'j'k'} \alpha - \frac{\delta_{\omega_0} \delta_{\omega} \theta_{ij}}{\delta_v^2}}}
\end{array} \right\} u_{ijk} = \eta \Rightarrow \eta = \tau \\
x'_{ijk} \\
e^{u_{ijk} + \delta_{\omega_0} \theta_{0ij} + \sqrt{\sum_{i'j'k'} \alpha - \frac{\delta_{\omega_0} \delta_{\omega} \theta_{ij}}{\delta_v^2}}}
\end{align*} \]

\[
\left\{ \begin{array}{l}
y_{ijk} - \left( x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i} \right) \\
e^{u_{ijk} + \delta_{\omega_0} \theta_{0ij} + \sqrt{\sum_{i'j'k'} \alpha - \frac{\delta_{\omega_0} \delta_{\omega} \theta_{ij}}{\delta_v^2}}}
\end{array} \right\} u_{ijk} = \eta \Rightarrow \eta = \beta \\
\theta_{0i} \\
e^{u_{ijk} + \delta_{\omega_0} \theta_{0ij} + \sqrt{\sum_{i'j'k'} \alpha - \frac{\delta_{\omega_0} \delta_{\omega} \theta_{ij}}{\delta_v^2}}}
\end{align*} \]

\[
\left\{ \begin{array}{l}
y_{ijk} - \left( x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i} \right) \\
e^{u_{ijk} + \delta_{\omega_0} \theta_{0ij} + \sqrt{\sum_{i'j'k'} \alpha - \frac{\delta_{\omega_0} \delta_{\omega} \theta_{ij}}{\delta_v^2}}}
\end{array} \right\} u_{ijk} = \eta \Rightarrow \eta = \alpha \\
\theta_{0ij} \\
e^{u_{ijk} + \delta_{\omega_0} \theta_{0ij} + \sqrt{\sum_{i'j'k'} \alpha - \frac{\delta_{\omega_0} \delta_{\omega} \theta_{ij}}{\delta_v^2}}}
\end{align*} \]

\[
\left\{ \begin{array}{l}
y_{ijk} - \left( x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i} \right) \\
e^{u_{ijk} + \delta_{\omega_0} \theta_{0ij} + \sqrt{\sum_{i'j'k'} \alpha - \frac{\delta_{\omega_0} \delta_{\omega} \theta_{ij}}{\delta_v^2}}}
\end{array} \right\} u_{ijk} = \eta \Rightarrow \eta = \delta_0 \\
\theta_{0ij} \\
e^{u_{ijk} + \delta_{\omega_0} \theta_{0ij} + \sqrt{\sum_{i'j'k'} \alpha - \frac{\delta_{\omega_0} \delta_{\omega} \theta_{ij}}{\delta_v^2}}}
\end{align*} \]

\[
\left\{ \begin{array}{l}
y_{ijk} - \left( x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i} \right) \\
e^{u_{ijk} + \delta_{\omega_0} \theta_{0ij} + \sqrt{\sum_{i'j'k'} \alpha - \frac{\delta_{\omega_0} \delta_{\omega} \theta_{ij}}{\delta_v^2}}}
\end{array} \right\} u_{ijk} = \eta \Rightarrow \eta = \delta_{\omega_0} \\
\theta_{0ij} + y_{ijk} - \left( x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i} \right) \\
e^{u_{ijk} + \delta_{\omega_0} \theta_{0ij} + \sqrt{\sum_{i'j'k'} \alpha - \frac{\delta_{\omega_0} \delta_{\omega} \theta_{ij}}{\delta_v^2}}}
\end{align*} \]

\[
\left\{ \begin{array}{l}
y_{ijk} - \left( x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i} \right) \\
e^{u_{ijk} + \delta_{\omega_0} \theta_{0ij} + \sqrt{\sum_{i'j'k'} \alpha - \frac{\delta_{\omega_0} \delta_{\omega} \theta_{ij}}{\delta_v^2}}}
\end{array} \right\} u_{ijk} = \eta \Rightarrow \eta = \delta_v \\
\theta_{0ij} + y_{ijk} - \left( x'_{ijk} \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i} \right) \\
e^{u_{ijk} + \delta_{\omega_0} \theta_{0ij} + \sqrt{\sum_{i'j'k'} \alpha - \frac{\delta_{\omega_0} \delta_{\omega} \theta_{ij}}{\delta_v^2}}}
\end{align*} \]

\[
\frac{\partial^2 \log(y_{ijk} \mid \theta_{0i}, \theta_{0ij}, \theta_{0ij})}{\partial \beta \partial \eta}
\]
\[
\begin{align*}
-2 & \left( \frac{y_{ijk} - (x_{ijk}'\beta + \delta_u\theta_{0ij} + \delta_0\theta_{oi})}{e^{u_{ijk}' + \frac{\delta_u\theta_{0ij}}{\delta_u} + \frac{\delta_0\theta_{oi}}{\delta_0}} + \frac{e^{z_{ijk}'\alpha - \delta_v\theta_{0ij}}}{\delta_v}} \right) x_{ijk}'u_{ijk}' & \eta = \tau \\
-2 & \left( \frac{x_{ijk}'}{e^{u_{ijk}' + \frac{\delta_u\theta_{0ij}}{\delta_u} + \frac{\delta_0\theta_{oi}}{\delta_0}} + \frac{e^{z_{ijk}'\alpha - \delta_v\theta_{0ij}}}{\delta_v}} \right) x_{ijk}' & \eta = \beta \\
-2 & \left( \frac{y_{ijk} - (x_{ijk}'\beta + \delta_u\theta_{0ij} + \delta_0\theta_{oi})}{e^{u_{ijk}' + \frac{\delta_u\theta_{0ij}}{\delta_u} + \frac{\delta_0\theta_{oi}}{\delta_0}} + \frac{e^{z_{ijk}'\alpha - \delta_v\theta_{0ij}}}{\delta_v}} \right) \left( \frac{\theta_{0ij} e^{z_{ijk}'\alpha} z_{ijk}'}{2 e^{z_{ijk}'\alpha - \delta_v^2/\delta_v}} \right) x_{ijk}' & \eta = \alpha \\
-2 & \left( \frac{\theta_{0ij}}{e^{u_{ijk}' + \frac{\delta_u\theta_{0ij}}{\delta_u} + \frac{\delta_0\theta_{oi}}{\delta_0}} + \frac{e^{z_{ijk}'\alpha - \delta_v\theta_{0ij}}}{\delta_v}} \right) x_{ijk}' & \eta = \delta_0 \\
-2 & \left( \frac{y_{ijk} - (x_{ijk}'\beta + \delta_u\theta_{0ij} + \delta_0\theta_{oi})}{e^{u_{ijk}' + \frac{\delta_u\theta_{0ij}}{\delta_u} + \frac{\delta_0\theta_{oi}}{\delta_0}} + \frac{e^{z_{ijk}'\alpha - \delta_v\theta_{0ij}}}{\delta_v}} \right) \left( \frac{e^{z_{ijk}'\alpha} - \delta_v^2/\delta_v}{\delta_v} \right) x_{ijk}' & \eta = \delta_v \\
-2 & \left( \frac{\theta_{0ij} + y_{ijk} - (x_{ijk}'\beta + \delta_u\theta_{0ij} + \delta_0\theta_{oi})}{e^{u_{ijk}' + \frac{\delta_u\theta_{0ij}}{\delta_u} + \frac{\delta_0\theta_{oi}}{\delta_0}} + \frac{e^{z_{ijk}'\alpha - \delta_v\theta_{0ij}}}{\delta_v}} \right) \left( \frac{-\delta_v\theta_{0ij} + \frac{\theta_{0ij}\delta_v^2}{\delta_v}}{\delta_v^3} + \frac{\theta_{0ij\delta_v^2}}{\delta_v} \right) x_{ijk}' & \eta = \delta_v \\
\end{align*}
\]

\[
\frac{\partial^2 \log(y_{ijk})}{\partial \alpha \partial \eta} = \left( \frac{y_{ijk} - (x_{ijk}'\beta + \delta_u\theta_{0ij} + \delta_0\theta_{oi})}{e^{u_{ijk}' + \frac{\delta_u\theta_{0ij}}{\delta_u} + \frac{\delta_0\theta_{oi}}{\delta_0}} + \frac{e^{z_{ijk}'\alpha - \delta_v\theta_{0ij}}}{\delta_v}} \right) \left( \frac{-z_{ijk}'\theta_{0ij} e^{z_{ijk}'\alpha}}{2 e^{z_{ijk}'\alpha - \delta_v^2/\delta_v}} \right) u_{ijk}', \quad \text{if } \eta = \tau
\]
\[ \frac{\partial^2 \log(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{\omega ij})}{\partial \alpha \partial \eta} = \left( \frac{x'_{ijk}}{e^{u'_{ijk} + \delta_{\omega ij} \delta_{\omega ij} + e^{z'_{ijk} \alpha - \delta_{\omega ij} / \delta_{\omega ij}}} + z'_{ijk} \theta_{\omega ij} e^{z'_{ijk} \alpha} \right) \left( -\frac{z'_{ijk} \theta_{\omega ij} e^{z'_{ijk} \alpha}}{2 \sqrt{e^{z'_{ijk} \alpha} - \delta^2_{\omega ij} / \delta^2_{\omega ij}}} \right), \quad \text{if } \eta = \beta \]

\[ \frac{\partial^2 \log(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{\omega ij})}{\partial \alpha \partial \eta} = \left( \frac{y_{ijk} - (x'_{ijk} \beta + \delta_{\omega ij} + \delta_{\omega ij}) \theta_{\omega ij} e^{z'_{ijk} \alpha} z'_{ijk}}{e^{u'_{ijk} + \delta_{\omega ij} \delta_{\omega ij} + e^{z'_{ijk} \alpha - \delta_{\omega ij} / \delta_{\omega ij}}} + z'_{ijk} \theta_{\omega ij} e^{z'_{ijk} \alpha} \right) \left( -\frac{z'_{ijk} \theta_{\omega ij} e^{z'_{ijk} \alpha}}{2 \sqrt{e^{z'_{ijk} \alpha} - \delta^2_{\omega ij} / \delta^2_{\omega ij}}} \right) \]

\[ + \left( 1 - \frac{y_{ijk} - (x'_{ijk} \beta + \delta_{\omega ij} + \delta_{\omega ij}) \theta_{\omega ij} e^{z'_{ijk} \alpha} z'_{ijk}}{e^{u'_{ijk} + \delta_{\omega ij} \delta_{\omega ij} + e^{z'_{ijk} \alpha - \delta_{\omega ij} / \delta_{\omega ij}}} + z'_{ijk} \theta_{\omega ij} e^{z'_{ijk} \alpha} \right) \left( -\frac{z'_{ijk} \theta_{\omega ij} e^{z'_{ijk} \alpha}}{2 \sqrt{e^{z'_{ijk} \alpha} - \delta^2_{\omega ij} / \delta^2_{\omega ij}}} + \frac{z'_{ijk} \theta_{\omega ij} e^{2z'_{ijk} \alpha} z'_{ijk}}{4 \left( e^{z'_{ijk} \alpha} - \delta^2_{\omega ij} / \delta^2_{\omega ij} \right)^{3/2}} \right), \quad \text{if } \eta = \alpha \]

