7-2015

# Mixed-Effects Location-Scale Models for Conditionally Normally Distributed Repeated-Measures Data

Ryan Walters
*University of Nebraska-Lincoln*, ryanwalters@huskers.unl.edu

# MIXED-EFFECTS LOCATION-SCALE MODELS FOR CONDITIONALLY NORMALLY DISTRIBUTED REPEATED-MEASURES DATA

by

Ryan W. Walters

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Psychology

Under the Supervision of Professors Lesa Hoffman and Jonathan Templin

Lincoln, Nebraska

July, 2015

# MIXED-EFFECTS LOCATION SCALE MODELS FOR CONDITIONALLY NORMALLY DISTRIBUTED REPEATED-MEASURES DATA

Ryan W. Walters, PhD

University of Nebraska, 2015

Advisors: Lesa Hoffman & Jonathan Templin

Hypotheses about psychological processes are most frequently dedicated to individual mean differences, but individual differences in variability are likely to be important as well. The mixed-effects location-scale model estimates individual differences in both mean level and variability in a single model, and represents an important advance in testing variability-related hypotheses. However, the mixed-effects location-scale model remains relatively novel to empirical scientists as statistical software is often handicapped by more complex models and a paucity of methodological studies exist examining the statistical properties of this model.

This dissertation investigates the mixed-effects location-scale model through the development of open-source software for its estimation and through simulation and empirical studies. First, the theoretical framework for the mixed-effects location-scale model is presented followed by a description of the Metropolis-Hastings algorithm developed to estimate this model. Then, two simulation studies are presented evaluating the power to detect and predict individual differences in variability as well as identify the consequences of model misspecification. Finally, results of an empirical analysis examining individual differences in mean level and variability of unstructured movements from a sample of older adults with and without probable mild Alzheimer's disease is presented.

Results of the power investigation simulation study indicated that the power to detect the scale-model random intercept variance and the effect of an individual-level predictor of residual variability increased with greater numbers of individuals and occasions, and that failing to detect the scale-model random intercept variance essentially precluded the detection of systematically varying fixed effects for an individual-level predictor of residual heterogeneity. Results of the misspecification simulation study indicated that misspecifying the location model and/or scale model for the residual variance had consequences only for fixed and random effects on the same side of the model. Finally, results of the empirical data analysis indicated individuals with probable mild Alzheimer's disease averaged less movement compared to healthy individuals, but did not differ in the variability of their unstructured movements.

In sum, this dissertation provides information useful to empirical scientists as they progress from study design through analysis, interpretation, and reporting for publication.

# DEDICATION

To Dominik (aka, The D Nugget; aka, Big Cat): I want you to know you can do anything you want. Don't let anyone, including yourself, try to convince you otherwise.

And, of course, To Mrs. LOP!: Our future will be here before we know it…

## ACKNOWLEDGEMENTS

A decisively humbling number of people have provided me support throughout my academic career. First and foremost, I want to acknowledge the incredible debt I owe Mike Fatis of Minnesota State University, Mankato who saw something in me that I did not see in myself, particularly in the presence of my department-assigned advisor who (almost) had me convinced I would amount to nothing. I also must acknowledge Wayne Mitchell of Missouri State University who significantly redirected my academic career by agreeing without hesitation to mentor me following my program change. Unbeknownst to him, with one nonchalant sentence he (with his ever-so-Wayne inflection and mannerism) convinced me that I can do anything I want; everything took off from there. Oh, and as you would have expected, Wayne, this is also one lengthy document!

This dissertation would not exist if not for the time and patience of my co-chairs Lesa Hoffman and Jonathan Templin (that's right, *co-chairs*; I want to clear the air that it is only an annoying technicality to state this any other way). Lesa, I cannot imagine where I would be if you had not saved me from treading water. For that, I truly owe you an unpayable debt! You are an outstanding mentor who had my best interests in mind even before I even started my doctoral training. I cannot overstate how much I appreciate the availability of your time and promptness of your feedback as well as your willingness to provide me with any available teaching and research opportunities; your influence can be easily seen in all my academic work. In light of all of that, however, what I appreciate most of all is your direct, no-nonsense approach; you will always be my favorite hard ass! Jonathan, just over two years ago you also began mentoring me simply because Lesa asked. I cannot overstate how much I appreciate your leap of faith as this dissertation

could not have been completed without you selflessly dedicating your time, opinions, and feedback. Through everything that has happened over the past two years, it was truly amazing that you always managed to find time for me in your hectic (understatement!) schedule and consistently maintained a positive mindset that allowed us to get shit done.

I need to thank Cal Garbin for dedicating the time to read my dissertation as well as his thoughts on the entire dissertation process. Cal, you reminded me that the dissertation should not be some daunting, dreaded chore; instead, the dissertation should be viewed as an incredible opportunity to learn and this type of indulgence rarely exists after graduating. What's more, you consistently demonstrated how to be a great teacher *for* your students. In observing your interactions with students, it was obvious how much you care about them. I have learned much from you; the approach and honesty you bring to teaching will continue to influence how I teach my own students. I am also grateful to Walt Stroup for the time he dedicated to this *bison* of a dissertation. Walt, your comments were pointed and insightful, permanently modifying how I think about and speak to mixed-effects models. This dissertation would not have been as statistically accurate without you dedicating the time to provide high-quality feedback.

I also must thank my family for their constant (and undeserved) support. Mrs. LOP!, since I began considering pursuing a PhD, you have provided me with unquestioning support. Nothing I have done could have been completed without you holding down all other areas of my life. I love you more than you know and look forward to seeing what our future holds. Dominik, it has been an amazing past 18 months watching you grow into your own little person and I am excited (well, not as visibly excited as you, but excited nonetheless…) see what your future holds. Nothing in life

worth fighting for will come easy; work hard, be strong enough to trust the right people, and you will accomplish anything you set your mind to. And too, I want both of you to know that I will spend the rest of my life trying to make up for all the time during my training in which I was physically *there* but mentally absent.

I also want to thank my parents. Mom, you are my educational role model. You taught me so much about persistence and demonstrated what it means to actually *earn* a degree. It was truly amazing to watch and you provided me with all the evidence I needed to know that it was possible to finish a degree while working full time with a family. Dad, you taught me the value of working my ass off, although, admittedly, it took me a while to figure out how to do that. As far back as I can remember, you always taught me about the value of getting a quality education; specifically, to do something inside so that I would not have to deal with uncontrollable elements such as bitter cold or sweltering heat.

Finally, I want to thank Amber Watts of the University of Kansas who graciously provided me with the data to analyze for the empirical analysis described in Chapter 5 as well as the time she dedicated to discuss the analysis and provide thoughtful comments on initial drafts. I want to acknowledge Paul Turner of Creighton University for encouraging me to pursue a PhD. I also need to thank Melody Scholl-Miller and Matt Kilbride of the University of Nebraska-Lincoln for the technological support by ensuring my proposal (which was essentially a virtual meeting) and dissertation defense went on without a single technological hiccup.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1: INTRODUCTION AND THEORETICAL FRAMEWORK

Hypotheses about psychological processes are frequently constructed to detect differences in the mean level of an outcome between groups or individuals (Hoffman, 2007). In this context, differential variability (i.e., heterogeneity) between groups with cross-sectional data, or between individuals with repeated-measures data has been viewed traditionally as a statistical nuisance that should be corrected, not as an interesting phenomenon for study; this is especially true for applications using the general linear model (e.g., regression or analysis of variance). Over the past two decades, this view has begun to change, especially with increased use of linear mixed-effects models and an increased ability to estimate more complex versions of these models. Empirical scientists are recognizing that important questions can be answered by examining differential variability alongside the traditional evaluation of mean responses.

To illustrate why hypotheses pertaining to differential variability are important, consider physical activity, which has been shown consistently to benefit cognitive performance and to decrease the risk of dementia (Ahlskog, Geda, Graff-Radford, & Petersen, 2011). Because most research hypotheses pertain to mean levels of physical activity, important information is absent regarding the day-to-day (or hour-to-hour) variability in physical activity. For example, older adults with Alzheimer's disease (AD) tend to report walking or completing household chores as their primary source of physical activity, and they tend to be sedentary when not engaged in these activities (Watts, Vidoni, Loskutove, Johnson, & Burns, 2013). Sedentary behavior creates a floor effect in the measurement of physical activity that necessarily decreases variability across occasions within the same individual. This decrease in variability could potentially bias

inferences from analyses that model mean levels of physical activity. Therefore, identifying factors, such as depression symptoms, that explain changes in the variability of physical activity would allow researchers to implement specific interventions to reduce variability (e.g., reducing depression symptoms), after which subsequent behavioral interventions could be used to produce *more consistent* (i.e., less variable) increases in physical activity to achieve potential health benefits.

Evaluating hypotheses about mean levels alongside hypotheses regarding differential variability in nested (or multi-level) data, such as when repeated occasions are nested within individuals, requires a mixed-effects (or multi-level) location-scale model. This model extends the traditional linear model to allow the explicit prediction of variability between and within individuals and allows additional random effects that represent individual differences in variability. However, the model is relatively novel in practice and preliminary methodological work to investigate its properties must be conducted before empirical scientists can use this method confidently in their research. Therefore, this dissertation will address several important methodological questions pertaining to the use of the mixed-effects location-scale model for conditionally normally distributed outcomes. It will also report the results of an empirical analysis showing the flexibility of this model to evaluate individual differences in both mean levels and variability.

This chapter contains the theoretical framework of the mixed-effects location-scale model and begins with a description of the single-level linear model, its assumptions, and a discussion of differential (or heterogeneous) variability alongside common statistical methods used by empirical scientists to correct heterogeneity. This is

followed by a description of the linear mixed-effects model with details of how this model has been used traditionally to account for and predict heterogeneity. Finally, the mixed-effects location-scale model is introduced with background of how the model has been used in practice.

Following the theoretical framework of the mixed-effects location-scale model, chapter 2 presents a complete example describing model-building approaches beginning with the single-level linear model and working through the mixed-effects location-scale model. Chapter 3 then provides details regarding Bayesian theory and the Markov chain Monte Carlo estimator used to estimate the mixed-effects location-scale models in this dissertation. This discussion is followed by two methodological studies, presented in chapter 4, which evaluate the power to detect and predict the scale-model random intercept variance as well as the consequences of misspecifying the location and/or scale model. Finally, in chapter 5, an empirical data analysis using the mixed-effects location-scale model is presented to compare individual differences in movement variability between individuals with and without probable Alzheimer's disease.

**The Single-Level Linear Model**

The single-level linear model (often termed *regression* or *analysis of variance*) is used to study the relationship between one continuous outcome and one or more predictor variables that are either continuously- or categorically-valued, assumed to be measured without error. Specifically, the model regresses the outcome onto a linear combination of predictor variables using ordinary least squares (OLS) to determine the extent to which the predictors minimize residual variance (Pedhazur, 1997), as shown in (1.1).

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_{p-1} X_{i,p-1} + e_i \qquad (1.1)$$

Here, $Y_i$ is the continuous outcome for individual $i$. $\beta_0$ represents the fixed intercept parameter (where the adjective *fixed* implies an effect applies equally to every individual in the sample) and is the average model-predicted outcome when all predictor values for individual $i$, $X_{i,1}$ to $X_{i,p-1}$, equal zero. Note that the subscript of the last predictor, $p - 1$, results in the total number of coefficients including the fixed intercept, to be $p - 1 + 1 = p$, which is used to maintain consistency with the matrix formulation of the model described below. Finally, $\beta_1$ to $\beta_{p-1}$ represent the fixed effect for a given predictor.

For a predictor $X_{i,k}$ measured on a continuous scale, the fixed effect is interpreted as the increase in $Y_i$ per one-unit increase in $X_{i,k}$. Note that $k$ is a generic index used to indicate a specific predictor, where $k = 1$ to $p - 1$. For a different predictor $X_{i,k'}$ measured on a categorical scale using reference coding (i.e., dummy coding; $X_{i,k'} = 0$ for a reference group and $X_{i,k'} = 1$ for some other group), the fixed effect represents the average difference in $Y_i$ between groups or between design conditions. Finally, $e_i$ represents the residual value (aka, error) for individual $i$ calculated as the difference between the observed outcome and model-predicted outcome resulting from the linear combination of the other terms in the right-hand side of (1.1).

The model in (1.1) is termed *single-level* because there are no additional dependencies (i.e., correlations) induced by the sampling design and, consequently, there is only one error term (i.e., the residual) for each individual $i$. Further, all predictors are modeled as fixed effects. As a final note, the name of model (1.1) will later be appended to note that it is the *location model*, as the linear combination of effects produces the model-based estimate of the mean of the conditional distribution of the data implied by (1.1).

The matrix form of the single-level linear model is shown in (1.2).

$$Y_i = \mathbf{X}_i\boldsymbol{B} + e_i \quad (1.2)$$

Note that the notation used in (1.2) is specific to individual $i$, and although not typical to traditional references for linear models (e.g., Pedhazur, 1997), this notation was used purposefully to map directly onto the linear mixed-effects model framework discussed below. Here, $Y_i$ is a 1 x 1 scalar representing the outcome for individual $i$, $\mathbf{X}_i$ is a 1 x $p$ row vector of observed predictor variable values $p$ for individual $i$, $\boldsymbol{B}$ is a $p$ x 1 column vector of fixed effects for each predictor $p$, and $e_i$ is a 1 x 1 scalar representing the residual value for individual $i$. Note that the first element of $\mathbf{X}_i$ is typically set to 1 to represent the value that multiplies the intercept parameter in the first element of $\boldsymbol{B}$, and that matrix $\boldsymbol{B}$ has no subscript $i$ because its fixed effects apply equally to all individuals.

**Assumptions of the single-level linear model.** The primary statistical assumptions of the single-level linear model are that the expected outcome, $Y_i$, conditional on the predictors in $\mathbf{X}_i$, is a linear function of one or more predictors, $E(Y_i|\mathbf{X}_i) = E(\mathbf{X}_i\boldsymbol{B}) = \mathbf{X}_i\boldsymbol{B}$. When estimated values of $\boldsymbol{B}$ are used (i.e., $\widehat{\boldsymbol{B}}$), the function is often referred to as $\mathbf{X}_i\widehat{\boldsymbol{B}} = \hat{Y}_i$. This assumption directly involves the mean for the conditional distribution of $Y_i$, more commonly known as the *location model* (aka, structural model or model for the means), which includes the observed predictor variables $\mathbf{X}_i$ and fixed effects $\boldsymbol{B}$ (e.g., intercept, slopes). More generally, the location model describes how average effects across individuals are used to obtain the estimated *model-predicted* outcome, $\hat{Y}_i = \mathbf{X}_i\widehat{\boldsymbol{B}}$. The remaining statistical assumptions directly involve the residual values, $e_i$, more commonly known as the *scale model* (aka, stochastic model or model for the variances), which assume that the residual values follow a normal

distribution with a mean of 0 and constant, homoscedastic variance, $e_i \sim N(0, \sigma_e^2)$, where $\sigma_e^2$ is the variance of the residual values. As the notation will change slightly for the location-scale models to come, here note that $\sigma_e^2$ is constant across all individuals $i$ (i.e., there is no subscript $i$).

When all assumptions are satisfied, the single-level linear model produces the most efficient (and consistent) *best linear unbiased estimates* (BLUE) of the fixed effects, $\widehat{\boldsymbol{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$, where $\mathbf{X}$ and $\mathbf{Y}$ are taken across all $n$ individuals and are of size $n$ x $p$ and $n$ x 1, respectively. Here, *best* implies the lowest squared error, *efficient* implies the estimated fixed effects are the most accurate (i.e., have the smallest variance), *consistent* implies the estimated fixed effects approach their true population values as sample size increases toward infinity, and *unbiased* implies the estimated fixed effects, averaged over repeated samples using the same sample size, represent the true population fixed effects (Pedhazur, 1997; Williams, Grajales, & Kurkiewicz, 2013).

***Linearity.*** The assumption of linearity implies that a straight line sufficiently represents the relationship in the sampled population. Linearity implies $Y_i$ is a linear function of fixed effects $\boldsymbol{B}$, not a linear function of predictor variables $\mathbf{X}_i$ (i.e., predictor variables can nonlinear, e.g., $X_i^2$), and also implies that the fixed effects are additive (Berry & Feldman, 1985). In the presence of non-linearity, fixed effect standard errors are biased and statistical power is reduced to the extent of the non-linearity, whereas the fixed effects themselves remain unbiased (Williams, Grajales, & Kurkiewicz, 2013).

***Independence.*** The assumption of independence of residuals implies that there is no correlation between residual values. That is, the residual value for one individual is assumed uncorrelated with the residual value from any other individual. In the presence

of non-independence, the fixed effects remain consistent and unbiased, but their standard errors are downwardly biased, increasing the probability of Type I errors (Blair, Higgins, Topping, & Mortimer, 1983).

Non-independence often results from a sampling design that has an underlying nested structure of clustered groups or repeated measures (e.g., students in the same school or occasions within the same individual, respectively). If a nested structure is inherent in the data, then the single-level linear model can be extended to a mixed-effects linear model that accounts for non-independence using random effects. This model is described in detail below. If nested structures are not inherent, non-independence due to unknown reasons can be evaluated graphically by plotting residuals in some chronological order (e.g., date) or spatial structure (e.g., Euclidean distance) or evaluated statistically using the Durbin-Watson bounds test (Durbin & Watson, 1950, 1951).

***Normality.*** The assumption that residuals are marginally normally distributed is required when testing the statistical significance of fixed effects in a linear model (Pedhazur, 1997). In the presence of non-normal residuals, fixed effect standard errors can become biased; however, the fixed effect estimates remain consistent and unbiased. Note that the assumption of normality does not apply to the *marginal* distribution of either the outcome $Y_i$ or any of the predictors (marginally or jointly) $\mathbf{X}_i$. However, by the assumption that residual values are marginally normally distributed, the *conditional* distribution of $Y_i$ is assumed to be normally distributed, $f(Y_i|X_i) \sim N(\mathbf{X}_i\boldsymbol{B}, \sigma_e^2)$. The marginal distribution of residuals can be evaluated subjectively using a histogram and/or normal probability plot (e.g., Q-Q plot) or evaluated objectively using statistical tests such as the Kolmogorov–Smirnov test (Kolmogorov, 1933; Smirnov, 1948).

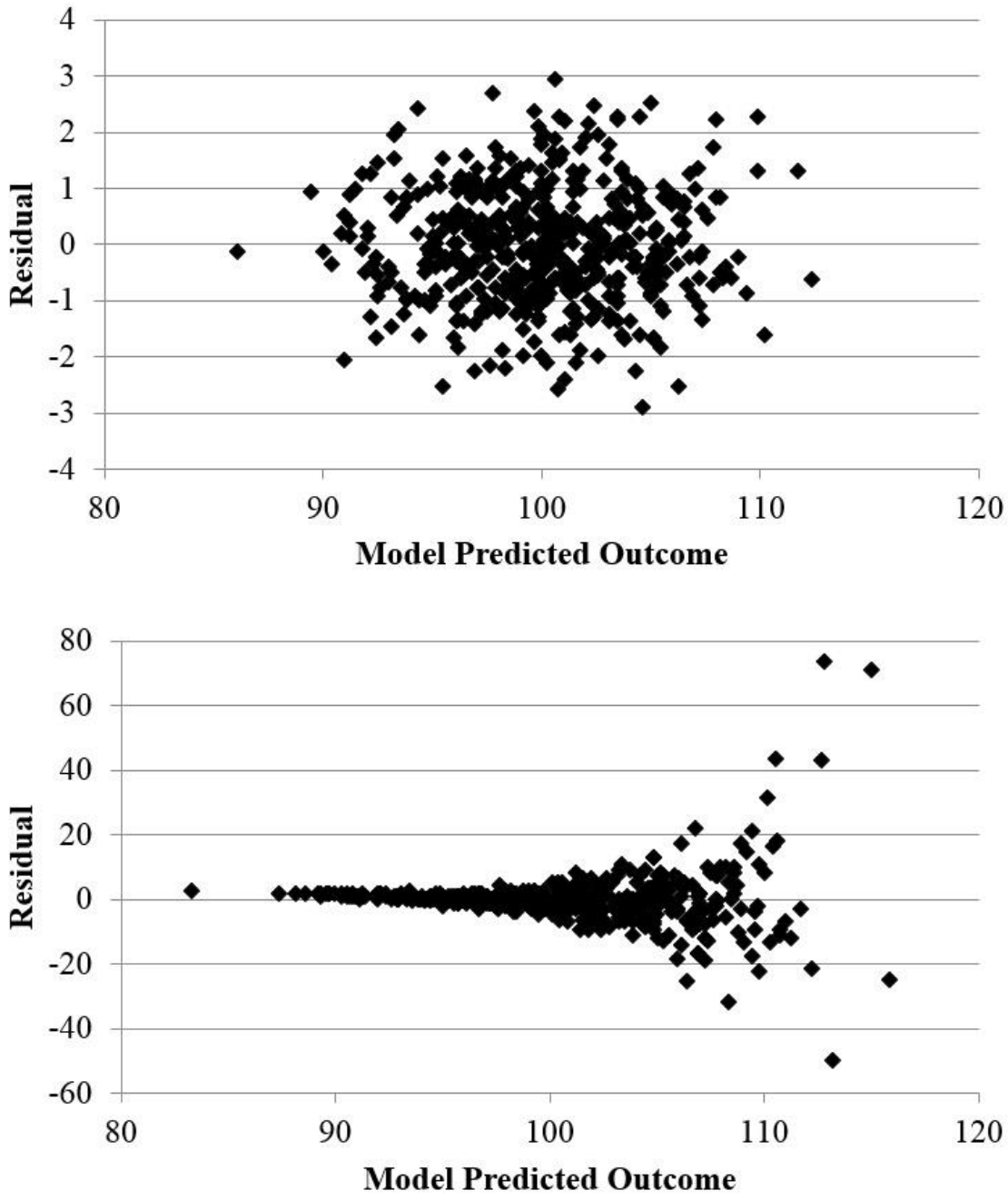


*Figure 1.1*. Homoscedastic variances (a) and heteroscedastic variances (b)

***Homoscedasticity.*** The assumption of homoscedasticity of residuals implies that residual values have a constant variance, $\sigma_e^2$, across all individuals (or, individuals' values of $\mathbf{X}_i$; see Figure 1.1a). Note that when predictor variables are categorical, this assumption is similar to the homogeneity of variance assumption from an analysis of

variance model. A violation of homoscedasticity is termed *heteroscedasticity* (or *heterogeneity of variance*), where residual variance increases or decreases with increases in predictor values $\mathbf{X}_i$ (see Figure 1.1b). In the presence of heteroscedasticity, fixed effects $\boldsymbol{B}$ remain consistent and unbiased, but because predictor values with larger residual values provide less information, standard errors become upwardly or downwardly biased depending on the value of $\mathbf{X}_i$ (Allison, 1999; Hayes & Cai, 2007).

*Detecting heteroscedasticity in the single-level linear model.* Heteroscedasticity can be detected subjectively using visual inspection or objectively using statistical tests. Heterogeneity is observed visually by inspecting a scatterplot of the model-predicted outcome or observed predictor values against the residual values (similar to Figures 1.1a and 1b). Often, however, heterogeneity may not be as pronounced as in Figure 1.1b, so an objective statistical test may be required. Although several tests are available, the Breusch-Pagan test and White's general test are used most commonly (Greene, 2002).

The Breusch-Pagan test (Breusch & Pagan, 1979) is a Lagrange multiplier (or score) test that calculates whether the residual variance estimated from the single-level linear model assuming homoscedasticity can be predicted by a log-linear combination of predictors as shown in (1.3).

$$\log(\sigma_i^2) = \alpha_0 + \mathbf{z}_i\boldsymbol{\alpha} \tag{1.3}$$

Here, $\sigma_i^2$ is the (log of the) estimated residual variance for individual $i$. $\mathbf{z}_i$ is the scale-model equivalent of the location-model design/predictor matrix, $\mathbf{X}_i$, and is an $n$ x $g$ matrix of $g$ predictor variables for individual $i$. $\boldsymbol{\alpha}$ is the scale-model equivalent to the column vector of linear model coefficients, $\boldsymbol{B}$, and is a $g$ x 1 column vector of fixed

effect coefficients, and $\alpha_0$ is the intercept parameter representing the (log of the) residual variance when $\mathbf{z}_i = 0$.

Two important characteristics of (1.3) are worth noting. First, the fact that $\sigma_i^2$ is based on a *log-linear combination* of predictor effects in $\mathbf{z}_i\boldsymbol{\alpha}$ is simply a convenient mathematical transformation that, once exponentiated, ensures the predicted residual variance remains greater than zero (Harvey, 1976). Second, the scale-model predictor variables in $\mathbf{z}_i$ are chosen explicitly by the empirical scientist and can include any, all, or none of the variables included in $\mathbf{X}_i$ from the location model. The test statistic for the Breusch-Pagan test is calculated by comparing the model in (1.3) to the null model that assumes homoscedasticity (i.e., $\boldsymbol{\alpha} = 0$). In large samples, this difference is distributed as a chi-square with degrees of freedom equal to the number of evaluated scale-model predictors, $\mathbf{z}_i$. A statistically significant test statistic indicates heterogeneous residual variance as $\boldsymbol{\alpha} \neq 0$.

White's general test (White, 1980) is considered a special case of the Breusch-Pagan test and is used when the empirical scientist has no specific hypothesis about the structure of residuals. The test also uses the model shown in (1.3), but with primary difference that $\mathbf{z}_i\boldsymbol{\alpha}$ is forced to include the log-linear combination of *all* possible predictors, their squares, and interactions. The test statistic and degrees of freedom are calculated exactly the same as the Breusch-Pagan test.

Although effective at identifying overall heterogeneity, both the Breusch-Pagan test and White's general test have significant shortcomings. First, because both tests are based on model comparisons, they fail to provide the unique effect for each scale-model predictor; thus, it may be difficult to determine which predictor in $\mathbf{z}_i$ is most important in

creating residual heteroscedasticity (although one could presumably program the Breusch-Pagan test manually in any statistics package using the heterogeneous variance model described below). Second, nonsignificant test results indicate the absence of linear heterogeneity, but do not explicitly rule out the potential for heterogeneity that is truly nonlinear. Third, both the Breusch-Pagan test and White's general test have been shown to produce Type I errors in the presence of residual non-normality (Koenker, 1981). In addition, in small samples, White's general test may have low statistical power as a large number of scale-model predictors necessarily results in many square and interaction effects effectively reducing (or completely eliminating) degrees of freedom.

If heteroscedasticity has been indicated either graphically or statistically, it is important to note that it can also result from a violation of several assumptions discussed previously. For example, a violation of the linearity assumption could result in model misspecification such as when missing an interaction effect (as indicated in Figure 1.1b) or nonlinear predictor effect, which can lead to heteroscedastic residual variances (Fox, 2008); this applies to both experimental studies and non-experimental studies (Bryk & Raudenbush, 1988). Further, ignoring the presence of nested structures and estimating a single-level linear model violates the assumption of independence of residuals. In this situation, heteroscedastic residuals result from failing to use a linear mixed-effects model to account for differential effects of a predictor variable (i.e., random slopes; see chapter 2). In addition, the use of an unreliable measure to obtain predictor values violates the assumption that predictors are measured without error. As a result, heteroscedasticity may result due to poor measurement precision at specific predictor values (Hayes & Cai,

2007). Taken together, it is critical that all other assumptions be satisfied when considering the legitimacy of observed heteroscedasticity.

*Correcting heteroscedasticity in the single-level linear model.* In the single-level linear model, heteroscedasticity has traditionally been viewed as an impediment to unbiased inference. Because the Breusch-Pagan test and White's general test do not explicitly correct for heterogeneity, methods have been developed to provide parameter estimates that are relatively unbiased in the presence of heteroscedasticity. Although many methods exist, for brevity, only variable transformation, robust standard errors, weighted least squares, and heterogeneous variance models will be discussed.

When heteroscedasticity results from a violation of the assumption of linearity, one of the traditional methods to reduce the effects of heteroscedasticity is to employ a variable transformation. Although many potential transformations are available, their overall purpose is similar—to reduce the effect of heteroscedasticity by decreasing the scale of the residual values. As an example, the natural log transformation may be applied to positively skewed dependent and/or predictor variables when the variance of residuals increases with increases in the model-predicted outcome. Although transformations may be effective in reducing inference biases due to heteroscedasticity, they have potential downsides. As previously described when discussing normality of residuals, bias due to heteroscedasticity is only minimized on the transformed model scale, so fixed effect standard errors on the data scale remain downwardly biased (Manning, 1998). In addition, the transformed model clouds interpretation due to non-linear and non-additive model scale effects (Bryk & Raudenbush, 1988).

An alternative to variable transformation is to use a robust standard error correction, accomplished by using variants of a sandwich (or empirical) estimator. The sandwich estimator uses a heteroscedasticity (or asymptotically) consistent covariance matrix (HCCM) that provides less biased estimates without transformation or restrictive distributional assumptions such as normality or homoscedasticity of residuals (Huber, 1967; White, 1980). Specifically, in typical linear model analyses where the assumption homoscedasticity of residuals is satisfied, the conditional variance of the outcome is estimated based on the estimated residual variance. However, in the presence of heteroscedasticity, the estimated residual variance will be an over- or under-estimate of conditional variance across predicted outcome values. In this situation, the sandwich estimator uses *observed squared residuals* to calculate the sampling variance—as opposed to using the estimated residual variance under the assumption of homoscedasticity—which decreases bias as individuals with large squared residuals are given less influence on the estimates of fixed effect standard errors (Kauermann & Carroll, 2001).

Five primary HCCM variants have been developed—HC0 introduced by White (1980), HC1, HC2, and HC3 developed by MacKinnon and White (1985), and HC4 developed by Cribari-Neto (2004). Note that HC0 is appropriate only for large samples and although HC1, HC2, and HC3 are asymptotically equivalent to HC0, they have better small samples properties and are less biased with homoscedastic residuals (MacKinnon & White, 1985). Further, Long and Ervin (2000) have shown that HC3 is most appropriate with samples of less than 250 individuals, especially when evaluating coefficients most responsible for creating heteroscedasticity, whereas Cribari-Neto (2004) has shown that

HC4 is less biased than HC3 in the presence of high leverage (i.e., outlier) values or when residuals are homoscedastic. Despite the advantage of maintaining data-scale interpretations, relaxing distributional assumptions comes at a price. The use of a sandwich estimator reduces downward bias in standard errors by increasing the sampling variance as appropriate, which can result in inefficient estimates and wider confidence intervals when compared to the standard error estimates produced by the linear model under the assumption of homoscedasticity (Kauermann & Carroll, 2001).

A third available method to correct bias resulting from heteroscedasticity is to use a different estimation technique such as weighted least squares (WLS). WLS transforms the heteroscedastic linear model into a homoscedastic model by giving more weight to individuals with greater estimated precision (Greene, 2002). That is, individuals with smaller residual values are treated as more influential because their estimates are, in theory, closer to the *true* location of the regression line. Notably, WLS can be used when the form of heteroscedasticity is known or unknown; however, the efficiency of WLS is improved dramatically in the former situation. In addition, weights could be chosen *a priori* based on theory, but such situations rarely occur. More frequently, weights are determined empirically by estimating the linear model, evaluating all predictor–residual plots to determine the cause of heteroscedasticity, regressing the squared residual values onto the predictor variables causing the heteroscedasticity, and calculating the weight(s) as the reciprocal of the squared model predicted value (Greene, 2002). If these steps are repeated until the parameter estimates stop changing (i.e., they converge), the procedure is termed *iteratively reweighted least squares*. Finally, it is important to note that WLS only produces consistent estimates when the model is correctly specified by assuming

that the true weight values are known exactly. Therefore, choosing incorrect weights, or weights from a misspecified heteroscedastic model, leads to decreased efficiency and increased bias in parameter estimates (Hayes & Cai, 2007).

One final method to correct bias due to heteroscedasticity is to use a *heterogeneous variance model*, which is directly applicable to experimental data that would typically be analyzed by analysis of variance (Bryk & Raudenbush, 1988). That is, a heterogeneous variance model is most applicable to the single-level model when heteroscedastic variances are observed across levels of a categorical predictor, such as when heterogeneity of variance is observed between treatment and control groups. The heterogeneous variance model does not correct or adjust for heteroscedasticity, but instead estimates a residual variance for each group $j$ (i.e., $\sigma_{e_j}^2$) and is analogous to estimating separate linear models using data from each individual group $j$. Therefore, by using group-specific standard errors, fixed effects inferences are less biased as the model ensures Type I error rates remain closer to the nominal level.

**The Linear Mixed-Effects Model**

As stated above, a violation of the independence of residuals assumption indicates a non-zero residual correlation between individuals, implying that the data potentially have some underlying nested structure. Because repeated occasions nested within an individual (probably) have a higher correlation than occasions from different individuals, and given this correlation is typically deemed non-trivial, a linear mixed-effects model must be used—a failure to do so increases the risk of inaccurate standard errors and misleading inference. For repeated measures data, the linear mixed-effects model explicitly models this correlation by partitioning between-individual variability from

residual variance via location-model random effects. In the repeated-measures context, location-model random effects represent individual-specific deviations from location-model fixed effects (i.e., individual differences; the effect is not the same for everyone in the sample). As an aside, the inclusion of location-model random effects is where the adjective *mixed-effects* originates; that is, a mixed-effects model includes both fixed and random effects.

Before continuing, it is worth noting that the linear mixed-effects models described in this dissertation apply directly to repeated-measures data, with repeated occasions at level 1 nested within individuals at level 2 (i.e., the linear mixed-effects model specified as a multi-level model). Further, the linear mixed-effects models described here are primarily concerned with modeling individual effects (i.e., the conditional model) as opposed to population-averaged effects (i.e., the marginal model). As such, much of the description of linear mixed-effects models follows from this sampling design (i.e., occasions nested within individuals). Yet it is important to note that linear mixed-effects models, as well as the methods proposed throughout, are not limited specifically to this type of data, and also can be applied to individuals nested within groups (see, for example, Raudenbush & Bryk, 2002 or Leckie, French, Charlton, & Browne, 2014).

The general form of the linear mixed-effects model for repeated-measures data is presented in (1.4).

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{B} + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i \tag{1.4}$$

Here, subscript $i$ is included for all matrices that vary across individuals and $n_i$ denotes the number of occasions in $\mathbf{Y}_i$ for individual $i$. $\mathbf{Y}_i$ is an $n_i$ x 1 column vector of observed

outcomes for $n$ occasions within individual $i$. $\mathbf{X}_i$ is an $n_i$ x $p$ design matrix of $p$ observed predictor variable values across the $n$ occasions within individual $i$. $\boldsymbol{B}$ is a $p$ x 1 column vector of fixed effects for the intercept and each of the $p - 1$ predictors. $\mathbf{Z}_i$ is an $n_i$ x $q$ design matrix of $q$ occasion-level predictors that have location-model random effects for $n$ occasions within individual $i$. In addition, $\mathbf{u}_i$ is a $q$ x 1 column vector of $q$ location-model random effect coefficients (i.e., deviations from the specific location-model fixed effects) for each individual $i$. Finally, $\mathbf{e}_i$ is an $n_i$ x 1 column vector of residual values for $n$ occasions within individual $i$. Note that for all models discussed in this dissertation, $\mathbf{Z}_i$ will include an $n_i$ x 1 column vector of ones as the first column to represent the random intercept (although this is not required by the linear-mixed effects model), all random effects will have analogous fixed effect terms in $\boldsymbol{B}$. Further, when $\mathbf{u}_i = \mathbf{0}$ for all individuals $i$, the linear mixed-effects model in (1.4) reduces to a single-level linear model in (1.2) (Littell, Milliken, Stroup, & Wolfinger, 1996).

In the linear mixed-effects model, the column vector of residual values for individual $i$, $\mathbf{e}_i$, is assumed to be multivariate normally distributed with a mean of 0 and with a positive semi-definite covariance matrix, $\mathbf{R}_i$, $\mathbf{e}_i \sim \mathbf{N}_{n_i}(\mathbf{0}, \mathbf{R}_i)$. $\mathbf{R}_i$ contains the variance and covariance of the distribution of residual values and is a positive semi-definite $n_i \times n_i$ matrix for the $n$ occasions for individual $i$. More commonly, mixed-model notation posits that $\mathbf{R}$ is constant between individuals (i.e., $\mathbf{R}_i = \mathbf{R}$), however, to remain consistent with notation appearing later in this dissertation, the subscript $i$ is included here. By definition, for $\mathbf{R}_i$ to be positive semi-definite, the determinant of $\mathbf{R}_i$ must be a non-negative value, or stated another way, $\mathbf{R}_i$ can be inverted, at minimum. For repeated-measures data, $n_i$ is the number of $n$ repeated occasions for individual $i$, which

may vary between individuals (e.g., due to missing data). Thus, $\mathbf{R}_i$ may be of different dimensions across individuals in the same sample. Further, note that when $n_i$ is large, $\mathbf{R}_i$ contains $\frac{n_i(n_i+1)}{2}$ unique elements which may lead to difficulty in having converged estimates if all possible covariances across the $n_i$ occasions are estimated uniquely.