\[ \frac{\partial^2 \log(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{\omega ij})}{\partial \alpha \partial \eta} = \left( \frac{\theta_{0i}}{e^{u'_{ijk} + \delta_{\omega ij} \delta_{\omega ij} + e^{z'_{ijk} \alpha - \delta_{\omega ij} / \delta_{\omega ij}}} + z'_{ijk} \theta_{\omega ij} e^{z'_{ijk} \alpha} \right) \left( -\frac{z'_{ijk} \theta_{\omega ij} e^{z'_{ijk} \alpha}}{2 \sqrt{e^{z'_{ijk} \alpha} - \delta^2_{\omega ij} / \delta^2_{\omega ij}}} \right), \quad \text{if } \eta = \delta_{0} \]
\[
\frac{\partial^2 \log(y_{ijk} | \theta_{0ij}, \theta_{0ij}, \theta_{0ij})}{\partial \alpha \partial \eta} = \\
\left( \frac{y_{ijk} - (x'_{ijk}\beta + \delta_u \theta_{0ij} + \delta_0 \theta_{0i})}{u'_{ijk} + \frac{\delta_u \theta_{0ij} + \delta_0 \theta_{0i}}{\delta_u}} \cdot \frac{\delta_{\theta_{0ij}} \delta_{\theta_{0ij}}}{\delta_u^2} \right) \cdot \left( -\frac{z'_{ijk} \theta_{0ij} e^{z'_{ijk}\alpha}}{2 \sqrt{e^{z'_{ijk}\alpha} - \delta_{\theta_{0ij}}^2}} \right) \\
+ \left( 1 - \frac{y_{ijk} - (x'_{ijk}\beta + \delta_u \theta_{0ij} + \delta_0 \theta_{0i})}{u'_{ijk} + \frac{\delta_u \theta_{0ij} + \delta_0 \theta_{0i}}{\delta_u}} \cdot \frac{\delta_{\theta_{0ij}} \delta_{\theta_{0ij}}}{\delta_u^2} \right) \cdot \frac{z'_{ijk} \theta_{\theta_{0ij}} e^{z'_{ijk}\alpha} \delta_{\theta_{0ij}}^2}{2 (e^{z'_{ijk}\alpha} - \delta_{\theta_{0ij}}^2)^{3/2}}, \text{ if } \eta = \delta_{\theta_{0ij}} \\
\frac{\partial^2 \log(y_{ijk} | \theta_{0ij}, \theta_{0ij}, \theta_{0ij})}{\partial \alpha \partial \eta} = \\
\left( \frac{\theta_{0ij}}{u'_{ijk} + \frac{\delta_u \theta_{0ij} + \delta_0 \theta_{0i}}{\delta_u}} \right) \cdot \frac{\delta_{u \theta_{0ij}}}{\delta_u^2} + \frac{y_{ijk} - (x'_{ijk}\beta + \delta_u \theta_{0ij} + \delta_0 \theta_{0i})}{u'_{ijk} + \frac{\delta_u \theta_{0ij} + \delta_0 \theta_{0i}}{\delta_u}} \cdot \frac{\theta_{\theta_{0ij}} \delta_{\theta_{0ij}}^2}{\delta_u^3} \cdot \frac{e^{z'_{ijk}\alpha} \delta_{\theta_{0ij}}^2}{\delta_u^2} \\
- \left( 1 - \frac{y_{ijk} - (x'_{ijk}\beta + \delta_u \theta_{0ij} + \delta_0 \theta_{0i})}{u'_{ijk} + \frac{\delta_u \theta_{0ij} + \delta_0 \theta_{0i}}{\delta_u}} \cdot \frac{\delta_{\theta_{0ij}} \delta_{\theta_{0ij}}}{\delta_u^2} \right) \cdot \frac{z'_{ijk} \theta_{\theta_{0ij}} e^{z'_{ijk}\alpha} \delta_{\theta_{0ij}}^2}{2 (e^{z'_{ijk}\alpha} - \delta_{\theta_{0ij}}^2)^{3/2}}, \text{ if } \eta = \delta_{u} 
\]

41
\[
\frac{\partial^2 \log(y_{ijk|\theta_{0i}, \theta_{0ij}, \theta_{0ij}})}{\partial \delta_0 \partial \eta} = \begin{cases} 
-2 \left( y_{ijk} - (x'_{ijk} \beta + \delta_0 \theta_{0ij} + \delta_0 \theta_{0i}) \right) \frac{\theta_{0i} u'_{ijk}}{e^{u_{ijk} + \frac{\delta_{u0} \theta_{0ij} + \sqrt{e^{z'_{ijk} \alpha - \delta_{u0} \delta_{\theta_{0ij}}}}}}}, & \eta = \tau \\
-2 \left( \frac{x'_{ijk}}{e^{u_{ijk} + \frac{\delta_{u0} \theta_{0ij} + \sqrt{e^{z'_{ijk} \alpha - \delta_{u0} \delta_{\theta_{0ij}}}}}}}, \right) \theta_{0i}, & \eta = \beta \\
-2 \left( \frac{y_{ijk} - (x'_{ijk} \beta + \delta_0 \theta_{0ij} + \delta_0 \theta_{0i})}{e^{u_{ijk} + \frac{\delta_{u0} \theta_{0ij} + \sqrt{e^{z'_{ijk} \alpha - \delta_{u0} \delta_{\theta_{0ij}}}}}}}, \right) \theta_{0i} \cdot \frac{\theta_{0ij} e^{z'_{ijk} \alpha} z'_{ijk}}{\sqrt{e^{z'_{ijk} \alpha} \delta_{\theta_{0ij}}^2}}, & \eta = \alpha \\
-2 \left( \frac{\theta_{0i}}{e^{u_{ijk} + \frac{\delta_{u0} \theta_{0ij} + \sqrt{e^{z'_{ijk} \alpha - \delta_{u0} \delta_{\theta_{0ij}}}}}}}, \right) \theta_{0i}, & \eta = \delta_0 \\
-2 \left( \frac{y_{ijk} - (x'_{ijk} \beta + \delta_0 \theta_{0ij} + \delta_0 \theta_{0i})}{e^{u_{ijk} + \frac{\delta_{u0} \theta_{0ij} + \sqrt{e^{z'_{ijk} \alpha - \delta_{u0} \delta_{\theta_{0ij}}}}}}}, \right) \theta_{0i} \left( \frac{\theta_{0ij}}{\sqrt{e^{z'_{ijk} \alpha} - \delta_{\theta_{0ij}}^2}} \right), & \eta = \delta_{u0} \\
-2 \left( \frac{\theta_{0ij}}{e^{u_{ijk} + \frac{\delta_{u0} \theta_{0ij} + \sqrt{e^{z'_{ijk} \alpha - \delta_{u0} \delta_{\theta_{0ij}}}}}}}, \right) + \frac{y_{ijk} - (x'_{ijk} \beta + \delta_0 \theta_{0ij} + \delta_0 \theta_{0i})}{e^{u_{ijk} + \frac{\delta_{u0} \theta_{0ij} + \sqrt{e^{z'_{ijk} \alpha - \delta_{u0} \delta_{\theta_{0ij}}}}}}}, \right) \theta_{0i} \left( \frac{\delta_{u0} \theta_{0ij} + \sqrt{e^{z'_{ijk} \alpha} - \delta_{\theta_{0ij}}^2}}{\delta_{u0}^2 e^{z'_{ijk} \alpha} \delta_{\theta_{0ij}}^2} \right), & \eta = \delta_{u} \\
\end{cases}
\]
\[
\frac{\partial^2 \log(y_{ijk} | \theta_{oi}, \theta_{oj}, \theta_{oi})}{\partial \delta_{uo} \partial \eta} = \left( \frac{x_{ijk}'}{e^{u_{ijk}^t + \delta_{uo} \theta_{oj} + \frac{\delta_{uo}}{\delta^2 v} \theta_{oi}}} \right) \left( \frac{\delta_{uo}}{\delta v} \right) \left( \frac{\delta_{uo}}{\delta^2 v} \right), \text{if } \eta = \tau.
\]

\[
\frac{\partial^2 \log(y_{ijk} | \theta_{oi}, \theta_{oj}, \theta_{oi})}{\partial \delta_{uo} \partial \eta} = \left( \frac{x_{ijk}'}{e^{u_{ijk}^t + \delta_{uo} \theta_{oj} + \frac{\delta_{uo}}{\delta^2 v} \theta_{oi}}} \right) \left( \frac{\delta_{uo}}{\delta v} \right) \left( \frac{\delta_{uo}}{\delta^2 v} \right), \text{if } \eta = \beta.
\]

\[
\frac{\partial^2 \log(y_{ijk} | \theta_{oi}, \theta_{oj}, \theta_{oi})}{\partial \delta_{uo} \partial \eta} = \left( \frac{y_{ijk} - (x_{ijk}^{-1} + \delta_v \theta_{oj} + \delta_{uo}) \theta_{oi} e^{z_{ijk}^t z_{ijk}'} \frac{\delta_{uo}}{\delta v}}{e^{u_{ijk}^t + \delta_{uo} \theta_{oj} + \frac{\delta_{uo}}{\delta^2 v} \theta_{oi}}} \right) \left( \frac{\delta_{uo}}{\delta v} \right) \left( \frac{\delta_{uo}}{\delta^2 v} \right) + \left( 1 - \frac{y_{ijk} - (x_{ijk}^{-1} + \delta_v \theta_{oj} + \delta_{uo}) \theta_{oi} e^{z_{ijk}^t z_{ijk}'} \frac{\delta_{uo}}{\delta v}}{e^{u_{ijk}^t + \delta_{uo} \theta_{oj} + \frac{\delta_{uo}}{\delta^2 v} \theta_{oi}}} \right) \left( \frac{z_{ijk}^t \theta_{oi} e^{z_{ijk}^t z_{ijk}'} \frac{\delta_{uo}}{\delta v}}{2 \left( e^{z_{ijk}^t z_{ijk}'} \delta_v^2 \right)^{3/2} \delta_{uo}^2} \right), \text{if } \eta = \alpha.
\]

\[
\frac{\partial^2 \log(y_{ijk} | \theta_{oi}, \theta_{oj}, \theta_{oi})}{\partial \delta_{uo} \partial \eta} = \left( \frac{\theta_{oi}}{e^{u_{ijk}^t + \delta_{uo} \theta_{oj} + \frac{\delta_{uo}}{\delta^2 v} \theta_{oi}}} \right) \left( \frac{\delta_{uo}}{\delta v} \right) \left( \frac{\delta_{uo}}{\delta^2 v} \right) + \left( 1 - \frac{y_{ijk} - (x_{ijk}^{-1} + \delta_v \theta_{oj} + \delta_{uo}) \theta_{oi} e^{z_{ijk}^t z_{ijk}'} \frac{\delta_{uo}}{\delta v}}{e^{u_{ijk}^t + \delta_{uo} \theta_{oj} + \frac{\delta_{uo}}{\delta^2 v} \theta_{oi}}} \right) \left( \frac{\theta_{oi} e^{z_{ijk}^t z_{ijk}'} \frac{\delta_{uo}}{\delta v}}{2 \left( e^{z_{ijk}^t z_{ijk}'} \delta_v^2 \right)^{3/2} \delta_{uo}^2} \right), \text{if } \eta = \delta_0.
\]
\[ \frac{\partial^2 \log(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{0ij})}{\partial \delta_{u0} \partial \eta} = \left( \frac{y_{ijk} - (\frac{u'_{ijk} + \delta_{u0} \theta_{0ij} + \delta_{v0} \theta_{0i}}{\sqrt{e^{\frac{2}{3} \frac{u'_{ijk} - \delta_{u0} \theta_{0i} + \delta_{v0} \theta_{0ij}}{\delta_{u0}^2}}} \right)^2 + 1 - \frac{y_{ijk} - (\frac{u'_{ijk} + \delta_{u0} \theta_{0ij} + \delta_{v0} \theta_{0i}}{\sqrt{e^{\frac{2}{3} \frac{u'_{ijk} - \delta_{u0} \theta_{0i} + \delta_{v0} \theta_{0ij}}{\delta_{u0}^2}})}}}{\frac{\theta_{0ij} \delta_{u0} / \delta_{u}^3}{\delta_{u0}} - \frac{\theta_{0ij} \delta_{u0} / \delta_{u}^3}{\delta_{u0}}} (e^{\frac{2}{3} \frac{u'_{ijk} - \delta_{u0} \theta_{0ij} + \delta_{v0} \theta_{0ij}}{\delta_{u0}^2}})^{3/2}, \right) 

\text{if } \eta = \delta_{u0}. \]

\[ \frac{\partial^2 \log(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{0ij})}{\partial \delta_{u0} \partial \eta} = \left( \frac{\theta_{0ij} - \theta_{0ij} \delta_{u0} / \delta_{u}^2}{\sqrt{e^{\frac{2}{3} \frac{u'_{ijk} - \delta_{u0} \theta_{0ij} + \delta_{v0} \theta_{0ij}}{\delta_{u0}^2}}}} + 1 - \frac{y_{ijk} - (\frac{u'_{ijk} + \delta_{u0} \theta_{0ij} + \delta_{v0} \theta_{0i}}{\sqrt{e^{\frac{2}{3} \frac{u'_{ijk} - \delta_{u0} \theta_{0ij} + \delta_{v0} \theta_{0ij}}{\delta_{u0}^2}}}})}{\frac{\theta_{0ij} \delta_{u0} / \delta_{u}^3}{\delta_{u0}} - \frac{\theta_{0ij} \delta_{u0} / \delta_{u}^3}{\delta_{u0}}} (e^{\frac{2}{3} \frac{u'_{ijk} - \delta_{u0} \theta_{0ij} + \delta_{v0} \theta_{0ij}}{\delta_{u0}^2}})^{3/2}, \right) \text{, if } \eta = \delta_{u}. \]
\[
\frac{\partial^2 \log(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{0ij})}{\partial \delta_v \partial \eta} = - \left( \frac{y_{ijk} - (x_{ijk}^t \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i})}{u_{ijk}^t + \delta v_0 \theta_{0ij} + \delta_0 \theta_{0i}} \right) \left( -\frac{\delta v_0 \theta_{0ij}}{\delta_v^2} + \frac{\theta_{0ij} \delta_0^3}{\delta v^2} \right) - 2 \theta_{0ij}, \text{ if } \eta = \tau.
\]

\[
\frac{\partial^2 \log(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{0ij})}{\partial \delta_v \partial \eta} = - \left( \frac{x_{ijk}^t}{u_{ijk}^t + \delta v_0 \theta_{0ij} + \delta_0 \theta_{0i}} \right) \left( -\frac{\delta v_0 \theta_{0ij}}{\delta_v^2} + \frac{\theta_{0ij} \delta_0^3}{\delta v^2} \right) - 2 \theta_{0ij}, \text{ if } \eta = \beta.
\]

\[
\frac{\partial^2 \log(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{0ij})}{\partial \delta_v \partial \eta}
= - \left( \frac{y_{ijk} - (x_{ijk}^t \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i})}{u_{ijk}^t + \delta v_0 \theta_{0ij} + \delta_0 \theta_{0i}} \right) \left( -\frac{\delta v_0 \theta_{0ij}}{\delta_v^2} + \frac{\theta_{0ij} \delta_0^3}{\delta v^2} \right) - 2 \theta_{0ij}
\]

\[
+ \left( \frac{y_{ijk} - (x_{ijk}^t \beta + \delta_v \theta_{0ij} + \delta_0 \theta_{0i})}{u_{ijk}^t + \delta v_0 \theta_{0ij} + \delta_0 \theta_{0i}} \right) \left( \frac{z_{ijk}^t \theta_{0ij} e^{z_{ijk}^t \alpha} \delta_0^2}{2 \left( e^{z_{ijk}^t \alpha} - \delta_0 \delta_0 \delta_0 \right)} \right)
\]

if \eta = \alpha.
\[
\frac{\partial^2 \log(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{oij})}{\partial \delta_u \partial \eta} = - \left( \frac{\theta_{oi}}{u'_{ijk} \tau + \delta_{uu} \theta_{oij} + \sqrt{e^2_{ijk} - \delta_{oij}^2}} \right) \left( \frac{\delta_{v0} \theta_{oij}}{\delta_0} + \frac{\theta_{oij} \delta_{v0}^2}{\delta_0} - 2 \theta_{0ij} \right), \text{if } \eta = \delta_0.
\]

\[
\frac{\partial^2 \log(y_{ijk} | \theta_{0i}, \theta_{0ij}, \theta_{oij})}{\partial \delta_u \partial \eta} = - \left( \frac{y_{ijk} - (x'_{ijk} \beta + \delta_u \theta_{0ij} + \delta_0 \theta_{0i})}{u'_{ijk} \tau + \delta_{uu} \theta_{oij} + \sqrt{e^2_{ijk} - \delta_{oij}^2}} \right) \left( \frac{\theta_{0ij}}{\delta_u} + \frac{\theta_{oij} \delta_{v0}^2}{\delta_u} - 2 \theta_{0ij} \right) + \left( 1 - \frac{y_{ijk} - (x'_{ijk} \beta + \delta_u \theta_{0ij} + \delta_0 \theta_{0i})}{u'_{ijk} \tau + \delta_{uu} \theta_{oij} + \sqrt{e^2_{ijk} - \delta_{oij}^2}} \right) \left( \frac{\theta_{0ij}}{\delta_u} + \frac{2 \theta_{oij} \delta_{v0}^3}{\delta_u^3} - \frac{\theta_{oij} \delta_{v0}^5}{\delta_u^5} \right), \text{if } \eta = \delta_{v0}.
\]
\[
\frac{\partial^2 \log(y_{ij}|\theta_{0i}, \theta_{ij}, \theta_{0ij})}{\partial \delta_v \partial \eta} = - \left( \frac{\theta_{0ij}}{u'_v + \delta_{uv} \theta_{0ij} + \sqrt{\epsilon_{ijk}^2 - \delta_{uv}^2 / \delta_v^2 \theta_{ij}}} \right) + \left( \frac{\delta_{uv} \theta_{0ij}}{\delta_v^3 \sqrt{\epsilon_{ijk}^2 - \delta_{uv}^2 / \delta_v^2 \theta_{ij}}} \right) - 2 \theta_{0ij} + \left( \left( \frac{4 \theta_{0ij} \delta_{uv}^2 / \delta_v^2 + \frac{\theta_{0ij} \delta_{uv}^4 / \delta_v^4}{\delta_v^3}}{\epsilon_{ijk}^2 - \delta_{uv}^2 / \delta_v^2} \right) - \frac{2 \delta_{uv} \theta_{0ij}}{\delta_v^3} \right)
\]

if \( \eta = \delta_v \).

EQ. II-28
M. **Numerical Integration**

The integration over \( \theta_{0i}, \theta_{0lj}, \theta_{0lj} \) can be approximately computed at any practical degree of accuracy by Gauss-Hermite Quadrature (Stroud, 1966). Gauss-Hermite quadrature approach is an approximation of integral value, usually stated as a weighted sum of function values at a specified number of quadrature points within the domain of integration. In the case of 3-dimensional integration over normal distribution \( I_3 \), the vector of optimal loads is denoted by \( B_q = (B_{q_1}, B_{q_2}, B_{q_3})' \) and the corresponding weights are \( W(B_q) = W(B_{q_1})W(B_{q_2})W(B_{q_3}) \). The first derivative of log likelihood with respect to \( \lambda \) is constructed by Gaussian quadrature approximation as

\[
\frac{\partial \log L}{\partial \lambda} = \sum_{i=1}^{n} \frac{1}{f(y_i)} \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} \left[ \sum_{q_3}^{m} \sum_{q_2}^{m} \sum_{q_1}^{m} \frac{f(y_i|B_{q_3})f(y_{ij}|B_{q_1}, B_{q_2}, B_{q_3})}{f(y_{ij}|B_{q_1})} LGf(B_{q_2}, B_{q_3})f(B_{q_1})W(B_q) \right] \quad \text{EQ. II-29}
\]

The second derivative of log likelihood with respect to
\[
\frac{\partial^2 \log L}{\partial \lambda \partial \eta} = \sum_{i=1}^{n} \frac{1}{f^2(y_i)} \left\{ \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} \left[ \sum_{q_1} \sum_{q_2} \sum_{q_3} f(y_i|B_{q_1}) f(y_{ij}|B_{q_1}, B_{q_2}, B_{q_3}) \frac{\partial \log \left( f(y_{ijk}|B_{q_1}, B_{q_2}, B_{q_3}) \right)}{\partial \lambda} f(B_{q_2}, B_{q_3}) f(B_{q_1}) W(B_{q}) \right] \right\}
\]

\[
\sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} \left[ \sum_{q_1} \sum_{q_2} \sum_{q_3} f(y_i|B_{q_1}) f(y_{ij}|B_{q_1}, B_{q_2}, B_{q_3}) \frac{\partial \log \left( f(y_{ijk}|B_{q_1}, B_{q_2}, B_{q_3}) \right)}{\partial \eta} f(B_{q_2}, B_{q_3}) f(B_{q_1}) W(B_{q}) \right] \]

\[
+ \frac{1}{f(y_i)} \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} \left[ \sum_{q_1} \sum_{q_2} \sum_{q_3} f(y_i|B_{q_1}) f(y_{ij}|B_{q_1}, B_{q_2}, B_{q_3}) \frac{\partial \log \left( f(y_{ijk}|B_{q_1}, B_{q_2}, B_{q_3}) \right)}{\partial \lambda} \right.
\]

\[
+ \frac{f(y_i|B_{q_1})}{f(y_{ij}|B_{q_1})} \frac{\partial f(y_{ij}|B_{q_1}, B_{q_2}, B_{q_3})}{\partial \eta} \frac{\partial \log \left( f(y_{ijk}|B_{q_1}, B_{q_2}, B_{q_3}) \right)}{\partial \lambda} f(B_{q_2}, B_{q_3}) f(B_{q_1}) \]

\[
- \frac{f(y_i|B_{q_1})}{f^2(y_{ij}|B_{q_1})} \frac{\partial f(y_{ij}|B_{q_1})}{\partial \eta} \frac{\partial \log \left( f(y_{ijk}|B_{q_1}, B_{q_2}, B_{q_3}) \right)}{\partial \lambda} \right\}
\]

The SAS PROC NLMIXED procedure applied the combination of maximum marginal likelihood method and empirical posterior estimate approach for model estimation, and utilize Gauss-Hermite quadrature to approximate the value of integral function.
III. SIMULATION RESULTS

A. Overview

Simulation process was carried out to validate the accuracy and reliability of the proposed three-level RL-RSS model. One hundred data sets, each with 80,000 observations (50 smoking events/occasions nested within each of two waves, wave is further nested within each of 800 subjects), were generated based on the proposed three-level RL-RSS model as specified below.

\[ y_{ijk} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + \beta_3 x_{3ijk} + v_{0ij} + v_{0i} + \epsilon_{ijk} \]

with the conditions:

\[ v_{0ij} \sim N(0, \delta_0^2), \quad v_{0i} \sim N(0, \delta_0^2), \quad \epsilon_{ijk} | \omega_i \sim N \left(0, \delta_{\epsilon_{ijk}}^2 \right), \quad \omega_i \sim N \left(0, \delta_{\omega_{ijk}}^2 \right) \]

\[
\begin{bmatrix}
  v_{0i} \\
  \omega_i \\
  v_{0ij}
\end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \delta_0^2 & \delta_{v\omega} & 0 \\ \delta_{v\omega} & \delta_{\omega_{ijk}}^2 & 0 \\ 0 & 0 & \delta_{\omega}^2 \end{bmatrix} \right)
\]

\[
\log \left(\delta_{\epsilon_{ijk}}^2 \right) = \tau_0 + \tau_1 x_{2ij} + \tau_2 x_{3ijk} + \omega_i, \quad \text{EQ. III-1}
\]

\[
\log \left(\delta_{\omega_{ij}}^2 \right) = \alpha_0 + \alpha_1 x_{1i}
\]

The \( y_{ijk} \) is the \( k \)th observed continuous outcome measures for the \( i \)th subject at the \( j \)th wave.

Three covariates, \( x_{1i}, x_{2ij}, \) and \( x_{3ijk} \) (\( x_{3ijk} \): level-1, occasion level; \( x_{2ij} \): level-2, wave level; \( x_{1i} \), level-3, subject level), were considered. The \( x_{3ijk} \) is a continuous covariate following normal distribution with mean = -0.5 and std = 0.5 at level-1; \( x_{2ij} \) is a level-2 dichotomized covariate with the values 0 and 1; and \( x_{1i} \) is a continuous normal-distributed variable with mean = 0.5 and std = 0.05 at the subject level. The \( \beta_1, \beta_2, \beta_3 \) are the fixed-effect regression coefficient parameters for \( x_i, x_{ij}, \) and \( x_{ijk} \) respectively. The random effect \( v_{0ij} \) indicates the random effect of subject \( i \) at wave \( j \), \( v_{0i} \)
represents the random effect for ith subject. The distribution of random effects \( v_{0i}, v_{0ij} \) are assumed to be normal distribution \( N(0, \delta_0^2) \) and \( N(0, \delta_0^2) \) respectively. \( \varepsilon_{ij} \) is the error term for the kth observation of subject i at wave j. The errors \( \varepsilon_{ijk} \) are assumed to be normal with zero mean and variance \( \delta_{\varepsilon_{ij}}^2 \). The error variance \( \delta_{\varepsilon_{ijk}}^2 \) is further modeled through a log-linear modeling structure 
\[
\log\left(\delta_{\varepsilon_{ijk}}^2\right) = \tau_0 + \tau_1 x_{2ij} + \tau_2 x_{3ijk} + \omega_i.
\]
Here, \( \tau \) coefficient parameters indicate the degree of covariates' influence on the WS variance. The random scale effect \( \omega_i \) is distributed as normal \( N(0, \delta_{\omega_i}^2) \). The log of random scale variance is further modeled as 
\[
\log(\delta_{\omega_i}^2) = \alpha_0 + \alpha_1 x_{1i}.
\]
The \( \alpha \) parameters are the regression coefficients in random scale variance model.