In addition, the column vector of location-model random effect values for individual $i$, $\mathbf{u}_i$, is assumed to be multivariate normally distributed with a mean vector of zero and positive semi-definite covariance matrix, $\mathbf{G}_i$, $\mathbf{u}_i \sim \mathbf{N}_q(\mathbf{0}, \mathbf{G}_i)$. $\mathbf{G}_i$ contains the variances and covariances of the distribution of random effect values and is a positive semi-definite $q$ x $q$ matrix, where $q$ is the number of occasion-level predictors with random effects in $\mathbf{Z}_i$. Note that when the linear mixed-effects model assumes constant $\mathbf{G}_i$ between individuals, then $\mathbf{G}_i = \mathbf{G}$; however, the subscript $i$ will be retained for homogeneous $\mathbf{G}$ to map directly onto the models discussed below. Further, note that random effect variances will be denoted as $\sigma^2_{u_r}$, where the subscript $r$ indexes a specific location-model random effect in $\mathbf{Z}_i$ (e.g., $\sigma^2_{u_0}$ indicates the location-model random intercept variance).

It is important to note that a linear mixed-effects model for repeated measures data partitions random effect variances from residual variance to account for correlation due to nested structures, but it does not *explain* variance in the outcome. That is, when interest is primarily on individual differences (i.e., a conditional model), the between-individual variability in $\mathbf{G}_i$ and within-individual variability in $\mathbf{R}_i$ are re-aggregated into the $\mathbf{V}_i$ matrix, calculated as shown in (1.5)

$$\mathbf{Var}(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{Z}_i) = \mathbf{V}_i = \mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i^T + \mathbf{R}_i, \tag{1.5}$$

where $\mathbf{V}_i$ is the covariance matrix for the multivariate normal distribution of $\mathbf{Y}_i$ conditional on $\mathbf{X}_i$ and $\mathbf{Z}_i$ for individual $i$, $f(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{Z}_i) \sim \mathbf{N}_{n_i}(\mathbf{X}_i\boldsymbol{B}, \mathbf{V}_i)$. Note that $\mathbf{Z}_i^T$ indicates the transpose of $\mathbf{Z}_i$.

**Assumptions of the linear mixed-effects model.** The primary statistical assumptions of the linear mixed-effects model are similar to the single-level linear model presented above, but with greater complexity due to multiple levels or sampling units of analysis. This section begins with a discussion of model specification concerns and then proceeds to describe the assumptions for both the location and scale models, with specific emphasis is placed on detecting and remedying heteroscedasticity at differing levels of analysis.

***Model specification.*** In the linear mixed-effects model, misspecification of the $\mathbf{X}_i$ in the location model, misspecification of the structure of $\mathbf{R}_i$, or misspecification of either the number of random effects or the structure of $\mathbf{G}_i$ can be detrimental to the underlying assumptions of the model relative to the severity of misspecification. With that said, although the greater complexity of the linear mixed-effects model magnifies the consequences of model misspecification when compared to the single-level linear model, a well-chosen linear mixed-effects model will be more accurate than a single-level linear model that fails to separate between-individual variability in $\mathbf{G}_i$ from within-individual variability in $\mathbf{R}_i$. Therefore, it is critical that the location and scale models are both specified as correctly (as least wrong) as possible. Because outcomes depend jointly on residuals and random effects, misspecification of the location model may result in incorrect residual values for individuals as well as inaccurate estimates of random effects (Hilden-Minton, 1995, section 4.1). Further, the distributions of unstandardized random

effects are necessarily different when $\mathbf{X}_i$ and $\mathbf{Z}_i$ contain different predictor variables; thus, it has been recommended that predictors within $\mathbf{Z}_i$ be subsumed within $\mathbf{X}_i$ (Verbeke & Molenberghs, 2000).

Finally, note that random effects can only be independent when there is no correlation beyond what can be explained systematically by the level-2 predictor variables in the location model. That is, there are no additional levels of nesting (e.g., individuals cannot be additionally nested within groups). This implies that location-model random effects have been specified correctly. Further, incorrectly specified location-model random effects (i.e., random slopes; see chapter 2) redistribute non-independent variance to lower-level variance components, in this case $\sigma_e^2$, which can potentially bias fixed effect standard errors and result in biased inferences (Snijders & Bosker, 2012).

***Location model assumptions.*** The assumptions of the location model include that the expected outcome $\mathbf{Y}_i$ is a linear function of one or more perfectly reliable predictors that can be measured at any level of the model, $E(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{Z}_i) = \mathbf{X}_i\boldsymbol{B} = \mathbf{Y}_i$. In addition, it is worth noting that, although not technically an assumption, for repeated-measures data in which occasions are nested within individuals, location-model random effects can only be estimated for level-1 predictors. That is, random effects at level 2 represent differences *between individuals*; therefore, additional random effects cannot be estimated for level-2, individual-level predictors because there is no additional level of nesting by which individuals could vary (e.g., individuals are assumed not to be additionally nested within groups).

***Residual assumptions.*** As stated above, the linear mixed-effects model assumes residual values to be multivariate normally distributed with a mean of zero and positive semi-definite covariance matrix $\mathbf{R}_i$, $\mathbf{e}_i \sim \mathbf{N}_{n_i}(\mathbf{0}, \mathbf{R}_i)$. In addition, the model *can* assume residual values to be independent and constant between individuals (although not technically required). More generally, $\mathbf{R}_i$ is simply assumed to have the correct covariance structure specified for the residuals within an individual (Littell et al., 1996). Many alternative covariance structures are available for $\mathbf{R}_i$, with the choice of structure depending on the data and study design, such that, within an individual, $\mathbf{R}_i$ can be structured so residuals are independent or correlated, with or without homoscedastic variances (see Littell, Pendergast, & Natarajan, 2000). Of particular note, the linear mixed effects model *can* estimate residuals within an individual to be independent and constant across repeated measurements, $\mathbf{R}_i = \sigma_e^2 \mathbf{I}_{n_i}$ (Laird & Ware, 1982), where $\mathbf{I}_{n_i}$ is an $n_i$ x $n_i$ identity matrix (i.e., diagonal matrix of ones) with $n_i$ representing the number of $n$ repeated occasions for individual $i$ (i.e., implies zero correlation between occasions).

With that said, it is an important distinction that specifying $\mathbf{R}_i$ to have a heterogeneous variance structure across occasions within an individual is not the same as *predicting* (or allowing) heterogeneity *between individuals* because in this instance heterogeneous $\mathbf{R}_i$ would consist of identical variance estimates for every individual. That is, the heterogeneity of residual variances in $\mathbf{R}_i$ would still be assumed to be constant across individuals (i.e., $\mathbf{R}_i = \mathbf{R}$), although this assumption can be relaxed through the use of a heterogeneous variance model described below.

***Random effect assumptions.*** The linear mixed-effects model assumes random effect values to be multivariate normally distributed with a mean of zero and positive

semi-definite covariance matrix $\mathbf{G}_i$, $\mathbf{u}_i \sim \mathbf{N}_q(\mathbf{0}, \mathbf{G}_i)$. Similar to $\mathbf{R}_i$, the model *can* assume random effect values to be constant over individuals, although this assumption can be relaxed by using the heterogeneous variance model described below.

As stated above, one reason to estimate the linear mixed-effects model is to account for non-independent residuals. For repeated-measures data, after accounting for between-individual differences via properly specified random effects, residuals can be assumed independent across individuals, although remaining non-independence can exist and can be tested empirically (i.e., if $\mathbf{R}_i$ is diagonal for all $i$). Further, the (conditional) linear mixed-effects model assumes that residuals and random effects are independent of each other (i.e., no covariance between residuals in $\mathbf{R}_i$ and random effects in $\mathbf{G}_i$), an assumption that provides the formal definition of what a *level* indicates when the linear mixed-effects model is specified as a multi-level model—the number of sets of *independent* variance components. Therefore, repeated-measures data has two *levels* because the level-1 residuals and level-2 random effects are independent of each other.

***Detecting heteroscedasticity in the mixed-effects model.*** In general, the linear mixed-effects model for repeated-measures data can assume that residuals and random effects have constant variance across individuals (Snijders & Bosker, 2012; i.e., $\mathbf{R}_i = \mathbf{R}$ and $\mathbf{G}_i = \mathbf{G}$, respectively). It is often assumed that each specific residual and random effect variance and covariance is homogeneous for all predictors across all levels of analysis. That is, both level-1 residuals and level-2 random effects are often assumed to have constant variance and covariance across all values of level-1 and level-2 predictors. This assumption is evaluated as detailed next.

*Detecting heterogeneous residual variances.* Assuming all random effects have been correctly specified, diagnosing heterogeneity in residual variances requires estimating a single-level linear model for each level-2 unit using only level-1 predictors, assuming sufficient level-1, within-unit sample sizes (Hilden-Minton, 1995). Estimating unit-specific models isolates the level-1 effects and ignores confounding due to level-2 predictors. All methods described above to detect heterogeneity in the single-level linear model can be used to test for residual heterogeneity within these unit-specific models.

Further, a general summary statistic has been proposed by Raudenbush and Bryk (1987, 2002) to indicate overall heterogeneity in level-1 residuals across level-2 units without connection to a specific predictor variable. This statistic is presented in (1.6).

$$ls_{\text{total}} = \frac{\sum_i [df_i \log(\sigma_i^2)]}{\sum_i df_i} \tag{1.6}$$

Here, $df_i$ is the number of occasions within an individual $i$ minus the number of level-1 predictors minus 1, and $\log(\sigma_i^2)$ is the natural log of the residual variance from a given unit-specific linear model. Using the quantity $ls_{\text{total}}$ from (1.6), a standardized dispersion measure, $d_i$, is calculated shown in (1.7).

$$d_i = \sqrt{\frac{df_i}{2}}[\log(\sigma_i^2) - ls_{\text{total}}] \tag{1.7}$$

Assuming the level-1 residuals are normally distributed, the summary statistic $H$ can then be calculated as the sum of squared $d_i$, as shown in (1.8).

$$H = \sum_i d_i^2 \tag{1.8}$$

The heterogeneity statistic, $H$, is distributed as a chi-square with degrees of freedom equal to the number of individuals minus 1, such that a statistically significant result

indicates significant heterogeneity of the level-1 residuals. Note that Raudenbush and Bryk (2002) recommend $H$ only be used when $df_i \geq 10$ because when $df_i < 10$ the null distribution of $H$ is not a chi-square. If, however, $df_i < 10$ for the majority of level-2 units, Snijders and Bosker (2012) have suggested a simulation approach that ultimately uses equations (1.6) through (1.8) to evaluate heteroscedasticity.

*Heteroscedasticity of random effect variances.* Due to the confounding of random effects by residuals as described above, evaluating heteroscedasticity of random effects is only recommended after the adequacy of residuals has been confirmed (Snijders & Bosker, 2012). Although the methods described above could be used to detect heteroscedasticity of random effects, the empirical Bayes estimates of random effects are affected by shrinkage to the overall mean; thus, it has been recommended that only the graphical methods be used when evaluating heteroscedasticity of random effects (Houseman, Ryan, and Coull, 2004).

***Correcting heteroscedasticity in the mixed-effects model.*** For repeated-measures data, non-constant, heterogeneous random effect variances can indicate that the model is missing a random effect in $\mathbf{Z}_i$ for a level-1 predictor or that $\mathbf{X}_i$ is missing an interaction effect between a level-1 and level-2 predictor (i.e., a cross-level interaction effect; see Raudenbush & Bryk, 2002). As a result, Bryk and Raudenbush (1988) have suggested that all level-1 predictors be considered random until proven otherwise, which, of course, assumes sufficient level-2 sample sizes with which to estimate all random effects. If the subsequent model fails to converge when estimating an additional random effect, a potential workaround may be to use the variable as a predictor of random slope variance

in a heterogeneous variance model or mixed-effects location scale model, as described below. However, this conjecture has not yet been studied.

In addition, if heterogeneity in random effects is observed across individuals (i.e., non-constant $\mathbf{G}_i$), and no additional level of analysis can be explicitly defined by the sampling design, then a violation of this assumption can be remedied by modeling an empirical $\mathbf{G}_i$ matrix at given values of predictors in a heterogeneous variance model or a mixed-effects location-scale model described below. That is, the $\mathbf{G}_i$ is estimated as unique to an individual (or, stated another way, heterogeneous between individuals) conditional on the values of specified predictor(s). Note that this option is most efficient computationally for a categorical predictor, but heterogeneity across values of a continuous predictor can also be estimated.

**Heteroscedasticity as an Interesting Phenomenon**

To this point, heteroscedasticity has been presented alongside the other location and scale model assumptions of the single-level and mixed-effects linear models specifically with the goal of detecting and correcting heteroscedasticity in an effort to produce less biased fixed effect inferences. That is, traditional corrections have treated heteroscedasticity as a nuisance that only introduces bias into the model, not as an interesting phenomenon to study. Although it has been shown that heteroscedasticity can result from misspecification of the location model due to omitted fixed effects and interactions (Raudenbush & Bryk, 2002), the identification of predictors of scale-model heteroscedasticity is crucial to indicating which location model predictors and interactions require further study. Therefore, in recent decades, empirical scientists have recognized the importance of heteroscedasticity (or variability, more generally) using

indices of level-1 variability as both outcomes and predictors. With that said, modeling heteroscedasticity as an outcome has generally not involved the use of repeated-measures data; instead, modeling heteroscedasticity as an outcome is often used in educational research settings with cross-sectional data/designs (see Raudenbush & Bryk, 1987 or Konstantopoulos, 2008, who both use a two-stage mixed-effects modeling approach). As a result, only a discussion of modeling heteroscedasticity as a predictor is provided below.

**Heteroscedasticity as a predictor.** Although cross-sectional designs can be modeled using residual (within-group) variability as a predictor (see Leckie, 2014 or Raudenbush & Bryk, 1987), using level-1 variability as a predictor of level-2 differences are more prominent in the analysis of repeated-measures data. They are frequently used in cognitive aging and health-related outcomes research using data obtained from measurement burst designs (Nesselroade & McCollam, 2000) or ecological momentary assessments (Stone & Shiffman, 2002) that yield many observations per individual. In these data, an index of variability termed *intra-individual variability* (IIV) is often calculated across repeated occasions using only an individual's own responses (Nesselroade & Ram, 2004). The IIV variable is then used as a level-2, between-individual predictor to evaluate the effect of being a more variable (or, less consistent) individual (Nesselroade & Ram, 2004; Ram, Rabbitt, Stollery, & Nesselroade, 2005). Note that the calculation of IIV at a minimum is a two-stage process in which variability estimates used in subsequent models were initially detrended to remove time-related effects (Rast & Zimprich, 2011).

One common metric of IIV is the *intra-individual standard deviation* ($ISD_i$), calculated as the standard deviation for each individual's own responses over the repeated occasions (Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000). Estabrook, Grimm, and Bowles (2012) hypothesize that $ISD_i$ is the most used IIV metric because it is relatively unaffected by extreme observations, ignores the order of responses, is simple to calculate, and is readily interpretable. As with any standard deviation, $ISD_i$ describes how much an individual varies with respect to their average level of performance, but is not adjusted for overall performance, such as floor or ceiling effects (Tractenberg & Pietrzak, 2011). As a result, a second metric termed the coefficient of variation ($CV_i$), which adjusts for overall performance, may be calculated as $CV_i = \frac{ISD_i}{\bar{x}_i}$, where $\bar{x}_i$ represents the individual's mean across the repeated occasions (Haldane, 1955; Hultsch et al., 2000; Tractenberg & Pietrzak, 2011).

IIV has been used often in cognitive aging and health research. For example, Hultsch et al. (2000) used a reaction time task to obtain several IIV metrics and found that greater IIV predicted lower performance on two memory tasks. With that said, IIV is not without limitations. First, predictors in single-level linear and linear mixed-effects models are assumed measured without error (i.e., perfectly reliable). This assumption will rarely be satisfied given variability generally represents error and an estimate of standard deviation may contain a significant amount of sampling error given it is typically based on only a few occasions per person. Second, IIV has been generally conceptualized only as short-term, transient fluctuation (Ram et al., 2005). Although some individuals may be more variable compared to others, IIV as a predictor implies that the repeated occasions used to calculate IIV are homoscedastic within an individual; however, if

heteroscedasticity is present and variability in response changes over occasions, the IIV metric becomes of questionable utility.

The limitations of traditional two-stage approaches using variability as a predictor highlight the need to employ a more parsimonious and powerful model to predict heteroscedasticity by extending the linear mixed-effects model to include heterogeneous scale model across individuals (Hoffman, 2007). These models include the *heterogeneous variance model*, which allows residual and random effect variances and covariances to differ across individuals based on specific scale-model predictor fixed effects, as well as the *mixed-effects location-scale model*, which extends the heterogeneous variance model to include additional random effects for the effect of lower-level predictors of residual variance and covariance.

**The Heterogeneous Variance Model**

Although similar in nature to the heterogeneous variance model described above in the context of the single-level linear model (see Bryk & Raudenbush, 1988), the heterogeneous variance model in the linear mixed-effects model framework serves two purposes. First, the model is used to ensure that more correct covariance estimates are used when calculating standard errors for the location-model fixed effects, thereby reducing inferential bias (Littell et al., 1996). Second, the model is used to test hypotheses related to differential variability by identifying the direction and magnitude by which predictor variables of residual and random effect variances and covariances produce heteroscedasticity. The heterogeneous variance model in a mixed model framework estimates the location model and heterogeneous scale model across levels of analysis, with location-model variance components based on a log-linear combination of

predictor variables (Hedeker, Mermelstein, & Demitras, 2008). Note that the natural log link function is used on the model scale to ensure predicted variances remain positive on the data scale (Harvey, 1976) and the inverse hyperbolic tangent link function is used to ensure random effect correlations stay bounded within the interval –1 to +1 (i.e., $\mathbf{G}_i$ remains positive semi-definite; Leckie et al., 2014).

***Predicting location-model random effect variances and covariances.*** Scale-model heterogeneity can be predicted for any location-model random effect variance or correlation in $\mathbf{G}_i$, shown in (1.9) through (1.11) using slightly different notation compared to what was used above (described shortly). Note that because these random effect variances and correlations are now conditional on a set of predictor values unique to individual $i$, $\mathbf{G}_i$ now *requires* the subscript $i$.

$$\log\left(\sigma^2_{u_{r,i}}\right) = \mathbf{A}_i^{u_r}\boldsymbol{\alpha}^{u_r} \tag{1.9}$$

$$\tanh^{-1}\left(\rho_{u_{r,i};u_{r',i}}\right) = \mathbf{A}_i^{u_r;u_{r'}}\boldsymbol{\alpha}^{u_r;u_{r'}} \tag{1.10}$$

where,

$$\rho_{u_{r,i};u_{r',i}} = \frac{\sigma_{u_{r,i};u_{r',i}}}{\sqrt{\sigma^2_{u_{r,i}}\sigma^2_{u_{r',i}}}}. \tag{1.11}$$

The new notation reflects the specific sets of predictors or effects for specific location-model random effect variances $u_r$ or correlations $u_r; u_{r'}$ as notated in the superscript, where the superscript's subscript $r$ indexes the specific location-model random effect. The superscript, therefore, is an index that denotes the element of $\mathbf{G}_i$ to which each design matrix, $\mathbf{A}_i^{u_r}$ or $\mathbf{A}_i^{u_r;u_{r'}}$, or parameter vector, $\boldsymbol{\alpha}^{u_r}$ or $\boldsymbol{\alpha}^{u_r;u_{r'}}$, refers. That is, the superscript indicates which $r$ location-model random effect is being estimated and that these random effects can be estimated using different sets of predictor variables. For

example, $\mathbf{A}_i^{u_0}$ indicates the specific set of predictor variables for the location-model random intercept variance, $\mathbf{A}_i^{u_1}$ indicates the set of predictors for an additional location-model random effect variance, and $\mathbf{A}_i^{u_0;u_1}$ indicates the set of predictors for their correlation.

Here, $\sigma^2_{u_{r,i}}$ is the (log of the) estimated level-2 random effect variance $r$ for individual $i$, $\rho_{u_{r,i};u_{r',i}}$ is the correlation between random effect $u_r$ and $u_{r'}$ for individual $i$, and $\sigma_{u_{r,i};u_{r',i}}$ is the covariance between random effect $u_r$ and $u_{r'}$ for individual $i$. Note that the correlation is used to allow the covariance to rescale as necessary based on the location-model random effect variances, $\sigma^2_{u_{r,i}}$ and $\sigma^2_{u_{r',i}}$, in which the inverse hyperbolic tangent link is used to ensure $\rho_{u_{r,i};u_{r',i}}$ remains bounded between –1 and +1 (i.e., $\mathbf{G}_i$ remains positive semi-definite). Note that when $\mathbf{G}_i$ is of order three or larger, appropriate link functions (e.g., log, $\tanh^{-1}$) are necessary but not sufficient to ensure that $\mathbf{G}_i$ remains positive definite (Leckie et al., 2014); although the Metropolis-Hastings algorithm developed for this dissertation uses methods that ensure $\mathbf{G}_i$ remains positive definite (see Barnard, McCulloch, & Meng, 2000).

In addition, $\mathbf{A}_i^{u_r}$ is a 1 x $a^{u_r}$ row vector and $\mathbf{A}_i^{u_r;u_{r'}}$ is a 1 x $a^{u_r;u_{r'}}$ row vector, containing $a^{u_r}$ and $a^{u_r;u_{r'}}$ scale-model predictor variables for individual $i$. Note that both $\mathbf{A}_i^{u_r}$ and $\mathbf{A}_i^{u_r;u_{r'}}$ are row vectors because they can only contain level-2, individual-level predictors (i.e., a constant value across all occasions; known as *time-invariant* predictors, as described in chapter 2), and for the models in this dissertation $\mathbf{A}_i^{u_r}$ and $\mathbf{A}_i^{u_r;u_{r'}}$ will have the first element be a 1 for all individuals to represent the intercept of the location-model random effect variance or correlation.

Finally, $\boldsymbol{\alpha}^{u_r}$ is an $a^{u_r}$ x 1 column vector and $\boldsymbol{\alpha}^{u_r;u_{r'}}$ is an $a^{u_r;u_{r'}}$ x 1 column vector of $a^{u_r}$ or $a^{u_r;u_{r'}}$ scale-model fixed effects for a specific location-model random effect variance $u_r$ or correlation $u_r; u_{r'}$ as notated in the superscript. Because fixed effects apply to all individuals equally, no subscript $i$ was included.

***Predicting residual variances and covariances.*** Similarly, scale-model heterogeneity can be predicted for all residual variances and correlations in $\mathbf{R}_i$ as shown in (1.12) through (1.14).

$$\log\left(\sigma^2_{e_{t,i}}\right) = \mathbf{T}_i^{e}\boldsymbol{\tau}^{e} \tag{1.12}$$

$$\tanh^{-1}\left(\rho_{e_{t,i};e_{t',i}}\right) = \mathbf{T}_i^{e_t;e_{t'}}\boldsymbol{\tau}^{e_t;e_{t'}} \tag{1.13}$$

where,

$$\rho_{e_{t,i};e_{t',i}} = \frac{\sigma_{e_{t,i};e_{t',i}}}{\sqrt{\sigma^2_{e_{t,i}}\sigma^2_{e_{t',i}}}}. \tag{1.14}$$

Here, $\sigma^2_{e_{t,i}}$ is the (log of the) estimated level-1 residual variance at occasion $t$ for individual $i$, $\rho_{e_{t,i};e_{t',i}}$ is the (inverse hyperbolic tangent of the) correlation between residual values for occasions $t$ and $t'$ for individual $i$, and $\sigma_{e_{t,i};e_{t',i}}$ is the covariance between residuals for occasions $t$ and $t'$ for individual $i$. The matrix $\mathbf{T}_i$ is for the residual variance, $\mathbf{T}_i^{e}$, or residual correlation, $\mathbf{T}_i^{e_t;e_{t'}}$, as notated in the superscript. Note that both $\mathbf{T}_i^{e}$ and $\mathbf{T}_i^{e_t;e_{t'}}$ are matrices because they can contain both level-1, occasion-level and/or level-2, individual-level predictors, and in the models described here will include a first column of ones to represent the intercept of location-model residual variances or correlations. Therefore, $\mathbf{T}_i^{e}$ is an $n_i$ x $c^{e}$ design matrix and $\mathbf{T}_i^{e_t;e_{t'}}$ is an $n_i$ x $c^{e_t;e_{t'}}$ design matrix, containing $c^{e}$ or $c^{e_t;e_{t'}}$ scale-model predictor variables for the $n$ occasions for

individual $i$. Finally, $\boldsymbol{\tau}^{e}$ is a $c^{e}$ x 1 column vector and $\boldsymbol{\tau}^{e_t;e_{t'}}$ is a $c^{e_t;e_{t'}}$ x 1 column vector of $c^{e}$ or $c^{e_t;e_{t'}}$ scale-model fixed effects for the (log of the) residual variance or residual correlation between occasion $t$ and $t'$, respectively, as notated in the superscript. Similar to above, no subscript $i$ is indicated.

***Interpretation of fixed effects.*** Scale-model fixed effects can be interpreted similarly to the location-model fixed effects interpreted above. For example, $\sigma^2_{u_{r,i}}$ equals the intercept variance, $\exp\left(\alpha_0^{u_r}\right)$, when all predictors $\mathbf{A}_i^{u_r} = 0$ and $\sigma^2_{u_{r,i}}$ increases for $\boldsymbol{\alpha}^{u_r} > 0$ (or decreases for $\boldsymbol{\alpha}^{u_r} < 0$) with one-unit increases in each element of $\mathbf{A}_i^{u_r}$. Location-model random effect correlations, as well as residual variances and residual correlations, follow an identical pattern of interpretation. Finally, regardless of superscript, when $\boldsymbol{\alpha} = \boldsymbol{\tau} = 0$, *excluding the intercept of the specific variance or correlation* (i.e., the first column of ones in all $\mathbf{A}_i$ and $\mathbf{T}_i$), the heterogeneous variance model is reduced to the linear mixed-effects model shown in (1.4).

Assuming a frequentist framework, the statistical significance of scale-model fixed effects can be evaluated by Wald $p$-values or model comparison using the likelihood ratio test (aka, deviance difference test) or information criteria (e.g., Akaike or Bayesian information criterion; see Littell et al., 1996). In a Bayesian framework, statistical significance is indicated by examining Bayesian confidence intervals or model comparisons based on the deviance information criterion (both detailed in chapter 3).

Finally, specifying heterogeneous $\mathbf{G}_i$ and/or $\mathbf{R}_i$ reduces standard error bias of location-model fixed effects by ensuring individuals are given the correct variance components in $\mathbf{G}_i$ and $\mathbf{R}_i$, which are then re-aggregated into $\mathbf{V}_i$ as shown in (1.15).

$$\mathbf{V}_i = \mathbf{Z}\mathbf{G}_i\mathbf{Z}^T + \mathbf{R}_i \quad (1.15)$$

***Brief example of the heterogeneous variance model.*** Because the notation for the scale model in (1.9) though (1.14) is (necessarily) complex, effects estimated by heterogeneous variance model can be solidified using a brief (albeit, overly simplistic) example; a much more detailed example is provided in chapter 2.

Consider a sample of 20 individuals where the outcome, $Y_{t,i}$, was observed at three occasions with no missing data, and a location model including one level-1, occasion-level predictor, $X_{t,i}$, one level-2, individual-level predictor, $\overline{X}_i$, and two location-model random effects, $u_{0,i}$ and $u_{1,i}$, representing the location-model random intercept and the location-model random slope for $X_{t,i}$, respectively, as well as their correlation, $\rho_{u_{0,i};u_{1,i}}$. Thus, $N$ = 20 (total number of individuals), $n_i$ = 3 (number of repeated occasions within individual $i$), $p$ = 3 (number of location-model fixed effects including the fixed intercept), and $q$ = 2 (number of location-model random effects). $\overline{X}_i$ is included as a predictor of both location-model random effect variances (i.e., in both $\mathbf{A}_i^{u_0}$ and $\mathbf{A}_i^{u_1}$), but not for their correlation (i.e., not in $\mathbf{A}_i^{u_0;u_1}$). Thus, $\mathbf{G}_i$ is modeled as unstructured and heterogeneous conditional on the values of $\overline{X}_i$. Further, the residual variance is predicted by $\overline{X}_i$ and $X_{t,i}$, which are both included in $\mathbf{T}_i^e$, with all residual correlations assumed to be zero. Thus, $\mathbf{R}_i$ are modeled as independent, but heterogeneous (given $X_{t,i}$) across repeated occasions, $\mathbf{R}_i = \sigma^2_{e_{t,i}}\mathbf{I}_{n_i}$.

This heterogeneous variance model has one linear predictor for the location model shown in (1.16), four linear predictors for the scale model shown in (1.17) through (1.20),

with independent residual correlations shown in (1.21). Note that the location-model uses multi-level, scalar notation in which $\gamma$ indicates a fixed effect (see chapter 2).

$$Y_{ti} = (\gamma_{00} + u_{0i}) + \gamma_{01}(\overline{X}_i) + (\gamma_{10} + u_{1i})(X_{t,i}) + e_{ti} \tag{1.16}$$

$$\log(\sigma^2_{u_{0,i}}) = \alpha_0^{u_0} + \alpha_1^{u_0}(\overline{X}_i) \tag{1.17}$$

$$\log(\sigma^2_{u_{1,i}}) = \alpha_0^{u_1} + \alpha_1^{u_1}(\overline{X}_i) \tag{1.18}$$

$$\tanh^{-1}(\rho_{u_{0,i};u_{1,i}}) = \alpha_0^{u_0;u_1} \tag{1.19}$$

$$\log(\sigma^2_{e_{t,i}}) = \tau_0^e + \tau_1^e(X_{t,i}) + \tau_2^e(\overline{X}_i) \tag{1.20}$$

$$\rho_{e_{t,i};e_{t',i}} = 0 \tag{1.21}$$

The estimated variances and correlation from the scale-model equations can be mapped directly onto $\mathbf{G}_i$ and $\mathbf{R}_i$, in which the correlation has been converted to a covariance in the off diagonal of $\mathbf{G}_i$, as shown in (1.22) through (1.26).

$$\mathbf{G}_i = \begin{bmatrix} \sigma^2_{u_{0,i}} & \sigma_{u_{0,i};u_{1,i}} \\ \sigma_{u_{1,i};u_{0,i}} & \sigma^2_{u_{1,i}} \end{bmatrix}, \tag{1.22}$$

where

$$\sigma^2_{u_{0,i}} = \exp\left(\alpha_0^{u_0} + \alpha_1^{u_0}\left(\overline{X}_i\right)\right) \tag{1.23}$$

$$\sigma_{u_{1,i};u_{0,i}} = \sigma_{u_{0,i};u_{1,i}} = \tanh(\alpha_0^{u_0;u_1})\sqrt{\exp\left(\alpha_0^{u_0} + \alpha_1^{u_0}\left(\overline{X}_i\right)\right)\exp\left(\alpha_0^{u_1} + \alpha_1^{u_1}\left(\overline{X}_i\right)\right)} \tag{1.24}$$

$$\sigma^2_{u_{1,i}} = \exp\left(\alpha_0^{u_1} + \alpha_1^{u_1}\left(\overline{X}_i\right)\right) \tag{1.25}$$

and

$$\mathbf{R}_i = \exp\left(\tau_0^e + \tau_1^e(X_{t,i}) + \tau_2^e(\overline{X}_i)\right)\mathbf{I}_{n_i}. \tag{1.26}$$

***Examples from the literature.*** Examples of heterogeneous variance models in published literature are sparse, suggesting that many empirical scientists have yet to

consider the value of hypotheses based on differential variability in addition to mean structures. However, their use has increased in recent years. For example, Hedeker and Mermelstein (2007) and Hedeker, Mermelstein, Berbaum, and Campbell (2009) examined variability in positive and negative affect and found less heterogeneity in both level-2 random intercept variance and level-1 residual variance as an adolescent's smoking experience increased. Almeida, Piazza, and Stawski (2009) found that level-1 residual variance of negative affect increased with baseline age for men, was stable across age for women, and that men had significantly greater residual variability compared to women as baseline age increased. In addition, Diehl and Hay (2010) determined that level-1 residual variance of negative affect was significantly lower in older individuals and in individuals who were coherent with respect to their own perceived self-concept across different social roles and situations. Finally, Schneider, Junghaenel, Keefe, Schwartz, Stone, and Broderick (2012) used a heterogeneous variance model to predict variability in level-1 residuals and level-2 random intercept variance in outcomes that included pain intensity, fatigue, happiness, and frustration using a sample of rheumatology patients. Results indicated that higher levels of depression predicted greater level-1 residual variability of pain, happiness, and frustration.

***Limitations of the heterogeneous variance model.*** Although the heterogeneous variance model can include predictors to address heteroscedasticity at any level of analysis, the model is not without limitations. First, the ability to predict heteroscedasticity is contingent on the variables collected and evaluated, as is the case for all statistical models. Therefore, inferences of location-model fixed effects may still be biased if the location model or heterogeneous scale model is misspecified by omitting

important predictors and interactions (the second simulation study in chapter 4 begins to study the potential for bias). Second, the heterogeneous variance model requires that all scale-model predictors be represented by fixed effects, such that the effect of scale-model predictors apply equally to all individuals. Therefore, level-1 residual variance is not allowed to vary across individuals beyond the effect of the predictors. This limitation can be addressed by including random effects in the scale model using a mixed-effects location-scale model, discussed next.

**The Mixed-Effects Location-Scale Model**

The heterogeneous variance model has the flexibility to include location-model random effects to model heterogeneity in mean levels across higher-level units, as well as scale-model predictor variables to estimate the average (i.e., fixed) effect of a predictor on heterogeneity of level-1 and/or level-2 variances. It is important to note that neither the location-model random effects nor the scale-model fixed effects explicitly account for individual differences in *variability* in the outcome. However, just as the location-model estimates can vary across higher-level units, so can the *variability* around an individual's mean response (Cleveland, Denby, & Liu, 2002). Therefore, the heterogeneous variance model above can be extended to include scale-model random effects using a *mixed-effects location-scale model* to estimate individual differences in both mean levels and variability of a given outcome in a single model (Hedeker et al., 2008). Note this model has also been termed a *double hierarchical generalized linear model* (DHGLM; Lee & Nelder, 2006; Lee & Noh, 2012), which for the models discussed in this dissertation would use the identity link function for the location model.

For repeated-measures data that have two levels of sampling units, scale-model random effects can be included for any *level-1 predictor of the residual variance* in $\mathbf{R}_i$, which is conceptually similar to the inclusion of location-model random effects. That is, random effects cannot be specified for level-2 predictors of residual variance, or for any predictor of location-model random effect variances in $\mathbf{G}_i$, because in repeated-measures data there are no higher-level units across which individuals could vary randomly. Therefore, only the scale model for the (log of the) residual variance in (1.12) is extended to include random effects, as shown in (1.27) using the same superscript notation as described for the heterogeneous variance model.

$$\log\left(\sigma^2_{e_{t,i}}\right) = \mathbf{T}_i^e\boldsymbol{\tau}^e + \mathbf{W}_i^e\boldsymbol{\omega}_i^e \tag{1.27}$$

Here, both $\mathbf{T}_i^e$ and $\boldsymbol{\tau}^e$ have identical dimensions and interpretations as described above for (1.12). $\mathbf{W}_i^e$ is an $n_i$ x $w^e$ matrix containing $w^e$ scale-model predictor variables from $\mathbf{T}_i^e$ that were specified to have random effects in the scale model for the residual variance, which for the models discussed here includes a vector of ones as the first column in the matrix to represent the intercept. In repeated-measures data, only the effects of level-1 predictors can deviate randomly across individuals, in which the dimensions of $\mathbf{W}_i^e$ indicate level-1 predictor values for the $n$ occasions for individual $i$. In addition, $\boldsymbol{\omega}_i^e$ is a $w^e$ x 1 column vector of deviations for individual $i$ from each of the $w^e$ fixed effects specified to be random in the scale model for the residual variance in $\mathbf{W}_i^e$. Finally, when $\boldsymbol{\omega}_i^e = 0$, the mixed-effects location-scale model is reduced to the heterogeneous variance model in (1.12).

The statistical assumptions of scale-model random effects are similar to location-model random effects discussed above. That is, because $\boldsymbol{\omega}_i^e$ represents between-

individual differences, the variance of any scale-model random effects, $\sigma^2_{\omega^e_{s,i}}$, is included in $\mathbf{G}_i$ alongside the location-model random effect variances, $\sigma^2_{u_{r,i}}$. Note that the subscript $s$ indexes the specific scale-model random effect in $\mathbf{W}^e_i$ (e.g., $\sigma^2_{\omega^e_{0,i}}$ indicates the random intercept of the residual variance for individual $i$). In addition, the correlation between location- and scale-model random effects for individual $i$, $\rho_{u_{r,i};\omega^e_{s,i}}$, could also be estimated and predicted in $\mathbf{G}_i$ (Hedeker et al., 2008) as shown below in (1.28) using a continuation of the brief example that was provided above.

$$\mathbf{G}_i = \begin{pmatrix} \sigma^2_{u_{0,i}} & \rho_{u_{0,i};u_{1,i}} & \rho_{u_{0,i};\omega^e_{0,i}} \\ \rho_{u_{1,i};u_{0,i}} & \sigma^2_{u_{1,i}} & \rho_{u_{1,i};\omega^e_{0,i}} \\ \rho_{\omega^e_{0,i};u_{0,i}} & \rho_{\omega^e_{0,i};u_{1,i}} & \sigma^2_{\omega^e_{0,i}} \end{pmatrix} \tag{1.28}$$

Note that scale-model random effect variances and correlations could also be predicted to be heterogeneous between individuals using methods identical to those described above for the location-model random effect variances and correlations.