The model in EQ. III-1 is re-parameterized in Cholesky form with standardized normal random effects as 
\[
y_{ijk} = x_{ijk}'\beta + v_{0ij} + v_{0i} + \varepsilon_{ijk} \\
= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + \beta_3 x_{3ijk} + \delta_{\omega_0} \theta_{0ij} + \delta_0 \theta_{0i} \\
+ e^{\frac{1}{2} \left( \tau_0 + \tau_1 x_{2ij} + \tau_2 x_{3ijk} + \frac{\delta_{\omega_0} \theta_{0ij}}{\delta_0} + \sqrt{e^{\alpha_0 + \alpha_1 x_{1i}} - \delta_{\omega_0}^2 \theta_{0ij}} \right)} e_{ijk} \\
\]
EQ. III-2

Three covariates were assumed to associate with the overall population mean; \( x_{2ij} \) and \( x_{3ijk} \) are the covariates in error variance model to affect WS within-wave (WS-WW) variance, and \( x_{1i} \) is associated with random scale variance. Not all covariates are included in error variance model or random scale variance model due to computation difficulty in SAS PROC NLMIXED given a large sample of data (80,000 observations in each dataset). The continuous covariate was generated from a normal distribution using RANNOR(seed) in SAS. The wave variable is uniform distributed by a do-loop. The number of observations is the same for each subject and each wave. The value of parameters and the distribution of random variables were specified in Table I.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>0.4</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.4</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.6</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.3</td>
</tr>
<tr>
<td>$\tau_0$</td>
<td>0.6</td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>-0.3</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>0.4</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.2</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>1</td>
</tr>
<tr>
<td>$\delta_{v^2}$</td>
<td>.1</td>
</tr>
<tr>
<td>$\delta_{v_0}$</td>
<td>0.2</td>
</tr>
<tr>
<td>$\delta_{\theta^2}$</td>
<td>0.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Variables</th>
<th>Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{1i}$</td>
<td>N(0.5,0.05)</td>
</tr>
<tr>
<td>$x_{2ij}$</td>
<td>Binomial(0.5)</td>
</tr>
<tr>
<td>$x_{3ijk}$</td>
<td>N(-0.5,0.5)</td>
</tr>
<tr>
<td>$\theta_{0i}, \theta_{0ij}, \theta_{oij}, e_{ijk}$</td>
<td>Independent N(0,1)</td>
</tr>
</tbody>
</table>
To evaluate the performance of the proposed three-level RL-RSS model, three models were fit for each simulated data set: three-level RL-RSS model (EQ. III-1, EQ. III-2), three-level RL-RS model (EQ. III-3) and three-level random intercept model (EQ. III-4) analyzed as result comparisons.

Model specification of three-level RL-RS model is expressed below. The notation follows three-level RL-RSS model in EQ. III-1.

\[
 y_{ijk} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + \beta_3 x_{3ijk} + v_{0ij} + v_{0i} + \varepsilon_{ijk}
\]

with the conditions:

\[
 v_{0ij} \sim N(0, \delta_v^2), \quad v_{0i} \sim N(0, \delta_v^2), \quad \varepsilon_{ijk} \sim N \left(0, \delta_{\varepsilon_{ijk}}^2\right)
\]

\[
 \omega_i \sim N \left(0, \delta_\omega^2\right), \quad \delta_{\omega_{ijk}}^2 = \tau_0 + \tau_1 x_{2ij} + \tau_2 x_{3ijk} + \omega_i
\]

\[
 \log \left( \delta_{\varepsilon_{ijk}}^2 \right) = \tau_0 + \tau_1 x_{2ij} + \tau_2 x_{3ijk} + \omega_i
\]

Model specification of three-level random intercept model (RL Model) is of the form with the same notation in EQ. III-1 and EQ. III-3.

\[
 y_{ijk} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + \beta_3 x_{3ijk} + v_{0ij} + v_{0i} + \varepsilon_{ijk}
\]

with the conditions:

\[
 v_{0ij} \sim N(0, \delta_v^2), \quad v_{0i} \sim N(0, \delta_v^2), \quad \varepsilon_{ijk} \sim N(0, \delta_\varepsilon^2)
\]

\[
 \log \left( \delta_\varepsilon^2 \right) = \tau_0
\]

\[
 EQ. III-3
\]

\[
 EQ. III-4
\]
B. **Evaluation Criteria**

Seven evaluation criteria were assessed to validate the model fitting. For each fitted model, the parameter estimates, standard errors (SE), biases, standardized biases (SB), RMSEs, 95% CI coverage rates, and the widths of the 95% CIs were reported. The evaluation standards are defined as follows:

Bias: Bias is the difference between the parameter estimate and the “true” value of parameter specified in the simulated dataset—i.e., \( Bias(\hat{\theta}, \theta) = E(\hat{\theta}) - \theta. \)

Standardized bias: SB is defined as the bias over the standard deviation—i.e., \( 100 \times \frac{E(\hat{\theta}) - \theta}{SE(\hat{\theta})} \), where SE represents the standard error associated with the 100 parameter estimates from the simulations (and is not the average of 100 standard errors of parameter estimates). It is the bias over one unit of estimation uncertainty across simulated data sets. A value that is greater than 100% or less than -100% suggests that the bias falls one standard error above or below the estimation variation. Demirtas (2004) recommended the desirable range of SB within -50% ≤ SB ≤ 50% as practically insignificant bias.

Root mean squared error: Mean squared error (MSE) is defined as \( Var(\hat{\theta}) + [Bias(\hat{\theta}, \theta)]^2. \) As an analogy to standard deviation, the square root of MSE yields the RMSE with the same units of parameter. An RMSE is arguably the best evaluation criterion, as it provides the combined information of accuracy (bias) and precision (variance). An RMSE is just standard deviation for an unbiased estimator.

Coverage rate for 95% CI: 95% CI can be constructed for each parameter in each simulated dataset. The 95% CI coverage rate is the proportion of the event that the true parameter value is within
the 95% CI. If the proposed model works well, the coverage rate would be close to the nominal rate, in this case, 0.95.

Average width (AW) of 95% CI: The width of 95% CI is the difference between the CI upper limit and lower CI limits. Higher coverage rate along with narrower CIs are preferred with the greater accuracy and power.

The following abbreviations are in use for the evaluation criteria: EST=estimate, SE=standard error, SB=standardized bias, RMSE=root mean squared error, COV=95% coverage rate, and AW=average width of 95% CI.

C. **Simulation Results**

Simulation study aims to assess the performance of three-level RL-RSS model when the real data satisfies the model assumption, and also to test how the two inappropriate three-level models (i.e., RL-RS and RL model) behave as a comparison purpose. The simulated data were generated under three-level RL-RSS assumption, as specified in EQ. III-2. Two continuous variables at the subject level and occasion level, and one categorical wave variable are considered. The results of the three models are summarized in Table II.

The parameter estimate, SE, bias, SB, RMSE, 95% coverage, and interval width are reported. The box-and-whisker diagrams are presented to compare the simulation results among three fitted models in Figure 1-12.
D. **Results of Three-level Random Intercept Model**

In the right panel of Table II, the summarized simulation results of three-level random intercept model (RL-model) are presented. The estimates of coefficient parameters in the mean model \((\beta_0, \beta_1, \beta_2, \text{ and } \beta_3)\) are almost identical to the true values with small bias in absolute values (less than 0.01). The SB of beta parameter estimates falls in the reasonable range (the absolute value of SB is around 100%). The RMSEs are acceptably small (RMSE <0.5) and 95% COV is fairly high (close to 95%).

As for the between-subject variances of random location effects, the parameters \(\delta_0^2\) and \(\delta_1^2\) are well-estimated as indicated by small biases and RMSEs, acceptable SBs and high CI coverages.

However, the random intercept model (RL-model) does not have a good performance on \(\tau_0\), the log scale intercept estimation of WS-WW variance. The estimate is way beyond the acceptable range of true value. A large bias, extremely high SB, and almost zero COV are observed. Ignoring the error variance model causes the problem of overestimation. The finding is consistent with Li’s 2012 simulation results. Lack of error variance model with random scale effect can lead to the poor quality of intercept parameter estimation in error variance model.

E. **Results of Three-level Random Location Scale Model**

In this model, the error variance is modeled by a log-linear structure, with the inclusion of covariates and random scale effect. Results in the middle panel of Table II show that this model performed well in the estimation of fixed location effects \((\beta_0, \beta_1, \beta_2, \text{ and } \beta_3)\) and random location variance \((\delta_0^2\text{ and } \delta_1^2)\), as indicated by small biases and RMSEs, acceptable SBs and high CI coverage.
The problematic parameter estimation occurs in RL model for $\tau_0$ estimation. The RL-RS model greatly improves the estimation results. The estimate is almost identical with true value with small biases, SB, RMSE, and high 95% COV.

Comparing to RL model, the additional new parameters in the RL-RS model includes fixed scale effects ($\tau_2, \tau_3$), the covariance of random location effect and random scale effect at the subject level ($\delta_{\nu_\omega}$), and log scale of random scale variance ($\alpha_0$). The estimates of $\tau_2, \tau_3$, and $\delta_{\nu_\omega}$ are very close to the true values. The absolute value of SB is around 100%, implying the bias is in the same magnitude of standard error of parameter estimates. The RMSEs are acceptably small (RMSE < 0.5) and 95% COV is fairly high (close to 95%).

However, the simulation results for the intercept of random scale variance in log scale ($\alpha_0$) are observed to be seriously overestimated. The estimate on average is 0.69, whereas the true value specified is 0.2. Most of the RL-RS models tend to overestimate the intercept of log random scale variance, as the SE of estimates is very low. The SB, the bias over SE, is extremely high (SB=10522.6). It implies that the bias is far beyond the variation range of estimation. The RMSE is close to the value of bias, as the SE is ignorable relative to great bias. The 95% COV is almost zero, suggesting that 95% CIs do not cover the true values at all. In summary, the estimate of $\alpha_0$ is considerably biased.

F. **Results of Random Location Scale Model with Modeling Random Scale Variance**

The three-level RL-RSS model was fit to the simulated datasets under the assumption of RL-RSS model. The left panel of Table II shows the simulation results from RL-RSS models. As indicated by small biases and RMSEs, acceptable SB, and nearly 95% COV, this model recovers the parameters of fixed location/scale effects and random location/scale effect very well. It also resolves the
overestimation problems of \( \tau_0 \) and \( \alpha_0 \) that occur in the RL model and RL-RS model respectively. The estimates of \( \tau_0 \) and \( \alpha_0 \) are very close to true values with small bias and RMSEs, acceptable SBs and high CI COV.

As for the new parameter in the RL-RSS model, the estimate of \( \alpha_1 \) is very close to the true values. The absolute value of SB is less than 50%. The RMSEs are acceptably small (RMSE=0.85) and 95% COV is close to 100%.

The box-and-whisker diagrams were plotted to graphically depict the distribution of each parameter estimate in three fitted models, as seen in Figure 1–

...Figure 1–Figure 4 displayed the results of fixed location effect parameters \( \beta_0, \beta_1, \beta_2 \), and \( \beta_3 \) in three fitted models (RL model, RL-RS model, RL-RSS model) under the simulation design. Figures 5–7 represents the same plots for fixed scale effect parameters \( \tau_0, \tau_1, \) and \( \tau_2 \). Figure 8 and Figure 9 are for \( \alpha_0, \alpha_1 \), the coefficient parameters in random scale variance model. The plots for variance parameters \( \delta^2_0, \delta^2_v \) and \( \delta_{vw} \) are displayed in Figure 10

respectively. The horizontal line in each plot represents the true value, and y-axis represents the minimum, 25th percentile, median, 75th percentile, and maximum values of parameter estimates. The diamond in the middle of box indicates the mean level of estimate. From left to right, the panels are separated for RL model, RL-RS model, RL-RSS model sequentially if the parameter appears in the three models, otherwise, the applicable models are displayed.

For the fixed location effect parameters \( (\beta_0, \beta_1, \beta_2, \beta_3) \), the estimates distribute equally well around the true values with their medians and means overlapping the true values. However, the estimated parameter values tend to vary or spread out more in RL model than RL-RS and RL-RSS
models, as seen the greater width of box-and-whisker in Figure 1Figure 4. The observation is particularly noticeable in the result of coefficient parameter $\beta_3$ (the observation-level covariate) in Figure 4. It implies that modeling WS variance and its subject-heteroscedasticity in the RL-RS model and RL-RSS model helps reduce the standard errors of the fixed-effects parameter estimates, and therefore improves the estimation precision by weighing down the large dispersions. The overestimation problem of $\tau_0$ prominently stands out in RL model. The estimate on average in the RL model is way above the true value 0.6. Moreover, the fact that the range of $\tau_0$ estimate (even the minimum value of $\tau_0$ estimate) is above the true level reinforces the overestimation issue in RL model.

The RL-RS overestimates the parameter $\alpha_0$, which is clearly shown in Table II. Even the minimum $\alpha_0$ estimate over 100 fitted RL-RS models is far beyond the true level.

The RL model also appears to be more spread out in the estimation of variance parameters ($\delta^2_0$, $\delta^2_0$) comparing to the RL-RS model and RL-RSS model. It shows that modeling error variance explicitly by covariates and random scale effect reduces the estimation variation and helps improve the precision of fixed location effects and between-subject variance.
### TABLE II

100 SIMULATION RESULTS UNDER 100 SIMULATED RL-RSS DATASETS

<table>
<thead>
<tr>
<th></th>
<th>RL-RSS Model</th>
<th>RL-RS Model</th>
<th>Random Intercept Model (RL Model)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>True Value</td>
<td>EST SE BIA SB RMS E 95% COV AW</td>
</tr>
<tr>
<td>( \beta_0 )</td>
<td>100</td>
<td>0.4</td>
<td>0.4 0.02 0 21.31 0.16 0.95 0.61</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>100</td>
<td>-0.4</td>
<td>-0.4 0.03 0 -8.75 0.31 0.93 1.22</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>100</td>
<td>0.6</td>
<td>0.59 0 -0.01 -149.59 0.03 0.98 0.13</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>100</td>
<td>-0.3</td>
<td>-0.3 0 26.07 0.01 0.92 0.02</td>
</tr>
<tr>
<td>( \tau_0 )</td>
<td>100</td>
<td>0.6</td>
<td>0.59 0.01 -0.01 -120.66 0.05 0.96 0.2</td>
</tr>
<tr>
<td>( \tau_1 )</td>
<td>100</td>
<td>-0.3</td>
<td>-0.3 0 13.84 0.02 0.93 0.08</td>
</tr>
<tr>
<td>( \tau_2 )</td>
<td>100</td>
<td>0.4</td>
<td>0.4 0 38.44 0.01 0.93 0.04</td>
</tr>
<tr>
<td>( \alpha_0 )</td>
<td>100</td>
<td>0.2</td>
<td>0.19 0.04 -0.01 -20.36 0.43 0.98 1.67</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>100</td>
<td>1</td>
<td>1 0.08 0 -4.64 0.85 0.99 3.32</td>
</tr>
<tr>
<td>( \delta_0 )</td>
<td>100</td>
<td>0.14</td>
<td>0.14 0 9.11 0.01 0.96 0.04</td>
</tr>
<tr>
<td>( \delta_1 )</td>
<td>100</td>
<td>0.1</td>
<td>0.1 0 191.37 0.01 0.97 0.02</td>
</tr>
<tr>
<td>( \delta_{\omega_0} )</td>
<td>100</td>
<td>0.2</td>
<td>0.2 0 32.3 0.03 0.93 0.11</td>
</tr>
</tbody>
</table>
Figure 1. The box-and-whisker diagram of $\beta_0$.

Figure 2. The box-and-whisker diagram of $\beta_1$. 
Figure 3. The box-and-whisker diagram of $\beta_2$.

Figure 4. The box-and-whisker diagram of $\beta_3$. 
Figure 5. The box-and-whisker diagram of $\tau_0$.

Figure 6. The box-and-whisker diagram of $\tau_1$. 
Figure 7. The box-and-whisker diagram of $\tau_2$.

Figure 8. The box-and-whisker diagram of $\alpha_0$. 
Figure 9. The box-and-whisker diagram of $\alpha_1$.

Figure 10. The box-and-whisker diagram of $\delta_0^2$. 
Figure 11. The box-and-whisker diagram of $\delta_v^2$.

Figure 12. The box-and-whisker diagram of $\delta_{vw}$. 
IV. DATA ANALYSIS

The application of three-level RL-RSS model was motivated by National Cancer Institute Adolescent Smoking Study (PO1 CA098262, PI: Robin J. Mermelstein).

A. Research Interest of Adolescent Smoking Study

The data from a natural history study of adolescent smoking motivated the application of the proposed RS-RSS model. A few studies have shown the fact that smoking can help relieve negative affect among adolescents (Kassel, 2003; Khantzian, 1997). The present adolescent smoking study aims to investigate the relationship between the acute mood changes and smoking by utilizing real-time EMA. Most of the empirical work has focused the examination of changes in the level of smoking-related mood change. The examination of mood variability in relation to smoking has rarely been studied, although the variation of mood also conveys important information of research interests and strengthens our understanding of the underlying mechanism of how smoking behavior or smoking dependency affects the mood. Of interest to the investigators is the degree of subject heterogeneity in the mood measures and their variability, over and beyond the exploration of the average mood level. Furthermore, research interests focus on whether certain covariates could explain two sources of heteroscedasticity in terms of mood, and more importantly mood variability. For instance, are males more homogeneous in smoking-related mood changes, or are males more similar to each other in terms of the ability to stabilize their moods? In the terminology of statistical field, random location effects are addressed to account for subject heterogeneity of mood change, while random scale effects adjust the similarity across subjects for their ability to stabilize their moods, that is, the heterogeneity of individual mood consistency. The random location effects are allowed to affect the mean response (to allow for subjects with different mood levels), and random scale effects have an impact on subject WS
variance (to allow for subjects with different levels of mood consistency). The degree of subject heterogeneity is further modeled by the potentially associated covariates in a log-linear representation.

B. **Description of Study**

The data used in this study were from a natural history study of adolescent smoking. Participants in this study were adolescents in 9th or 10th grade at baseline. A total of 461 students, of whom 55% were females, completed the baseline measurement. The majority of the participants were White (about 57% White enrolled). There were 20% Hispanic, 16% Black, and 7% of other race. On a screening questionnaire six to eight weeks prior to the baseline, the participants reported that they had smoked at least one cigarette in their lifetime, and about 57.6% had smoked at least one cigarette in the past month.

To access the smoking behavior or dependency of adolescents, the multi-method approach was used, including self-report questionnaires, a week-long time/event sampling method via hand-held computers (EMA), and detailed surveys. Each adolescent carried a hand-held computer all the time during the data collection period of seven consecutive days. Each adolescent was trained to respond to the random prompts from his/her computer, as well as to record the smoking events (i.e., initiate a data collection interview) in conjunction with smoking episodes. The random prompts and the self-initiated smoking records were mutually exclusive, and no smoking occurred during random prompts. Each entry of the hand-held computer was dated and time-stamped. Following the baseline assessment, additional EMA sessions at 6-, 15-, 24-, and 60-month follow-ups, for a total of five EMA measurement waves, were completed by the subjects. Full-likelihood estimation method was used for model analysis. The assumption of MAR mechanism (Rubin, 1976) was made; that is, the missing data
related to covariates or observed dependent variables were ignorable (Laird, 1988). In longitudinal studies, ignorable nonresponse under the MAR mechanism (Rubin, 1976) implies that the missing data depends solely on observed data.

In this study, we are interested in comparing mood within subjects who had smoking events across measurement waves. We hence restrain our analysis cohort to the subjects who had two or more waves of data, and at each wave the subjects had at least two smoking events. This reduces the total to 4727 smoking events for 158 subjects, among which 65, 30, 33, and 30 subjects provided data at two, three, four, and five waves, respectively. The number of subjects obtained from the measurement wave taken at baseline, 6-months, 15-months, 24-months, and 5-years were 126, 93, 95, 101, and 87, respectively, with the average number of smoking events equaling 6.90 (range: 2 to 42), 7.53 (2 to 32), 9.74 (2 to 43), 10.14 (2 to 49), and 13.90 (2 to 64) at these same five waves, respectively.

C. **Outcome Measures: Positive Affect and Negative Affect**

This study considered two outcomes of mood: measures of the subject’s negative affect (NA) and positive affect (PA), respectively at smoking episodes. Each of these two measures consisted of the average value of several individual mood items. The values were rated using a scale from 1 to 10 with a larger number indicating a high level of mood attribute. For example, the following items were used in PA measurement to access subjects’ positive mood.

1. I felt happy,
2. I felt relaxed,
3. I felt cheerful,
4. I felt confident,
5. I felt accepted by others.

Similarly, NA consisted of the following items:

6. I felt sad,
7. I felt stressed,
8. I felt angry,
9. I felt frustrated,
10. I felt irritable.

Each participant rated his/her mood on two assessments, one “before” and the other “now after smoking” smoking episodes.

Considering the five items of the “before” with “now after smoking” in PA mood assessments, Cronbach’s alpha equaled 0.78 (0.78), 0.80 (0.81), 0.83 (0.83), 0.86 (0.86) and 0.89 (0.89) at baseline, 6-, 9-, 24-, and 60-months, respectively. Similarly, in terms of the NA mood assessments, Cronbach’s alpha equaled 0.81 (0.81), 0.86 (0.86), 0.88 (0.88), 0.89 (0.89) and 0.92 (0.92) at baseline, 6-, 9-, 24-, and 60-months, respectively.

As the interest of this study is on the mood change related to smoking, the difference between “before” and “now after smoking,” calculated as now through before, is used as the outcome measure. It is shortly represented as ΔPA/ΔNA throughout the work.
D. **Covariate Measure**

**Gender:** The variable gender is coded as 0 for female, and 1 for male.

**Wave:** The wave variable represents the time points of baseline, and the 4 follow-ups. The codes are 0 for baseline, 1 for 6-months follow up, 2.5 for 15 months follow up, 4 for 24-months follow up, and 10 for 5-year follow up.

**NDSS:** A shortened 10-item youth-specific version of Nicotine Dependence Syndrome Scale (Shiffman, 2004) was used to assess the smoking dependence. Using the shortened version of NDSS was based on a previous study (Sterling, 2009), which conducted psychometric analysis on an adolescent sample. The included 10 items primarily reflect the Drive/Tolerance subscale items of the original NDSS, where each item has the Likert-scale responses from 1=not true at all to 4=very true. For two items, one about smoking in the morning and one about craving a cigarette, 0 (“I don't smoke in the morning”) and 0 (“I don't crave cigarettes”) was also a response option. Only the participants who had tried smoking at least a puff once in their lifetimes completed this scale.

Several symptoms of dependence, including smoking to avoid withdrawal symptoms, craving, and increasing smoking to achieve similar effects (tolerance), are assessed by the NDSS. The average of all items are used for overall NDSS scale scores with a higher score reflecting a higher level of dependence.

E. **Model Setup**

In the present study, the interest focused on the degree to which covariates could explain $\Delta PA/\Delta NA$ variation, and more importantly, the heterogeneity of $\Delta PA/\Delta NA$ variation across different
subjects, over and above the exploration of population mean. Explained by the proposed model EQ. II-19, \( \tau \) indicates the influence of covariates on \( \Delta PA/\Delta NA \) variation, and \( \alpha \) indicates how the covariates affect the subject heterogeneity of \( \Delta PA/\Delta NA \) variation. The \( \beta \) is well known as the influence of covariates on the mean response of \( \Delta PA/\Delta NA \), and \( v_{0ij}, v_{0l} \) represents how \( \Delta PA/\Delta NA \) differs by subjects in terms of means. The \( v_{0ij}, v_{0l} \) are random effects on the population mean (location), while \( \omega_{ij} \) is a random effect that influences an individual’s variance (scale). Hedeker et al. (2008) allows the WS variation to vary across subjects, above and beyond the contribution of covariates to population mean and variation using the data of two cluster levels, namely, \( \log(\delta_{\varepsilon ij}^2) = u_{ij}' \tau + \omega_{ij} \). The proposed RL-RSS model further allows random scale variance \( \delta_{\omega ij k}^2 \) (the variance of \( \omega_{ij} \)) to be modeled by a log-linear structure and affected by the covariates. Three cluster levels of data are illustrated as an example.

\[
\log \left( \delta_{\varepsilon ij k}^2 \right) = u_{ij k}' \tau + \omega_{ij k}
\]

\[
\log \left( \delta_{\omega ij k}^2 \right) = z_{ij k}' \alpha
\]

The selected covariates are wave, gender, and smoking dependence assessment score (NDSS), as described above, where gender is a subject-level covariate (level-3), wave and NDSS are a wave-level covariate (level-2).

F. **Comparison Results**

Li’s 2012 three-level RL-RS model is extensively developed based on the regular three-level random effect model (RL model) noted in EQ. IV-1. The error variance \( \varepsilon_{ijk} \) is simply assumed to be a
constant $\delta_\varepsilon^2$ in RL model. The proposed three-level RL-RSS model (noted in EQ. II-19) can be further viewed as an extension of Li’s RL-RS model. In RL-RS model noted in EQ. IV-2., the variance of random scale effect $\delta_\omega^2$ is assumed to be a constant, not varied by the covariates.