When location- and scale-model random effects are included in the mixed-effects location-scale model, location-model random effects $\mathbf{u}_i$ and scale-model random effects $\boldsymbol{\omega}^e_i$ are aggregated into $\mathbf{m}_i$ which is an $m_{\mathbf{u}_i+\boldsymbol{\omega}^e_i}$ x 1 column vector of all location- and scale-model random effects, which is assumed multivariate normally distributed with a mean of zero and estimated covariance matrix $\mathbf{G}_i$, $\mathbf{m}_i \sim \mathbf{N}_m(\mathbf{0}, \mathbf{G}_i)$. Finally, the statistical significance of scale-model random effects is evaluated by model comparisons using the likelihood ratio test or information criteria (e.g., AIC, BIC) in the frequentist framework, or by deviance information criterion in a Bayesian framework.

***Examples from the literature.*** Scale-model random effects have been considered in the statistical literature using both frequentist and Bayesian frameworks, but have been

used far less frequently in practice by empirical scientists (Cleveland et al., 2002; Hedeker et al., 2008). This is troubling given that neglecting scale-model random effects potentially results in a misspecified model that has been shown to result in downwardly-biased residual variance fixed effect standard errors, especially for level-2 predictors, an issue similar in nature to ignoring location-model random effects (Leckie, 2014; Leckie et al., 2014). Therefore, estimating a heterogeneous variance model without initially evaluating for scale-model random effects via mixed-effects location-scale model could result in spurious inferences, which could lead to Type I errors and inappropriate delegation of resources to identify the source of observed heterogeneity.

With that said, the mixed-effects location-scale model is being used more frequently due to recent instructional papers (e.g., Hedeker et al., 2008; Hedeker & Mermelstein, 2012; Leckie et al., 2014; Lee & Noh, 2012; Rast, Hofer, & Sparks, 2012) and software tutorials (Hedeker & Nordgren, 2013; Leckie, 2014; Li, Bruyneel, & Lesaffre, 2014; Rast et al., 2012). Examples of several empirical studies are presented next.

Hedeker et al. (2008) modeled both level-2 random intercept variance and level-1 residual variance for positive and negative affect using data from 461 students who each averaged approximately 30 observations. Results indicated a statistically significant scale-model random intercept variance, and statistically significant correlation between location- and scale-model random intercepts. Hedeker and Mermelstein (2012) extended the location model from Hedeker et al. (2008) to include random linear time effects, but continued to model $\mathbf{G}_i$ as constant between individuals (i.e., $\mathbf{G}_i = \mathbf{G}$); results were similar

for the scale-model random intercept and correlation between location- and scale-model random intercepts.

In addition, Rast and Zimprich (2011) estimated residual variance in reaction time, holding $\mathbf{G}_i$ constant, using data from 335 individuals who averaged 20 reaction time trials, and found significant individual differences in residual variability and significant location- and scale-model random intercept correlation where higher average reaction times was associated with greater average variability. Rast et al. (2012) estimated residual variance in positive and negative affect, holding $\mathbf{G}_i$ constant, using seven consecutive measurement days across 178 individuals, and found significant individual differences in variability for both outcomes. Further, individuals who were less variable in positive or negative affect had less variable responses to daily stressors such as having an argument with others.

Li and Noh (2014) evaluated eye-tracking data from 43 non-schizophrenic and 43 schizophrenic individuals, holding $\mathbf{G}_i$ constant. Results indicated that schizophrenics had significantly greater residual variability in eye tracking compared to non-schizophrenic individuals.

Finally, the mixed-effects location-scale model has also been applied to cross-sectional data in education settings, as Leckie et al. (2014) estimated within-school residual variance in mathematics achievement using school sector (i.e., public vs. private) student SES as predictors and constant $\mathbf{G}_i$ between schools. They also included scale-model random effects for the residual variance intercept and student SES, finding significant between-school variability in both effects.

## Chapter Summary

The purpose of this chapter was to provide the theoretical framework of the mixed-effects location-scale model applied to conditionally normally distributed repeated-measures data. The chapter began by presenting the single-level linear model, its assumptions, and a discussion of heterogeneous variance alongside the common statistical methods used frequently by empirical scientists to account for or correct heterogeneity. This was followed by a description of the linear mixed-effects model that detailed how this model has been used traditionally to account for and predict heterogeneity of variance between-individuals in $\mathbf{G}_i$ and $\mathbf{R}_i$. Finally, the mixed-effects location-scale model was introduced, along with a background of how the model has been used in practice. Using the theoretical framework detailed in this chapter, chapter 2 will now provide a complete model-building example of the mixed-effects location-scale model.

# CHAPTER 2: A MODEL-BUILDING EXAMPLE OF THE MIXED-EFFECTS LOCATION-SCALE MODEL FOR CONDITIONALLY NORMALLY DISTRIBUTED REPEATED-MEASURES DATA

Chapter 1 detailed the theoretical framework of the mixed-effects location-scale model for a conditionally normally distributed repeated-measures data beginning with the single-level linear model and continuing through the mixed-effects location-scale model. In this chapter, an explicit model-building example of the mixed-effects location-scale model is presented using repeated-measures data that is similar (but not identical) to the empirical data analysis presented in chapter 5. The chapter example highlights the application of the mixed-effects location-scale model and presents model equations in multi-level, scalar form (i.e., level-1 occasions nested within level-2 individuals), alongside details regarding model comparisons as well as the inclusion of random effects, predictors at both levels of analysis, and predictor interactions.

The example below follows the model-building procedures and recommendations typically encountered in the literature (see Leckie et al., 2014; Hedeker et al., 2008), in which the location model is assumed properly specified *before* including scale-model random effects and predictors. It is important to note that the order in which these effects are included in the model has not been validated empirically and there is no consensus as to their order of entry into a model (see Bryk & Raudenbush, 1988, who recommend properly specifying the scale model before estimating effects in the location model); the second simulation study presented in chapter 4 begins to address this issue. With that said, the example will begin by modeling the location-model random intercept, followed by location-model fixed effects, and then location-model random effects for level-1

predictors. This is followed by a similar procedure for the scale model, whereby the scale-model random intercept is estimated initially, followed by fixed effects for the variances and correlations in $\mathbf{R}_i$, random effects for level-1 predictors within $\mathbf{R}_i$, and finally, level-2 predictors as fixed effects for variances and correlations in $\mathbf{G}_i$.

**Description of the Chapter Example**

For the example in this chapter, consider a repeated-measures design that measured physical activity in a sample of 100 older adults, of which 50 have probable mid Alzheimer's disease (AD) and 50 do not. Physical activity was measured using an accelerometer and quantified by vector magnitude, a composite metric of triaxial movement in the medio-lateral, antero-posterior, and vertical planes (i.e., front-to-back, side-to-side, and rotational movements, respectively; or, for the Postural Restoration savvy, sagittal, frontal, and transverse planes, respectively), which can range from 0 to a maximum dictated by the length of the occasion (known as an epoch). Each individual wore the accelerometer during waking hours over a 24-hour period; real-time accelerometer data was binned into 60-minute epochs. Thus, the dependent variable is the observed vector magnitude summed every 60 minutes over the course of 1 day, $VM_{t,i}$, where $t$ denotes the specific occasion (epoch) for individual $i$. Note that because the number of waking hours can vary between individuals, so can the number of total observations per individual (e.g., 10 awake hours = 10 observations; 16 awake hours = 16 observations). The primary independent variable of interest is AD status, $AD_i$, which is a binary level-2 predictor for individual $i$, where 0 = no AD and 1 = AD. Two covariates are also modeled. The first is the individual's years of formal education, $Ed_i$, which is a continuous, level-2, individual-level variable collected at baseline. $Ed_i$ was centered at a

value of 12 (i.e., a high school graduate; $Ed_i - 12$, where 0 indicates 12 years of formal education) to ensure the intercept value remains meaningful and to allow for meaningful interpretation of potential interaction effects. Note that a value of 12 was chosen arbitrarily for this example; in practice the centering value should be any value believed to be meaningful. The second covariate indicates whether the individual was alone or with others at any time during the epoch, $Alone_{t,i}$, which is a binary, level-1, occasion-level variable, where 0 = not alone and 1 = alone.

This example maps closely, but not identically, onto the empirical data analysis presented in chapter 5 that models vector magnitude as the outcome with years of formal education as one of several individual-level predictors. Note that the time-varying predictor indicating whether an individual was alone or not at a given occasion is a hypothetical covariate that is not included in chapter 5. It was included here to illustrate the flexibility of the mixed-effects location-scale model.

**The (Unconditional) Single-Level Linear Model**

It was stated in chapter 1 that repeated-measures data most likely have a non-zero correlation between occasions within the same individual, the extent of which can be quantified directly using the linear mixed-effects model. However, this procedure is most accessible when considered as an extension of the (unconditional) single-level linear model shown in (2.1). Note that this is an *unconditional* model because the location model only includes the intercept.

$$VM_{t,i} = \beta_0 + e_{t,i} \tag{2.1}$$

The subscript $t, i$ is required in this single-level model to identify the correct vector magnitude for an individual at a given occasion; thus, $VM_{t,i}$ is the vector magnitude at

specific to occasion $t$ for individual $i$, $\beta_0$ is the location-model fixed intercept representing average vector magnitude across all observations, and $e_{t,i}$ is the residual value at occasion $t$ for individual $i$ representing the difference between each vector magnitude at occasion $t$ for individual $i$ and the fixed intercept, $e_{t,i} = VM_{t,i} - \beta_0$.

The unconditional single-level linear model in (2.1) assumes uncorrelated residual values and constant variance across individuals and observations. However, repeated-measures data tends to violate the assumption of independence as repeated occasions nested within an individual will usually have one or more sources non-trivial correlation, each of which must be accounted for using a linear mixed-effects model.

**The Linear Mixed-Effects Model**

As described in chapter 1, the (conditional) linear mixed-effects model explicitly accounts for residual correlation due to nesting by partitioning residual variance into between- and within-individual variance components via location-model random effects. After including location-model random effects in the model, residuals are once again assumed independent *between* individuals. To account initially for the proportion of total variance in vector magnitude that is between individuals, a new location-model variance component is estimated, termed *random intercept variance*. The *unconditional random intercept model* is shown in (2.2) using multi-level, scalar notation because the model now includes distinct within- and between-individual levels of analysis; a simple visual depiction of this model presented for two individuals is also provided in Figure 2.1.

$$\begin{aligned} &\text{Level 1:} \quad VM_{t,i} = \beta_{0,i} + e_{t,i} \\ &\text{Level 2:} \quad \beta_{0,i} = \gamma_{00} + u_{0,i} \\ &\text{Combined:} \quad VM_{t,i} = (\gamma_{00} + u_{0,i}) + e_{t,i} \end{aligned} \tag{2.2}$$

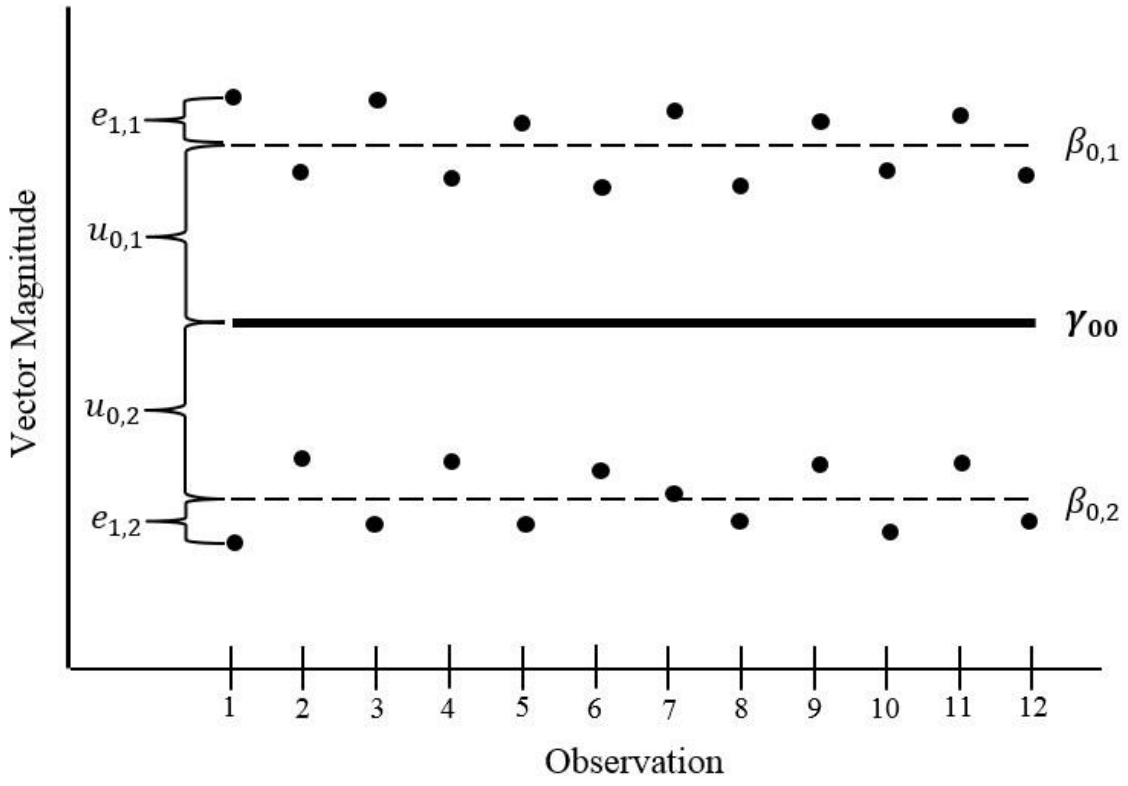


*Figure 2.1*. The unconditional random intercept model

Here, although the level-1 model appears similar to (2.1), this model now describes *within-individual* variation in vector magnitude as a function of the intercept specific to individual $i$, $\beta_{0,i}$ (dashed lines in Figure 2.1), and a residual deviation from that intercept specifically at occasion $t$ for individual $i$, $e_{t,i}$. Note that $\beta_{0,i}$ at level 1 is simply a placeholder for the effects included in the level-2, *between-individual* model, and is a function of the location-model fixed intercept, $\gamma_{00}$, representing the grand mean of the means across all individuals (the solid line in Figure 2.1), and the location-model random intercept, $u_{0,i}$, represents the constant *deviation* from the fixed intercept for individual $i$. These level-2 effects are substituted for $\beta_{0,i}$ at level 1 to create the combined equation.

Note that because $\gamma_{00}$ is a fixed effect, it applies to every individual in the sample (thus, there is no subscript $i$) and serves as the reference point for quantifying the

location-model random effect $u_{0,i}$. This can be observed in Figure 2.1 as the fixed intercept $\gamma_{00}$ underestimates the actual observed intercept for individual 1 (i.e., $u_{0,1} > 0$) and overestimates the observed intercept for individual 2 (i.e., $u_{0,2} < 0$).

Considering the multi-level, scalar format of (2.2) and the matrix representation of the linear mixed-effects model presented in (1.4), $\mathbf{Y}_i$ is an $n_i$ x 1 column vector holding the $VM_{t,i}$ outcomes, where $n_i$ is the number of $n$ repeated occasions within individual $i$. Further, because only the location-model fixed intercept is currently being modeled, $\mathbf{X}_i$ is an $n_i$ x 1 column vector of ones for the $n$ repeated occasions within individual $i$, and $\boldsymbol{B}$ is a scalar holding the estimate of the fixed intercept, $\gamma_{00}$. In addition, given that the intercept is also the only location-model random effect in this model, $\mathbf{Z}_i$ is an $n_i$ x 1 column vector of ones for the $n$ repeated occasions within individual $i$, and $\mathbf{u}_i$ is a scalar holding the deviation from the location-model fixed intercept specific to individual $i$, $u_{0,i}$. Finally, $\mathbf{e}_i$ is an $n_i$ x 1 column vector holding the residual values, $e_{t,i}$, across the $n$ repeated occasions within individual $i$.

In addition, as described in detail in chapter 1, $\mathbf{G}_i$ holds the random effect variances for each individual $i$ at level 2. Because the random intercept for individual $i$, $u_{0,i}$, is the only location-model random effect in this model and because it is currently assumed constant between individuals (i.e., $\mathbf{G}_i = \mathbf{G}$), $\mathbf{G}$ is a scalar holding the location-model random intercept variance estimate, $\sigma^2_{u_0}$, applicable to all individuals (i.e., no subscript $i$). Further, $\mathbf{R}_i$ holds the residual variance estimates across the $n$ repeated occasions at level 1 for individual $i$. Assuming residuals to be independent and constant between- and within-individuals, $\mathbf{R}_i = \sigma^2_e \mathbf{I}_{n_i} = \mathbf{R}$, $\mathbf{R}$ is a diagonal matrix holding the constant residual variance estimate, $\sigma^2_e$, that is applicable to all observations across all

individuals (i.e., no subscript $t, i$). Finally, $\mathbf{V}_i$ is the $n_i$ x $n_i$ covariance matrix for the multivariate normal distribution of $\mathbf{Y}_i$ conditional on $\mathbf{X}_i$ and $\mathbf{Z}_i$ for individual $i$, which in this example is also assumed constant between individuals, such that $\mathbf{V}_i = \mathbf{ZGZ}^T + \mathbf{R} = \mathbf{V}$, calculated as shown below (note that thus far in this example, $\mathbf{V}$ is compound symmetric, i.e., equal variances; equal covariances).

$$\mathbf{V} = \mathbf{ZGZ}^T + \mathbf{R} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \left[\sigma_{u_0}^2\right] [1 \quad 1 \quad \cdots \quad 1] + \begin{bmatrix} \sigma_e^2 & 0 & \cdots & 0 \\ 0 & \sigma_e^2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_e^2 \end{bmatrix}$$

$$= \begin{bmatrix} \left(\sigma_{u_0}^2 + \sigma_e^2\right) & \sigma_{u_0}^2 & \cdots & \sigma_{u_0}^2 \\ \sigma_{u_0}^2 & \left(\sigma_{u_0}^2 + \sigma_e^2\right) & \sigma_{u_0}^2 & \vdots \\ \vdots & \sigma_{u_0}^2 & \ddots & \sigma_{u_0}^2 \\ \sigma_{u_0}^2 & \cdots & \sigma_{u_0}^2 & \left(\sigma_{u_0}^2 + \sigma_e^2\right) \end{bmatrix}$$

The primary purpose of estimating the unconditional random intercept model in (2.2) is to quantify the proportion of variability in vector magnitude that is specifically between individuals (i.e., how much does $\mathbf{G}$ contribute to $\mathbf{V}$), obtained by calculating the unconditional intra-class correlation (ICC) as shown in (2.3).

$$\text{ICC} = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_e^2} \tag{2.3}$$

When specifying a conditional model with $\mathbf{G}$ and $\mathbf{R}$, the ICC ranges from 0 to 1, with larger values indicating more between-individual variability (note that ICC can range from –1 to +1 if an $\mathbf{R}$-only, marginal model is estimated including a compound symmetry correlation instead). Further, the fixed intercept and location-model random intercept variance can be used to calculate a 95% random effects confidence interval, $\gamma_{00} \pm$

$1.96\sqrt{\sigma_{u_0}^2}$, which indicates the expected variability in the intercept values for 95% of individuals in the sample (Snijders & Bosker, 2012).

Finally, regardless of whether the frequentist or Bayesian framework was used, model comparisons determine whether the random intercept is needed. Because (2.1) can be obtained by constraining $\sigma_{u_0}^2$ to be zero in (2.2), these models are considered *nested*. Thus, using a frequentist framework, the statistical significance of the location-model random intercept variance (i.e., whether the ICC is significantly different than 0) is determined using the likelihood ratio test or information criteria such as Akaike information criterion (AIC) or Bayesian information criterion (BIC; see Littell et al., 1996), whereas a Bayesian framework requires model comparisons using the deviance information criterion (DIC), where smaller values indicate the more appropriate model (see chapter 3 for full description of DIC).

**The inclusion of predictors.** Adding predictors to the linear mixed-effects model requires a greater amount of care compared to the single-level linear model, as haphazardly adding predictors without considering the levels of analysis can produce incorrect fixed effect estimates and interpretations. Therefore, this section will highlight the complexity of including predictors in the linear mixed-effect model by discussing time-invariant predictors, time-varying predictors, additional location-model random effects, and, finally, a brief comment on interaction effects.

***Time-invariant predictors.*** Predictors measured only once per individual or averaged across an individual's occasions are termed *time-invariant* predictors because the predictor is a constant within an individual across their repeated occasions. Therefore, time-invariant predictors are level-2, individual-level predictors that explain between-

individual differences. Note that the adjective *time-invariant* is used throughout this dissertation to remain consistent with the literature on longitudinal designs (e.g., see Hedeker & Gibbons, 2006; Hoffman, 2014; Singer & Willett, 2003); however, an alternate adjective could be *occasion-invariant* as replacing *time* with *occasion* may reduce confusion regarding models that are explicitly estimating change over time, which is not the case for any of the models described in this dissertation.

Because time-invariant predictors carry between-individual effects, the procedures to include and interpret these predictors are similar to the single-level linear model. The primary difference between the two models, however, is that in the linear mixed-effects model, between-individual variability was partitioned into random intercept variance, $\sigma_{u_0}^2$; therefore, time-invariant predictors explain level-2, random intercept variance.

Continuing with the example, consider two time-invariant predictors—years of formal education, $Ed_i$, and AD status, $AD_i$, both measured for individual $i$ prior to receiving their accelerometer. Because $Ed_i$ is a continuous variable, it was centered at a value of 12 (i.e., a high-school graduate; as described in the example description above). Further, $AD_i$ remained uncentered because 0 was already meaningful as it indicates an individual without AD. Because time-invariant predictors carry between-individual effects, and because they modify fixed intercept values, they are added to the level-2 model for the intercept as shown in (2.4).

$$\begin{aligned} &\text{Level 1:} \quad VM_{t,i} = \beta_{0,i} + e_{t,i} \\ &\text{Level 2:} \quad \beta_{0,i} = \left(\gamma_{00} + u_{0,i}\right) + \gamma_{01}(AD_i) + \gamma_{02}(Ed_i - 12) \\ &\text{Combined:} \quad VM_{t,i} = \left(\gamma_{00} + u_{0,i}\right) + \gamma_{01}(AD_i) + \gamma_{02}(Ed_i - 12) + e_{t,i} \end{aligned} \tag{2.4}$$

Here, all previous effects interpreted for (2.2) remain identical, except that the location-model fixed intercept, $\gamma_{00}$, now represents the average vector magnitude specifically for a high school graduate without AD. The new level-2 location-model fixed effect for AD status, $\gamma_{01}$, represents the average difference in vector magnitude between individuals with and without AD, where $\gamma_{01} > 0$ indicates individuals with AD have greater average vector magnitude and $\gamma_{01} < 0$ indicates individuals without AD have greater average vector magnitude. In addition, the location-model fixed effect for years of education, $\gamma_{02}$, represents the average difference in vector magnitude for every additional year of formal education. As stated above, both of these effects explain level-2 random intercept variance.

Finally, reconsidering the matrix formulation provided in (1.4), both level-2 predictor variables $AD_i$ and centered $Ed_i$ are added to $\mathbf{X}_i$, which is now an $n_i$ x 3 matrix, and $\boldsymbol{B}$ is now a 3 x 1 column vector holding the additional location-model fixed effects for $AD_i$ and $Ed_i$, $\gamma_{01}$ and $\gamma_{02}$, respectively. All other matrices remain identical as described above.

***Time-varying predictors.*** Predictors that were measured concurrently alongside the outcome are termed *time-varying* predictors (or, alternatively, *occasion-varying* predictors). Including time-varying predictors in the linear mixed-effects model adds significant complexity to the analysis because time-varying predictors generally contain both level-2, between-individual and level-1, within-individual variability. That is, although an individual's level of a given predictor may fluctuate across occasions at level 1, some individuals will *average* higher or lower levels of the predictor when compared

to other individuals at level 2. As a result, it is possible that the two levels of the time-varying predictor could have differential effects on the outcome.

In addition, because time-varying predictors typically contain both level-2, between-individual and level-1, within-individual variability, including a time-varying predictor in a linear mixed-effects model without its between-individual counterpart will result in a weighted combination of the level-1 and level-2 effects called a *convergence effect* (see Sliwinski, Hoffman, & Hofer, 2010). The convergence effect assumes the level-1 and level-2 effects of the time-varying predictor are equal, an assumption that difficult to satisfy in practice (Neuhaus & Kalbfleisch, 1998). The extent to which these two effects are not equal will result in disproportional variance explained, and can even result in an increasing variance component if the level-1 and level-2 effects have opposite signs (Hoffman, 2014). Note that a convergence effect will often more closely represent the within-individual effect simply because level-1 occasions typically outnumber higher-level units (Raudenbush & Bryk, 2002).

Whether the effect of a time-varying predictor represents a convergence effect can be determined by calculating an ICC using the time-varying predictor *as an outcome* in an unconditional random intercept model similar to (2.2). Although seemingly odd, the outcome and time-varying predictor are both measured at all occasions, so there is no difference between them other than the side of the equal sign where these variables are located. Although time-varying predictors typically contain some proportion of between- and within-individual variability, convergence effects are not possible when the ICC = 0 or + 1 (and/or –1 in a marginal linear mixed-effects model) because all the variability in the time-varying predictor is either within- or between-individuals, respectively.

As an example of a time-varying predictor that may inappropriately assume convergence effects, consider the occasion-level variable indicating whether the individual was alone or with others during a given epoch, $Alone_{t,i}$. In general, although an individual will fluctuate as to whether they are alone or not across occasions at level 1, some individuals will be alone more often compared to other individuals at level 2 (e.g., family visits more often). Therefore, including time-varying $Alone_{t,i}$ in the linear mixed-effects model by itself would result in a weighted combination of these within- and between-individual effects, explaining both level-1, residual variance *and* level-2, random intercept variance simultaneously. To resolve this problem, $Alone_{t,i}$ could be explicitly partitioned into two variables that represent the unique level-1 and level-2 effects, using the procedures described next.

The partitioning procedure and the subsequent interpretation of the partitioned between-individual effect are dependent on whether *person-mean centering* or *grand-mean centering* was used (note that person-mean centering is termed *group-mean centering* in cross-sectional studies; see Raudenbush & Bryk, 2002). Although interpretation resulting from person-mean centering (aka, *variable centering*) is more intuitive for repeated-measures data compared to grand-mean centering, person-mean centering does not lend itself directly to the interpretation of binary predictors such as $Alone_{t,i}$. Thus, grand-mean centering is used throughout this example. Hoffman (2014, chapter 8) provides a complete description of how person-mean centering is used to partition time-varying predictors in longitudinal data, whereas Enders and Tofighi (2007) provide a focused description of using group-mean centering in cross-sectional data. Although grand-mean centering implies the use of the grand mean in centering decisions,

in general, this form of centering simply results in predictors being centered at some meaningful constant (that could, in fact, be zero). Thus, grand-mean centering could instead be synonymously termed *constant centering*.

To explicitly partition the between- and within-individual effects of $Alone_{t,i}$ using grand-mean centering, the proportion of occasions at which the individual was alone is calculated for each individual $i$, $\overline{Alone}_i$, which is a constant (i.e., time-invariant predictor) within each individual that carries between-individual variability (as indicated by subscript $i$) explaining level-2, location-model random intercept variance. For example, if an individual had 10 repeated occasions and was alone for 7 of them, the proportion being alone is 7/10 = 0.70 or 70%. Note that $\overline{Alone}_i$ could be centered at any meaningful proportion; however, in the example that follows, it will be left uncentered (i.e., $\overline{Alone}_i - 0$). By including both $Alone_{t,i}$ and $\overline{Alone}_i$ in the linear mixed-effects model, the fixed effect for $Alone_{t,i}$ represents the pure level-1 effect of being alone or not at each occasion. The fixed effect for $\overline{Alone}_i$ then represents the *difference* between the level-1 and level-2 effects (aka, a *contextual effect*; Raudenbush & Bryk, 2002) and is an explicit test of whether $Alone_{t,i}$ modeled by itself would have erroneously assumed a convergence effect. That is, given that a convergence effect assumes the level-1, within-individual and level-2, between-individual effects are equal, a statistically significant contextual effect indicates that the level-1 and level-2 effects are significantly different. The contextual effect is interpreted as the incremental change to the vector magnitude resulting from a one-unit increase in the proportion of total occasions at which the individual was alone *over and above* the effect of being alone at the current occasion. That is, the contextual effect indicates whether being alone at a given occasion has as

large of an effect if the individual is alone more often than others during the study period. Further, to obtain the level-2, between-individual effect, the level-1, within-individual and contextual effects are summed (i.e., within + contextual = between); this level-2, between-individual effect will also be interpreted below.

Continuing with the example, consider including both $Alone_{t,i}$ and $\overline{Alone_i}$ in the location model as shown in (2.5).

$$\begin{aligned}
&\text{Level 1:} \quad VM_{t,i} = \beta_{0,i} + \beta_{1,i}\left(Alone_{t,i}\right) + e_{t,i} \\
&\text{Level 2:} \quad \beta_{0,i} = \left(\gamma_{00} + u_{0,i}\right) + \gamma_{01}(AD_i) + \gamma_{02}(Ed_i - 12) + \\
&\qquad\qquad \gamma_{03}\left(\overline{Alone_i}\right) \\
&\qquad\qquad \beta_{1,i} = \gamma_{10} \\
&\text{Combined:} \quad VM_{t,i} = \left(\gamma_{00} + u_{0,i}\right) + \gamma_{01}(AD_i) + \gamma_{02}(Ed_i - 12) + \\
&\qquad\qquad \gamma_{03}\left(\overline{Alone_i}\right) + \gamma_{10}\left(Alone_{t,i}\right) + e_{t,i}
\end{aligned} \tag{2.5}$$

Here, all previous effects interpreted for (2.2) and (2.4) remain identical, with the location-model fixed intercept, $\gamma_{00}$, now representing the average vector magnitude specifically for a high school graduate without AD who is currently alone and who was alone at all occasions. $\beta_{1,i}$ is a new level-1 placeholder for the location-model fixed effect for $Alone_{t,i}$, $\gamma_{10}$, that represents the average difference in vector magnitude between individuals who were alone at a given occasion compared to individuals who were not *after controlling for the proportion of occasions at which an individual was alone*. In this model, $\gamma_{10}$ is a pure level-1 effect that explains level-1 residual variance. Further, $\gamma_{03}$ is the location-model fixed *contextual effect* for $\overline{Alone_i}$ representing the difference between the level-1, within-individual and level-2, between-individual effects. More specifically,

$\gamma_{03}$ represents the incremental change to vector magnitude that results from an individual being alone for an increased proportion of occasions *after controlling for whether the individual is alone or not at a given occasion*. Finally, as stated above, the pure level-2, between-individual effect is the sum of the level-1, within-individual and contextual effects, $\gamma_{10} + \gamma_{03}$, which represents the average change in vector magnitude per one-unit increase in the proportion of occasions for which the individual reported being alone.

Finally, adding to the matrix formulation of the linear mixed-effects model provided in (1.4), both $\overline{Alone_i}$ and $Alone_{t,i}$ are added to $\mathbf{X}_i$, which is now an $n_i$ x 5 matrix, and $\boldsymbol{B}$ is now a 5 x 1 column vector holding the additional location-model fixed effects for $\overline{Alone_i}$ at level 2, $\gamma_{03}$, and $Alone_{t,i}$ at level 1, $\gamma_{10}$. All other matrices remain identical as described in the discussion of time-invariant predictors.

***Including additional random effects.*** The only random effect discussed thus far was the location-model random intercept, $u_{0,i}$. However, with repeated-measures data, additional location-model random effects (e.g., random slopes) can be specified for any level-1 predictor to the extent that the level-2 variability is non-zero and there are enough level-2 units (Snijders & Bosker, 2012). That is, for repeated-measures data, random slopes can only be estimated for time-varying predictors, and if necessary, ensure the location-model fixed-effect standard errors are less wrong (or, more correct). A location-model random slope indicates the within-individual effect of a time-varying predictor differs randomly across individuals (i.e., a heterogeneous effect). Note that in repeated-measures data, location-model random slopes cannot be included for level-2 predictors because there are no higher-level units over which the individuals could vary randomly.

With this in mind, a location-model random slope could only be estimated for the level-1 partition of the time-varying predictor $Alone_{t,i}$. Thus, extending the example to include this location-model random slope is shown in (2.6) and also presented visually in Figure 2.2.

$$
\begin{aligned}
&\text{Level 1:} \quad VM_{t,i} = \beta_{0,i} + \beta_{1,i}\left(Alone_{t,i}\right) + e_{t,i} \\
&\text{Level 2:} \quad \beta_{0,i} = \left(\gamma_{00} + u_{0,i}\right) + \gamma_{01}(AD_i) + \gamma_{02}(Ed_i - 12) + \\
&\qquad\qquad \gamma_{03}\left(\overline{Alone_i}\right) \\
&\qquad\qquad \beta_{1,i} = \left(\gamma_{10} + u_{1,i}\right) \\
&\text{Combined:} \quad VM_{t,i} = \left(\gamma_{00} + u_{0,i}\right) + \gamma_{01}(AD_i) + \gamma_{02}(Ed_i - 12) + \\
&\qquad\qquad \gamma_{03}\left(\overline{Alone_i}\right) + \left(\gamma_{10} + u_{1,i}\right)\left(Alone_{t,i}\right) + e_{t,i}
\end{aligned}
\tag{2.6}
$$

This model specifically evaluates whether the effect of being alone at a given occasion is the same for all individuals—whether the fixed linear effect of being alone is appropriate to describe everyone in the sample. Note that this model requires that the location-model, level-1, fixed effect for being alone, $\gamma_{10}$, is retained in the model regardless of its statistical significance. That is, random effects represent deviations from fixed effects, and without the fixed effect there would be nothing from which the random effect could deviate. Thus, the level-1 effect of being alone, $\beta_{1,i}$, is now represented at level 2 by the location-model fixed effect of being alone, $\gamma_{10}$, indicating the average difference in vector magnitude between being alone or not (the solid line in Figure 2.2), and the individual-specific deviation from the location-model fixed effect of being alone, $u_{1,i}$. Note in Figure 2.2 that $\gamma_{10}$ is an underestimate for individual 1 (i.e., upper dashed line; $u_{1,1} > 0$) and an overestimate for individual 2 (i.e., lower dashed line; $u_{1,2} < 0$).

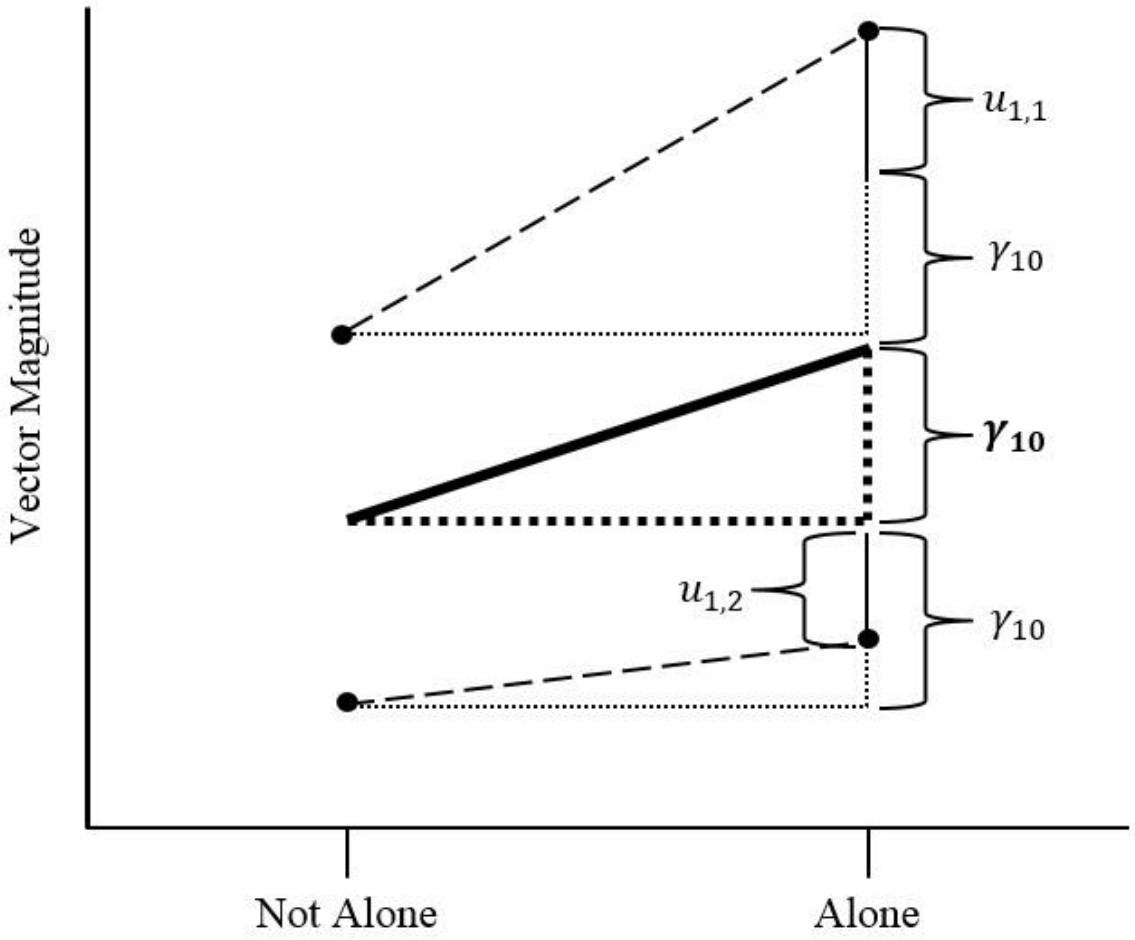


*Figure 2.2*. Random effect for level-1 effect of being alone or not

In adding these new effects to the matrix formulation provided in (1.4), $Alone_{t,i}$ is also included in $\mathbf{Z}_i$, which is now an $n_i$ x 2 matrix of predictors that have location-model random effects, and $\mathbf{u}_i$ is now a 2 x 1 column vector additionally holding the deviation from the fixed $Alone_{t,i}$ effect for individual $i$, $u_{1,i}$. All other matrices remain identical as described in the discussion of time-varying predictors.

Similar to the random intercept variance, the random $Alone_{t,i}$ slope variance, $\sigma^2_{u_1}$, is estimated as constant between individuals and included in $\mathbf{G}$ alongside the correlation between the random intercept and random $Alone_{t,i}$ slope, $\rho_{u_0;u_1}$ and $\rho_{u_1;u_0}$ (used to obtain covariances), as shown below in (2.7).

$$\mathbf{G} = \begin{pmatrix} \sigma^2_{u_0} & \rho_{u_0;u_1}\sqrt{\sigma^2_{u_0}\sigma^2_{u_1}} \\ \rho_{u_1;u_0}\sqrt{\sigma^2_{u_1}\sigma^2_{u_0}} & \sigma^2_{u_1} \end{pmatrix} \quad (2.7)$$

Further, the location-model fixed effect for $Alone_{t,i}$ and its location-model random slope variance can be used to calculate a 95% random effects confidence interval, $\gamma_{10} \pm 1.96\sqrt{\sigma^2_{u_1}}$, which indicates the expected variability in the difference between being alone or not for 95% of individuals in the sample (Snijders & Bosker, 2012).

***A brief comment on interaction effects.*** The linear mixed-effects model can include interaction effects composed of any combination of level-1 and level-2 predictors. However, the inclusion of interaction effects that involve level-1 predictors as well as the identification of the specific variance component the interaction effect explains is complex due to the potential for assumed convergence effects, and is therefore beyond the scope of this chapter (see Hoffman 2014). Briefly, the interaction of pure level-1 predictors explains level-1, residual variance, the interaction of pure level-2 predictors explains level-2, random intercept variance, and the interaction of a pure level-1 predictor and a pure level-2 predictor (i.e., a *cross-level interaction*) explains level-1, residual variance if the level-1 effect *is not* random or it explains level-2 random slope variance if the level-1 effect *is* random.

When considering the matrix formulation provided in (1.4), because interactions are multiplicative, the product of any predictors involved in the interaction (e.g., $AD_i * \overline{Alone}_i$) would produce an additional column in $\mathbf{X}_i$, with the location-model fixed effect for the interaction included in an additional row of $\boldsymbol{B}$. Further, it is possible to include a

pure level-1 interaction as a location-model random effect, which would additionally modify $\mathbf{Z}_i$ and $\mathbf{u}_i$ accordingly.

**The Heterogeneous Variance and Mixed-Effects Location-Scale Models**

As described in chapter 1, the heterogeneous variance model allows scale-model predictors of residual and/or random effect variances to be heterogeneous between individuals using *fixed effects*. That is, the heterogeneous variance model has the flexibility to include location-model random effects to estimate heterogeneity in mean level across individuals as well as a log-linear combination of scale-model predictor variables to estimate the average (i.e., fixed) effect of a predictor on level-1 and/or level-2 variances. It is important to note that neither the location-model random effects nor the scale-model fixed effects explicitly model the *random individual differences in the variability* of the outcome. Thus, the mixed-effects location-scale model is required to include scale-model random effects representing these individual differences.

**The scale-model random intercept.** Although the theoretical framework in chapter 1 described the heterogeneous variance model prior to mixed-effects location-scale model, it is important to estimate scale-model random effects in the same order as location-model random effects. That is, estimating a heterogeneous variance model (i.e., scale-model fixed effects) absent of scale-model random effects has been shown to result in downwardly biased standard errors for scale-model fixed effects of residual variance, especially for level-2, between-individual predictors (Leckie, 2014; Leckie et al., 2014).

Following this recommendation, in the example below the mixed-effects location-scale model is used initially to estimate an unconditional scale-model random intercept model for the residual variance to determine whether differences in residual variance

exist between individuals. Thus, assuming the location model in (2.6), the scale model for the residual variance is shown in (2.8) using multi-level, scalar notation similar to the location models above. Note that residual values are still assumed independent within individuals in which residual correlations are constrained to be zero, $\mathbf{R}_i = \sigma^2_{e_i}\mathbf{I}_{n_i}$. Therefore, the subscript $e$ for the (log of the) residual variance in (2.8) does not require an additional subscript for occasion $t$ because there is only one estimated residual variance across all occasions for each individual.

$$\begin{aligned} &\text{Level 1:} && \log\left(\sigma^2_{e_i}\right) = \tau^e_{0,i} \\ &\text{Level 2:} && \tau^e_{0,i} = \delta^e_{00} + \omega^e_{0,i} \\ &\text{Combined:} && \log\left(\sigma^2_{e_i}\right) = \delta^e_{00} + \omega^e_{0,i} \end{aligned} \tag{2.8}$$

Here, the level-1 scale model now describes *within-individual* variation in estimated residual variance as a function of the residual variance specific to individual $i$, $\tau^e_{0,i}$, which is simply a placeholder for the effects included in the level-2, *between-individual* model. Specifically, $\tau^e_{0,i}$ is a function of the scale-model fixed intercept for the residual variance, $\delta^e_{00}$, representing the grand mean of the (log of the) residual variance estimates across all individuals, and the scale-model random intercept, $\omega^e_{0,i}$, representing the constant *deviation* from the fixed intercept for the residual variance for individual $i$. These level-2 effects are substituted for $\tau^e_{0,i}$ at level 1 to create the combined equation. Similar to the location models above, because $\delta^e_{00}$ is a fixed effect, it applies to every individual in the sample (i.e., there is no subscript $i$) and serves as the reference point for quantifying the scale-model random effect for the residual variance for each individual $i$, $\omega^e_{0,i}$.

When considering the matrix formulation of the mixed-effects location-scale model shown in (1.27), both $\mathbf{T}^e_i$ and $\mathbf{W}^e_i$ are $n_i$ x 1 column vectors of ones for the $n$

occasions for individual $i$, whereas $\boldsymbol{\tau}^e$ and $\boldsymbol{\omega}_i^e$ are scalars holding the fixed intercept of the (log of the) residual variance, $\delta_{00}^e$, and the random deviation from the fixed intercept of the residual variance for each individual $i$, $\omega_{0,i}^e$, respectively. The variance of $\omega_{0,i}^e$ is indicated by $\sigma_{\omega_0^e}^2$, and represents the scale-model random intercept variance.

Further, $\boldsymbol{\omega}_i^e$ is subsumed within $\mathbf{m}_i$ alongside location-model random effects, $\mathbf{u}_i$. Therefore, when considering both location- and scale-model random effects, $\mathbf{m}_i$ is a 3 x 1 column vector where the first and second rows hold location-model random effects, $u_{0,i}$ and $u_{1,i}$, and the third row holds the scale-model random effect, $\omega_{0,i}^e$, as shown in (2.9).

$$\mathbf{m}_i = \begin{bmatrix} \mathbf{u}_i \\ -- \\ \boldsymbol{\omega}_i^e \end{bmatrix} = \begin{bmatrix} u_{0,i} \\ u_{1,i} \\ \omega_{0,i}^e \end{bmatrix} \tag{2.9}$$

Note that all elements in $\mathbf{G}$ remain constant between individuals (i.e., no subscript $i$) given the unconditional scale model for all variances and correlations (used to obtain covariances) in $\mathbf{G}$, as shown below in (2.10).

$$\mathbf{G} = \begin{pmatrix} \sigma_{u_0}^2 & \rho_{u_0;u_1}\sqrt{\sigma_{u_0}^2\sigma_{u_1}^2} & \rho_{u_0;\omega_0^e}\sqrt{\sigma_{u_0}^2\sigma_{\omega_0^e}^2} \\ \rho_{u_1;u_0}\sqrt{\sigma_{u_1}^2\sigma_{u_0}^2} & \sigma_{u_1}^2 & \rho_{u_1;\omega_0^e}\sqrt{\sigma_{u_1}^2\sigma_{\omega_0^e}^2} \\ \rho_{\omega_0^e;u_0}\sqrt{\sigma_{\omega_0^e}^2\sigma_{u_0}^2} & \rho_{\omega_0^e;u_1}\sqrt{\sigma_{\omega_0^e}^2\sigma_{u_1}^2} & \sigma_{\omega_0^e}^2 \end{pmatrix} \tag{2.10}$$

A visual depiction of a mixed-effects location-scale model is shown in Figure 2.3. Here, the distribution of the dashed lines around the location-model fixed intercept, $\gamma_{00}$ (i.e., the solid line), indicate between-individual variability, modeled by including location-model random effects, $u_{0,i}$. The distribution of an individual's data (i.e., the dots) around their own dashed lines indicates within-individual variability, $\sigma_{e_i}^2$,

whereas $\sigma_e^2$ is the fixed residual variance (i.e., $\exp(\delta_{00}^e)$; the average residual variance across these two individuals). Individual differences in residual variance for a given individual $i$, $\omega_{0,i}^e$, represent the difference between $\sigma_{e_i}^2$ and $\sigma_e^2$. Note that individual 1 has more residual variability compared to individual 2; thus, the fixed residual variance estimate underestimates the residual variance for individual 1 (i.e., $\omega_{0,1}^e = \sigma_{e_1}^2 - \sigma_e^2 > 0$) and overestimates the residual variance for individual 2 (i.e., $\omega_{0,2}^e = \sigma_{e_2}^2 - \sigma_e^2 < 0$).

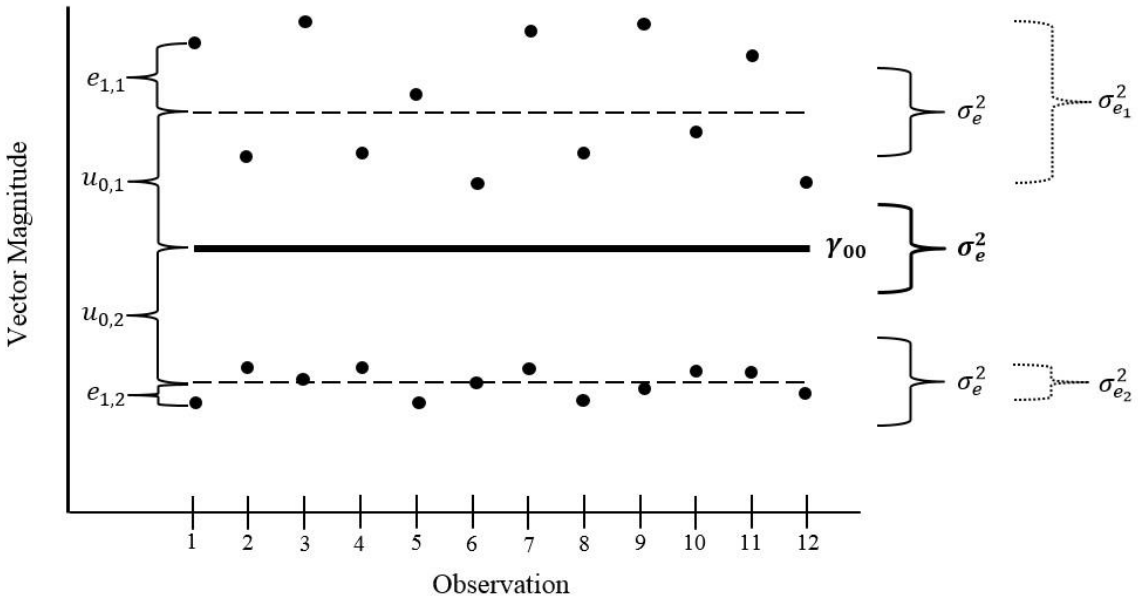


*Figure 2.3*. Visual depiction of the mixed-effects location-scale model

**Including scale-model predictors.** It has been generally accepted that scale-model predictors can only be included at their own level or lower (Hedeker et al., 2008; Snijders & Bosker, 2012). For example, in repeated-measures data, level-2 random effect variances are predicted only by level-2 predictors, whereas level-1 residual variances could be predicted by either level-1 or level-2 predictors. However, Hedeker and Nordgren (2013) have updated this perspective stating that a level-1 variable could be used to predict level-2 variance components by making *the individual* at level-2 more or

less heterogeneous across level-1 occasions. This suggestion is controversial as their software MIXREGLS is the only software available to estimate these effects. Thus, until more research is conducted, it is assumed that predictors can be used to explain heterogeneity at their own level or lower. More specifically, regardless of superscript, although the design matrix for the scale-model fixed effects of level-2 variance components in $\mathbf{A}_i$, described in (1.9) and (1.10), as well as the design matrix for the scale-model fixed effects of level-1 residual variances in $\mathbf{T}_i$, described in (1.12) and (1.13), can contain level-2 predictors, only $\mathbf{T}_i$ can contain time-varying predictors. In addition, it is assumed the procedures for including time-varying predictors in the location model apply directly to the scale model for the residual variance. Therefore, throughout the examples below, variable partitioning and grand-mean centering are used to prevent erroneously assuming convergence effects as described above.

***Predicting residual variances and correlations.*** Continuing with the example, consider the scale model for the residual variance in (2.8) that includes $AD_i$, $Alone_{t,i}$, and $\overline{Alone_i}$ as predictors of within-individual heterogeneity in $\mathbf{R}_i$, presented below in (2.11).

$$\begin{aligned}
&\text{Level 1:} \quad \log\left(\sigma^2_{e_{t,i}}\right) = \tau^e_{0,i} + \tau^e_{1,i}\left(Alone_{t,i}\right) \\
&\text{Level 2:} \quad \tau^e_{0,i} = \left(\delta^e_{00} + \omega^e_{0,i}\right) + \delta^e_{01}(AD_i) + \delta^e_{02}\left(\overline{Alone_i}\right) \\
&\qquad\qquad \tau^e_{1,i} = \delta^e_{10} \\
&\text{Combined:} \quad \log\left(\sigma^2_{e_{t,i}}\right) = \left(\delta^e_{00} + \omega^e_{0,i}\right) + \delta^e_{01}(AD_i) + \delta^e_{02}\left(\overline{Alone_i}\right) + \\
&\qquad\qquad \delta^e_{10}\left(Alone_{t,i}\right)
\end{aligned} \tag{2.11}$$

Although $\mathbf{R}_i$ is now predicted to be heterogeneous between individuals and across occasions, the residual values are still assumed independent within individuals, $\mathbf{R}_i =$

$\sigma^2_{e_{t,i}}\mathbf{I}_{n_i}$. Thus, the subscript $e$ for the (log of the) residual variance in (2.11) now requires the additional subscript for occasion $t$ because, given the time-varying predictor $Alone_{t,i}$, the estimated residual variance is allowed to vary across the $t$ occasions for each individual $i$.

In (2.11), the fixed intercept included in the scale model for the residual variance, $\delta^e_{00}$, now represents the (log of the) residual variance specifically for an individual without AD who is currently alone and who was alone at all occasions. $\tau^e_{1,i}$ is a new level-1 placeholder of the scale-model fixed effect for $Alone_{t,i}$, $\delta^e_{10}$, that represents the difference in (the log of the) residual variance between individuals who were alone at a given occasion compared to individuals who were not, after controlling for the proportion of occasions at which an individual was alone. Further, $\delta^e_{02}$ is the scale-model fixed contextual effect (included to prevent erroneously assuming convergence effects for reasons discussed above) representing the difference between the level-1, within-individual and level-2, between-individual effects, interpreted as the incremental difference in (the log of the) residual variance for a one-unit increase in the proportion of occasions an individual was alone after controlling for whether the individual was alone or not at a given occasion. In addition, the level-2, between-individual effect is the sum of the level-1 and contextual effects for being alone, $\delta^e_{10} + \delta^e_{02}$, which represents the difference in (the log of the) residual variance for a one-unit increase in the proportion of occasions at which the individual reported being alone. Further, $\delta^e_{01}$ is the scale-model fixed effect representing the difference in (the log of the) residual variance between individuals with and without AD. Finally, the predicted residual variance for individual $i$

at occasions $t$ is calculated by exponentiating the combined equation (excluding scale-model random effect $\omega_{0,i}^{e}$), as shown below in (2.12).

$$\mathbf{R}_i = \exp\left(\delta_{00}^{e} + \delta_{01}^{e}(AD_i) + \delta_{02}^{e}\left(\overline{Alone}_i\right) + \delta_{10}^{e}\left(Alone_{t,i}\right)\right)\mathbf{I}_{n_i} \tag{2.12}$$

When considering the matrix representations of the scale model for residual variance in (1.27), $AD_i$, $Alone_{t,i}$, and $\overline{Alone}_i$ are all added to $\mathbf{T}_i^{e}$, which is now a $n_i$ x 4 matrix, and $\boldsymbol{\tau}^{e}$ is now a 4 x 1 column vector holding the additional scale-model fixed effects $\delta_{01}^{e}$, $\delta_{02}^{e}$, and $\delta_{10}^{e}$.

***Random effects for scale-model predictors of residual variance.*** Similar to the location model, when modeling repeated-measures data, scale-model random effects (i.e., random slopes) can only be specified for level-1, time-varying predictors included in the scale model for the residual variance (time-varying predictors cannot be included in the scale model for random effect variances and correlations in data with two levels of nesting). Therefore, in this example, only the effect of $Alone_{t,i}$, $\delta_{10}^{e}$, could vary randomly between individuals.

The procedure to include $Alone_{t,i}$ as an additional random effect in the scale model for the residual variance is identical to the procedure for including additional random effects in the location model, as shown in (2.13).

$$\begin{aligned}
&\text{Level 1:} \quad \log\left(\sigma_{e_{t,i}}^{2}\right) = \tau_{0,i}^{e} + \tau_{1,i}^{e}\left(Alone_{t,i}\right)\\
&\text{Level 2:} \quad \tau_{0,i}^{e} = \left(\delta_{00}^{e} + \omega_{0,i}^{e}\right) + \delta_{01}^{e}(AD_i) + \delta_{02}^{e}\left(\overline{Alone}_i\right)\\
&\qquad\qquad \tau_{1,i}^{e} = \left(\delta_{10}^{e} + \omega_{1,i}^{e}\right)\\
&\text{Combined:} \quad \log\left(\sigma_{e_{t,i}}^{2}\right) = \left(\delta_{00}^{e} + \omega_{0,i}^{e}\right) + \left(\delta_{10}^{e} + \omega_{1,i}^{e}\right)(AD_i) +\\
&\qquad\qquad \delta_{02}^{e}\left(\overline{Alone}_i\right) + \delta_{10}^{e}\left(Alone_{t,i}\right)
\end{aligned} \tag{2.13}$$

Here, the level-1 effect of being alone at a given occasion, $\tau_{1,i}^{e}$, is now represented at level 2 by the fixed effect of being alone at a given occasion, $\delta_{10}^{e}$, and the deviation from this fixed effect, $\omega_{1,i}^{e}$. More specifically, $\omega_{1,i}^{e}$ indicates that the difference in (the log of the) residual variance between occasions in which the individual was alone or not, $\delta_{10}^{e}$, does not describe all individuals equally. That is, compared to other individuals, being alone may have resulted in a greater increase in residual variability compared to being not alone.

Adding these new effects to the matrix formulation in (1.27), $\mathbf{W}_i^e$ is now an $n_i$ x 2 matrix of predictors that have scale-model random effects, and $\boldsymbol{\omega}_i^e$ is a 2 x 1 column vector now holding the individual-specific deviation from the scale-model fixed effect of $Alone_{t,i}$, $\omega_{1i}^{e}$. Similar to above, $\boldsymbol{\omega}_i^e$ will be subsumed within $\mathbf{m}_i$ alongside location-model random effects, $\mathbf{u}_i$. Thus, $\mathbf{m}_i$ is now a 4 x 1 column vector holding all location- and scale-model random effects.

Finally, the scale-model random $Alone_{t,i}$ slope variance, $\sigma_{\omega_1^e}^2$, is also included in $\mathbf{G}$ alongside the other location- and scale-model random effect variances and correlations (used to obtain covariances), as shown in (2.14).

$$\mathbf{G} = \begin{pmatrix} \sigma_{u_0}^2 & \rho_{u_0;u_1}\sqrt{\sigma_{u_0}^2\sigma_{u_1}^2} & \rho_{u_0;\omega_0^e}\sqrt{\sigma_{u_0}^2\sigma_{\omega_0^e}^2} & \rho_{u_0;\omega_1^e}\sqrt{\sigma_{u_0}^2\sigma_{\omega_1^e}^2} \\ \rho_{u_1;u_0}\sqrt{\sigma_{u_1}^2\sigma_{u_0}^2} & \sigma_{u_1}^2 & \rho_{u_1;\omega_0^e}\sqrt{\sigma_{u_1}^2\sigma_{\omega_0^e}^2} & \rho_{u_1;\omega_1^e}\sqrt{\sigma_{u_1}^2\sigma_{\omega_1^e}^2} \\ \rho_{\omega_0^e;u_0}\sqrt{\sigma_{\omega_0^e}^2\sigma_{u_0}^2} & \rho_{\omega_0^e;u_1}\sqrt{\sigma_{\omega_0^e}^2\sigma_{u_1}^2} & \sigma_{\omega_0^e}^2 & \rho_{\omega_0^e;\omega_1^e}\sqrt{\sigma_{\omega_0^e}^2\sigma_{\omega_1^e}^2} \\ \rho_{\omega_1^e;u_0}\sqrt{\sigma_{\omega_1^e}^2\sigma_{u_0}^2} & \rho_{\omega_1^e;u_1}\sqrt{\sigma_{\omega_1^e}^2\sigma_{u_1}^2} & \rho_{\omega_1^e;\omega_0^e}\sqrt{\sigma_{\omega_1^e}^2\sigma_{\omega_0^e}^2} & \sigma_{\omega_1^e}^2 \end{pmatrix} \quad (2.14)$$

Note that all elements in $\mathbf{G}$ remain constant between individuals (i.e., no subscript $i$) given the unconditional scale model for variances and correlations in $\mathbf{G}$. Further, similar

to location-model random slope variances, scale-model random slope variances can be explained by including cross-level interactions in the scale model for the residual variance that include the random level-1 predictor and a level-2 predictor. Interaction effects involving time-varying predictors are as complex for the scale model as they are for the location model due the potential of erroneously assuming convergence effects. Therefore, an explicit example of a cross-level interaction is beyond the scope of this chapter (see Hoffman 2014).

***Predicting location-model random effect variances and correlations.*** Moving on to the prediction of random effect variances and correlations in **G**, assuming the location model in (2.6), consider extending the scale model for random effect variances and correlations to include $AD_i$ and $\overline{Alone}_i$ as predictors of between-individual heterogeneity in both the location-model random intercept and location-model random $Alone_{t,i}$ slope variances, as presented below in (2.15) and (2.16), respectively (Note that **G** now requires the subscript $i$; $\mathbf{G}_i$). Further, the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, and scale-model random $Alone_{t,i}$ slope variance, $\sigma^2_{\omega_1^e}$, will be estimated but not predicted to be heterogeneous across individuals (i.e., no subscript $i$) as shown in (2.17) and (2.18), as will all random effect correlations, $\rho_{u_{0,i};u_{1,i}}$, $\rho_{u_{0,i};\omega_0^e}$, $\rho_{u_{0,i};\omega_1^e}$, $\rho_{u_{1,i};\omega_0^e}$, $\rho_{u_{1,i};\omega_1^e}$, and $\rho_{\omega_0^e;\omega_1^e}$ shown in (2.19) through (2.24). Note that Leckie et al. (2014) have suggested that the inverse hyperbolic tangent link used for random effect correlations is no longer sufficient to ensure $\mathbf{G}_i$ remains positive definite given $\mathbf{G}_i$ is larger than 2 x 2; however, this link is necessary to ensure that the range of scale-model fixed effects remains unbounded.

$$\log\left(\sigma^2_{u_{0,i}}\right) = \alpha_0^{u_0} + \alpha_1^{u_0}(AD_i) + \alpha_2^{u_0}\left(\overline{Alone}_i\right) \tag{2.15}$$

$$\log\left(\sigma^2_{u_{1,i}}\right) = \alpha_0^{u_1} + \alpha_1^{u_1}(AD_i) + \alpha_2^{u_1}\left(\overline{Alone}_i\right) \tag{2.16}$$

$$\log\left(\sigma^2_{\omega_0^e}\right) = \alpha_0^{\omega_0^e} \tag{2.17}$$

$$\log\left(\sigma^2_{\omega_1^e}\right) = \alpha_0^{\omega_1^e} \tag{2.18}$$

$$\tanh^{-1}\left(\rho_{u_{0,i};u_{1,i}}\right) = \alpha_0^{u_{0,i};u_{1,i}} \tag{2.19}$$

$$\tanh^{-1}\left(\rho_{u_{0,i};\omega_0^e}\right) = \alpha_0^{u_{0,i};\omega_0^e} \tag{2.20}$$

$$\tanh^{-1}\left(\rho_{u_{0,i};\omega_1^e}\right) = \alpha_0^{u_{0,i};\omega_1^e} \tag{2.21}$$

$$\tanh^{-1}\left(\rho_{u_{1,i};\omega_0^e}\right) = \alpha_0^{u_{1,i};\omega_0^e} \tag{2.22}$$

$$\tanh^{-1}\left(\rho_{u_{1,i};\omega_1^e}\right) = \alpha_0^{u_{1,i};\omega_1^e} \tag{2.23}$$

$$\tanh^{-1}\left(\rho_{\omega_0^e;\omega_1^e}\right) = \alpha_0^{\omega_0^e;\omega_1^e} \tag{2.24}$$

The interpretation of scale-model fixed effects are identical to those for location-model fixed effects, $\alpha_0^{u_0}$ represents the (log of the) location-model random intercept variance for an individuals without AD (i.e., when $AD_i = 0$) who was alone at every occasion (i.e., when $\overline{Alone}_i = 0$). In addition, $\alpha_1^{u_0}$ is the scale-model fixed effect representing the difference in (the log of the) location-model random intercept variance between individuals with and without AD, and $\alpha_2^{u_0}$ is the scale-model fixed effect representing the difference in (the log of the) location-model random intercept variance for a one-unit increase in the proportion of occasions at which the individual was alone. Therefore, the location-model random intercept variance specifically for *individuals without AD* who were alone at all occasions is given by (2.25),

$$\sigma^2_{u_{0,i}} = \exp\left(\alpha_0^{u_0}\right) \tag{2.25}$$

whereas the location-model random intercept variance specifically for *individuals with AD* who were alone at all occasions is given by (2.26).

$$\sigma^2_{u_{0,i}} = \exp\left(\alpha_0^{u_0} + \alpha_1^{u_0}\right) \tag{2.26}$$

Following exponentiation, a similar interpretation is used for the location-model random $Alone_{t,i}$ slope variance, $\sigma^2_{u_{1,i}}$, conditional on scale-model fixed effects $\alpha_0^{u_1}$ and $\alpha_1^{u_1}$. Exponentiating the fixed effects for the scale-model random intercept variance and scale-model random $Alone_{t,i}$ slope variance, $\alpha_0^{\omega_0^e}$ and $\alpha_0^{\omega_1^e}$, respectively, will provide the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, and scale-model random $Alone_{t,i}$ slope variance, $\sigma^2_{\omega_1^e}$.

By contrast, although the correlations were estimated as constant between individuals, any covariance with a location-model random effect will be heterogeneous, as shown below in (2.27) to (2.32).

$$\sigma_{u_{0,i};u_{1,i}} = \tanh\left(\alpha_0^{u_{0,i};u_{1,i}}\right)\sqrt{\exp\left(\alpha_0^{u_0} + \alpha_1^{u_0}(AD_i) + \alpha_2^{u_0}\left(\overline{Alone}_i\right)\right)\exp\left(\alpha_0^{u_1} + \alpha_1^{u_1}(AD_i) + \alpha_2^{u_1}\left(\overline{Alone}_i\right)\right)} \tag{2.27}$$

$$\sigma_{u_{0,i};\omega_0^e} = \tanh\left(\alpha_0^{u_{0,i};\omega_0^e}\right)\sqrt{\exp\left(\alpha_0^{u_0} + \alpha_1^{u_0}(AD_i) + \alpha_2^{u_0}\left(\overline{Alone}_i\right)\right)\exp\left(\alpha_0^{\omega_0^e}\right)} \tag{2.28}$$

$$\sigma_{u_{0,i};\omega_1^e} = \tanh\left(\alpha_0^{u_{0,i};\omega_1^e}\right)\sqrt{\exp\left(\alpha_0^{u_0} + \alpha_1^{u_0}(AD_i) + \alpha_2^{u_0}\left(\overline{Alone}_i\right)\right)\exp\left(\alpha_0^{\omega_1^e}\right)} \tag{2.29}$$

$$\sigma_{u_{1,i};\omega_0^e} = \tanh\left(\alpha_0^{u_{1,i};\omega_0^e}\right)\sqrt{\exp\left(\alpha_0^{u_1} + \alpha_1^{u_1}(AD_i) + \alpha_2^{u_1}\left(\overline{Alone}_i\right)\right)\exp\left(\alpha_0^{\omega_0^e}\right)} \tag{2.30}$$

$$\sigma_{u_{1,i};\omega_1^e} = \tanh\left(\alpha_0^{u_{1,i};\omega_1^e}\right)\sqrt{\exp\left(\alpha_0^{u_1} + \alpha_1^{u_1}(AD_i) + \alpha_2^{u_1}\left(\overline{Alone}_i\right)\right)\exp\left(\alpha_0^{\omega_1^e}\right)} \tag{2.31}$$

$$\sigma_{\omega_0^e;\omega_1^e} = \tanh\left(\alpha_0^{\omega_0^e;\omega_1^e}\right)\sqrt{\exp\left(\alpha_0^{\omega_0^e}\right)\exp\left(\alpha_0^{\omega_1^e}\right)} \tag{2.32}$$

Thus, when considering heterogeneous location-model random intercept and random $Alone_{t,i}$ slope variances, $\mathbf{G}_i$ is modified slightly to include the subscript $i$ when appropriate, as shown below in (2.33).

$$\mathbf{G}_i = \begin{pmatrix} \sigma^2_{u_{0,i}} & \rho_{u_{0,i};u_{1,i}}\sqrt{\sigma^2_{u_{0,i}}\sigma^2_{u_{1,i}}} & \rho_{u_{0,i};\omega^e_0}\sqrt{\sigma^2_{u_{0,i}}\sigma^2_{\omega^e_0}} & \rho_{u_{0,i};\omega^e_1}\sqrt{\sigma^2_{u_{0,i}}\sigma^2_{\omega^e_1}} \\ \rho_{u_{1,i};u_{0,i}}\sqrt{\sigma^2_{u_{1,i}}\sigma^2_{u_{0,i}}} & \sigma^2_{u_{1,i}} & \rho_{u_{1,i};\omega^e_0}\sqrt{\sigma^2_{u_{1,i}}\sigma^2_{\omega^e_0}} & \rho_{u_{1,i};\omega^e_1}\sqrt{\sigma^2_{u_{1,i}}\sigma^2_{\omega^e_1}} \\ \rho_{\omega^e_0;u_{0,i}}\sqrt{\sigma^2_{\omega^e_0}\sigma^2_{u_{0,i}}} & \rho_{\omega^e_0;u_{1,i}}\sqrt{\sigma^2_{\omega^e_0}\sigma^2_{u_{1,i}}} & \sigma^2_{\omega^e_0} & \rho_{\omega^e_0;\omega^e_1}\sqrt{\sigma^2_{\omega^e_0}\sigma^2_{\omega^e_1}} \\ \rho_{\omega^e_1;u_{0,i}}\sqrt{\sigma^2_{\omega^e_1}\sigma^2_{u_{0,i}}} & \rho_{\omega^e_1;u_{1,i}}\sqrt{\sigma^2_{\omega^e_1}\sigma^2_{u_{1,i}}} & \rho_{\omega^e_1;\omega^e_0}\sqrt{\sigma^2_{\omega^e_1}\sigma^2_{\omega^e_0}} & \sigma^2_{\omega^e_1} \end{pmatrix} \quad (2.33)$$

Finally, when considering the matrix formulations of the scale model for location-model random effects in (1.9) and (1.10), $\mathbf{A}_i^{u_0}$ and $\mathbf{A}_i^{u_1}$ are now 1 x 3 row vectors given the addition of the level-2 predictors $AD_i$ and $\overline{Alone}_i$, whereas $\mathbf{A}_i^{\omega^e_0}$, $\mathbf{A}_i^{\omega^e_1}$, $\mathbf{A}_i^{u_0;u_1}$, $\mathbf{A}_i^{u_0;\omega^e_0}$, $\mathbf{A}_i^{u_0;\omega^e_1}$, $\mathbf{A}_i^{u_1;\omega^e_0}$, $\mathbf{A}_i^{u_1;\omega^e_1}$, and $\mathbf{A}_i^{\omega^e_0;\omega^e_1}$ all remain scalars equal to 1 given that the scale-model random effect variances and all correlations were not predicted to be heterogeneous between individuals. Regarding the scale-model fixed effects for the level-2 variances and correlations, $\boldsymbol{\alpha}^{u_0}$ and $\boldsymbol{\alpha}^{u_1}$ are now both 3 x 1 column vectors holding the additional scale-model fixed effect for $AD_i$ and $\overline{Alone}_i$, $\alpha_1^{u_0}$ and $\alpha_2^{u_0}$, and $\alpha_1^{u_1}$ and $\alpha_2^{u_1}$, respectively, whereas $\boldsymbol{\alpha}^{\omega^e_0}$, $\boldsymbol{\alpha}^{\omega^e_1}$, $\boldsymbol{\alpha}^{u_0;u_1}$, $\boldsymbol{\alpha}^{u_0;\omega^e_0}$, $\boldsymbol{\alpha}^{u_0;\omega^e_1}$, $\boldsymbol{\alpha}^{u_1;\omega^e_0}$, $\boldsymbol{\alpha}^{u_1;\omega^e_1}$, and $\boldsymbol{\alpha}^{\omega^e_0;\omega^e_1}$ each remain scalars holding fixed intercepts representing the scale-model random intercept variance, $\alpha_0^{\omega^e_0}$, scale-model random $Alone_{t,i}$ slope variance, $\alpha_0^{\omega^e_1}$, and the correlations between the random effects, $\alpha_0^{u_{0,i};u_{1,i}}$, $\alpha_0^{u_{0,i};\omega^e_0}$, $\alpha_0^{u_{0,i};\omega^e_1}$, $\alpha_0^{u_{1,i};\omega^e_0}$, $\alpha_0^{u_{1,i};\omega^e_1}$, and $\alpha_0^{\omega^e_0;\omega^e_1}$.

**Chapter Summary**

The purpose of this chapter was to present an explicit model-building example of the mixed-effects location-scale model that mapped directly onto the theoretical

framework detailed in chapter 1 and mapped closely onto the empirical data analysis presented in chapter 5. The example presented in this chapter followed the model-building procedures and recommendations typically encountered in the literature, in which the location model was assumed properly specified before including scale-model random effects and predictors. With that order in mind, this chapter first included the location-model random intercept followed by location-model fixed effects, and then location-model random effects for a level-1, within-individual predictor. Then a similar procedure followed for the scale model, with the scale-model random intercept estimated initially, followed by scale-model fixed effects for the variances and correlations in both $\mathbf{G}_i$ and $\mathbf{R}_i$, and finally, random effects of a level-1, within-individual predictor included in the scale model for the residual variance.

Although the model-building sequence used in this chapter has been followed traditionally, methodological literature has only provided suggestions regarding the order in which effects should be modeled. Thus, the second simulation study presented in chapter 4 will detail the consequences misspecifying the location and/or scale model have on location- and scale-model fixed effects. Prior to the methodological studies in chapter 4, however, explicit details regarding the estimation of the mixed-effects location-scale model are presented next in chapter 3, which describes a newly developed Markov chain Monte Carlo algorithm to estimate this model.

# CHAPTER 3: MCMC ESTIMATION OF THE MIXED-EFFECTS LOCATION-SCALE MODELS FOR A CONDITIONALLY NORMALLY DISTRIBUTED OUTCOME FOR REPEATED MEASURES DATA

Chapter 1 presented the theoretical framework for the mixed-effects location-scale model and chapter 2 illustrated a complete example of the model alongside a discussion of relevant concepts pertaining to the model building process. In this chapter, the estimation of the mixed-effects location-scale model is presented, beginning with an introduction to current software available for estimation and its limitations. This discussion is followed by an overview of Bayes theorem, Markov chain Monte Carlo (MCMC) methods including estimation and convergence, and concludes with the technical details of the estimation algorithm used to estimate the mixed-effects location-scale models in this dissertation.