Three-level random effect model (RL model):

$$y_{ijk} = x'_{ijk} \beta + v_{0ij} + v_{0i} + \varepsilon_{ijk}$$

with the conditions:

$$\begin{bmatrix} v_{0i} \\
v_{0ij} \end{bmatrix} \sim N \begin{pmatrix} 0 & 0 \\ 0 & \delta_0^2 \end{pmatrix}, \quad \varepsilon_{ijk} \sim N(0, \delta_\varepsilon^2)$$

$$EQA. IV-1$$

Three-level RL-RS model:

$$y_{ijk} = x'_{ijk} \beta + v_{0ij} + v_{0i} + \varepsilon_{ijk}$$

with the conditions:

$$v_{0ij} \sim N(0, \delta_v^2), \quad v_{0i} \sim N(0, \delta_0^2), \quad \varepsilon_{ijk} | \omega_i \sim N \left(0, \delta_{\varepsilon_{ijk}}^2 \right), \quad \omega_i \sim N(0, \delta_\omega^2)$$

$$\begin{bmatrix} v_{0i} \\
v_{0ij} \\
\omega_i \end{bmatrix} \sim N \begin{pmatrix} 0 & 0 & 0 \\ 0 & \delta_v^2 & \delta_{\omega_v} \\ 0 & \delta_{\omega_v} & \delta_\omega^2 \end{pmatrix}$$

$$log \left(\delta_{\varepsilon_{ijk}}^2\right) = u'_{ijk} \tau + \omega_i$$

$$EQA. IV-2$$
Visually, Figure 13 shows the RL model illustrated by EMA data without considering the error variance modeling (noted by EQ. IV-1),
Figure 14 illustrates the RL-RS model with the error variance model and random scale effects (i.e., EQ. IV-2), and
Figure 15 represents the RL-RSS model (i.e., EQ. II-19) that additionally models random scale variance on the basis of three level RL-RS model.

We present artificial data in these figures to better demonstrate model features. In the three figures, the circle dot in the left panel of figures represents the outcome measure $y_{ijk}$ at wave $j$/occasion $k$ for subject $i$. The population mean of $y$ across all subjects is depicted with black solid horizontal lines, and the average of $y_{ijk}$ are presented as dotted horizontal lines (subject $i=1$ colored in red, and subject $i=2$ colored in blue). Each individual subject has a dotted line for his/her average in the real dataset. For the illustration purpose, we plot only two subjects ($i=1$ and $i=2$) and two waves (wave $j=0$ and wave $j=1$) as the representative cases. The slanted solid black lines represent the population
time-trends (averaged over subjects), while the slanted dotted lines represent the time-trends for the two subjects.

The right panel of the figures provides the information of data dispersion or variation. Each diamond dot represents the data dispersion at wave $j$ for subject $i$ (red or blue diamonds) or the average of data variation across all subjects at wave $j$ (black diamonds). A higher diamond indicates a greater level of WS variation at the specific wave, conversely, low diamond dot reflects less WS variability. The solid line linking two diamond dots is indicative of time trend that the WS variation change across two waves. Red and blue solid lines in the right panel indicates the time trend of two subjects, and black solid line is for the average time trend of all subjects in terms of data dispersion.

Considering
Figure 13 first, the black solid horizontal line in the left panel at Wave 0 (i.e., baseline) corresponds to the population intercept $\beta_0$, and the black slanted solid line represents the population slope $\beta_1$. Covariates are allowed to affect this mean response by either raising or lowering the slanted solid line (main effect) or changing its slope (time interaction) in the usual way. The dotted lines of the two subjects at Wave 0 are indicative of subjects’ random intercept $u_{0i}$, which reflects the deviation of a subject from the mean response at Wave 0. In the figure, subject $i=1$ is above population mean with the positive value of $u_{01}$ while subject $i=2$ is below the mean line with the negative value of $u_{02}$. For an easy representation, $u_{01}$ is a positive value, while $u_{01}$ is a negative value to indicate their deviation from the population mean. The BS intercept variance ($\delta^2_0$) represents subject heterogeneity of random intercept effect at Wave 0: the closer the dotted lines are, the less heterogeneity occurs among subjects (if the dotted lines are close together then there is not much subject heterogeneity, conversely if the dotted lines are spread out then more heterogeneity is indicated). Similarly, the BS slope variance ($\delta^2_{\beta_j}$) reflects subject heterogeneity of the dotted trend lines. In
Figure 13, these slopes in the left panel vary between the two subjects, which corresponds to the notion that subjects vary in their time trajectories. Finally, the WS variance ($\delta^2_{\varepsilon_{ij}}$) is a parameter that indicates the degree of variation of a person’s data points around each of their horizontal dotted lines.
In Figure 13, as no error model is specified in the regular mixed-effect model (RL model, EQ. IV-1), the scale of WS variance (the degree of the dots spread-out, $\delta_e^2$) is assumed to be an invariant constant for all subjects and waves. The diamond dots in the right panel of
Figure 13 equal to a fixed value for two subjects in two waves, which indicate the constant WS variance ($\delta_{\epsilon_{ij}}^2$). The time trend of data variability is a flat line with no change across waves for both subjects and population average because of the constant attribute of error variance.
Figure 14 illustrates the concept of the random scale effect and error variance modeling inherent in RL-RS model EQ. IV-2. Notice that the dispersion of the observations around each of the horizontal dotted lines varies. At both Waves 0 and 1, more WS variance is observed for subject \(i=1\) than the subject \(i=2\), that is, \(\delta^2_{\varepsilon_{10}} > \delta^2_{\varepsilon_{20}}\) and \(\delta^2_{\varepsilon_{11}} > \delta^2_{\varepsilon_{21}}\). The right panel of the figure clearly illustrates the greater level of WS variance in subject \(i=1\) than subject \(i=2\) in both waves, shown by two red diamond dots higher above the blue diamond dots. This disparate WS variation across subjects is exactly what the random scale effect \(\omega\) captures, and the variance of random scale effect, \(\delta^2_\omega\), indicates the degree of subject heterogeneity in the WS variance. Notice also that the WS variance for both subjects is lessened at Wave 1 relative to Wave 0. This illustrates that the covariate Wave (and their coefficients \(\tau\)) can influence the WS variance \(\delta^2_{\varepsilon_{ijk}}\), as expressed in EQ. IV-2. Three parallel falling-trend lines shown in
the right panel indicate that the coefficient of Wave, \( \tau_1 \) would be negative (i.e., as Wave increases, the WS variance diminishes for two subjects and population trend). Covariates of the WS variance could be wave-varying variables like Wave, or subject-varying variables like gender, in which case the WS variation (across all occasions) of males would be more/less relative to females.

There is a homogenous time-trend of WS variation across all subjects in Figure 14. The subject heterogeneity of WS variance is invariant over time. In another words, all subjects share the same attribute of WS variance with the invariant slope \( \tau_1 \), and the random effect \( \omega_i \) does not change across wave. Besides, \( \delta^2_\omega \), the variance of random effect \( \omega_i \), is not affected by wave or any other relevant covariates. However, our proposed RL-RSS as in EQ. II-19 allows the possibilities.
As illustrated in the right panel of Figure 15, the slanted blue line connected by two blue diamond dots (for subject \(i=2\)) is steeper than the red line (for subject \(i=1\)), showing that subject \(i=2\) has a more rapid decline of WS variance over waves than subject 1. The time trend of WS variance \(\tau\) differs from subjects. The random effect \(\omega_{ij}\) is dedicated to capture the subject heterogeneity in the time-trend of WS variance. The random scale effect \(\omega_{ij}\) is allowed to vary by wave, and in turn produces the possibility of disparate time trend slope of WS variance. If we consider the WS variance \(\log\left(\delta^2_{\varepsilon_{ijk}}\right)\) as dependent variable in error model, \(\log\left(\delta^2_{\varepsilon_{ijk}}\right) = u'_{ijk}\tau + \omega_i\) in EQ. IV-1 can be treated as a simple random intercept model for scale (error variance), whereas \(\log\left(\delta^2_{\varepsilon_{ijk}}\right) = u'_{ijk}\tau + \omega_{ij}\) in EQ. II-19 can be viewed as a random slope.
model with random slope scale effect. The right panel of Figure 15 is an alternative way to illustrate the error variance model. The red and blue diamonds correspond to subject i’s error variance at wave j, denoted by $\delta^2_{\varepsilon ij}$. At both Wave 0 and Wave 1, subject $i=1$ is characterized with higher-than-average error variance, whereas subject $i=2$ has lower error variance relative to the population average. It reveals that subjects vary in error variance and random scale effect $\omega_i$ should necessarily be considered. In addition, the time trend of error variance over waves, represented by $\tau$, also varies by subjects. Subject $i=1$ apparently has a higher slope $\tau$ than Subject $i=2$, which addresses the need of random slope effect $\omega_{ij}$ and corresponds to the notion that subjects vary in time trajectory in terms of error variance. The variance $\delta^2_{\omega_{ijk}}$ associated with random
effect $\omega_{ij}$, reflecting the degree of subject heterogeneity in the WS variance, is not assumed to be a constant

Figure 15 or EQ. II-19. At each wave, the dispersion/spread-out of red and blue diamond is indicative of $\delta_{\omega_{ijk}}^2$, given that only two subjects are concerned. If diamond dots are close to each other, then there is not much subject heterogeneity in WS error variance (BS-WS variance $\delta_{\omega}^2$ is small); conversely if diamond dots are spread out, then more heterogeneity is present with big $\delta_{\omega}^2$. In
Figure 14, the degree of subject heterogeneity in $\delta^2_{\xi_{ij}}$ is identical across waves, that is, $\delta^2_{\omega}$ does not vary by subject or wave.
Figure 13. Illustration of random intercept model.
Figure 14. Illustration of random location scale model.
G. **Analysis Results**

The descriptive analysis of outcome measures $\Delta PA/\Delta NA$ and covariates are presented in TABLE III. The covariate NDSS is clustered at wave-level. Wave is also a wave-level variable, and gender is a subject-level covariate. An average 30 smoking events per person were reported, and 6 smoke events are recorded per person per wave in analysis dataset. One hundred fifty-eight subjects
with two or more waves and at least two events in each wave were included in the analysis. There are 869, 700, 925, 1024, and 1209 smoking episodes recorded at baseline, 6 months, 12 months, 24 months, and 60 months respectively. The counts, mean, standard deviation, minimum, and maximum are provided for continuous measures by wave. The counts and percentages are provided for dichotomized measures by each wave. Three-level RL model expressed in EQ. IV-1 was estimated as a basic comparison model, where ΔPA and ΔNA were separately modeled.
TABLE IV lists the results of the RL model including the covariates gender, wave, and NDSS. For each model, the parameter estimate, standard error of estimate, the number of subjects, the number of observations, and -2 log likelihood value are provided. In two RL models, the BS variance $\hat{\delta}_0^2$, WS between-wave (WS-BW) variance $\hat{\delta}_{\text{BW}}^2$, and WS variance $\hat{\delta}_c^2$ are highly significant. Significant BS variance $\hat{\delta}_0^2$ indicates that subjects do vary in their levels of smoking-related mood changes. It shows that the subject heterogeneity over the level of $\Delta$PA and $\Delta$NA is present at different waves. The WS error variance $\hat{\delta}_c^2$ accounts for the majority of variance, explaining approximately 75% total variation of outcome measures. It motivates the need to further investigate WS-WW variance like modeling the variance by potentially associated covariates.

As for the mean model, the intercept is highly significant in RL $\Delta$PA model. This indicates that smoking had a beneficial effect by increasing positive affect change ($\hat{\beta}_0=.379$, p=.006 for in
TABLE IV ΔPA model), but not for negative affect change ($\hat{\beta}_0 = -0.213, p=0.120$ for in
**TABLE IV** ANA model) when all covariates equal zero. Wave had a diminishing effect on smoking-related PA mood change ($\hat{\beta}_1 = -0.057$, $p = 0.003$).
TABLE IV ΔPA model), but no significant effect on NA. Namely, as time increased the smoking-related benefit to positive affect change decreased. Neither gender nor NDSS is significant, showing that neither smoking group nor NDSS has the effect on smoking-related mood change.