## Current Software to Estimate the Mixed-Effects Location-Scale Model

Most well-known, commercial statistical software packages (e.g., HLM, MLwiN, Mplus, SAS, SPSS, Stata) have the capability to estimate and predict heterogeneous variances in both $\mathbf{G}_i$ and $\mathbf{R}_i$. However, their utility is limited when attempting to estimate *scale-model* random effects. What follows below is a brief discussion of available software that uses likelihood-based methods in a frequentist framework to estimate the mixed-effects location-scale model. This is followed by a brief discussion of available software that uses the more flexible and powerful MCMC estimation algorithms.

**Software using maximum likelihood or restricted maximum likelihood estimation.** HLM, MLwiN, Mplus, SAS, SPSS, and Stata use either maximum likelihood (ML) or restricted/residual maximum likelihood (REML) estimation (see Hartley & Rao,

1967 for ML, Patterson & Thompson, 1971 for REML) to estimate heterogeneous variances in both $\mathbf{G}_i$ and $\mathbf{R}_i$. ML estimates (known as MLEs) have been shown to be asymptotically consistent, asymptotically normal, and efficient, which means that as sample size increases, the estimated values converge onto their population values, the estimates converge onto a normal distribution, and that no other estimator produces smaller standard errors, respectively (Harville, 1977).

The primary purpose of ML is to obtain a set of parameters that maximize the (log) likelihood function, $f(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{Z}_i) \sim \mathbf{N}_{n_i}(\mathbf{X}_i\boldsymbol{B}, \mathbf{V}_i)$, by estimating location-model fixed effects and variance components in a single model. Although ML performs well with large sample sizes, estimates of variance components will be downwardly biased because ML does not account for the estimation of the location-model fixed effects (Harville, 1977; Patterson & Thompson, 1974). For this reason, REML estimation has become much more of a gold standard in the frequentist framework. REML is used to produce less biased variance estimates by maximizing the (log) likelihood *of the residuals* and accounting for the uncertainty from the estimation of location-model fixed effects (Patterson & Thompson, 1971).

With that said, of the software mentioned above using likelihood-based methods, only the NLMIXED procedure within SAS software (SAS Institute Inc., 2011) can estimate scale-model random effects (note, NLMIXED only implements ML). With that said, although NLMIXED can theoretically estimate any combination of location- and scale-model random effects, NLMIXED typically iterates slowly, often needs precise starting values to converge, and may require more computational memory than is available (Hedeker et al., 2008). As a result, Hedeker and Nordgren (2013) developed

their MIXREGLS software specifically to estimate the mixed-effects location-scale model for conditionally normally distributed response variables (also using only ML). Although MIXREGLS is a free, stand-alone program that can also be called directly within SAS, R, or Stata (see Hedeker & Nordgren, 2013 for SAS and R and Leckie, 2014 for Stata), it has limited flexibility given it only estimates the location- and scale-model random intercept variance (i.e. no location- or scale-model random slopes). As a result of these limitations, a more powerful and flexible estimation method must be employed.

**Software that uses Markov chain Monte Carlo estimation.** In contrast to software using ML, Bayesian methods using MCMC sampling algorithms can provide researchers with the ability to estimate complex mixed-effects location-scale models (i.e., that produced non-positive definite $\mathbf{G}_i$ or convergence errors within NLMIXED or that could not be estimated in MIXREGLS). Two commonly used MCMC algorithms include Metropolis-Hastings (Metropolis & Ulam, 1949; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) and the Gibbs sampler (Geman & Geman, 1984).

The e-Stat estimation engine within the Stat-JR statistics add-on for the MLwiN software package (Browne, Charlton, Michaelides, Parker, Cameron, Szmaragd, et al., 2013; Charlton, Michaelides, Parker, Cameron, Szmaragd, Yang, et al., 2013) allows the estimation of multiple location- and scale-model random effects using either Metropolis-Hastings or the Gibbs sampler. An example using Stat-JR to estimate the mixed-effects location-scale model in an educational research setting using cross-sectional data has been provided by Leckie et al. (2014). In addition, WinBUGS and JAGS syntax has been presented to estimate a mixed-effects location-scale model, syntax that could be amended

to include any number of location- and/or scale-model random effects (Li et al., 2014; Lunn, Thomas, Best, & Spiegelhalter, 2000; Plummer, 2003, 2012; Rast et al., 2012).

Although Stat-JR is a flexible package available to estimate the mixed-effects location-scale model, it is only free to empirical scientists and students affiliated with an academic institution based in the United Kingdom (UK). Therefore, MLwiN may be prohibitively expensive for anyone working outside of the UK. Indeed, as of January 2015, Stat-JR was approximately $547 for a single-user license and $304 for a PhD student license. As a result, a novel MCMC estimator using the Metropolis-Hastings algorithm was developed for this dissertation using R software (R Core Team, 2013) to estimate the mixed-effects location-scale model. Notably, this algorithm allows for any number of location- and scale-model random effects. Before describing the specifics of this algorithm, however, it is important to discuss the Bayes theorem underlying MCMC estimation.

**Overview of Bayes' Theorem for Continuous Outcomes**

Bayes' theorem (Bayes, 1763) has many formulations for continuous outcomes. Constant across formulations, however, is the required use of probability density functions (i.e., continuous distributions) that allow probabilities to be calculated over a *range of values* given that the probability of observing one exact value within the distribution is zero. In general, continuous distributions are defined by $f(\cdot)$; thus, Bayes' theorem for continuous events is shown in (3.1).

$$f(b|a) = \frac{f(a|b)f(b)}{f(a)} \tag{3.1}$$

Before providing an interpretation of the individual elements in (3.1), it is important to note that for any continuous distribution to be a proper probability density function, it

must satisfy the *non-negativity rule* and *normalization rule*. That is, the distribution must be non-negative, $f(\cdot) \geq 0$, and must integrate to 1, $\int_{-\infty}^{\infty} f(\cdot)d\cdot = 1$, respectively (see Lynch, 2007).

Further, applying the *law of total probability* to (3.1), where $f(a) = \int f(a|b)f(b)db$, Bayes' theorem can be re-formulated as shown in (3.2), which has a form similar to that published originally by Bayes (1763). Note that because continuous distributions are density functions, the integral must be taken across the sample space of the distribution.

$$f(b|a) = \frac{f(a|b)f(b)}{\int_{-\infty}^{\infty} f(a|b)f(b)db} \tag{3.2}$$

Here, $f(b|a)$ is the conditional posterior distribution of observing $b$ given $a$, $f(a|b)$ is the conditional distribution of observing $a$ given $b$ (aka, the *likelihood* of the data), $f(b)$ is the distribution of the prior, and $\int_{-\infty}^{\infty} f(a|b)f(b)db$ is the marginal distribution or normalizing constant. Further, by the normalization rule, $\int_{-\infty}^{\infty} f(a|b)f(b)db = 1$. Thus, Bayes' theorem can be additionally re-formulated as shown in (3.3), where $\propto$ indicates *proportional to*.

$$f(b|a) \propto f(a|b)f(b) \tag{3.3}$$

Here, the posterior conditional distribution is proportional to the likelihood of the data multiplied by the prior distribution. Note that for the remainder of this dissertation, $f(a|b)$ will be referred to as the *likelihood of the data* and $f(b)$ will be referred to as the *likelihood of the prior*.

Because it is often difficult to calculate the normalizing constant due to the non-identifiable form of the posterior distribution and the fact that multiple integrals may be

involved, the proportional relationship in (3.3) serves as the basis for the sampling procedures that underlie MCMC methods (Rupp, Dey, & Zumbo, 2004). Further, because MCMC methods are used to sample observations with respect to the posterior distribution, by sampling enough observations, the characteristics of the posterior distribution (e.g., mean and variance) can be approximated, which serves to approximate the location- and scale-model fixed and random effects of interest. These methods are where the discussion turns next.

## Markov Chain Monte Carlo (MCMC) Methods

The MCMC process was first described by mathematical physicists Metropolis and Ulam (1949) and Metropolis et al. (1953) who needed a method to integrate complex functions. This section will begin with a brief discussion of the theory of Markov chains as well as Monte Carlo methods followed by details of the MCMC process that includes prior specification, estimation, autocorrelation reduction, convergence, and the evaluation of model fit.

**Markov chain theory and Monte Carlo methods.** Markov chains use probability theory and operate on quasi-random processes that create sequentially autocorrelated values. A Markov chain is a sequence of $T$ random states (e.g., $\kappa_1$, $\kappa_2$, $\kappa_3$, $\cdots$, $\kappa_T$) and begins with an event at a specific point, with all outcomes within the event referred to as the sample space (Bolstad, 2007). The chain is defined by its *transition probability* (aka, *transition kernel*), which is the probability that the current state, $\kappa$, moves to another state, $\kappa^*$, in a single step (Chib & Greenberg, 1995). A transition probability can be thought of as a conditional probability defined as the probability of moving to a particular next state given the current state, $p(\kappa^*|\kappa)$. A

random event is a Markov process if the transition probabilities between different events in the sample space depend only on the event's current state (Walsh, 2004). Thus, a subsequent event, $\kappa^*$, should only be predicted by the current event, $\kappa$. Note that because the future event is predicted from the first event, they are necessarily correlated—a more specific term is *autocorrelated*.

As a simplistic example, the Markov process beings with an arbitrarily chosen starting value within the parameter's sample space. Then, a second value is chosen dependent on the first value. Subsequently, a third value in the chain is chosen that is dependent on second, but not on the first. Further, a fourth value is chosen that is dependent on the third, but not the first or second, and so on. This process is repeated for a finite number of samples or iterations (Gallager, 1996). That is, the Markov process will continue to sample values until the transition duration (i.e., number of samples or iterations) is satisfied and the process is consistently sampling from the posterior distribution of interest known as the *stationary distribution*. The stationary distribution is identified when the predicted state is independent of the original state (i.e., defined by an autocorrelation equal to 0; Walsh, 2004). However, convergence is truly unknown in practice and can only be evidenced, not proven, using a variety of statistical and graphical methods described below.

Finally, a Monte Carlo method is a broad term that describes a computer simulation approach to solving problems that employ a sequence of randomly generated numbers (Ertekin & Grossman, 2008). The original purpose of the Monte Carlo method was to compute complex integrals (Walsh, 2004).

**The specifics of MCMC estimation.** Whenever the functional form of the posterior distribution is unknown or presents computational difficulty, MCMC methods are often employed (Rupp et al., 2004). MCMC methods approximate the posterior distribution based on Bayes theorem in (3.3) via Monte Carlo computer simulation using the theory of Markov chains to randomly sample from some posterior distribution of interest. In general, MCMC simulations begin with initial parameter estimates defined by the researcher or sampled from the appropriate posterior distribution, and then successively sample values depending on the values previously sampled (Kim & Bolt, 2007; Rupp et al., 2004). The MCMC process is considered complete when estimates converge to a stationary posterior distribution.

**Specifying prior distributions.** The prior distribution plays a fundamental role in Bayesian statistics and MCMC estimation. The characteristics of prior distributions (e.g., the mean and variance for a normal distribution) are typically specified based on one of three goals: 1) to describe existing knowledge, 2) to describe belief in the absence of existing knowledge, and 3) to have the prior contribute little, if any, information (Pullenayegum & Thabane, 2009).

Prior specification is required for all model parameters, thus allowing known information to be incorporated into model estimation (Kim & Bolt, 2007). Because selection of a prior distribution is subjective, the characteristics of all prior distributions must be explicitly delineated before analysis is conducted. Further, it is important to note that misspecified prior distributions are a nonissue as long as the posterior distribution is a proper density function (i.e., it satisfies the non-negativity and normalization rules; Rupp et al., 2004).

Finally, the prior distribution has the potential to increase or decrease the influence of the observed data, with the strength of the prior controlled through *hyperparameters* that describe the distribution of the prior (e.g., the variance of the prior distribution; Kim & Bolt, 2007). Strong priors that reduce the influence of the data are termed *informative priors* because they are specified to have small variances. By contrast, highly *noninformative* (or *vague*) *priors* have large variances that allows the collected data to have as much influence on the posterior distribution as possible. However, note that prior specification can be problematic, as theoretical justification for all priors is subjective, and even the use of noninformative priors could contribute more (or less) information than expected, which can over- (or under-) power the influence of your data (known as the *prior-data conflict*, Evans & Moshonov, 2006).

**The Metropolis-Hastings algorithm and the Gibbs sampler.** The Metropolis algorithm was developed originally by mathematical physicists Metropolis and Ulam (1949) and Metropolis et al. (1953) who were having difficulty integrating complex functions. This algorithm was subsequently generalized by Hastings (1970) using an arbitrary transition probability function, defined as the probability that a current state moves to a subsequent state given the current state.

In general, the Metropolis-Hastings (MH) algorithm is known as a *rejection sampler* (Rupp et al., 2004). The algorithm begins with an initial value, $\theta$, that is selected from the sample space of the prior distribution and represents the current value. Then, the algorithm randomly samples a candidate value, $\theta^*$, from an appropriately specified jumping (or, *candidate-generating*) distribution, $Q(\theta^*|\theta)$, defined as the probability (or likelihood) of selecting the candidate value, $\theta^*$, given the current value, $\theta$ (Chib &

Greenberg, 1995). Next, a MH ratio of the densities ($r_{\mathrm{MH},\theta}$) is calculated based on the candidate and current values to determine when a candidate value should be accepted or rejected. A general form of this ratio using a Bayesian framework to predict parameter $\theta$ is shown in (3.4).

$$r_{\mathrm{MH},\theta} = \frac{f(\mathbf{Y}_i|\theta^*)f(\theta^*)Q(\theta|\theta^*)}{f(\mathbf{Y}_i|\theta)f(\theta)Q(\theta^*|\theta)} \tag{3.4}$$

Here, $\theta$ is the current value of the parameter, $\theta^*$ is the candidate value of the parameter, and $\mathbf{Y}_i$ is the $n_i$ x 1 column vector of observed outcome data for individual $i$. The numerator of the ratio consists of $f(\mathbf{Y}_i|\theta^*)$, which is the likelihood of the observed data, $\mathbf{Y}_i$, given the candidate value, $\theta^*$, $f(\theta^*)$ is the likelihood of the prior distribution for the candidate value, $\theta^*$, and $Q(\theta|\theta^*)$ is the likelihood of drawing the current value, $\theta$, given the candidate value, $\theta^*$ (i.e., the candidate-generating distribution). Similarly, the denominator consists of $f(\mathbf{Y}_i|\theta)$ which is the likelihood of the observed data, $\mathbf{Y}_i$, given the current value, $\theta$, $f(\theta)$ is the likelihood of the prior distribution for the current value, $\theta$, and $Q(\theta^*|\theta)$ is the likelihood of drawing the candidate value, $\theta^*$, given the current value, $\theta$ (again, the candidate-generating distribution). Note that $f(\mathbf{Y}_i|\theta^*)$ and $f(\mathbf{Y}_i|\theta)$ may be required to be summed or multiplied over individuals depending on whether likelihoods are on the log scale or not. Further, when the candidate-generating distributions are symmetric, $Q(\theta|\theta^*) = Q(\theta^*|\theta)$, these distributions can be factored from the MH ratio (Patz & Junker, 1999).

Accepting or rejecting the candidate value is determined with probability (3.5),

$$min(r_{\mathrm{MH}}, 1) \tag{3.5}$$

where $r_{\mathrm{MH}}$ was defined in (3.4). This accept–reject process continues for a large number of $T$ iterations to form the Markov chain and is considered complete when the algorithm

samples consistently (i.e., converges) from a stationary distribution, which for Bayesian inference should approximate the posterior density function, $f(b|a)$ (Patz & Junker, 1999).

Finally, the Gibbs sampler was introduced in the context of image sampling by Geman and Geman (1984). The Gibbs sampler is best viewed as a special case of the MH algorithm in which all candidate values are accepted unconditionally (Rupp et al., 2004). The Gibbs sampling process begins with an initial value, $\theta$, and then samples iteratively from univariate conditional distributions that include both the data as well as all previously and subsequently sampled values. This sequence is continued until the Gibbs sampler converges onto a stationary distribution, which is ideally equal to the posterior distribution. It is important to note that using the Gibbs sampler requires computing the normalizing constant defined in (3.2), which can be a difficult task for complex integrals; using a rejection-sampling algorithm, such as the MH algorithm, circumvents the need for these calculations (Patz & Junker, 1999).

**The tuning of candidate-generating distributions.** It is important to note that the MH algorithm has lower efficiency (i.e., longer convergence times) with candidate acceptance rates below 15% or above 50% (Roberts & Rosenthal, 2001). Further, in situations where posterior and candidate distributions are normal (as in this dissertation), Roberts, Gelman, and Gilks (1997) recommend a 45% acceptance rate to achieve the greatest efficiency (i.e., the least number of iterations) of the MH algorithm. Appropriate acceptance rates can be achieved by *tuning* the variances of candidate distributions. That is, a series of preliminary iterations are used to evaluate the acceptance rate and adjust the variance of candidate distributions accordingly. Iterations contribute to the Markov chain

only after tuning is complete (Chib & Greenberg, 1995). Tuning of candidate variances is conducted as shown in (3.6).

$$\sigma^2_{\theta^*_{\text{new}}} = \frac{\sigma^2_\theta \cdot \phi^{-1}(p_{\text{optimal}}/2)}{\phi^{-1}(p_{\text{current}}/2)} \tag{3.6}$$

Here, $\sigma^2_{\theta^*_{\text{new}}}$ is the new variance of the candidate-generating distribution for the parameter of interest, $\sigma^2_\theta$ is the variance of the current candidate-generating distribution, $\phi^{-1}$ indicates the inverse cumulative density function of the standard normal distribution, $p_{\text{optimal}}$ is the optimal acceptance rate, and $p_{\text{current}}$ is the current acceptance rate.

**Burn-in period and thinning.** As stated above, the stationary distribution is obtained when the sampled values are independent of the starting state (i.e., no autocorrelation). By definition, independence implies no relationship or correlation between states; however, by nature of the sequential process of a Markov chain, adjacent values in the chain will share some positive and problematic autocorrelation. Thus, it has been recommended that posterior estimates based on a small number of sampled states not be trusted because the initial states will inevitably be influenced by their autocorrelation with the starting state (Kim & Bolt, 2007). Therefore, common practice is to dismiss a number of initial sampled states, often referred to as the *burn-in period*; posterior estimates are based on the sampled states following burn-in. There is no consensus on the number of initial sampled states to burn, although estimates typically range from the first 500 to 5000 values depending on the model, the data, and the structure of the algorithm itself (Kim & Bolt, 2007). Note that Raftery and Lewis (1992) suggested using an empirical method where the required burn-in period is the number of states necessary to achieve an estimated autocorrelation of 0.

Another strategy for reducing autocorrelation is to use *thinning*, where, for example, only the tenth value in the Markov chain is evaluated (Kim & Bolt, 2007). As the thinning distance is increased, the autocorrelation should decrease. Note that when thinning is used, the length of the Markov chain must be increased dramatically as a result of a drastic decrease in thinned sampled states. That is, if considering every tenth sampled state from an MCMC chain using 1000 iterations, only 100 (i.e., 1000/10) sampled states will be considered in estimates of the posterior distribution.

**Determining convergence of the Markov chain.** Assuming a proper posterior density, the MCMC algorithm should converge onto the stationary distribution given enough iterations; however, the MH algorithm will continue sampling candidate values continuously until the Markov chain is terminated by the researcher (Spiegelhalter, Thomas, Best, & Lunn, 2003). Thus, determining convergence can be a subjective process and is always completed 100% post hoc. Several empirical methods for determining convergence have been developed, but each is not without limitations. Two statistical methods, one developed by Geweke (1992) and the other developed by Gelman and Rubin (1992) are often used (Cowles & Carlin, 1996). Graphical methods have also been developed, in which parameter estimates are graphed against the iteration number.

Geweke (1992) developed a diagnostic test that compares the parameter mean across a set of earlier sampled states to the mean across later sampled states; convergence is indicated when the two means are similar. It is suggested that, for each parameter in the model, the mean of the first 10% of iterations (after burn-in) be compared to the last 50% of iterations. The Geweke test statistic is the difference of the two means divided by the estimated standard error (similar to a $z$-test). For a given parameter, a statistically

significant mean difference (i.e., $z < -1.96$ or $z > 1.96$) provides evidence that the chain has not converged. Note, however, that the standard error is largely affected by the number of iterations; therefore, Type I errors are possible, resulting in converged chains that have significantly different means.

Alternatively, the Gelman and Rubin criterion compares the variances within and between multiple chains for each parameter (with divergent starting values) and calculates a variance ratio statistic, $\hat{R}$, for each parameter similar to an $F$-test in analysis of variance (Gelman & Rubin, 1992). Convergence is likely when $\hat{R} = 1$; however, Gelman and Hill (2007) suggested using $\hat{R} \leq 1.5$ as a cut-off for convergence. Although designed for multiple chains, a single chain can be partitioned to create multiple chains, at which point the Gelman and Rubin criterion is evaluated (Plummer, Best, Cowles, & Vines, 2006). Criticisms of this criterion include that it is inefficient, relies too heavily on the researcher to specify an exact target distribution, and that a single chain with adequate burn-in may be more efficient than multiple chains (Cowles & Carlin, 1996).

Convergence can be evidenced graphically by examining *trace plots* (aka, *sampling history plots*). Examples of two trace plots are presented in Figure 3.1. Convergence is suggested by trace plots that level off (or snake) around the mean of the parameter estimate. Thus, the top plot of Figure 3.1 serves as an example of probable convergence, whereas the bottom plot of Figure 3.1 serves as an example of probable non-convergence. Note that convergence is defined by a *range* of sampled estimates, not by a single estimate. Thus, a narrower range equates to less variability, which is more indicative of convergence (Kass, Carlin, Gelman, & Neal, 1998). Further, problematic estimates can also be identified quickly by multi-modal histograms of sampled values.

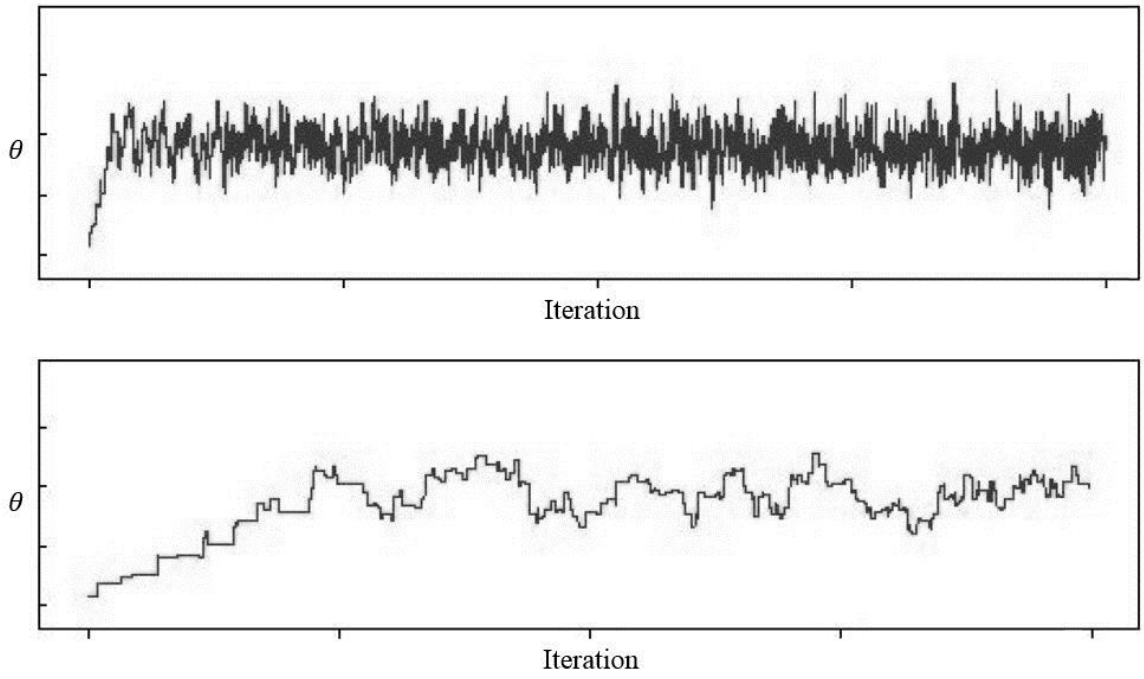


*Figure 3.1*. Trace plots for parameter $\theta$ showing convergence and non-convergence

It is important to note that in many situations the trace plot may not converge around an estimate as quickly as shown in top plot of Figure 3.1. In fact, convergence may not occur during a set number of iterations. Several factors directly affecting convergence include the selected prior distributions, the likelihood of the data, and/or initial estimates (Pullenayegum & Thabane, 2009). Further, Kim and Bolt (2007) indicated the algorithm employed might also affect the rate of convergence. For example, if the MH algorithm frequently rejects candidate values, the rate of convergence will inevitably be slower. In these situations, the length of the chain may need to be increased considerably.

**Intervals, significance, and model comparison.** In a Bayesian analysis and MCMC estimation, each parameter is sampled from a distribution (i.e., all parameters are considered random). Therefore, when convergence evidence is strong, the estimate for a given parameter is determined by summary statistics of the posterior distribution of the

Markov chain. More specifically, the estimate for a given parameter is often defined by the mean or the median of the posterior distribution (Chib & Greenberg, 1995). Because this estimate is relatively uninformative in isolation, Bayesian confidence intervals can be placed around the estimate using several available methods. Two of the more common intervals include a symmetric *credible interval*, in which the 95% interval is based on the 2.5 and 97.5 percentiles of the posterior distribution for the parameter of interest, or the *highest posterior density* interval, which captures 95% of the posterior distribution regardless of symmetry (Gill, 2008). Note that although the highest posterior density interval is more flexible, for unimodal distributions that are approximately symmetric, the two intervals will be similar (Gill, 2008).

Although Bayesian *p*-values have been proposed (see Gelman, Meng, & Stern, 1996), traditionally the significance of a given parameter is indicated by an credible or HPD interval that excludes zero. Further, Bayesian hypothesis testing can be conducted by evaluating whether the null value (which may or may not be zero) lies outside the interval.

With that said, there may be occasions when evaluating significance via any interval is theoretically inappropriate, such as when evaluating the significance of location- or scale-model variance components. In these situations, overall model comparisons are required. In the context of MCMC estimation, several are available, of which the Bayes factor (BF), pseudo-Bayes factor, and deviance information criterion (DIC) will be discussed.

The BF compares two models using a ratio of the marginal likelihoods of the data (i.e., the normalizing constant) from each model as shown in (3.7).

$$\text{BF} = \frac{\int_{\theta_1} f(\mathbf{Y}_i|\theta_1) f(\theta_1) d\theta_1}{\int_{\theta_2} f(\mathbf{Y}_i|\theta_2) f(\theta_2) d\theta_2} \tag{3.7}$$

A BF ≥ 1 indicates model 1 is preferred with the extent to which BF < 1 providing evidence against model 1 (Gill, 2008). Although BF comparisons are straightforward to interpret, their computation is difficult in practice given the difficulty in calculating the normalizing constant (Patz & Junker, 1999). Thus, the BF comparison is often approximated using a pseudo-Bayes factor called *conditional predictive ordinate* (CPO; Gelfand, Dey, & Chang, 1992). The CPO reduces computational complexity by only requiring the likelihood of the observed data given the current parameter, $f(\mathbf{Y}_i|\theta)$, as shown in (3.8)

$$\text{CPO}^{-1} = \frac{1}{T}\sum_{1}^{T} \frac{1}{f(\mathbf{Y}_i|\theta)}, \tag{3.8}$$

where $T$ is the total number of sampled values in the chain for parameter $\theta$.

One final model comparison is the deviance information criterion (DIC), shown in (3.9). The DIC attempts to identify the most parsimonious model by weighing model fit against model complexity and is similar conceptually to AIC or BIC from the frequentist framework; thus, the better fitting model has a lower DIC.

$$\text{DIC} = -2\log\big(f(\mathbf{Y}_i|\theta)\big) + 2p_D \tag{3.9}$$

Here, $-2\log\big(f(\mathbf{Y}_i|\theta)\big)$ is the –2 log-likelihood of the data given the model parameters (i.e., the deviance), and $p_D$ is the effective number of parameters that quantifies the penalty for complexity by correcting the natural bias of the deviance to prefer the model with more parameters, as shown in (3.10)

$$p_D = \overline{D(\theta)} - D\big(\overline{\theta}\big), \tag{3.10}$$

where $\overline{D(\theta)}$ is the sum of the posterior mean of the deviances as shown in (3.11)

$$\overline{D(\theta)} = \frac{1}{T}\sum_{1}^{T} -2\,log\big(f(\mathbf{Y}_i|\theta_{\text{current}})\big), \tag{3.11}$$

and $D\big(\overline{\theta}\big)$ is the deviance evaluated at the posterior mean of the parameter, $\overline{\theta}$, as shown in (3.12).

$$D\big(\overline{\theta}\big) = -2\,log\left(f\big(\mathbf{Y}_i|\overline{\theta}\big)\right) \tag{3.12}$$

**Concluding remarks.** This section presented specifics regarding how Bayes' theorem is used throughout MCMC estimation using the MH algorithm. Given the flexibility of the MH algorithm, while at the same time considering the limitations of ML or REML estimation for complex models described above, estimating the mixed-effects location-scale model using the MH algorithm provides obvious benefits. Therefore, the next section fully describes the MH algorithm used to estimate all mixed-effects location-scale models in this dissertation, providing details regarding the likelihood functions for the observed data as well as specifics of all prior and candidate-generating distributions.

## The Metropolis-Hastings Algorithm to Estimate the Mixed-Effects Location-Scale Model

To begin the MH estimation algorithm, starting values for location-model fixed and random effects as well as the residual variance were identified using a linear mixed-effects model; starting values for scale-model fixed and random effects were set to zero. Below are the steps of the MH algorithm defining the value of $r_{\text{MH}}$ from (3.4) using the distributions from which candidate values were drawn as well as the distributions of the prior for each parameter. Prior to initiation of the chain, candidate variances for all

parameters were tuned using (3.6) to approximate an optimal acceptance rate of 45% (Roberts et al., 1997).

Each step of the algorithm is represented below by each subsequent equation, where all parameters not being considered are held constant at their *current values*. The distributions are consolidated so that the parameter being considered in a given step within a given iteration is the only conditional value. Further, note that subscript $k$ is used for all model parameters as a generic index to indicate a given element within the specific parameter vector of interest, and that, when relevant, notation for all effects is mapped directly onto the multi-level, scalar notation used throughout chapter 2 (i.e., elements of $\boldsymbol{B} = \gamma_k$; elements of $\boldsymbol{\tau}^e = \delta_k^e$). Thus, as an example, although the distribution of the data given the entire mixed-effects location-scale model is given by $f(\mathbf{Y}_i|\boldsymbol{B}, \mathbf{u}_i)$, if only one element of $\boldsymbol{B}$, say $\gamma_k$, is being evaluated, then its distribution is denoted as $f(\mathbf{Y}_i|\gamma_k)$ to simplify the text.

**Location-model fixed effects.** Using the multi-level, scalar notation provided in chapter 2, the elements of $\boldsymbol{B}$ are represented by $\gamma_k$, as shown initially in (2.2). Thus, location-model fixed effects are the elements of the $p$ x 1 column vector $\boldsymbol{B}$, with individual location-model fixed effects within this vector denoted as $\gamma_k$, where $k$ = 0 to $p$. The values of the $\gamma_k$ parameters were updated individually using the MH ratio described generically in (3.4), and modified specifically for location-model fixed effects as shown in (3.13).

$$r_{\mathrm{MH},\gamma} = \frac{f(\mathbf{Y}_i|\gamma_k^*)f(\gamma_k^*)Q(\gamma_k|\gamma_k^*)}{f(\mathbf{Y}_i|\gamma_k)f(\gamma_k)Q(\gamma_k^*|\gamma_k)} \tag{3.13}$$

Here, $f(\mathbf{Y}_i|\gamma_k^*)$ is the likelihood of the observed data for individual $i$, $\mathbf{Y}_i$, given the candidate value of the $k^{\text{th}}$ location-model fixed effect, $\gamma_k^*$. The likelihood of the observed

data was calculated as the product of the individual likelihoods across all individuals' $t$ repeated occasions using the multivariate normal distribution shown in (3.14).

$$f(\mathbf{Y}_i|\gamma_k^*) = \prod_1^N \left( \frac{1}{(2\pi)^{\frac{t}{2}}|\mathbf{V}_i|^{\frac{1}{2}}} e^{-(\mathbf{Y}_i-(\mathbf{X}_i\boldsymbol{B}+\mathbf{Z}_i\mathbf{u}_i))^T \mathbf{V}_i^{-1}((\mathbf{Y}_i-(\mathbf{X}_i\boldsymbol{B}+\mathbf{Z}_i\mathbf{u}_i))/2)} \right) \quad (3.14)$$

Here, $\mathbf{X}_i\boldsymbol{B} + \mathbf{Z}_i\mathbf{u}_i$ is the location-model mean vector and $\mathbf{V}_i$ is the covariance matrix for the multivariate normal distribution of $\mathbf{Y}_i$ conditional on $\mathbf{X}_i$ and $\mathbf{Z}_i$ for individual $i$, $f(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{Z}_i) \sim \mathbf{N}_{n_i}(\mathbf{X}_i\boldsymbol{B}, \mathbf{V}_i)$, as defined by (1.5).

In addition, $f(\gamma_k^*)$ is the likelihood of the prior distribution for the candidate value of the $k^{\text{th}}$ location-model fixed effect, $\gamma_k^*$. An uninformative prior was used for all $\gamma_k^*$, sampled from the univariate normal distribution shown in (3.15).

$$f(\gamma_k^*) \sim N(0, 10000) \quad (3.15)$$

Finally, $Q(\gamma_k|\gamma_k^*)$ represents the candidate-generating distribution and is the likelihood of drawing the current value of the $k^{\text{th}}$ location-model fixed effect, $\gamma_k$, given the candidate value of the $k^{\text{th}}$ location-model fixed effect, $\gamma_k^*$. All location-model fixed effects were drawn from the univariate normal distribution shown in (3.16)

$$Q(\gamma_k|\gamma_k^*) \sim N\left(\gamma_k^*, \sigma_{\gamma_k}^2\right), \quad (3.16)$$

with the candidate-generating variance for the $k^{\text{th}}$ location-model fixed effect, $\sigma_{\gamma_k}^2$, tuned to achieve candidate acceptance rates of 45% as described above.

Calculating the denominator of $r_{\text{MH},\gamma}$ followed a similar process. Here, the likelihood of the observed data for individual $i$ given the current value of the $k^{\text{th}}$ location-model fixed effect, $f(\mathbf{Y}_i|\gamma_k)$, the likelihood of the prior distribution for the current value of the $k^{\text{th}}$ location-model fixed effect, $f(\gamma_k)$, and the candidate-generating

distribution, $Q(\gamma_k^*|\gamma_k)$, each were obtained by substituting $\gamma_k$ for $\gamma_k^*$ (and vice versa, as necessary) in (3.14), (3.15), and (3.16), respectively.

**Location-model random effects.** Location-model random effects are the elements of the $q$ x 1 column vector $\mathbf{u}_i$, with location-model random effect values for individual $i$ denoted as $u_{k,i}$, where $k$ = 0 to $q$. Values of $u_{k,i}$ were updated individually for each individual $i$ using the MH ratio shown in (3.17).

$$r_{\mathrm{MH},u_i} = \frac{f(\mathbf{Y}_i|u_{k,i}^*)f(u_{k,i}^*)Q(u_{k,i}|u_{k,i}^*)}{f(\mathbf{Y}_i|u_{k,i})f(u_{k,i})Q(u_{k,i}^*|u_{k,i})} \tag{3.17}$$

Here, $f(\mathbf{Y}_i|u_{k,i}^*)$ is the likelihood of the observed data for individual $i$, $\mathbf{Y}_i$, given the candidate value of the $k^{\text{th}}$ location-model random effect for individual $i$, $u_{k,i}^*$. The likelihood for each individual $i$, $f(\mathbf{Y}_i|u_{k,i}^*)$, was calculated using the same multivariate normal distribution as shown on the right side of (3.14), where $u_{k,i}^*$ was used to form $\mathbf{u}_i$.

In addition, $f(u_{k,i}^*)$ represents the likelihood of the prior distribution for the candidate value of the $k^{\text{th}}$ location-model random effect for individual $i$, $u_{k,i}^*$, which was sampled from the univariate normal distribution shown in (3.18)

$$f(u_{k,i}^*) \sim N(0, \sigma_{u_{k,i}}^2), \tag{3.18}$$

where $\sigma_{u_{k,i}}^2$ is the variance of the $k^{\text{th}}$ location-model random effect for individual $i$ obtained from current $\mathbf{G}_i$.

Finally, $Q(u_{k,i}|u_{k,i}^*)$ represents the candidate-generating distribution and is the likelihood of drawing the current value of the $k^{\text{th}}$ location-model random effect for individual $i$, $u_{k,i}$, given the candidate value of the $k^{\text{th}}$ location-model random effect for

individual $i$, $u^*_{k,i}$. All location-model random effects were sampled from the univariate normal distribution shown in (3.19)

$$Q\left(u_{k,i}|u^*_{k,i}\right) \sim N\left(u^*_{k,i}, \sigma^2_{u_{k,i}}\right), \tag{3.19}$$

with the candidate-generating variance for the $k^{\text{th}}$ location-model random effect for individual $i$, $\sigma^2_{u_{k,i}}$, tuned to achieve candidate acceptance rates of 45%.