Table V lists the results of RL-RS model for the dependent variables ΔPA and ΔNA adjusting gender, wave, and NDSS, of which the covariates are identical with the model setup in RL model. The results for the mean model in RS-RL models are similar as those in RL models. Intercept is highly significant with positive estimate in ΔPA model, suggesting that smoking enhances positive affect change in the condition of all zero covariates. Wave still remains a diminishing effect on smoking-related PA mood change ($\hat{\beta}_1 = -.039, p = .009$ in Table V ΔPA model), but no significant effect is observed on ΔNA. Gender and NDSS still have no effect on mood change.

In terms of the error variance model, Wave has a significant effect in reducing variation for both ΔPA ($\hat{\tau}_2 = -.225, p < .001$ in Table V ΔPA model) and ΔNA ($\hat{\tau}_2 = -.170, p < .001$ in Table V ΔNA model); showing that the variation of smoking-related mood change diminished across time. Gender and NDSS score has no effect on the variation of smoking-related mood change (p values of $\hat{\tau}_2$ and $\hat{\tau}_3$ are all >.05). It reflects individual’s mood change variation is not affected by gender or the level of smoking dependence (NDSS).

Turning to the variance estimates, the BS variance estimate $\delta_0^2$, WS-BW variance estimate $\delta_0^2$ remains highly significant in RL-RS models. The random scale variance estimate $\delta_0^2$ is observed to be highly significant in all RL-RS models, which indicates the non-ignorable degree of between-subject WS variance. In another words, subjects act in heterogeneous behavior (viewed as between-subject disparity) to stabilize or be consistent with their mood change after smoking. We describe the consistency of smoking-related mood change within a subject as a WS variation. As to the covariance,
for both outcomes, the association between random location effect and random scale effect is seen to be significant. For ΔPA it is positive ($\hat{\delta}_{\hat{\omega}}^2=0.198$, p<0.001 in Table V), which indicates that subjects with ΔPA level have greater ΔPA mood change variation. Conversely, for ΔNA, this covariance is negative ($\hat{\delta}_{\hat{\omega}}^2=-.206$, p<.001 in Table V), which suggests that subjects with higher ΔNA levels (smoking-related negative mood changes) exhibit less ΔNA mood change variation. As the outcomes are the change of scores after smoking minus before smoking, the scale variance is reduced as the change score levels go toward zero (lower ΔPA change and higher ΔNA change). It is worth noting that zero is not the lower boundary of these change-scores, which varied from -9 to 9, and so these correlations do not necessarily reflect a floor effect of measurement.

The results of the RL-RSS model are presented in Table VI. Similar results are observed in the mean model. Intercept is highly significant with a positive estimate in the ΔPA model and with a negative value in the ΔNA model, suggesting that smoking enhances a positive affect mood and reduces a negative affect mood. Wave still remains a diminishing effect on smoking-related PA mood change ($\hat{\beta}_1=-.041$, p=.006 in Table VI ΔPA model), but no significant effect on NA. Gender and NDSS are not associated with ΔPA/ΔNA with p values >.05.

In terms of the error variance model, Wave has a consistent significant effect in reducing variation for both ΔPA ($\hat{\tau}_2=-.223$, p<.001 in Table VI ΔPA model) and ΔNA ($\hat{\tau}_2=-.171$, p<.001 in Table VI ΔNA model); suggesting that the variation in smoking-related mood change diminished across time. Non-significant Gender and NDSS reveals no association with mood change variation.

As for random scale variance model, nonsignificant intercept in ΔPA/ΔNA model suggests that no subject heterogeneity is present in the ΔPA/ΔNA variation among females at baseline—in other words, the variation of smoking-related mood change does not differ from female participants at
baseline. Gender has a significant effect in Table VI ΔNA model ($\hat{\alpha}_1=.629$, $p=.005$). Males exhibit a greater heterogeneity of ΔNA variation relative to females, that is, males are less similar with each other in the ability to stabilize their smoking-related mood changes relative to female cohorts. In a brief sum, the significant positive estimate indicates a greater level of subject heterogeneity in terms of WS mood change variation/consistency, while the significant negative estimate conversely represents less subject heterogeneity or more subject homogeneity in an individual’s ΔPA/ΔNA variation. The NDSS is significant in the ΔPA model and marginally significant in the ΔNA model ($\hat{\alpha}_3=-.296$, $p=.016$ in ΔPA model, $\hat{\alpha}_3=-.247$, $p=.095$ in the ΔNA model). It suggests that higher dependence on nicotine lessens subject heterogeneity in both ΔPA and ΔNA variations. A high nicotine-dependent population tends to be more homogeneous in mood change variation than are low-dependency cohorts.

A likelihood ratio test has been conducted to compare the fit of three types of models for both ΔPA/ΔNA outcomes. In the comparison between the RL-RS model and the RL model, the $p$ values computed in the likelihood ratio test are highly significant for the ΔPA/ΔNA models as seen in the bottom of Table VI. Thus, the RL model in null hypothesis is rejected in favor of the alternative RL-RS model at the significant level of .001. To distinguish the RL-RSS model from the RL-RS model, the similar likelihood ratio test was performed. The highly significant $p$ values ($p<.01$) demonstrate that RL-RSS should be considered as a better fit model relative to the RL-RS model ($p$ values are displayed in the bottom of Table VI for both the ΔPA and ΔNA models). When the RL-RSS model is compared to the RL model, the results are even more significant. In sum, the proposed three-level RL-RSS model is the best model to fit the smoking study EMA data, compared with the three-level RL-RS model and RL model.
TABLE III
DESCRIPTIVE ANALYSIS OF OUTCOME MEASUREMENTS AND COVARIATES

<table>
<thead>
<tr>
<th>Year</th>
<th>Baseline</th>
<th>0.5</th>
<th>1.25</th>
<th>2</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔPA</td>
<td>N</td>
<td>869</td>
<td>700</td>
<td>925</td>
<td>1024</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.59</td>
<td>0.51</td>
<td>0.26</td>
<td>0.34</td>
</tr>
<tr>
<td>Std</td>
<td></td>
<td>1.70</td>
<td>1.54</td>
<td>1.37</td>
<td>1.31</td>
</tr>
<tr>
<td>Min</td>
<td></td>
<td>-8.60</td>
<td>-3.80</td>
<td>-7.00</td>
<td>-5.40</td>
</tr>
<tr>
<td>Max</td>
<td></td>
<td>9.00</td>
<td>9.00</td>
<td>8.80</td>
<td>6.40</td>
</tr>
<tr>
<td>ΔNA</td>
<td>N</td>
<td>869</td>
<td>700</td>
<td>925</td>
<td>1024</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>-0.37</td>
<td>-0.47</td>
<td>-0.28</td>
<td>-0.36</td>
</tr>
<tr>
<td>Std</td>
<td></td>
<td>1.79</td>
<td>1.66</td>
<td>1.43</td>
<td>1.36</td>
</tr>
<tr>
<td>Min</td>
<td></td>
<td>-9.00</td>
<td>-8.80</td>
<td>-8.60</td>
<td>-8.60</td>
</tr>
<tr>
<td>Max</td>
<td></td>
<td>9.00</td>
<td>9.00</td>
<td>7.20</td>
<td>4.80</td>
</tr>
<tr>
<td>NDSS</td>
<td>N</td>
<td>869</td>
<td>695</td>
<td>925</td>
<td>1024</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>2.37</td>
<td>2.56</td>
<td>2.73</td>
<td>2.68</td>
</tr>
<tr>
<td>Std</td>
<td></td>
<td>0.84</td>
<td>0.76</td>
<td>0.65</td>
<td>0.75</td>
</tr>
<tr>
<td>Min</td>
<td></td>
<td>0.80</td>
<td>0.00</td>
<td>0.00</td>
<td>0.80</td>
</tr>
<tr>
<td>Max</td>
<td></td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>3.90</td>
</tr>
</tbody>
</table>

Gender

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>47</th>
<th>54</th>
<th>57</th>
<th>47</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td></td>
<td>52.38</td>
<td>50.54</td>
<td>56.84</td>
<td>56.44</td>
</tr>
<tr>
<td>Male</td>
<td>60</td>
<td>49.46</td>
<td>43.16</td>
<td>43.56</td>
<td>45.98</td>
</tr>
</tbody>
</table>
### TABLE IV
THREE-LEVEL RANDOM EFFECT LOCATION MODEL (RL MODEL) OF POSITIVE AFFECT AND NEGATIVE AFFECT

MAXIMUM LIKELIHOOD ESTIMATES (P value)

<table>
<thead>
<tr>
<th></th>
<th>ΔPA</th>
<th>ΔNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 Log Likelihood</td>
<td>16171</td>
<td>16795</td>
</tr>
<tr>
<td>Maximum Observations Per Subject</td>
<td>149</td>
<td>149</td>
</tr>
<tr>
<td>Observations Used</td>
<td>4722</td>
<td>4722</td>
</tr>
<tr>
<td>Subjects</td>
<td>158</td>
<td>158</td>
</tr>
</tbody>
</table>

Mean Model $y_{ijk} = x_{ijk}'\beta + u_{0ij} + v_{0i} + \epsilon_{ijk}$

<table>
<thead>
<tr>
<th></th>
<th>Estimate (P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept $\hat{\beta}_0$</td>
<td>0.379 (0.006)</td>
</tr>
<tr>
<td>Gender $\hat{\beta}_1$</td>
<td>0.139 (0.183)</td>
</tr>
<tr>
<td>Wave $\hat{\beta}_2$</td>
<td>-0.057 (0.003)</td>
</tr>
<tr>
<td>NDSS $\hat{\beta}_3$</td>
<td>0.076 (0.136)</td>
</tr>
<tr>
<td>WS Variance $\hat{\delta}_\epsilon^2$</td>
<td>1.620 (0.000)</td>
</tr>
<tr>
<td>BS Variance $\hat{\delta}_0^2$</td>
<td>0.241 (0.000)</td>
</tr>
<tr>
<td>WS-BW Variance $\hat{\delta}_{ij}^2$</td>
<td>0.219 (0.000)</td>
</tr>
</tbody>
</table>
### TABLE V
THREE-LEVEL RANDOM LOCATION RANDOM SCALE MODEL (RL-RS MODEL) OF POSITIVE AFFECT AND NEGATIVE AFFECT

<table>
<thead>
<tr>
<th></th>
<th>ΔPA</th>
<th>ΔNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 Log Likelihood</td>
<td>14971</td>
<td>15185</td>
</tr>
<tr>
<td>Maximum Observations Per Subject</td>
<td>149</td>
<td>149</td>
</tr>
<tr>
<td>Observations Used</td>
<td>4722</td>
<td>4722</td>
</tr>
<tr>
<td>Subjects</td>
<td>158</td>
<td>158</td>
</tr>
</tbody>
</table>

Mean Model $y_{ijk} = x'_{ijk} \beta + v_{0ij} + u_{0i} + \varepsilon_{ijk}$

<table>
<thead>
<tr>
<th></th>
<th>Estimate (P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept $\hat{\beta}_0$</td>
<td>0.359 (0.002)</td>
</tr>
<tr>
<td>Gender $\hat{\beta}_1$</td>
<td>0.109 (0.192)</td>
</tr>
<tr>
<td>Wave $\hat{\beta}_2$</td>
<td>-0.039 (0.009)</td>
</tr>
<tr>
<td>NDSS $\hat{\beta}_3$</td>
<td>0.057 (0.157)</td>
</tr>
</tbody>
</table>

Variance Model $\log \left( \delta_{\varepsilon_{ijk}}^2 \right) = u'_{ijk} \tau + \omega_i$

<table>
<thead>
<tr>
<th></th>
<th>Estimate (P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept $\hat{\tau}_0$</td>
<td>0.540 (0.001)</td>
</tr>
<tr>
<td>Gender $\hat{\tau}_1$</td>
<td>0.203 (0.201)</td>
</tr>
<tr>
<td>Wave $\hat{\tau}_2$</td>
<td>-0.225 (0.000)</td>
</tr>
<tr>
<td>NDSS $\hat{\tau}_3$</td>
<td>0.065 (0.186)</td>
</tr>
<tr>
<td>BS-WS Variance $\delta^2_{\omega}$</td>
<td>0.821 (0.000)</td>
</tr>
<tr>
<td>Covariance $\delta_{u\omega}$</td>
<td>0.198 (0.000)</td>
</tr>
<tr>
<td>BS Variance $\delta^2_{0}$</td>
<td>0.138 (0.000)</td>
</tr>
<tr>
<td>WS-BW Variance $\delta^2_{0j}$</td>
<td>0.105 (0.000)</td>
</tr>
</tbody>
</table>