Calculating the denominator of $r_{\text{MH},u_i}$ followed a similar process. Specifically, the likelihood of the observed data for individual $i$ given the current value of the $k^{\text{th}}$ location-model random effect for individual $i$, $f\left(\mathbf{Y}_i|u_{k,i}\right)$, the likelihood of the prior distribution for the current value of the $k^{\text{th}}$ location-model random effect for individual $i$, $f\left(u_{k,i}\right)$, and the candidate-generating distribution, $Q\left(u^*_{k,i}|u_{k,i}\right)$, each were obtained by substituting $u_{k,i}$ for $u^*_{k,i}$ (and vice versa, as necessary) in (3.14), (3.18), and (3.19), respectively.

**Scale-model fixed effects for level-2 variance components.** Scale-model fixed effects for level-2 variance components in $\mathbf{G}_i$ are the elements of the $a^{u_r}$ x 1 column vector $\boldsymbol{\alpha}^{u_r}$, with individual scale-model fixed effects for level-2 variance components denoted by $\alpha_k^{u_r}$, where $k$ = 0 to $a^{u_r}$. Note that $u_r$ includes both location- and scale-model random intercept variances in $\mathbf{G}_i$. Values of $\alpha_k^{u_r}$ were estimated on the log scale and updated individually using the MH ratio presented in (3.20).

$$r_{\text{MH},\alpha^{u_r}} = \frac{f\left(\mathbf{m}_i|\alpha_k^{u_r{}^*}\right)f\left(\alpha_k^{u_r{}^*}\right)Q\left(\alpha_k^{u_r}|\alpha_k^{u_r{}^*}\right)}{f\left(\mathbf{m}_i|\alpha_k^{u_r}\right)f\left(\alpha_k^{u_r}\right)Q\left(\alpha_k^{u_r{}^*}|\alpha_k^{u_r}\right)} \tag{3.20}$$

As described in chapter 1, $\mathbf{m}_i$ is a $m_{\mathbf{u}_i+\boldsymbol{\omega}_i^{et}}$ x 1 column vector holding all location- and scale-model random effects. Thus, $f\left(\mathbf{m}_i|\alpha_k^{u_r{}^*}\right)$ is the likelihood of the *deviations* from

the $k^{\text{th}}$ location-model fixed effect, or $k^{\text{th}}$ fixed effect included in the scale model for the residual variance, for individual $i$, $\mathbf{m}_i$, given candidate $\mathbf{G}_i$, $\mathbf{G}_i^*$, for individual $i$ calculated from the (exponentiated) log-linear combination of $k$ candidate level-2, scale-model fixed effects, $\alpha_k^{u_r{}^*}$, as defined in (1.9). The likelihood for individual $i$ was calculated using the multivariate normal distribution shown in (3.21), with a mean (vector) of zero and variances and covariances indicated by $\mathbf{G}_i^*$.

$$f\left(\mathbf{m}_i|\alpha_k^{u_r{}^*}\right) = \prod_{1}^{N}\left(\frac{1}{(2\pi)^{\frac{t}{2}}\left|\mathbf{G}_i^*\right|^{\frac{1}{2}}} e^{-(\mathbf{m}_i-\mathbf{0})^T\mathbf{G}_i^{*-1}((\mathbf{m}_i-\mathbf{0})/2)}\right) \tag{3.21}$$

In addition, $f\left(\alpha_k^{u_r{}^*}\right)$ is the likelihood of the prior density for the candidate value of the $k^{\text{th}}$ level-2, scale-model fixed effect, $\alpha_k^{u_r{}^*}$. An uninformative prior was used for all $\alpha_k^{u_r{}^*}$, which were sampled from the univariate normal distribution shown in (3.22).

$$f\left(\alpha_k^{u_r{}^*}\right) \sim N(0, 10000) \tag{3.22}$$

Further, $Q\left(\alpha_k^{u_r}|\alpha_k^{u_r{}^*}\right)$ represents the candidate-generating distribution and is the likelihood of drawing the current value for the $k^{\text{th}}$ level-2, scale-model fixed effect, $\alpha_k^{u_r}$, given the candidate value of the $k^{\text{th}}$ level-2, scale-model fixed effect, $\alpha_k^{u_r{}^*}$. All level-2, scale-model fixed effects were sampled from a truncated normal distribution, shown in (3.23), with lower bound, $b$, and upper bound, $h$, determined before sampling using the lower-upper decomposition, to ensure $\mathbf{G}_i$ remains positive definite (see Barnard et al., 2000)

$$Q\left(\alpha_k^{u_r}|\alpha_k^{u_r{}^*}\right) = \frac{1}{\sqrt{\sigma^2_{\alpha_k^{u_r}}(Z)}}\phi(\xi), \tag{3.23}$$

where

$$\xi = \frac{\alpha_k^{u_r*} - \alpha_k^{u_r}}{\sqrt{\sigma^2_{\alpha_k^{u_r}}}} \tag{3.24}$$

$$\phi(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{\xi^2}{2}\right) \tag{3.25}$$

$$Z = \Phi(\varphi) - \Phi(\lambda) \tag{3.26}$$

$$\Phi(x) = \int_{-\infty}^{x} \phi(x)dx \tag{3.27}$$

$$\varphi = \frac{h - \alpha_k^{u_r}}{\sqrt{\sigma^2_{\alpha_k^{u_r}}}} \tag{3.28}$$

$$\lambda = \frac{b - \alpha_k^{u_r}}{\sqrt{\sigma^2_{\alpha_k^{u_r}}}}, \tag{3.29}$$

with the candidate-generating variance for the $k^{\text{th}}$ level-2, scale-model fixed effect, $\sigma^2_{\alpha_k^{u_r}}$, tuned to achieve candidate acceptance rates of 45%.

Calculating the denominator of $r_{\text{MH},\alpha_k^{u_r}}$ followed a similar process. Specifically, the likelihood of the *deviations* from the $k^{\text{th}}$ location-model fixed effect, or fixed effect included in the scale model for the residual variance, for individual $i$, $\mathbf{m}_i$, given current $\mathbf{G}_i$ for individual $i$ calculated from the (exponentiated) log-linear combination of $k$ current level-2, scale-model fixed effects, $\alpha_k^{u_r}$, $f\left(\mathbf{m}_i|\alpha_k^{u_r}\right)$, was calculated using the right-hand side of (3.21). Further, the likelihood of the prior distribution for the current value of the $k^{\text{th}}$ level-2, scale-model fixed effect, $f\left(\alpha_k^{u_r}\right)$, as well as the candidate-generating density, $Q\left(\alpha_k^{u_r*}|\alpha_k^{u_r}\right)$, was obtained by substituting $\alpha_k^{u_r}$ for $\alpha_k^{u_r*}$ (and vice versa, as necessary) using (3.22) and (3.23) through (3.29), respectively.

Finally, as stated in chapter 1, covariances in $\mathbf{G}_i$ were modeled as correlations where fixed effects are elements of the $a^{u_r;u_{r'}}$ x 1 column vector $\boldsymbol{\alpha}^{u_r;u_{r'}}$, with individual scale-model fixed effects for the correlation between level-2 variance components denoted by $\alpha_k^{u_r;u_{r'}}$. Similar to above, note that $u_r;u_{r'}$ included all correlations between location- and scale-model random effects in $\mathbf{G}_i$. Values were estimated on the inverse hyperbolic tangent scale and updated individually using the MH ratio shown in (3.30).

$$r_{\mathrm{MH},\alpha^{u_r;u_{r'}}} = \frac{f\left(\mathbf{m}_i|\alpha_k^{u_r;u_{r'}*}\right) f\left(\alpha_k^{u_r;u_{r'}*}\right) Q\left(\alpha_k^{u_r;u_{r'}}|\alpha_k^{u_r;u_{r'}*}\right)}{f\left(\mathbf{m}_i|\alpha_k^{u_r;u_{r'}}\right) f\left(\alpha_k^{u_r;u_{r'}}\right) Q\left(\alpha_k^{u_r;u_{r'}*}|\alpha_k^{u_r;u_{r'}}\right)} \tag{3.30}$$

The procedures used to estimate the fixed effects for correlation between level-2 variance components in $\mathbf{G}_i$ were identical to the procedures described above for level-2 variance components, of course, only after substituting $\alpha_k^{u_r;u_{r'}}$ for $\alpha_k^{u_r}$, $\alpha_k^{u_r;u_{r'}*}$ for $\alpha_k^{u_r*}$, and $\sigma^2_{\alpha_k^{u_r}}$ for $\sigma^2_{\alpha_k^{u_r;u_{r'}}}$ as needed in (3.24) through (3.29). Therefore, explanations and interpretations are nearly identical and are therefore not re-presented.

**Scale-model fixed effects for the residual variance.** All mixed-effects location-scale models were estimated assuming $\mathbf{R}_i$ was heterogeneous between individuals, but constant and independent within individuals, $\mathbf{R}_i = \sigma^2_{e_i}\mathbf{I}_{n_i}$. Thus, similar to the description in chapter 2, correlations between residual values in $\mathbf{R}_i$ were constrained to be zero and only one residual variance was estimated for each individual $i$. Thus, the subscript indicating occasion $t$ was not necessary.

Using the multi-level, scalar notation provided in chapter 2, the elements of $\boldsymbol{\tau}^e$ are represented by $\delta_k^e$, as shown initially in (2.8). Therefore, scale-model fixed effects for the residual variance in $\mathbf{R}_i$ are the elements of the $c^e$ x 1 column vector $\boldsymbol{\tau}^e$, with individual

scale-model fixed effects for the residual variance denoted by $\delta_k^e$, where $k$ = 0 to $c^e$. Values of $\delta_k^e$ were estimated on the log scale and updated individually using the MH ratio shown in (3.31).

$$r_{\mathrm{MH},\delta^e} = \frac{f\left(\mathbf{Y}_i|\delta_k^{e*}\right)f\left(\delta_k^{e*}\right)Q\left(\delta_k^e|\delta_k^{e*}\right)}{f(\mathbf{Y}_i|\delta_k^e)f(\delta_k^e)Q(\delta_k^{e*}|\delta_k^e)} \tag{3.31}$$

Here, $f\left(\mathbf{Y}_i|\delta_k^{e*}\right)$ is the likelihood of the observed data for individual $i$, $\mathbf{Y}_i$, given the candidate $\mathbf{R}_i$, $\mathbf{R}_i^*$, for individual $i$ calculated from the (exponentiated) log-linear combination of $k$ candidate scale-model fixed effects for the residual variance, $\delta_k^{e*}$. The likelihood for individual $i$ was calculated using the multivariate normal distribution as shown on the right-hand side of (3.14), where $\delta_k^{e*}$ was used to form candidate $\mathbf{R}_i^*$ that was used in the calculation of $\mathbf{V}_i$, as shown in (1.5).

In addition, $f\left(\delta_k^{e*}\right)$ is the likelihood of the prior distribution for the candidate value of the $k^{\mathrm{th}}$ scale-model fixed effect for the residual variance, $\delta_k^{e*}$. An uninformative prior was used for all $\delta_k^{e*}$ parameters, which were sampled from the univariate normal distribution shown in (3.32).

$$f\left(\delta_k^{e*}\right) \sim N(0, 10000) \tag{3.32}$$

Finally, $Q\left(\delta_k^e|\delta_k^{e*}\right)$ represents the candidate-generating distribution and is the likelihood of drawing the current value for the $k^{\mathrm{th}}$ scale-model fixed effect for the residual variance, $\delta_k^e$, given the candidate value for the $k^{\mathrm{th}}$ scale-model fixed effect for the residual variance, $\delta_k^{e*}$. All scale-model fixed effects for the residual variance were drawn from the univariate normal distribution shown in (3.33)

$$Q\left(\delta_k^e|\delta_k^{e*}\right) \sim N\left(\delta_k^{e*}, \sigma_{\delta_k^e}^2\right), \tag{3.33}$$

with the candidate-generating variance for the $k^{\text{th}}$ scale-model fixed effect for the residual variance, $\sigma^2_{\delta^e_k}$, tuned to achieve candidate acceptance rates of 45%.

Calculating the denominator of $r_{\text{MH},\delta^e}$ followed a similar process. Specifically, the likelihood of the observed data given current $\mathbf{R}_i$ for individual $i$ based on the scale-model fixed effects for the residual variance, $f(\mathbf{Y}_i|\delta^e_k)$, the likelihood of the prior distribution for the current value of the $k^{\text{th}}$ scale-model fixed effect for the residual variance, $f(\delta^e_k)$, and the candidate-generating distribution, $Q(\delta^{e*}_k|\delta^e_k)$, each were obtained by substituting $\delta^{e*}_k$ for $\delta^e_k$ (and vice versa, as necessary) in (3.14), (3.32), and (3.33), respectively.

**Scale-model random effects.** Scale-model random effects are elements of the $w^e$ x 1 column vector $\boldsymbol{\omega}^e_i$, with individual scale-model random effects denoted as $\omega^e_{k,i}$, where $k$ = 0 to $w^e$. Values of $\omega^e_{k,i}$ were estimated on the log scale and updated individually for each individual $i$ using the MH ratio shown in (3.34).

$$r_{\text{MH},\omega^e_i} = \frac{f(\mathbf{Y}_i|\omega^{e\,*}_{k,i})f(\omega^{e\,*}_{k,i})Q(\omega^e_{k,i}|\omega^{e\,*}_{k,i})}{f(\mathbf{Y}_i|\omega^e_{k,i})f(\omega^e_{k,i})Q(\omega^{e\,*}_{k,i}|\omega^e_{k,i})} \tag{3.34}$$

Here, $f(\mathbf{Y}_i|\omega^{e\,*}_{k,i})$ is the likelihood of the observed data for individual $i$, $\mathbf{Y}_i$, given the candidate value of the $k^{\text{th}}$ scale-model random effect for individual $i$, $\omega^{e\,*}_{k,i}$. The likelihood for each individual was calculated using the multivariate normal distribution on the right-hand side of (3.14), where $\omega^{e\,*}_{k,i}$ modified $\mathbf{G}_i$ with the addition of the $k$ scale-model random effect variance components (and additional covariances) as well as $\mathbf{R}_i$ by modifying the residual variance for a given individual, both of which were used in the calculation of $\mathbf{V}_i$, as shown in (1.5).

In addition, $f\left(\omega_{k,i}^{e}{}^{*}\right)$ represents the likelihood of the prior distribution for the candidate value for the $k^{\text{th}}$ scale-model random effect, $\omega_{k,i}^{e}{}^{*}$, which was sampled from the univariate normal distribution shown in (3.35)

$$f\left(\omega_{k,i}^{e}{}^{*}\right) \sim N\left(0, \sigma_{\omega_{k,i}^{e}}^{2}\right), \tag{3.35}$$

where $\sigma_{\omega_{k,i}^{e}}^{2}$ is the variance of the $k^{\text{th}}$ location-model random effect for individual $i$ obtained from current $\mathbf{G}_i$.

Finally, $Q\left(\omega_{k,i}^{e}|\omega_{k,i}^{e}{}^{*}\right)$ represents the candidate-generating distribution and is the likelihood of drawing the current value of the $k^{\text{th}}$ scale-model random effect for individual $i$, $\omega_{k,i}^{e}$, given the candidate value of the $k^{\text{th}}$ scale-model random effect for individual $i$, $\omega_{k,i}^{e}{}^{*}$. All scale-model random effects were sampled from the univariate normal distribution shown in (3.36)

$$Q\left(\omega_{k,i}^{e}|\omega_{k,i}^{e}{}^{*}\right) \sim N\left(\omega_{k,i}^{e}{}^{*}, \sigma_{\omega_{k,i}^{e}}^{2}\right), \tag{3.36}$$

with the candidate-generating variance for the $k^{\text{th}}$ scale-model random effect for individual $i$, $\sigma_{\omega_{k,i}^{e}}^{2}$, tuned to achieve candidate acceptance rates of 45%.

Calculating the denominator of $r_{\text{MH},\omega_i^e}$ followed a similar process. Specifically, the likelihood of the observed data for individual $i$ given the current value of the $k^{\text{th}}$ scale-model random effect for individual $i$, $f\left(\mathbf{Y}_i|\omega_{k,i}^{e}\right)$, the likelihood of the prior distribution for the current value of the $k^{\text{th}}$ scale-model random effect for individual $i$, $f\left(\omega_{k,i}^{e}\right)$, and the candidate-generating distribution, $Q\left(\omega_{k,i}^{e}{}^{*}|\omega_{k,i}^{e}\right)$, each were obtained by substituting $\omega_{k,i}^{e}{}^{*}$ for $\omega_{k,i}^{e}$ (and vice versa, as necessary) in (3.14), (3.35), and (3.36), respectively.

**Limitations of the algorithm.** Although the MH algorithm presented in this chapter is flexible, it is not without limitations. First, the algorithm is limited to conditionally normally distributed outcomes; however, this does not preclude the use of variable transformation to approximate link functions from generalized linear models prior to estimating the model (e.g., natural log transformation = log link). Second, although residual variances in $\mathbf{R}_i$ were allowed to be heterogeneous between individuals and across occasions within an individual, residual values were always assumed independent within an individual (although some quantity of within-individual correlation is being captured by the location-model random effects). That is, residuals were always assumed to have a correlation of zero because the algorithm does not (yet) allow the estimation of alternative covariance structures in $\mathbf{R}_i$ (e.g., AR1, Toeplitz).

## Chapter Summary

The purpose of this chapter was to detail the current software and estimation algorithms available to estimate the mixed-effect location-scale model. The chapter began with an introduction to current software available for ML and MCMC estimation alongside their limitations. This was followed by an overview of Bayes theorem for continuous outcomes as well as the background information regarding MCMC methodology including discussions of estimation and convergence evaluation. Finally, the technical details of the MH algorithm used to estimate all mixed-effects location-scale models within this dissertation were presented. This algorithm was used to conduct the methodological studies to be reported chapter 4, which estimate the statistical power to detect and predict the scale-model random intercept variance and begin to study the consequences of alternative strategies for modeling location- and scale-model fixed and

random effects. The algorithm was also used to conduct the empirical data analysis in chapter 5.

# CHAPTER 4: THE POWER TO DETECT AND PREDICT THE SCALE-MODEL RANDOM INTERCEPT VARIANCE AND THE CONSEQUENCES OF ALTERNATIVE STRATEGIES FOR MODELING LOCATION- AND SCALE-MODEL FIXED AND RANDOM EFFECTS

Literature regarding the mixed-effects location-scale model has tended to focus on model estimation and interpretation (Cleveland, Denby, & Liu, 2000; Hedeker & Nordgren, 2013; Leckie et al., 2014; Lee & Nelder, 2006; Pugach, Hedeker, & Mermelstein, 2014; Rast et al., 2012). As a result, a paucity of methodological studies exists pertaining specifically to the mixed-effects location-scale model. In educational and social sciences literature, it appears that the only methodological studies to date have evaluated the effect of omitting the scale-model random intercept variance on inferences for fixed effects included in the scale model for the residual variance using educational data (Leckie, 2014; Leckie et al., 2014). The results of these studies indicated that Type I error rates for level-1 and level-2 fixed effects included in the scale model for the residual variance, as well as for their cross-level interaction, increased as the number of level-1 and level-2 units increased, and that Type I error rates were most pronounced for level-2 predictors (as high as 54% with 250 schools and 100 students per school). Thus, erroneously omitting scale-model random effects could result in incorrect inference for hypotheses specific to predicting individual differences in outcome variability, which could lead to (incorrect) final location and/or scale models that contain spuriously significant effects.

With this in mind, this chapter presents two simulation studies that begin to address both of these concerns. The first simulation study provides power curves to detect

and predict the scale-model random intercept variance representing individual differences in outcome variability, and the second simulation study provides empirical scientists with information regarding the consequences of alternative strategies for modeling location- and scale-model fixed and random effects. A discussion of the results follows each simulation study; the chapter concludes with a discussion of the limitations and directions for future methodological research.

## Simulation Study I: The Power to Detect and Predict Scale-Model Random Intercept Variance

Given that Leckie (2014) and Leckie et al. (2014) have shown that erroneously omitting significant scale-model random intercept variance can drastically increase Type I error rates for level-2 fixed effects included in the scale model for the residual variance, it is important for empirical scientists to know *a priori* when scale-model random intercept variance can be detected. Along these same lines, if the scale-model random intercept variance cannot be detected, it is important to know whether inferences can be trusted for fixed effects of level-2 predictors included in the scale model for the residual variance.

To date, it appears that no methodological study has been conducted specifically evaluating the study design characteristics as well as model parameters that likely affect the power to detect the scale-model random intercept variance. Therefore, in the first simulation study, power curves were calculated to identify the sample of individuals at level 2, and number of repeated occasions within an individual at level 1, required to detect the scale-model random intercept variance based on pseudo-randomly sampled scale-model fixed and random intercepts, the correlation between location- and scale-

model random intercepts, and the effect of a level-2 predictor included in both the location and scale models. In addition, for models in which the scale-model random intercept variance was detected, this study evaluated the power to detect the effect of a level-2 predictor included in the scale model for the residual variance; for models in which the scale-model random intercept variance could not be detected, this study also evaluated the Type I error rate for a level-2 predictor included in the scale model for the residual variance.

**Data-generating mixed-effects location-scale model.** In the data-generating mixed-effects location-scale model for this simulation study, only the residual variance was specified to be heterogeneous between individuals (i.e., $\mathbf{R}_i$), with the location- and scale-model random intercept variance, and their correlation, specified as homogeneous between individuals (i.e., $\mathbf{G}_i = \mathbf{G}$; no subscript $i$ necessary).

The data-generating location model is shown in (4.1), which included a fixed intercept, $\gamma_{00}$, random intercept, $u_{0,i}$, and the fixed effect for the level-2, time-invariant predictor, $\gamma_{01}$.

$$\begin{aligned} &\text{Level 1:} \quad Y_{t,i} = \beta_{0,i} + e_{t,i} \\ &\text{Level 2:} \quad \beta_{0,i} = \left(\gamma_{00} + u_{0,i}\right) + \gamma_{01}(X_i) \\ &\text{Combined:} \quad Y_{t,i} = \left(\gamma_{00} + u_{0,i}\right) + \gamma_{01}(X_i) + e_{t,i} \end{aligned} \tag{4.1}$$

The scale model for the residual variance is shown in (4.2), which included the fixed intercept for the residual variance, $\delta_{00}^{e}$, scale-model random intercept, $\omega_{0,i}^{e}$, and the fixed effect for the level-2, time-invariant predictor, $\delta_{01}^{e}$, which was the same level-2 predictor included in the location model in (4.1).

$$\begin{aligned} &\text{Level 1:} \quad \log\left(\sigma_{e_i}^2\right) = \tau_{0,i}^{e} \\ &\text{Level 2:} \quad \tau_{0,i}^{e} = \left(\delta_{00}^{e} + \omega_{0,i}^{e}\right) + \delta_{01}^{e}(X_i) \\ &\text{Combined:} \quad \log\left(\sigma_{e_i}^2\right) = \left(\delta_{00}^{e} + \omega_{0,i}^{e}\right) + \delta_{01}^{e}(X_i) \end{aligned} \tag{4.2}$$

Here, $\mathbf{R}_i$ was predicted to be heterogeneous between individuals, but residual values were assumed to be independent within an individual, such that the correlations between residuals were constrained to be zero, $\mathbf{R}_i = \sigma_{e_i}^2 \mathbf{I}_{n_i}$. Note that the subscript $e$ for the residual variance estimate did not require an additional subscript for occasion $t$ because there was only one estimated residual variance for each individual $i$. That is, there were no time-varying predictors in the scale model for the residual variance, so the residual variance was not estimated to vary across occasions within an individual.

Finally, the scale models for the level-2 random effect variances and correlation in $\mathbf{G}$ were assumed constant between individuals. That is, the level-2 scale models included only fixed intercepts for the location-model random intercept, $\alpha_0^{u_0}$, scale-model random intercept, $\alpha_0^{\omega_0^e}$, and the location-scale model random intercept correlation, $\alpha_0^{u_0;\omega_0^e}$, as shown below in (4.3).

$$\begin{aligned} \log\left(\sigma_{u_0}^2\right) &= \alpha_0^{u_0} \\ \log\left(\sigma_{\omega_0^e}^2\right) &= \alpha_0^{\omega_0^e} \\ \tanh^{-1}\left(\rho_{u_0;\omega_0^e}\right) &= \alpha_0^{u_0;\omega_0^e} \end{aligned} \tag{4.3}$$

**The sampling distributions for study design and model parameters.**

Considering the paucity of methodological studies and small number of empirical studies using the mixed-effects location-scale model, most study design and model parameters were sampled from a range of values that could be reasonably expected from typical

repeated-measures data, with a uniform distribution used frequently to ensure appropriate coverage of potential effects and to increase applicability.

***Individuals and occasions.*** The number of individuals at level 2 and repeated occasions at level 1 were sampled from ranges that mirror what could be reasonably expected in traditional IIV studies (e.g., Nesselroade & Salthouse, 2004; Rast et al., 2012; Schmiedek, Lövdén, & Lindenberger, 2009), with the number of individuals sampled as shown in (4.4)

$$N_{\text{individuals}} \sim U(25{,}200) \tag{4.4}$$

and the number of repeated occasions within an individual sampled as shown in (4.5).

$$n_i \sim U(5{,}50) \tag{4.5}$$

Both of these sampled values were rounded to the nearest whole number and both simulation studies assumed complete data. Taken together, this allowed the total number of observations to range from 125 to 5,000 within the simulated datasets (which will be known for the remainder of this dissertation as *replications*).

***The total residual variance.*** Although the scale-model random intercept variance can be conceptualized as a proportion of *total* residual variance, this proportion is not calculated as directly as the ICC defined in (2.3). Briefly, Hedeker et al. (2008) used Cholesky factorization to standardize the *unconditional* location- and scale-model random intercept variances. Following standardization, the unconditional scale-model random intercept variance can be viewed directly as a proportion of total residual variance as shown in (4.6).

$$\sigma_{e_i}^2 = \exp\left(\delta_{00}^{e} + \frac{1}{2}\left(\exp\left(\alpha_0^{\omega_0^e}\right)\right)\right) \tag{4.6}$$

Here, $\sigma^2_{e_i}$ is the total residual variance, $\delta^e_{00}$ is the fixed intercept for the (log of the) residual variance, and $\alpha_0^{\omega_0^e}$ is the fixed intercept for the (log of the) scale-model random intercept variance as defined in (4.3). It might be slightly confusing to see what appears to be double exponentiation of $\alpha_0^{\omega_0^e}$ in (4.6). The initial exponentiation of $\alpha_0^{\omega_0^e}$ is required to ensure the estimate of the scale-model random intercept variance remains positive (i.e., variance estimates are predicted on the log scale; thus, $\exp\left(\alpha_0^{\omega_0^e}\right) = \sigma^2_{\omega_0^e}$). The second exponentiation is required to convert the total residual variance estimate (determined by $\delta^e_{00}$ and $\sigma^2_{\omega_0^e}$), which is also on the log scale, back onto the variance scale.

Accordingly, $\delta^e_{00}$ and $\alpha_0^{\omega_0^e}$ were sampled on the log scale as shown in (4.7) and (4.8), respectively.

$$\delta^e_{00} \sim U(0.59, 1.93) \tag{4.7}$$

$$\alpha_0^{\omega_0^e} \sim U(-5, -0.11) \tag{4.8}$$

The range of values for both $\delta^e_{00}$ and $\alpha_0^{\omega_0^e}$ were based on previous empirical studies of Hedeker et al. (2008), Rast and Zimprich (2011), and Rast et al. (2012). The fixed intercept for the residual variance, $\delta^e_{00}$, mapped directly onto variance-scale estimates ranging between approximately 1.80 and 6.89 after exponentiation, whereas the fixed intercept for the scale-model random intercept variance, $\alpha_0^{\omega_0^e}$, mapped onto variance-scale estimates ranging from approximately 0 to 0.90 after exponentiation.

In addition, when considering the calculation of total residual variance in (4.6), the range of values in (4.7) and (4.8) resulted in the proportion of total residual variance due specifically to the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, to range from

approximately 0.00% (when $\alpha_0^{\omega_0^e} = -5$ regardless of the value of $\delta_{00}^e$) to 43.27% (when $\alpha_0^{\omega_0^e} = -0.11$ and $\delta_{00}^e = 0.59$).

***The location-model random intercept variance and the heterogeneous ICC.*** To explicitly determine the proportion of variance in the outcome at level 1 and level 2, the (log of the) location-model random intercept variance, $\alpha_0^{u_0}$, was fixed to 1 as shown in (4.9).

$$\alpha_0^{u_0} = 1 \tag{4.9}$$

Because the mixed-effects location-scale model allows scale-model predictors of all variances and correlations in $\mathbf{G}_i$ and $\mathbf{R}_i$, as well as individual differences in outcome variability via the scale-model random intercept variance, the ICC will necessarily be heterogeneous between individuals (i.e., as such the ICC will now also require the subscript $i$; $ICC_i$). Therefore, the notation for the ICC shown previously in (2.3) is no longer applicable and must be updated specifically for the mixed-effects location-scale model using notation described in (1.9) and (1.27). The matrix formulation for a generic model is shown in (4.10).

$$\text{ICC}_i = \frac{\exp\left(\mathbf{A}_i^{u_0}\boldsymbol{\alpha}^{u_0}\right)}{\exp\left(\mathbf{A}_i^{u_0}\boldsymbol{\alpha}^{u_0}\right) + \exp\left(\mathbf{T}_i^e\boldsymbol{\tau}^e + \frac{1}{2}\left(\exp(\mathbf{W}_i^e\boldsymbol{\omega}_i^e)\right)\right)} \tag{4.10}$$

As described previously in chapter 1, $\mathbf{A}_i^{u_0}$ is a *1* x $a^{u_0}$ row vector containing the specific set of $a^{u_0}$ level-2, individual-level predictor variables of the location-model random intercept variance for individual $i$, and $\boldsymbol{\alpha}^{u_0}$ is $a^{u_0}$ x 1 column vector containing the estimated fixed effect for each of these $a^{u_0}$ level-2 predictors. Further, $\mathbf{T}_i^e$ is a $n_i$ x $c^e$ matrix containing the $c^e$ level-1, occasion-level and/or level-2, individual-level predictors

of the residual variance across the $n$ occasions for individual $i$, and $\boldsymbol{\tau}^e$ is the $c^e$ x 1 column vector containing the estimated fixed effect for each of these $c^e$ level-1 or level-2 predictors. Finally, $\mathbf{W}_i^e$ is an $n_i$ x $w^e$ matrix containing $w^e$ level-1 predictor variables included in the scale model for the residual variance specified to have random effects, and $\boldsymbol{\omega}_i^e$ is a $w^e$ x 1 column vector of deviations for individual $i$ from each of the $w^e$ scale-model fixed effects specified to be random.

Specifically for the data-generating location and scale models described above, the location-model random intercept variance was unconditional and therefore assumed constant between individuals. Thus, $\mathbf{A}_i^{u_0}$ is a scalar for each individual $i$ containing the value of 1 to indicate the intercept, and $\boldsymbol{\alpha}^{u_0}$ is a scalar containing the (log of the) location-model random intercept variance, $\alpha_0^{u_0}$. In addition, $\mathbf{T}_i^e$ is an $n_i$ x 2 matrix containing a value of 1 to represent the intercept as well as the value of level-2 predictor $X_i$, and $\boldsymbol{\tau}^e$ is a 2 x 1 column vector containing the (log of the) fixed intercept for the residual variance and the (log of the) fixed effect for the level-2 predictor $X_i$. Finally, given that the scale-model random intercept was the only random effect included in the scale model for the residual variance, and that it was assumed constant across individuals (i.e., no predictors), $\mathbf{W}_i^e$ is a scalar for each individual $i$ containing the value of 1 to indicate the intercept, and $\boldsymbol{\omega}_i^e$ is a scalar containing the (log of the) scale-model random intercept variance, $\alpha_0^{\omega_0^e}$.

In addition, the inclusion of the scale-model random intercept variance requires that multi-level notation be used for the scale model for the residual variance, as described in (2.8) and (2.11). Therefore, the heterogeneous $\text{ICC}_i$ for the data-generating model varies based on the fixed intercept for the residual variance, $\delta_{00}^e$, and the fixed

effect for the level-2 predictor included in the scale model for the residual variance, $\delta_{01}^{e}$, as shown in (4.11).

$$\mathrm{ICC}_i = \frac{\exp\left(\alpha_0^{u_0}\right)}{\exp\left(\alpha_0^{u_0}\right) + \exp\left(\left(\delta_{00}^{e} + \delta_{01}^{e}(X_i)\right) + \frac{1}{2}\left(\exp\left(\alpha_0^{\omega_0^{e}}\right)\right)\right)} \tag{4.11}$$

Therefore, with $\alpha_0^{u_0}$ fixed to 1, the range of values for $\delta_{00}^{e}$ and $\alpha_0^{\omega_0^{e}}$, defined by (4.7) and (4.8), respectively, was chosen specifically to ensure the unconditional heterogeneous $\mathrm{ICC}_i$ (i.e., excluding $\delta_{01}^{e}$) ranged between approximately 0.20 and 0.60 on the data scale, which was considered reasonable for repeated-measures data.

***The correlation between location- and scale-model random intercepts.*** The correlation between the location- and scale-model random intercepts, $\rho_{u_0;\omega_0^{e}}$, was sampled as fixed to either 0.00 or 0.50 determined by the Bernoulli distribution shown in (4.12).

$$\rho_{u_0;\omega_0^{e}} \sim B(1,0.50) \tag{4.12}$$

Here, the probability of success was set at 0.50 resulting in an approximately equal number of replications within each condition; sampled value of 0 indicated a correlation of 0.00, whereas a sampled value of 1 indicated a correlation of 0.50. A negative correlation was excluded from this simulation study because it was expected that any effects for the positive correlation would be mirrored for a negative correlation.

***The value of the level-2, time-invariant predictor.*** One level-2, time-invariant predictor, $X_i$, was included in both the location model and scale model for the residual variance, sampled from a standard normal distribution as shown in (4.13).

$$X_i \sim N(0,1) \tag{4.13}$$

***The effect for the level-2 predictor included in the location model.*** The effect for the level-2 predictor of the location-model random intercept variance was sampled from a zero-inflated Poisson distribution with a mean of 0.60 with a 0.20 probability of being an extra zero, as shown in (4.14).

$$\gamma_{01} \sim ZIP(0.60,0.20) \tag{4.14}$$

This distribution was chosen because it forced predictor effects to be positive, allowing a clear indication of whether a misspecified location model or scale model for the residual variance increased or decreased the predictor's effect. Further, it was expected that any effects for the positive parameter estimate would be mirrored for a negative estimate.

An initial simulation study of 100 replications using an unconditional, homogeneous ICC of 0.50, where $\exp(\alpha_0^{u_0}) = \exp(\delta_{00}^{e}) = 2.72$, $\exp\left(\alpha_0^{\omega_0^e}\right) = 0$, and $X_i \sim N(0,1)$, resulted in an effect size (i.e., pseudo-$R^2$) distribution with a median reduction of location-model random intercept variance of 0.03, IQR [0.01,0.19], which was considered reasonable. Pseudo-$R^2$ was calculated as shown in (4.15), where *unconditional* and *conditional* imply that $X_i$ was excluded and included in the location model, respectively.

$$\text{pseudo}-R^2_{u_0} = \frac{\exp(\alpha_0^{u_0})_{\text{unconditional}} - \exp(\alpha_0^{u_0})_{\text{conditional}}}{\exp(\alpha_0^{u_0})_{\text{unconditional}}} \tag{4.15}$$

***The effect for the level-2 predictor included in the scale model for the residual variance.*** The effect for the level-2 predictor included in the scale model for the residual variance, $\delta_{01}^{e}$, explains scale-model random intercept variance and was sampled from a slightly more conservative zero-inflated Poisson distribution, with a mean of 0.25 and a 0.20 probability of being an extra zero, as shown in (4.16).

$$\delta_{01}^{e} \sim ZIP(0.25,0.20) \tag{4.16}$$

An initial simulation study of 100 replications using an unconditional heterogeneous ICC of 0.47, where $\exp\left(\alpha_0^{u_0}\right) = \exp(\delta_{00}^{e}) = 2.72$, $\exp\left(\alpha_0^{\omega_0^{e}}\right) = 0.25$, and $X_i \sim N(0,1)$, resulted in an pseudo-$R^2$ distribution with a median reduction of scale-model random intercept variance of 0.04, IQR [0.01,0.14], that was consistent with results of Hedeker et al. (2008), Rast and Zimprich (2011), and Rast et al. (2012). Here, pseudo-$R^2$ was calculated similarly to (4.15), with $\alpha_0^{\omega_0^{e}}$ substituted for $\alpha_0^{u_0}$; however, it is important to note that it has not been determined whether pseudo-$R^2$ for the reduction in scale-model random intercept variance behaves in a similar fashion to pseudo-$R^2$ for reduction of location-model random effect variances.

**Data generation.** All data were simulated using SAS v. 9.4 (SAS Institute Inc., Cary, NC). A total of 4,000 individual replications were simulated for this study (i.e., 500 replications per sampled parameter). All parameters were sampled pseudo-randomly using a seed based on the computer's system clock (i.e., seed = 0), thus creating a different seed for each replication.

**Estimated model sequence.** Four separate models were estimated for each replication, each of which only modified the scale model for the residual variance such that all level-2 random effects were considered homogeneous as shown in (4.3). It is generally not recommended to estimate scale-model random effects before estimating location-model random effects because individual mean differences would not have been partitioned out of residual variance via location-model random effects. As a result, these unaccounted for individual mean differences could result in Type I errors when evaluating the necessity of scale-model random effects. That is, scale-model random

effects (i.e., individual differences in residual variability) could be artifacts of existing, un-modeled individual mean differences. However, more research is needed on this topic. Therefore, to ensure proper evaluation of scale model random effects, the location model shown in (4.1) remained constant in all four estimated models, in which individual mean differences in the location-model fixed intercept, $\gamma_{00}$, were quantified by the location-model random intercept variance, $\exp(\alpha_0^{u_0}) = \sigma_{u_0}^2$, and explained by the level-2 fixed effect for $X_i$, $\gamma_{01}$. Table 4.1 provides the effects included in each mixed-effects location-scale model; what follows below is a complete description of how the scale model was specified to change across the four models estimated in this study.

Table 4.1
*The Effects Included in the Mixed-Effects Location-Scale Models for the First Simulation Study*

| | Location Model | | | Scale Model for Level-2 Variance Components | | | Scale Model for Level-1 Residual Variance | | |
|---|---|---|---|---|---|---|---|---|---|
| | Fixed Effects | | Random Effects | Fixed Effects | | | Fixed Effects | | Random Effects |
| | $\gamma_{00}$ | $\gamma_{01}$ | $u_{0,i}$ | $\alpha_0^{u_0}$ | $\alpha_0^{\omega_0^e}$ | $\alpha_0^{u_0;\omega_0^e}$ | $\delta_{00}^e$ | $\delta_{01}^e$ | $\omega_{0,i}^e$ |
| Model 1 | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| Model 2 | ● | ● | ● | ● | | | ● | | |
| Model 3 | ● | ● | ● | ● | ● | ● | ● | | ● |
| Model 4 | ● | ● | ● | ● | | | ● | ● | |

*Note.* A ● in a given column indicates that the effect is included in the model. $\gamma_{00}$ = location-model fixed intercept. $\gamma_{01}$ = location-model fixed effect for level-2 predictor $X_i$. $u_{0,i}$ = location-model random intercept for individual $i$. $\alpha_0^{u_0}$ = log of the scale-model fixed intercept for the location-model residual variance. $\alpha_0^{\omega_0^e}$ = log of the scale-model fixed intercept for the scale-model random intercept. $\alpha_0^{u_0;\omega_0^e}$ = inverse hyperbolic tangent of the scale-model fixed intercept for the correlation between the location- and scale-model random intercepts. $\delta_{00}^e$ = log of the scale-model fixed intercept for the residual variance. $\delta_{01}^e$ = log of the scale-model fixed effect for level-2 predictor $X_i$. $\omega_0^e$ = scale-model random intercept for individual $i$.

***Model 1.*** The first estimated model included the true scale model for the residual variance shown in (4.2), in which individual differences in residual variance were estimated via $\omega_{0,i}^e$, with the level-2 fixed effect, $\delta_{01}^e$, allowing residual variance to be

heterogeneous across values of $X_i$, and to explain the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$.

***Model 2.*** The second model, shown in (4.17), estimated an unconditional scale model for the residual variance by omitting the fixed effect for the level-2 predictor, $\delta_{01}^e$, and further assumed that residual variance was homogeneous between individuals by omitting the scale-model random intercept, $\omega_{0,i}^e$. Because this scale model for the residual variance was a typical empty (or unconditional) model, the subscript $i$ was not technically required for any effect in this model.

$$\begin{aligned} &\text{Level 1:} \quad \log\left(\sigma_{e_i}^2\right) = \tau_{0,i}^e \\ &\text{Level 2:} \quad \tau_{0,i}^e = \delta_{00}^e \\ &\text{Combined:} \quad \log\left(\sigma_{e_i}^2\right) = \delta_{00}^e \end{aligned} \tag{4.17}$$

***Model 3.*** The third estimated model allowed individual differences in residual variance, but did not attempt to predict why these individual differences existed by omitting the fixed effect for the level-2 predictor, $\delta_{01}^e$, as shown in (4.18).

$$\begin{aligned} &\text{Level 1:} \quad \log\left(\sigma_{e_i}^2\right) = \tau_{0,i}^e \\ &\text{Level 2:} \quad \tau_{0,i}^e = \left(\delta_{00}^e + \omega_{0,i}^e\right) \\ &\text{Combined:} \quad \log\left(\sigma_{e_i}^2\right) = \left(\delta_{00}^e + \omega_{0,i}^e\right) \end{aligned} \tag{4.18}$$

***Model 4.*** The fourth estimated model retained the fixed effect for the level-2 predictor, $\delta_{01}^e$, but omitted the scale-model random intercept, $\omega_{0,i}^e$, as shown in (4.19). Therefore, this model assumed no random variation (i.e., individual differences) in residual variance, but instead assumed residual variance to vary systematically with (i.e., as a deterministic function of) predictor $X_i$.

$$\begin{aligned} &\text{Level 1:} \quad \log\left(\sigma_{e_i}^2\right) = \tau_{0,i}^e \\ &\text{Level 2:} \quad \tau_{0,i}^e = \delta_{00}^e + \delta_{01}^e(X_i) \\ &\text{Combined:} \quad \log\left(\sigma_{e_i}^2\right) = \delta_{00}^e + \delta_{01}^e(X_i) \end{aligned} \tag{4.19}$$

**Model estimation via MCMC.** All mixed-effects location-scale models in the first simulation study were estimated using the MCMC estimator based on the Metropolis-Hastings algorithm described in chapter 3 that was developed and run in R v. 3.1.0 (R Core Development Team, 2014). The four models described in the previous section were estimated for each of the 4,000 replications (i.e., 16,000 total estimated models) using the Crane supercomputer within the Holland Computing Center of the University of Nebraska-Lincoln.

The estimation procedure was identical for all models. Although true values for all model parameters were known, to better simulate real-world conditions, start values for the location-model fixed effects, $\gamma_{00}$ and $\gamma_{01}$, the location-model random intercept variance, $\sigma_{u_0}^2$, and the residual variance, $\sigma_e^2$ (which served as a proxy for the fixed intercept of the residual variance, $\delta_{00}^e$), were based on preliminary estimation of a traditional linear mixed-effects model using the lme4 package in R developed by Bates, Mächler, Bolker, and Walker (2014). Start values for the effect for the level-2 predictor included in the scale model for the residual variance, $\delta_{01}^e$, the scale-model random intercept variance, $\sigma_{\omega_0^e}^2$, and the correlation between the location- and scale-model random intercepts, $\rho_{u_0;\omega_0^e}$, were all set to 0.

Prior to initiation of the Markov chain, candidate-generating distributions for all parameters (for all replications) were tuned as described in chapter 3 using 20 tuning chains of 50 iterations to achieve an optimal acceptance rate of 45%. Following tuning,

the Markov chain for every replication, regardless of the sampled parameter values, was initially specified to sample 5,000 iterations. However, initial testing determined that the MCMC estimator was not consistently meeting convergence criteria for replications simulated to have fewer than 50 individuals. As a result, the Markov chain for any replication with fewer than 50 individuals was specified to sample 7,500 iterations, which consistently met convergence criteria.

The burn-in period consisted of the first 1,000 iterations, regardless of the number of iterations, with a thinning interval of 1 (i.e., no thinning); convergence was assessed empirically by Geweke's diagnostic test (Geweke, 1992) and the Gelman and Rubin criterion (Gelman & Rubin, 1992), both calculated using the CODA package in R developed by Plummer et al. (2006). As described in detail in chapter 3, Geweke's diagnostic test is a $z$-test, where $z < -1.96$ or $z > 1.96$ indicated the parameter failed to meet convergence criteria, whereas the Gelman and Rubin criterion, $\hat{R}$, is essentially an $F$-test, where $\hat{R} \leq 1.5$ was used to indicate convergence. Because both criteria have the potential to indicate non-convergence as a Type I error, convergence was defined as satisfying at least one criterion. For parameters that failed to converge, visual inspection of trace plots and posterior distributions was conducted for a pseudo-random sample of approximately 10% of the failed parameters across models.

Finally, the estimate for each parameter was based on the mean of the posterior distribution provided by the Markov chain. Parameter recovery and accuracy were calculated for the model 1 (i.e., the true model). Recovery estimates were calculated as the proportion of the 4,000 replications in which the estimated 95% credible interval for a given parameter contained the true sampled value. The accuracy of parameter estimates

was evaluated using signed bias calculated as the difference between the estimated parameter, $\hat{\theta}$, and the true value of the parameter, $\theta$, as shown in (4.20).

$$\text{Bias} = \hat{\theta} - \theta \tag{4.20}$$

Signed bias estimates were calculated for each parameter using the mean bias for a given parameter across the 4,000 replications, presented alongside their 95% confidence interval. The larger the absolute value of the bias estimate, the larger the distance between the estimated and true parameter value, with a positive bias estimate indicating the parameter was overestimated.

**Estimated model comparisons.** First, to determine the power to detect the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, the deviance information criterion (DIC) from model 2 and model 3 were compared directly, with lower DIC values indicating improved model fit. DIC was calculated as described in (3.9) through (3.12), and this model-based comparison was akin to indicating that $\sigma^2_{\omega_0^e}$ was greater than zero. Note that model 1 (i.e., the true model) was not evaluated here because this analysis assumed the traditional model-building approach in which random effects were evaluated prior to fixed effects (i.e., model 1 included the fixed effect for the level-2 predictor, whereas model 3 did not).

Given that all sample characteristics and model parameters were sampled on continuous scales (except the correlation between location- and scale-model random intercepts), the number of individuals, $N$, the number of repeated occasions, $n_i$, the fixed effect for the level-2 predictor included in the location model, $\gamma_{01}$, the fixed intercept for the residual variance, $\delta^e_{00}$, and the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, were each partitioned into 20 bins allowing for more reliable estimation of their univariate power

curves. The (univariate or marginal) power to detect $\sigma^2_{\omega^e_0}$ was then calculated for each bin as the proportion of replications where the DIC from model 3 was lower than model 2 for each sampled parameter. The univariate power curves were constructed by connecting calculated power across bins; no smoothing was used in order to prevent masking the observed change in power across bins.

In addition, because the univariate power curves held the other sampled parameters at their average values across models, they may have captured power too simplistically. Thus, to determine whether certain combinations of sampled parameters moderated the power to detect the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, interactions between all sampled parameters were evaluated. Given the unknown functional form for all sampled parameters, generalized additive models (GAM; see Hastie & Tibshirani, 1986) were considered initially to evaluate whether the detection of $\sigma^2_{\omega^e_0}$ (as determined by lower DIC between models 2 and 3) was moderated by the sample parameters. However, the local scoring algorithm would not converge for any GAM that included $\sigma^2_{\omega^e_0}$. A second approach proposed a series of multivariable logistic regression models predicting significant $\sigma^2_{\omega^e_0}$ evaluating for interactions between the binned sampled parameters. However, quasi-complete separation occurred upon inclusion of any interaction effect. This indicated that the detection of $\sigma^2_{\omega^e_0}$ was almost perfectly predicted by the moderation of the sampled parameters included in the interaction effect. Therefore, in the reported analysis, multivariable logistic regression models were estimated to predict significant $\sigma^2_{\omega^e_0}$ in which all sampled parameters were modeled as continuous. These models assumed a linear functional form (i.e., linearity in the logit) between each

continuous predictor and the logit (or log-odds) of detecting $\sigma^2_{\omega^e_0}$. This assumption was tested via the Box-Tidwell test (Box & Tidwell, 1962), and was violated for all sampled parameters. Thus, a liberal $p < .20$ was used as impetus to examine power curves for a specific two-way interaction effects.

Second, using only replications in which the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, estimated by model 3 was detected, the power to detect the effect of the level-2 predictor included in the scale model for the residual variance, $\delta^e_{01}$, was calculated using $\delta^e_{01}$ estimated from model 1 (i.e., the true model). Statistical significance for $\delta^e_{01}$ was indicated by its 95% credible interval, defined as the region between the 2.5 and 97.5 percentiles of the posterior distribution excluding 0, as described in chapter 3. When calculating the power to detect $\delta^e_{01}$ estimated from model 1, all sampled parameters were partitioned into 20 bins (except the correlation between location- and scale-model random intercepts) and the power to detect $\delta^e_{01}$ within each bin was calculated as the proportion of replications for which $\delta^e_{01}$ was statistically significant. A univariate power curve for each sampled parameter was constructed by connecting calculated power across bins, with no smoothing parameter.

Similar to the univariate power curves for detecting $\sigma^2_{\omega^e_0}$, the univariate power curves for detecting $\delta^e_{01}$ may have been overly simplistic; thus, a series of multivariable logistic regression models were used (again due non-convergence of the local scoring algorithm from the GAM and quasi-separation when using binned sampled parameters in logistic regression) to determine which additional sampled parameters moderated the power to detect $\delta^e_{01}$. Because the Box-Tidwell test indicated linearity of the logit was

violated for each sampled parameter, $p < .20$ was used to indicate whether power curves should be examined for a specific two-way interaction effect.

In addition, given that the data were available, a replication of Leckie (2014) and Leckie et al. (2014) was also conducted. Here, using only replications in which the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, was detected and the effect of the true sampled value of the level-2 predictor included in the scale model for the residual variance, $\delta^e_{01}$, was 0, the Type I error rate for $\delta^e_{01}$ estimated by model 3 was calculated as the proportion of these replications for which $\delta^e_{01}$ estimated by model 3 was (spuriously) statistically significant, presented alongside its 95% Clopper-Pearson confidence interval (the Clopper-Pearson interval is an exact confidence interval that has been shown to have greater coverage probability than either the Wald or Agresti-Coull intervals; see Agresti & Coull, 1998 and Clopper & Pearson, 1934).

Finally, for replications in which the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, was not detected, the proportion of replications was determined in which the true sampled value of the level-2 predictor included in the scale model for the residual variance, $\delta^e_{01}$, was 0. Further, using only the replications for which the true sampled value of $\delta^e_{01}$ was 0, Type I error rates were calculated for $\delta^e_{01}$ estimated from model 4, presented alongside its 95% Clopper-Pearson confidence interval.

**Results of simulation study I.** The results of the first simulation study are provided below. Convergence criteria and signed bias are discussed initially, followed by a description of the power to detect the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, and the power to detect the effect for the level-2 predictor included in the scale model for the

residual variance, $\delta_{01}^{e}$. Finally, Type I error rates are discussed for $\delta_{01}^{e}$ when $\sigma_{\omega_0^e}^2$ could not be detected for replications in which the true sampled value of $\delta_{01}^{e}$ was zero.

Table 4.2
*Total Number and Proportion of Parameters Satisfying Geweke's and/or Gelman and Rubin's Convergence Criterion (N = 4,000)*

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| Location Model | | | | |
| $\gamma_{00}$ | 3942 (98.55) | 3951 (98.78) | 3945 (98.63) | 3960 (99.00) |
| $\gamma_{01}$ | 3908 (97.70) | 3956 (98.90) | 3933 (98.33) | 3934 (98.35) |
| Scale Model | | | | |
| $\delta_{00}^{e}$ | 3979 (99.48) | 3996 (99.90) | 3982 (99.55) | 3993 (99.83) |
| $\delta_{01}^{e}$ | 3990 (99.75) | - | - | 3995 (99.88) |
| $\alpha_0^{u_0}$ | 3953 (98.83) | 3971 (99.28) | 3963 (99.08) | 3955 (98.88) |
| $\alpha_0^{u_0;\omega_0^e}$ | 3963 (99.08) | - | 3728 (93.20) | - |
| $\alpha_0^{\omega_0^e}$ | 3975 (99.38) | - | 3986 (99.65) | - |

*Note.* Data are presented as number (%). $\gamma_{00}$ = location-model fixed intercept. $\gamma_{01}$ = location-model fixed effect for level-2 predictor $X_i$. $u_{0,i}$ = location-model random intercept for individual $i$. $\alpha_0^{u_0}$ = log of the scale-model fixed intercept for the location-model residual variance. $\alpha_0^{\omega_0^e}$ = log of the scale-model fixed intercept for the scale-model random intercept. $\alpha_0^{u_0;\omega_0^e}$ = inverse hyperbolic tangent of the scale-model fixed intercept for the correlation between the location- and scale-model random intercepts. $\delta_{00}^{e}$ = log of the scale-model fixed intercept for the residual variance. $\delta_{01}^{e}$ = log of the scale-model fixed effect for level-2 predictor $X_i$. $\omega_0^e$ = scale-model random intercept for individual $i$.

***Convergence evaluation.*** The proportion of parameters across replications satisfying Geweke's and/or the Gelman and Rubin convergence criterion are presented in Table 4.2. Results indicated that at least one convergence criterion was met for a very high proportion of replications across all four models. The only exception was the (inverse hyperbolic tangent of the) correlation between the location- and scale-model random intercepts, $\alpha_0^{u_0;\omega_0^e}$, estimated by model 3 that had a convergence rate of 93.20% (which was still considered acceptable, however). This smaller convergence rate was hypothesized to result from the omission of the effect for the level-2 predictor in the scale model for the residual variance, $\delta_{01}^{e}$, which necessarily increased estimated scale-model random intercept variance, $\sigma_{\omega_0^e}^2$ (as the missing $\delta_{01}^{e}$ would have explained, or reduced,

$\sigma^2_{\omega_0^e}$). This hypothesis was supported given that the convergence rate for $\alpha_0^{u_0;\omega_0^e}$ estimated by model 1, which included $\delta^e_{01}$, was 99.08%.

With that said, of the 272 replications in which $\alpha_0^{u_0;\omega_0^e}$ estimated by model 3 failed to converge, a pseudo-random sample of 27 trace plots (i.e., ≈10%) were examined, with each trace plot having one subtle peak or valley, which may have prevented convergence criterion from being met. An archetype for this observation is presented in Figure 4.1 for replication 134.

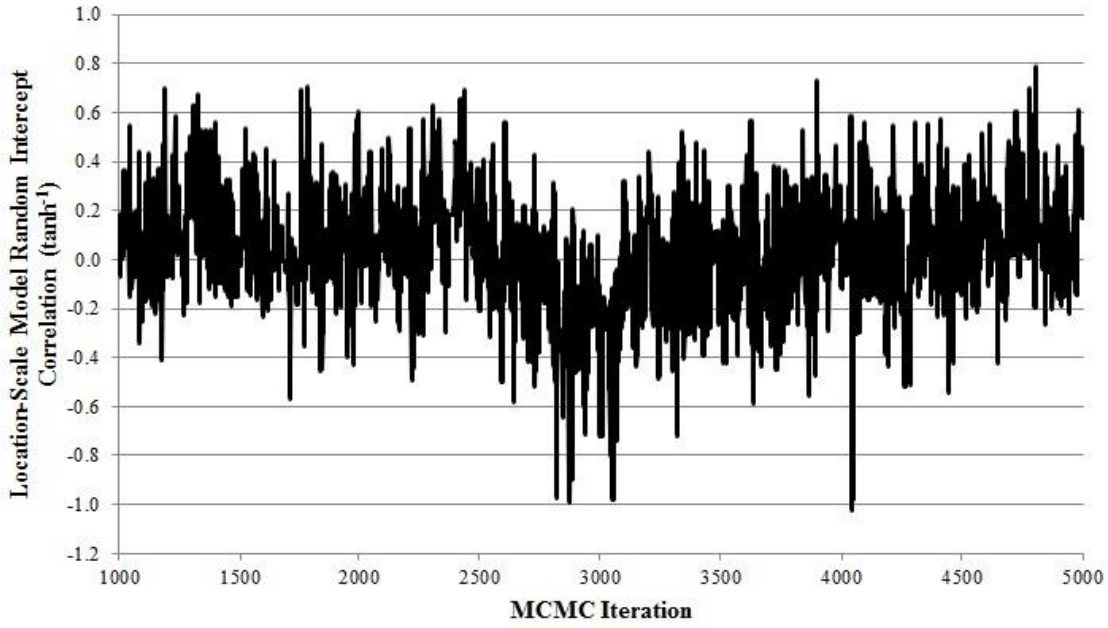


*Figure 4.1.* Trace plot from replication 134 for the correlation between the location- and scale-model random intercepts ($\tanh^{-1}$ scale).

***Parameter recovery and accuracy.*** Parameter recovery was indicated by the proportion of replications for which the estimated 95% credible interval contained the true sampled parameter value, whereas the accuracy of parameter estimates was indicated by signed bias; both are presented for the model 1 (i.e., the true model) in Table 4.3. Model 1 recovered the location-model random intercept variance and most scale-model

fixed and random effects with 95% Clopper-Pearson intervals that included expected error rates (i.e., 95% recovery indicates a 5% error rate). However, the 95% credible interval for $\gamma_{00}$ and $\gamma_{01}$ included the true sampled parameter for approximately 90% of replications, which was unexpectedly low given the starting values for these parameters estimated by the traditional linear mixed-effects model were nearly identical to final model estimates from the mixed-effects location-scale model; signed bias estimates were 0.00 for both $\gamma_{00}$ and $\gamma_{01}$. Therefore, given the ideal starting values and absence of signed bias for both $\gamma_{00}$ and $\gamma_{01}$, it was hypothesized that the MCMC algorithm may have sampled from highly leptokurtic posterior distributions (i.e., small posterior standard deviations), which resulted in extremely narrow 95% credible intervals that may have just excluded the true sampled parameter.

Table 4.3
*Parameter Recovery and Signed Bias for Parameters Estimated by Model 1 (i.e., the true model; N = 4,000)*

| | Recovery | Signed Bias |
|---|---|---|
| Location Model | | |
| $\gamma_{00}$ | 3650 (91.25) | 0.00 [–0.01,0.00] |
| $\gamma_{01}$ | 3524 (88.10) | 0.00 [0.00,0.00] |
| Scale Model | | |
| $\delta_{00}^{e}$ | 3779 (94.48) | 0.00 [–0.01,0.00] |
| $\delta_{01}^{e}$ | 3758 (93.95) | 0.00 [0.00,0.00] |
| $\alpha_{0}^{u_0}$ | 3813 (95.33) | 0.03 [0.02,0.03] |
| $\alpha_{0}^{u_0;\omega_0^e}$ | 3262 (81.55) | –0.05 [–0.06,–0.05] |
| $\alpha_{0}^{\omega_0^e}$ | 3714 (92.85) | 0.14 [0.12,0.16] |

*Note*. Recovery presented as frequency (percent). Signed bias presented as mean [95% CI]. $\gamma_{00}$ = location-model fixed intercept. $\gamma_{01}$ = location-model fixed effect for level-2 predictor $X_i$. $u_{0,i}$ = location-model random intercept for individual $i$. $\alpha_{0}^{u_0}$ = log of the scale-model fixed intercept for the location-model residual variance. $\alpha_{0}^{\omega_0^e}$ = log of the scale-model fixed intercept for the scale-model random intercept. $\alpha_{0}^{u_0;\omega_0^e}$ = inverse hyperbolic tangent of the scale-model fixed intercept for the correlation between the location- and scale-model random intercepts. $\delta_{00}^{e}$ = log of the scale-model fixed intercept for the residual variance. $\delta_{01}^{e}$ = log of the scale-model fixed effect for level-2 predictor $X_i$. $\omega_0^e$ = scale-model random intercept for individual $i$.

In addition, the 95% credible interval for the (inverse hyperbolic tangent of the) correlation between the location- and scale-model random intercepts, $\alpha_0^{u_0;\omega_0^e}$, included the true sampled parameter for only 81.55% of replications, and bias estimates indicated that model 1 underestimated the true parameter of $\alpha_0^{u_0;\omega_0^e}$ by an average of 0.05 (or 4.88% on the variance scale; $[1 - \exp(-0.05)] * 100$). Both results were hypothesized to result from the overestimation of the (log of the) scale-model random intercept variance, $\alpha_0^{\omega_0^e}$, by an average of 15.03% (i.e., $[\exp(0.14) - 1] * 100$), which affected the estimated correlation between location- and scale-model random intercepts. Further, although the overestimation of $\alpha_0^{\omega_0^e}$ by 15.03% appears significant, it represented an average difference between the true sampled value and the model-estimated value of 0.011 on the variance-scale (mean $\sigma^2_{\omega_0^e}$ of true sampled values = 0.076 vs. mean $\sigma^2_{\omega_0^e}$ estimated by model 1 = 0.087). Therefore, considering $\alpha_0^{\omega_0^e}$ had a convergence rate of 99.38% and recovery of 92.95%, the overestimation of $\alpha_0^{\omega_0^e}$ was not considered to bias the power estimates presented below; however, this overestimation was noted when appropriate.

***The power to detect the scale-model random intercept variance.*** Considering all 4,000 replications, the average decrease in DIC between models 2 and 3 was 695.53, 95% CI [644.12, 746.95], with an overall empirical power rate to detect the scale-model random intercept variance of 89.93%, 95% CI [88.95%, 90.84%], $n_{\text{significant}}$ = 3,597.

Univariate (or marginal) power curves based on the detection the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, were calculated across bins for the number of individuals, the number of repeated occasions, the fixed effect of the level-2 predictor included in the location model, and the fixed intercept for the residual variance. Each

univariate power curve was essentially flat, with an empirical power rate of approximately 90% across bins (as stated above, the average empirical power rate to detect $\sigma^2_{\omega^e_0}$ was 89.93%). An exemplar of the functional form of this power curve is presented in Figure 4.2 for the (log of the) fixed intercept for the residual variance estimated by model 3. Overall, the flat function forms of these power curves at the average power to detect $\sigma^2_{\omega^e_0}$ were hypothesized to result from each power curve being marginalized across the remaining sampled parameters.

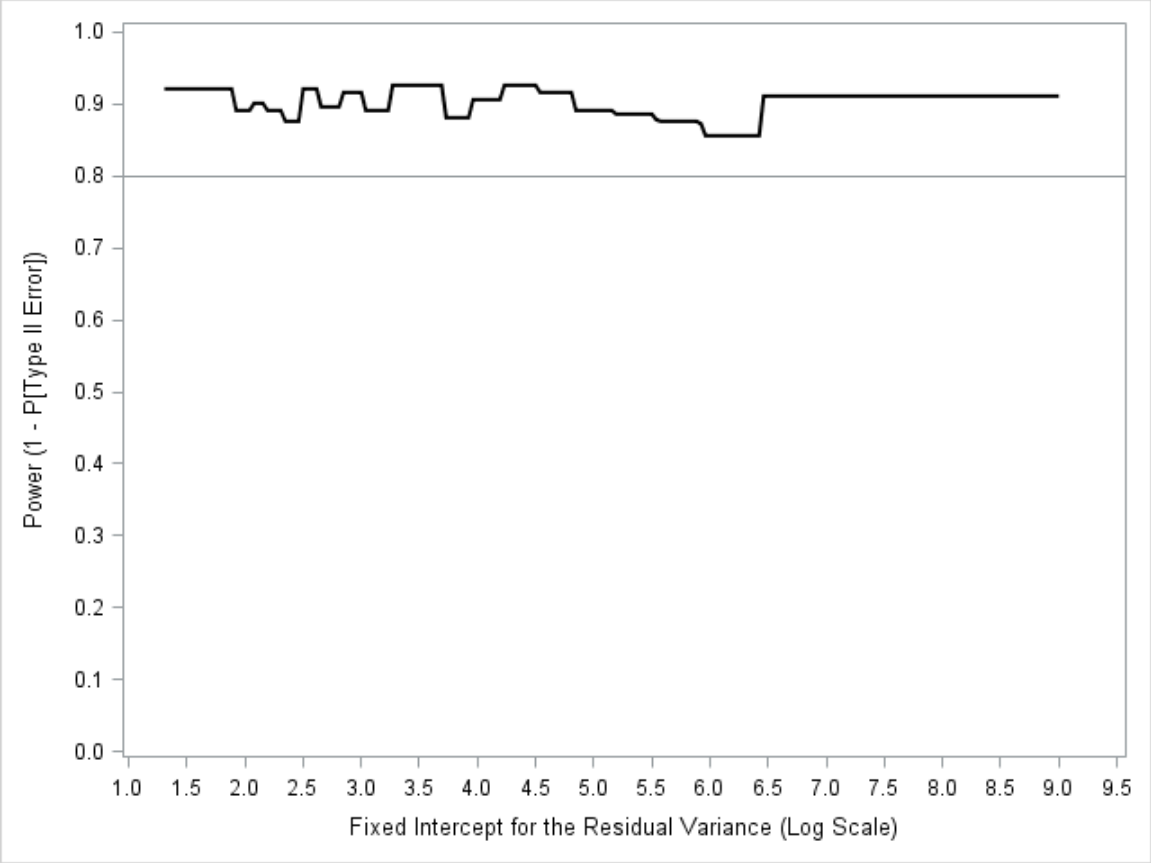


*Figure 4.2*. Power curve to detect scale-model random intercept variance by the fixed intercept for the residual variance (on the log scale)

Similarly, for the correlation between the location- and scale-model random intercepts, an empirical power rate was approximately 90% for both conditions.

Specifically, when this correlation was sampled to be 0, the power to detect $\sigma^2_{\omega_0^e}$ was 89.78%, 95% CI [88.39%,91.05%] and 90.08%, 95% CI [88.66%,91.38%] when this correlation was sampled to be 0.50; no statistically significant difference in power was indicated between the sampled correlation conditions, $X^2$ (1, $N$ = 4,000) = 0.10, $p$ = 0.75.

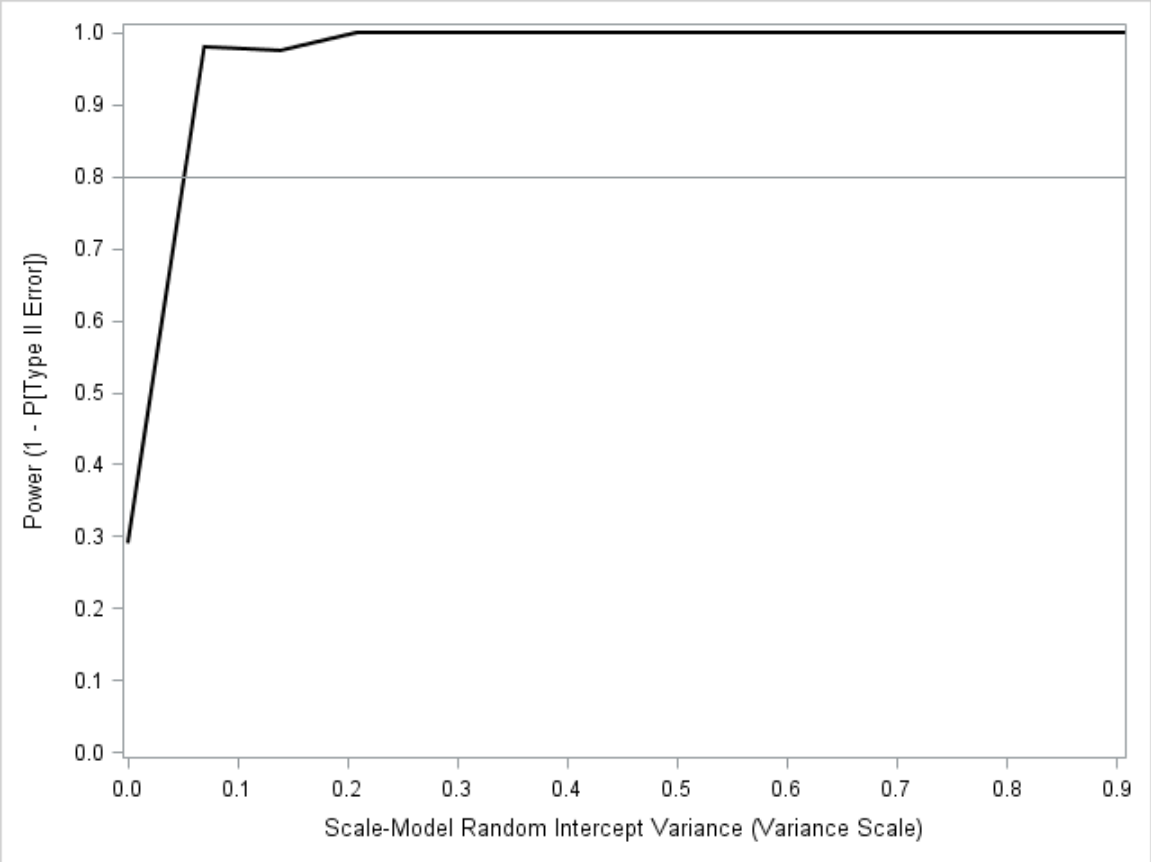


*Figure 4.3*. Power to detect scale-model random intercept variance (on variance scale)

By contrast, Figure 4.3 presents the univariate power curve to detect the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, by the amount of $\sigma^2_{\omega_0^e}$ estimated by model 3, in which the $x$-axis is on the variance scale, not the log scale. Here, the power to detect $\sigma^2_{\omega_0^e}$ increases as estimated $\sigma^2_{\omega_0^e}$ increases from 0.00 to 0.10 before reaching 100% power when $\sigma^2_{\omega_0^e}$ was estimated by model 3 to be at least 0.15.

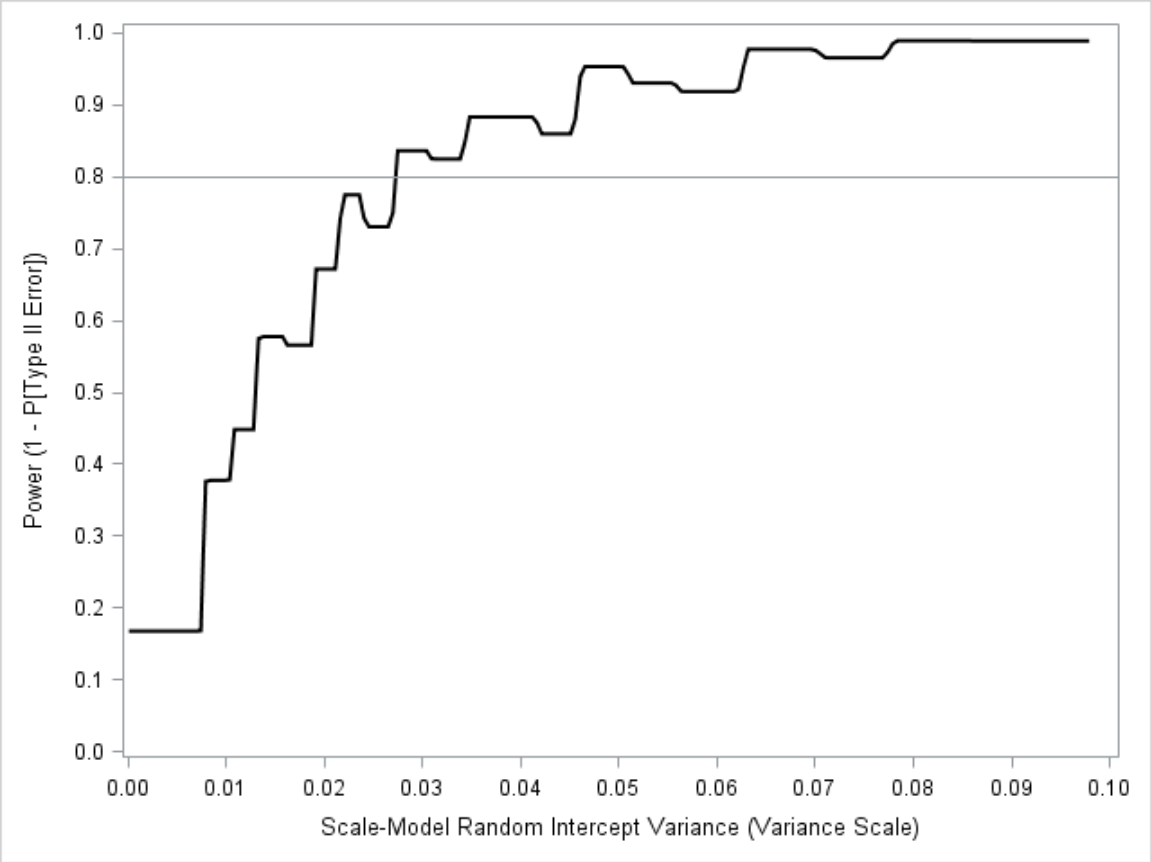


*Figure 4.4.* Power to detect scale-model random intercept variance (on variance scale) showing power increases from 0.00 to 0.10 ($n$ = 1,696)

Figure 4.4 presents the univariate power curve to detect the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, by $\sigma^2_{\omega_0^e}$ ranging from 0.00 to 0.10 ($n$ = 1,696). Here, 80% power to detect $\sigma^2_{\omega_0^e}$ is achieved when $\sigma^2_{\omega_0^e}$ was estimated to be approximately 0.03 (or, when taking into account the 15% overestimation, 80% power was achieved when $\sigma^2_{\omega_0^e}$ was approximately 0.026). This small variance estimate to achieve 80% power to detect $\sigma^2_{\omega_0^e}$ was hypothesized to result specifically from the averaging of the four other sampled parameters across replications. Therefore, a multivariable logistic regression model was estimated in which the sampled parameters were modeled on their continuous scales (due non-convergence of the local scoring algorithm from the GAM and quasi-separation

when using binned sampled parameters) to determine which sampled parameters estimated by model 3 moderated the relationship between $\sigma^2_{\omega^e_0}$ estimated by model 3 and whether $\sigma^2_{\omega^e_0}$ was detected via DIC model comparison. The results of the final multivariable logistic regression model are presented in Table 4.4, in which only replications for which $\sigma^2_{\omega^e_0}$ was less than 0.10 were considered ($n$ = 1,696). Note that both the number of individuals and the number of occasions were centered at their lowest possible sampled values of 25 and 5, respectively, and that $\sigma^2_{\omega^e_0}$ was left uncentered given 0 was possible (centering in the final model did not affect the inference of the two-way interaction effects). Further, the effect of $\sigma^2_{\omega^e_0}$ is reported for a 0.01-unit increase given $\sigma^2_{\omega^e_0}$ ranged from 0.00 to 0.10 in these models.