Likelihood ratio test vs. RL model (P value) <.001 <.001
<table>
<thead>
<tr>
<th>MAXIMUM LIKELIHOOD ESTIMATES (P value)</th>
<th>ΔPA</th>
<th>ΔNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 Log Likelihood</td>
<td>14959</td>
<td>15141</td>
</tr>
<tr>
<td>Maximum Observations Per Subject</td>
<td>149</td>
<td>149</td>
</tr>
<tr>
<td>Observations Used</td>
<td>4722</td>
<td>4722</td>
</tr>
<tr>
<td>Subjects</td>
<td>158</td>
<td>158</td>
</tr>
</tbody>
</table>

Mean Model $y_{ijk} = x_{ijk}'\beta + v_{0ij} + v_{0i} + \epsilon_{ijk}$ Estimate (P value)

| Intercept $\hat{\beta}_0$ | 0.335 (0.004) | -0.248 (0.010) |
| Gender $\hat{\beta}_1$    | 0.120 (0.157) | -0.013 (0.860) |
| Wave $\hat{\beta}_2$      | -0.041 (0.006) | 0.017 (0.137) |
| NDSS $\hat{\beta}_3$      | 0.066 (0.108) | -0.051 (0.127) |

Variance Model $\log(\delta^2_{\epsilon_{ijk}}) = u'_{ijk}\tau + \omega_{ij}$

| Intercept $\hat{\tau}_0$ | 0.520 (0.001) | 0.893 (0.000) |
| Gender $\hat{\tau}_1$    | 0.191 (0.206) | 0.167 (0.399) |
| Wave $\hat{\tau}_2$      | -0.223 (0.000) | -0.171 (0.000) |
| NDSS $\hat{\tau}_3$      | 0.074 (0.136) | -0.083 (0.104) |

Random Scale Effect Model $\log(\delta^2_{\omega_{ij}v}) = z_{ijk}'\gamma$ Estimate (P value)

| Intercept $\hat{\gamma}_0$ | 0.371 (0.224) | 0.401 (0.268) |
| Gender $\hat{\gamma}_1$    | -0.210 (0.358) | 0.629 (0.005) |
| Wave $\hat{\gamma}_2$      | 0.286 (0.165) | 0.115 (0.534) |
| NDSS $\hat{\gamma}_3$      | -0.296 (0.016) | -0.247 (0.095) |
| Covariance $\delta_{\nu\omega}$ | 0.181 (0.000) | -0.195 (0.000) |
| BS Variance $\delta^2_{\delta}$ | 0.141 (0.000) | 0.090 (0.000) |
| WS-BW Variance $\delta^2_{\delta} | 0.106 (0.000) | 0.034 (0.009) |

Likelihood ratio test vs. RL model (P value) | <.001 | <.001 |
Likelihood ratio test vs. RL-RS model (P value) | 0.008 | <.001 |
H. **Example SAS Codes**

Below is a sample of SAS program syntax that was used to run the three-level RL-RSS model for the analysis in Table VI in SAS PROC NLMIXED procedure. The parameters to be estimated along with the observed variables are displayed in Table VII.

The PROC NLMIXED statement in SAS invokes NLMIXED procedure. The dataset ANAL is analyzed by the proposed three-level RL-RSS model as specified in Table VI. The data set ANAL should be formatted in long structure, of which one observation represents a smoking event at one wave for one subject. The GCONV option specifies a relative gradient convergence criterion as 1E-12. The PARMS statement lists the names of parameters and specifies their initial values. The choice of initial values is critical in PROC NLMIXED procedure. Inappropriate initial values usually lead to the failure of convergence. In order to find a reasonable initial value for the parameters, the simpler models like the RL-RS model or the RL model are highly recommended to run first. The analysis results from simpler models would be a good choice to optimize the initial values of the RL-RSS model. The parameters not listed in the PARMS statement are assigned a starting value of 1 by default.

The MODEL statement is the mechanism for specifying the conditional distribution of the data given the random effects. The outcome variable PA or NA is followed by a tilde sign and normal distribution with the mean Z and variance VARE. The mean model Z is predicted by the potential covariates with the addition of random location effects U0, D1, D2, D3, D4, and D5. The population mean of PA or NA is associated with three covariates GENDER, WAVE, and NDSS. The variables BGENDER, BWAVE, and BNDSS correspond to the coefficient parameters respectively, where B0 is the intercept parameter, and U0 is the random location effect at subject level, noted by \( u_{0i} \) in Table VI. The random location effect at wave level, noted by \( u_{0ij} \), differs from waves. The dummy variables W1-
W5 were created per the algorithm in Table VII. The design of dummy variables for each wave allows random location effects to vary in waves. The error variance VARE is modeled by a log-linear representation including the same covariates as in the mean model (GENDER, WAVE, and NDSS). The fixed scale parameters T0, TGENDER, TWAVE, TNDSS correspond to the intercept, and coefficient parameters for GENDER, WAVE and NDSS respectively in error variance model. The random scale variance \( \delta_{\omega_{ijk}}^2 \) (variable in VW0), the variance of random scale effect \( \omega_{ij} \) (variable in W0), is further modeled by a log linear structure and associated with the same three covariates. The coefficient parameters are named as A0, AGENDER, AWAVE, and ANDSS in order. The RANDOM statement defines the random effects \( u_{0i} \) (variable in VU0), \( \omega_{ij} \) (variable in W0), and \( u_{0ij} \) (variables in D1-D5). The RANDOM statement consists of a list of the random effect variables followed by a tilde and the distribution for random effects, and then a SUBJECT ID variable. As described previously, all random effects are assumed normal distributed with mean zero. The variance parameters to be estimated are listed in Table VII, which are the random location variance \( \delta_{0}^2 \) (variable in VU0), \( \delta_{\omega}^2 \) (variable in VWAVE) and the covariance between random location effect and random scale effect at subject-level \( \delta_{\nu\omega} \) (CVW).

The following are examples of SAS Codes:

```
PROC NLMIXED DATA=ANAL GCONV=1E-12;
PARMS
  B0=-0.2967 BWAVE=0.000811 BGENDER=-0.00303 BNDSS=-0.01378
  EVAR0=0.1650 EWAVE=-0.1809 EGENDER=0.1704 ENDSS=-0.2551
  SVAR0=-0.0465 SWAVE=0.01809 SGENDER=0.4041 SNDSS=-0.4933
VU0=0.1 CVW=0.2 VWAVE=0.1;
```
Z = B0 + BGENDER*GENDER + Bwave*WAVE + BNDSS*NDSS + U0 + D1*W1 + D2*W2 + D3*W3 + D4*W4 + D5*W5;

VARE = EXP(T0 + TGENDER*GENDER + TWAVE*WAVE + TNDSS*NDSS + W0);

VW0 = EXP(A0 + AGENDER*GENDER + AWAVE*WAVE + ANDSS*NDSS);

MODEL PA ~ NORMAL(Z, VARE);

RANDOM U0 W0 D1 D2 D3 D4 D5 ~ NORMAL([0,0,0,0,0,0,0], [VU0, CVW, VW0, 0, 0, VWAVE, 0, 0, 0, VWAVE, 0, 0, 0, VWAVE, 0, 0, 0, 0, VWAVE, 0, 0, 0, 0, 0, VWAVE]) SUBJECT=ID;

RUN;
<table>
<thead>
<tr>
<th>Parameters to be estimated</th>
<th>Outcome variables</th>
<th>Covariate variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>In RL-RSS model specification</td>
<td>In SAS PROC NLMIXED program</td>
<td></td>
</tr>
<tr>
<td>Intercept $\beta_0$</td>
<td>B0</td>
<td>PA, NA</td>
</tr>
<tr>
<td>Gender $\beta_1$</td>
<td>BGENDER</td>
<td></td>
</tr>
<tr>
<td>Wave $\beta_2$</td>
<td>B WAVE</td>
<td></td>
</tr>
<tr>
<td>NDSS $\beta_3$</td>
<td>BNDSS</td>
<td></td>
</tr>
<tr>
<td>Intercept $\tau_0$</td>
<td>A0</td>
<td></td>
</tr>
<tr>
<td>Gender $\tau_1$</td>
<td>TGENDER</td>
<td></td>
</tr>
<tr>
<td>Wave $\tau_2$</td>
<td>T WAVE</td>
<td></td>
</tr>
<tr>
<td>NDSS $\tau_3$</td>
<td>TNDSS</td>
<td></td>
</tr>
<tr>
<td>Intercept $\alpha_0$</td>
<td>A0</td>
<td></td>
</tr>
<tr>
<td>Gender $\alpha_1$</td>
<td>AGENDER</td>
<td></td>
</tr>
<tr>
<td>Wave $\alpha_2$</td>
<td>A WAVE</td>
<td></td>
</tr>
<tr>
<td>NDSS $\alpha_3$</td>
<td>ANDSS</td>
<td></td>
</tr>
<tr>
<td>Covariance $\delta_{\nu \omega}$</td>
<td>CVW</td>
<td></td>
</tr>
<tr>
<td>BS Variance $\delta_0^2$</td>
<td>VU0</td>
<td></td>
</tr>
<tr>
<td>WS-BW Variance $\delta_0^2$</td>
<td>V WAVE</td>
<td></td>
</tr>
</tbody>
</table>
V. CONCLUSIONS

In longitudinal or clustered studies, the assumption made on the consistent variance within subjects and the homogeneous variance across subjects may be violated. Random subject effects can further be correlated with error terms. The methods for collecting EMA data usually yield thirty or forty observations per subject. Such a large number of repeated observations within subjects offers a greater opportunity in heterogeneous variance modeling than conventional longitudinal studies.

By treating the observations WS-WW, a three-level mixed-effects location scale model separates between/WS variation, and allows covariates to influence those variances, thereby providing a more comprehensive examination for the analysis of EMA data.

However, the existing methodology fails to investigate the effects of covariates on subject heterogeneity of data dispersion (within-cluster data dispersion is noted as error variance statistically). For such a purpose, we propose a three-level mixed-effects random location scale model with modeling random scale variance (RL-RSS model) in this dissertation. Different from the existing methods, this model allows covariates to influence both error variance and random scale variance through a log-linear representation.

The error variance varies across subjects through a subject-level normally distributed random scale effect, above and beyond the contribution of covariates on error variance. The subject-level random scale effect and random location effect are allowed to correlate with each other. Parameter estimation was based on the combination of the MML method and the EB method. An iterative Newton-Raphson solution was used to maximize the log likelihood, and
multidimensional Gauss-Hermite quadrature is used to numerically approximate integral values. A SAS program via PROC NLMIXED using adaptive quadrature was developed to fit the proposed model.

The data from EMA Adolescent Smoking Study are used to illustrate the application of the proposed model. In this study, adolescents responded to smoking events from hand-held computers. Their positive and negative mood changes were collected before and on/after smoking events. The data were collected for seven consecutive days and several times a day during the collection period.

One aspect of the research interest focuses on the degree to which covariates could explain WS variation for ΔPA and ΔNA and also account for subject heterogeneity of data variation, over and above the influence of covariates on the mean response. The covariates included gender, wave, and NDSS score. A three-level clustering data structure, level-1 smoking events/occasions nested within level-2 waves nested within level-3 subjects, was used in the data analysis. Three models (RL-RSS model, RL-RS model, and RL model) were fit to the data for the comparison purpose, and likelihood ratio tests showed that the proposed three-level RL-RS model was superior to the other two three-level models.

A simulation process was carried out to validate the accuracy and reliability of the proposed three-level RL-RSS model. One hundred simulations, each with 80,000 observations (50 smoking events/occasions nested within each of 2 waves, wave is further nested within each of the 800 subjects), were generated under the three-level RL-RSS model with three covariates (either continuous or dichotomous). Three models: a three-level random intercept model, a three-level RL-RS model, and a three-level RL-RSS model were fitted on the simulated data. These
covariates were specified to have effects on variances as well as the overall mean. The simulation results show that RL-RSS resolves the intercept overestimation of random scale variance occurring in RL-RS model. Thus ignoring the random scale variance model may lead to the poor quality of intercept parameter estimation of random scale variance. The RL model is an even worse choice with overestimated intercept of error variance model and less reliable variance parameter estimates.
CITED LITERATURE


<table>
<thead>
<tr>
<th>NAME</th>
<th>Hong Li</th>
</tr>
</thead>
</table>
| EDUCATION | BS Food Science, Jiangnan University, Wuxi, China, 1998.  
MS Bioengineering, Chongqing University, Chongqing, China, 2004.  
MS Biostatistics, University of Illinois at Chicago, Chicago, 2009.  
| TEACHING EXPERIENCE | Department of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago: Model Selection. |
| PROFESSIONAL EXPERIENCE | Statistician, Rush University Medical Center, 2009–2013.  