Table 4.4
*Logistic Regression Results Predicting whether the Scale-Model Random Intercept Variance was Detected as Estimated by Model 3 (n = 1,696)*

| | | | 95% Credible Interval | |
|---|---|---|---|---|
| | Log-Odds | SE | Lower | Upper |
| Intercept | –7.31 | 0.59 | –8.46 | –6.16 |
| $N_{individuals}$ (0=25) | **0.03** | 0.00 | 0.02 | 0.04 |
| $n_i$ (0=5) | 0.01 | 0.01 | –0.02 | 0.04 |
| $\sigma^2_{\omega^e_0}$ | **0.56** | 0.12 | 0.33 | 0.78 |
| $\sigma^2_{\omega^e_0} * N_{individuals}$ | **0.01** | 0.00 | 0.00 | 0.01 |
| $\sigma^2_{\omega^e_0} * n_i$ | **0.09** | 0.01 | 0.07 | 0.11 |

*Note*. Bold font indicates $p < .20$. $N_{individuals}$ = number of individuals. $n_i$ = number of repeated occasions within an individual. $\sigma^2_{\omega^e_0}$ = scale-model random intercept variance from model 3.

Results indicated that the effect of $\sigma^2_{\omega^e_0}$ estimated by model 3 was moderated by the number of individuals and the number of repeated occasions; no other two-way or higher interactions achieved $p < 0.20$. Interpretation of specific effects from this final model is not provided due to the violation of linearity of the logit for all sampled parameters (i.e., non-linear functional form as indicated by the Box-Tidwell test). In

general, however, the power to detect $\sigma^2_{\omega_0^e}$ by the amount of $\sigma^2_{\omega_0^e}$ increased with increases in either the number of individual or the number of repeated occasions. In addition, the power curves presented below for the two-way interactions are based on the binned versions of the sampled parameters, not continuous versions; thus, these power curves are simply *approximations* of the two-way interaction effects indicated by the multivariable logistic regression analysis.

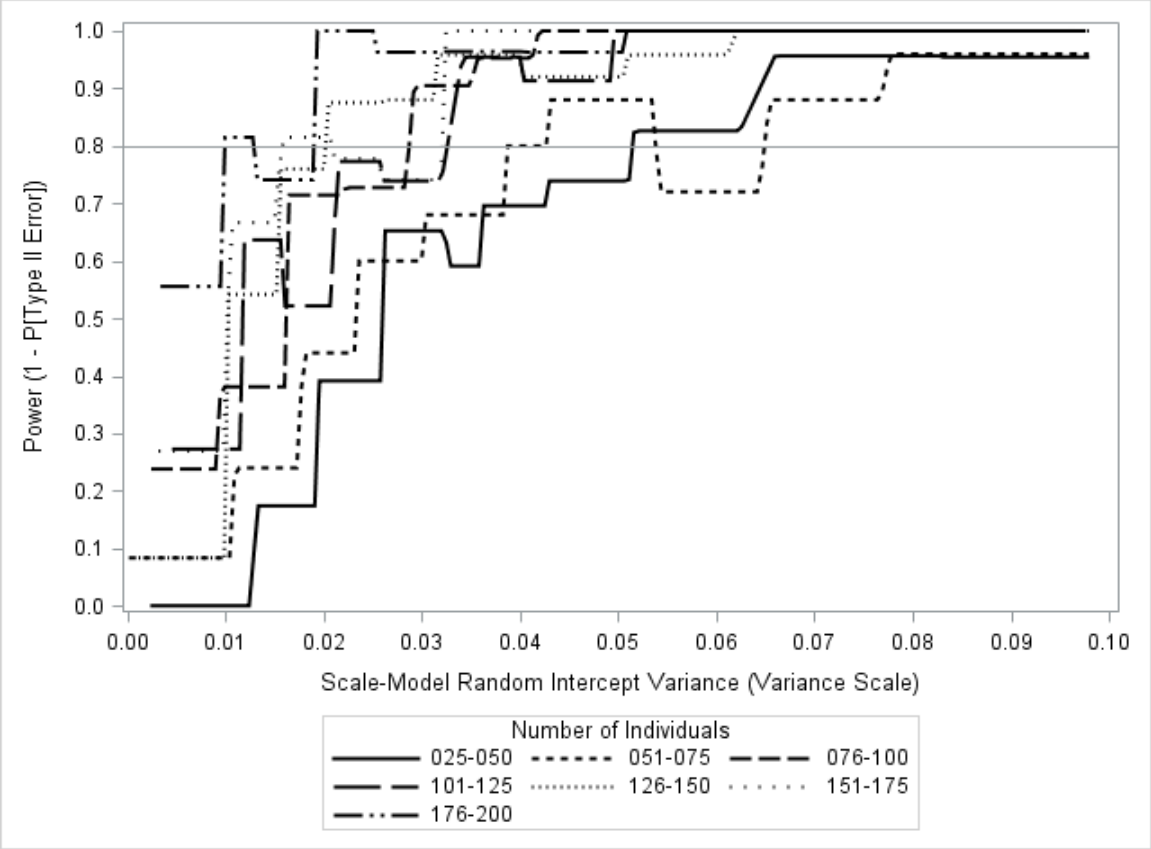


*Figure 4.5*. Power to detect the scale-model random intercept variance by the number of individuals ($n$ = 1,696)

An approximation of the interaction effect between the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, and the number of individuals is shown graphically by the power curves presented in Figure 4.5. Here, the number of individuals was binned into groups of 25 to prevent congestion of power curves. This resulted in 7 total power curves, which

indicated that the power to detect $\sigma^2_{\omega_0^e}$ increased more quickly for larger sample sizes. For example, sample sizes ranging from 25 to 50 achieved 80% power to detect $\sigma^2_{\omega_0^e}$ when $\sigma^2_{\omega_0^e}$ was approximately 0.055 (or 0.047 when taking into account 15% overestimation observed by the true model), whereas sample sizes ranging from 176 to 200 achieved 80% power to detect $\sigma^2_{\omega_0^e}$ when $\sigma^2_{\omega_0^e}$ was approximately 0.02 (or 0.017 when taking into account 15% overestimation by the true model).

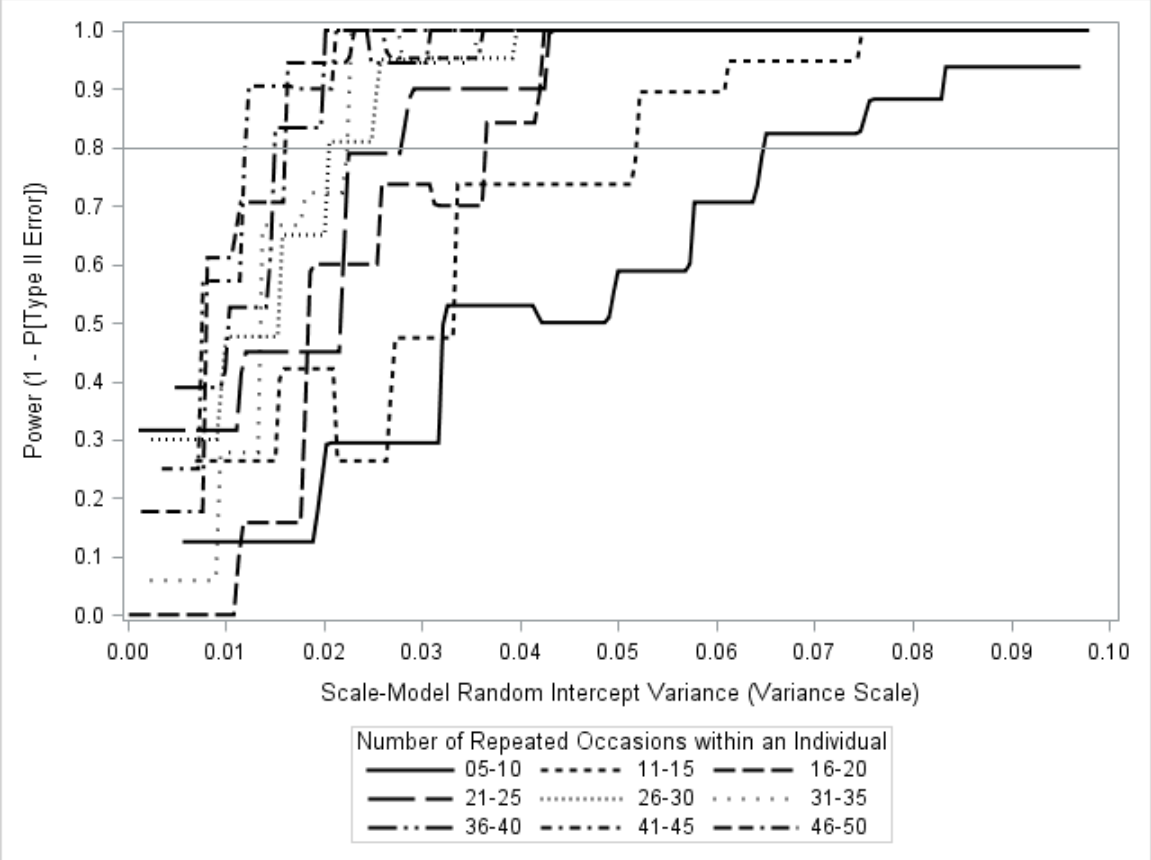


*Figure 4.6.* Power to detect the scale-model random intercept variance by the number of repeated occasions

An approximation of the interaction effect between scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, and the number of repeated occasions within an individual is shown graphically by the power curves in Figure 4.6. Here, repeated occasions were binned into

groups of 5 to prevent congestion of the power curves; 9 power curves were calculated, in which the power to detect $\sigma^2_{\omega_0^e}$ increased more quickly with more repeated occasions. For example, for 5 to 10 occasions, 80% power to detect $\sigma^2_{\omega_0^e}$ was achieved when $\sigma^2_{\omega_0^e}$ was approximately 0.065 (or 0.055 when considering the 15% overestimation by the true model), whereas for 46 to 50 occasions, 80% power to detect $\sigma^2_{\omega_0^e}$ was achieved when $\sigma^2_{\omega_0^e}$ was approximately 0.015 (or 0.013 when considering 15% overestimation by the true model).

***The power to detect the effect for a level-2 predictor included in the scale model for the residual variance.*** Using only replications in which the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, was significant ($n$ = 3,597), the empirical power rate for the fixed effect for the level-2 predictor in the scale model for the residual variance, $\delta^e_{01}$, was approximately 26.69%, 95% CI [25.25%,28.17%], $n_{\text{significant}}$ = 960. This result was expected given that the zero-inflated Poisson distribution from which $\delta^e_{01}$ was sampled resulted in a majority of zeros being sampled ($n_{zeros}$ = 2,733; 75.98%). In addition, for replications in which $\sigma^2_{\omega_0^e}$ was detected and $\delta^e_{01}$ was sampled to be 0 ($n$ = 2,733), the Type I error rate for $\delta^e_{01}$ estimated by the model 1 (i.e., the true model) was 3.51%, 95% CI [2.85%,4.27%], which was close to the expected 5%.

As a replication of Leckie (2014) and Leckie et al. (2014), considering replications where $\sigma^2_{\omega_0^e}$ was detected and where the true value of $\delta^e_{01}$ was sampled to be 0 ($n$ = 2,733), the Type I error rate from model 4 that erroneously omitted $\sigma^2_{\omega_0^e}$ was 14.82%, 95% CI [13.51%,16.21%]. This finding supported the results of Leckie (2014) and Leckie et al. (2014), albeit with a smaller Type I error rate than they reported, and indicated that

erroneously omitting $\sigma^2_{\omega_0^e}$ considerably increases Type I error rates for level-2 fixed effects included in the scale model for the residual variance.

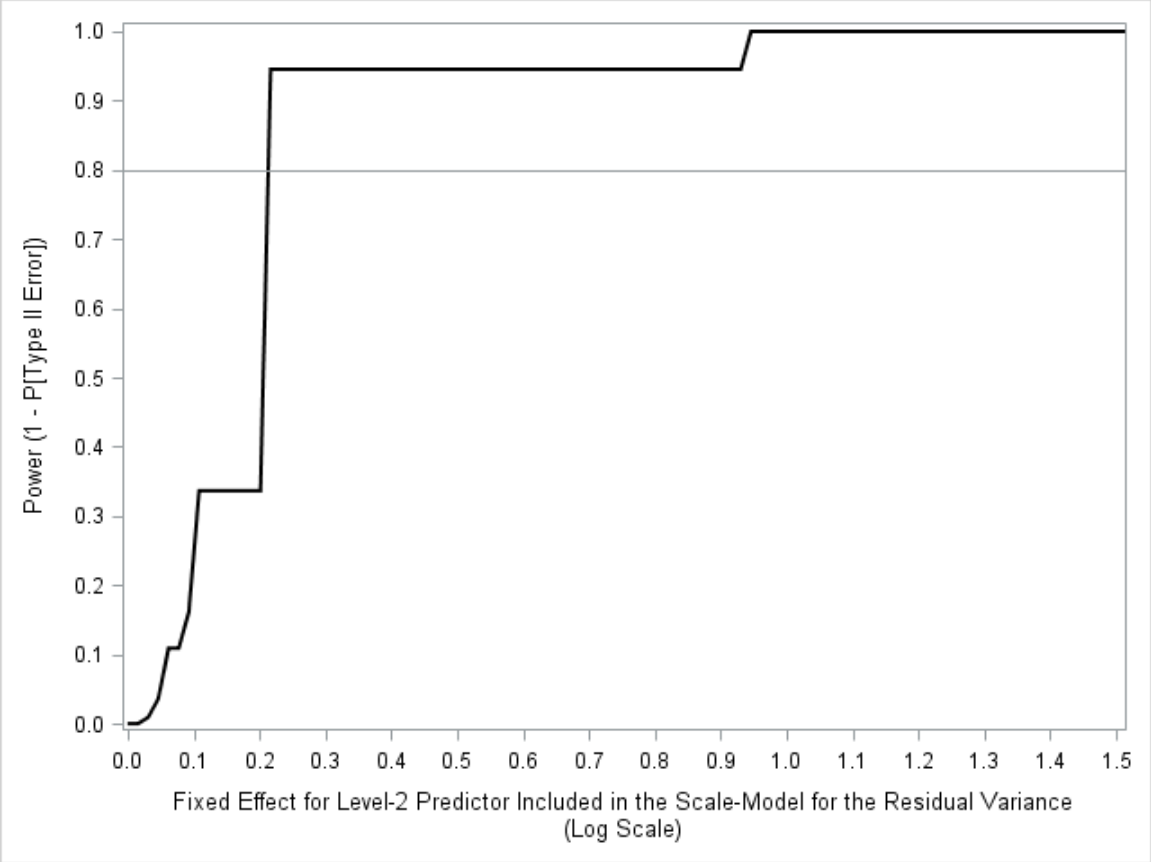


*Figure 4.7.* Power to detect a level-2 predictor of scale-model residual variance by the effect of the predictor estimated by model 1 ($n$ = 3,597)

Considering only replications in which the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, was detected ($n$ = 3,597), the estimated univariate power curve for the binned fixed effect of the level-2 predictor included in the scale model for the residual variance, $\delta^e_{01}$, estimated by model 1 is shown in Figure 4.7 (in which the $x$-axis is the actual parameter estimate for $\delta^e_{01}$ on the log scale). The parameter estimate of $\delta^e_{01}$ had inadequate statistical power when $\delta^e_{01}$ was less than 0.25, and near 100% power when $\delta^e_{01}$ was greater than or equal to 0.25.

The parameter estimate for the fixed effect of the level-2 predictor included in the scale model for the residual variance, $\delta_{01}^{e}$, was presented initially due to the unknown behavior of pseudo-$R^2$ calculated for the reduction of scale-model random intercept variance, $\sigma_{\omega_0^e}^2$. With that in mind, pseudo-$R^2$ was calculated by substituting $\alpha_0^{\omega_0^e}$ for $\alpha_0^{u_0}$ in (4.15) by comparing the reduction in $\sigma_{\omega_0^e}^2$ estimated by models 1 and 3. The univariate power curve across binned values of $\delta_{01}^{e}$ is shown in Figure 4.8.

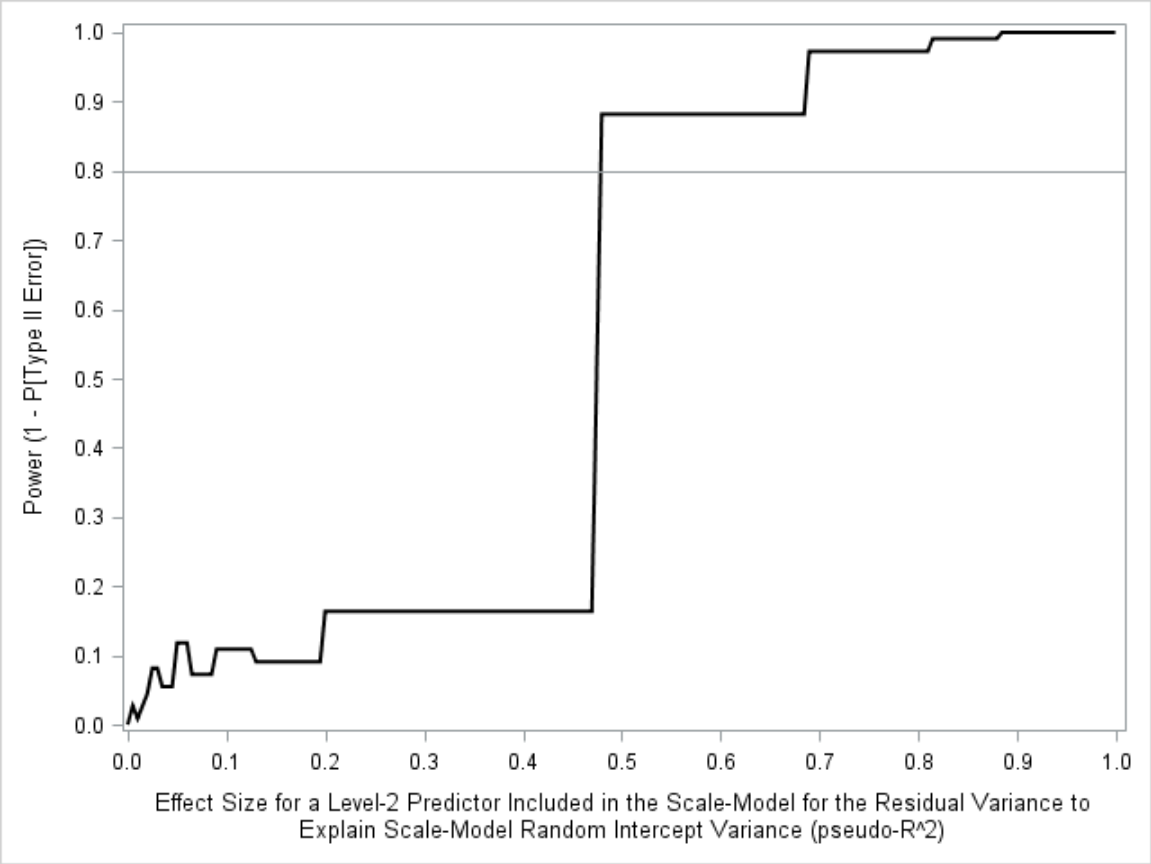


*Figure 4.8.* Power to detect scale-model fixed effect for level-2 predictor by the predictor's effect size estimate

Results indicated a median reduction of scale-model random intercept variance of 0.13, IQR [0.02,0.88], in which the parameter estimate of $\delta_{01}^{e}$ had inadequate statistical power when $\delta_{01}^{e}$ explained less than 50% of scale-model random intercept variance, $\sigma_{\omega_0^e}^2$,

and near 100% power when $\delta_{01}^{e}$ explained more than 50% of $\sigma_{\omega_0^e}^2$. Although this power curve is similar in functional form to Figure 4.7, this result does not provide explicit confirmation of the consistency of calculating pseudo-$R^2$ for $\sigma_{\omega_0^e}^2$; thus, all power curves presented below are based on the binned parameter estimate of $\delta_{01}^{e}$ estimated by model 1.

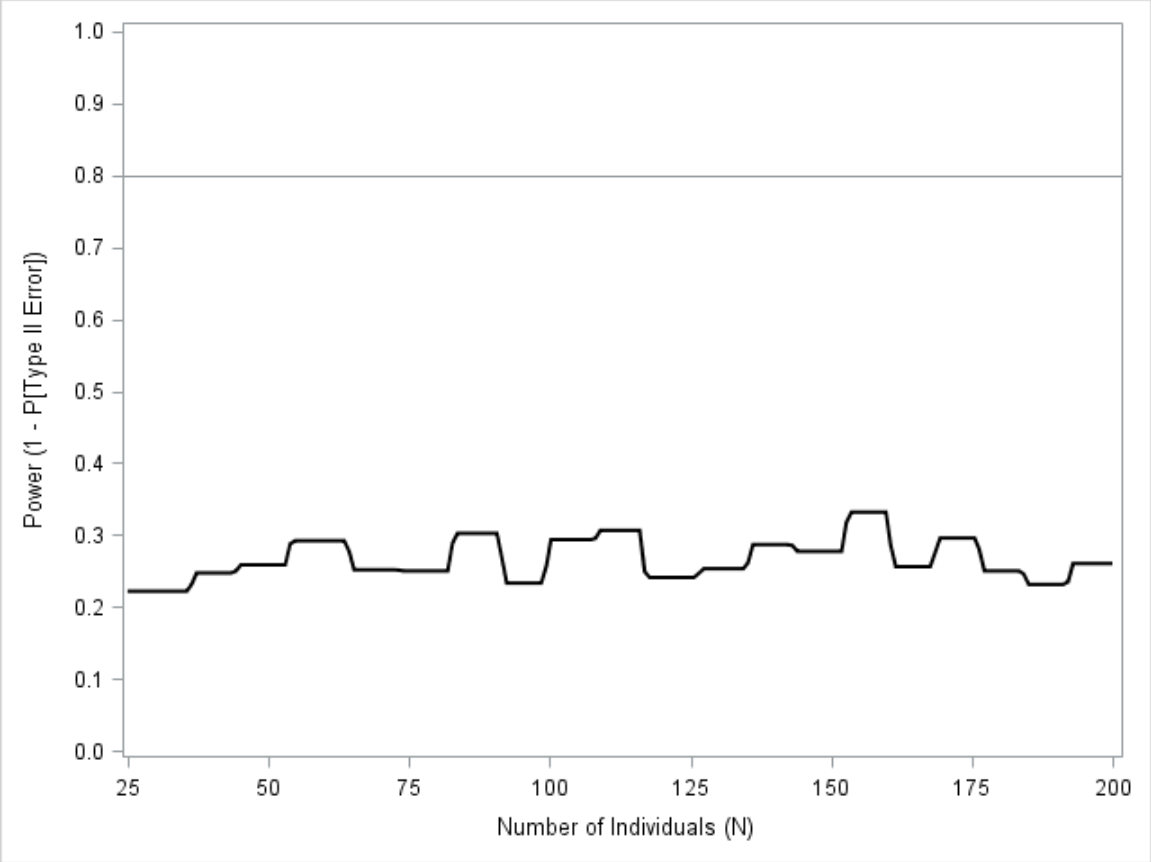


*Figure 4.9.* Power to detect a level-2 predictor of scale-model residual variance by the number of individuals

The remaining power curves for the fixed effect of the level-2 predictor included in the scale model for the residual variance, $\delta_{01}^{e}$, were calculated across binned values of the number of individuals, the number of repeated occasions, the fixed intercept for the residual variance, and the scale-model random intercept variance. All power curves were nearly identical, with empirical power rates of approximately 25% across bins for these

sampled parameters (as stated above, the average empirical power rate to detect $\delta_{01}^{e}$ was 26.69%). An exemplar of this functional form is presented for the number of individuals in Figure 4.9. Similar results were indicted for the correlation between the location- and scale-model random intercepts, $\rho_{u_0 w_0}$; when $\rho_{u_0 w_0}$ was sampled to be 0 the empirical power to detect $\delta_{01}^{e}$ was 27.15%, 95% CI [25.12%,29.17%], and empirical power was 26.20%, 95% CI [24.15%,28.34%] when $\rho_{u_0 w_0}$ was sampled to be 0.50; no statistically significant difference in the empirical power rate was indicated between conditions, $X^2$ (1, $N$ = 3,597) = 0.41, $p$ = 0.52. When considered together, similar to the univariate power curves for detecting the scale-model random intercept variance, the empirical power estimates consistently near 25% representing the average power to detect $\delta_{01}^{e}$ was hypothesized to result from the univariate power curves being marginalized across the other sampled parameters.

Because the univariate power curve presented in Figure 4.7 for the fixed effect of the level-2 predictor included in the scale model for the residual variance, $\delta_{01}^{e}$, assumed all sampled parameters were averaged across replications, the power to detect $\delta_{01}^{e}$ may have been estimated too simplistically. Given that power increased for estimated $\delta_{01}^{e} <$ 0.25, jumped vertically for estimated $\delta_{01}^{e} \approx 0.25$, and then appeared to plateau near 100% for estimated $\delta_{01}^{e} \geq 0.25$, a series of piecewise logistic regression models were estimated to determine which continuous sampled parameters moderated whether $\delta_{01}^{e}$ estimated by model 1 was significant. Similar to the logistic model for the scale-model random intercept variance, all sampled parameters were modeled on continuous scales as a result of the non-convergence of the local scoring algorithm from the GAM and quasi-separation using binned sampled parameters. Of the 3,597 replications in which the scale-

model random intercept variance, $\sigma^2_{\omega^e_0}$, was detected, 2,724 (75.73%) estimated the fixed effect for the level-2 predictor included in the scale model for the residual variance, $\delta^e_{01}$, to be less than 0.25 (*M* = –0.00, *SD* = 0.06, *min* = –0.38, *max* = 0.24), whereas 873 (24.27%) replications estimated $\delta^e_{01} \geq 0.25$ (*M* = 1.14, *SD* = 0.41, *min* = 0.25, *max* = 3.10). The results of the final piecewise logistic regression model are presented in Table 4.5. For the final piecewise logistic model, the total number of individuals and the number of repeated occasions within an individual were centered at their lowest sampled values (i.e., 25 and 5, respectively), whereas both $\delta^e_{01}$ and scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, were left uncentered given zero was a possible sampled value. Further, the effect of both $\delta^e_{01}$ and $\sigma^2_{\omega^e_0}$ are reported for a 0.01-unit increase.

Table 4.5
*Piecewise Logistic Regression Results Predicting whether the Scale-Model Fixed Effect for a Level-2 Predictor was Detected from Model 1 with a Breakpoint at 0.25 (n = 3,597)*

| | | | 95% CI | |
|---|---|---|---|---|
| | Log-Odds | SE | Lower | Upper |
| Intercept: $\delta^e_{01} < 0.25$ | –23.28 | 3.97 | –31.06 | –15.51 |
| Intercept: $\delta^e_{01} \geq 0.25$ | –12.34 | 3.04 | –18.30 | –6.38 |
| $\delta^e_{01} < 0.25$ | –0.10 | 0.18 | –0.46 | 0.25 |
| $\delta^e_{01} \geq 0.25$ | 0.10 | 0.12 | –0.13 | 0.34 |
| $N_{individuals}$ (0=25) | **0.21** | 0.03 | 0.15 | 0.27 |
| $n_i$ (0=5) | **0.43** | 0.09 | 0.26 | 0.61 |
| $\sigma^2_{\omega^e_0}$ | **–0.10** | 0.01 | –0.13 | –0.07 |
| $(\delta^e_{01} < 0.25) * N_{individuals}$ | **0.01** | 0.00 | 0.01 | 0.01 |
| $(\delta^e_{01} < 0.25) * n_i$ | **0.01** | 0.01 | 0.00 | 0.02 |
| $(\delta^e_{01} < 0.25) * \sigma^2_{\omega^e_0}$ | **–0.34** | 0.07 | –0.47 | –0.22 |

*Note*. Bold font indicates $p < .20$. $\delta^e_{01}$ = log of the scale-model fixed effect for level-2 predictor $X_i$. $N_{individuals}$ = number of individuals. $n_i$ = number of repeated occasions within an individual. $\sigma^2_{\omega^e_0}$ = scale-model random intercept variance from model 3.

Two-way interactions were indicated between $\delta^e_{01} < 0.25$ and the number of individuals, occasions, and the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$; no

interaction effects were indicated for $\delta_{01}^{e} \geq 0.25$ and no additional two-way or higher interactions were indicated. Interpretations for effects in this final model are not provided due to the non-linear functional form for all sampled parameters as indicated by the Box-Tidwell test. In general, however, the power to detect $\delta_{01}^{e} < 0.25$ increased with increases in the number of individual or the number of repeated occasions and decreased with increases in $\sigma_{\omega_0^e}^2$. In addition, power curves for the two-way interactions are based on the binned versions of the sampled parameters, not continuous versions, and thus are simply approximations of the interaction effects indicated by the final piecewise model.

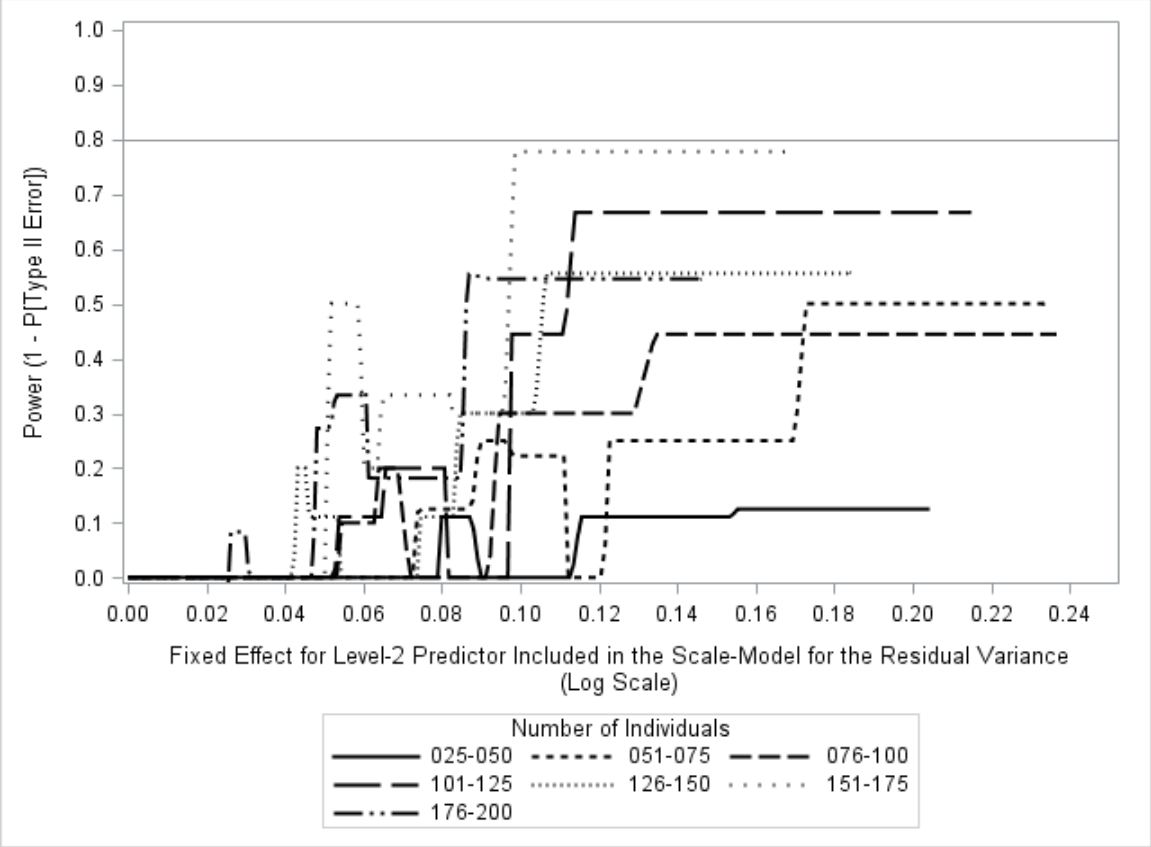


*Figure 4.10.* Power to detect the scale-model fixed effect for the level-2 predictor by the number of individuals

An approximation of the two-way interaction between the fixed effect of the level-2 predictor included in the scale model for the residual variance, $\delta_{01}^{e}$, estimated to be < 0.25, and the number of individuals is shown graphically in Figure 4.10. Here, the number of individuals was binned into groups of 25 to prevent congestion of the power curves; this resulted in 7 total power curves. In general, the power to detect smaller values of $\delta_{01}^{e}$ increased more quickly as the number of individuals increased; however, even replications with samples ranging from 175 to 200 individuals failed to achieve 80% statistical power when $\delta_{01}^{e}$ < 0.25.

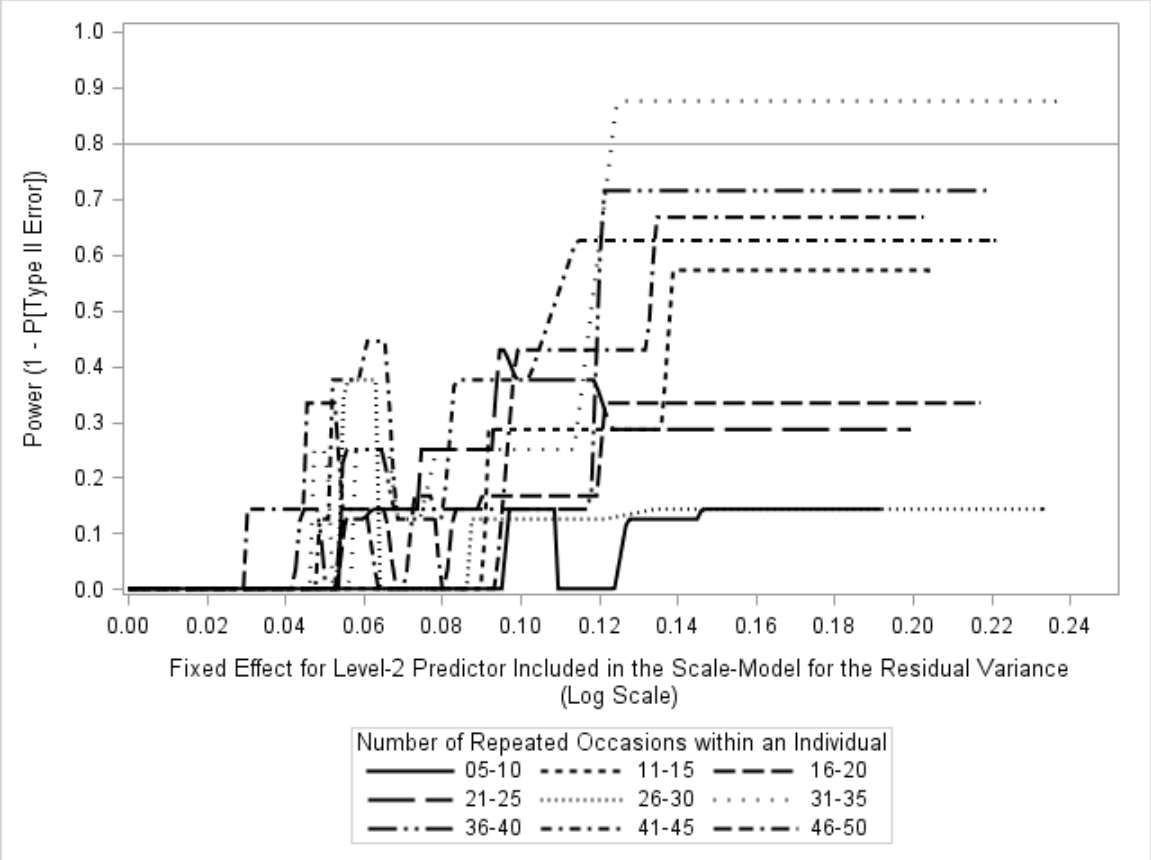


*Figure 4.11.* Power to detect the scale-model fixed effect for the level-2 predictor by the number of repeated occasions

A similar finding was observed for the approximation of the two-way interaction between $\delta_{01}^{e}$ < 0.25 and the number of repeated occasions within an individual shown Figure 4.11. Here, the power to detect smaller values of $\delta_{01}^{e}$ increased more quickly as the number of occasions increased. Only occasions ranging from 31 to 35 achieved 80% power when $\delta_{01}^{e}$ < 0.25; however, this was likely a spurious result given the significant increase in power between $\delta_{01}^{e} = 0.11$ and $\delta_{01}^{e} = 0.12$ for this specific range of occasions.

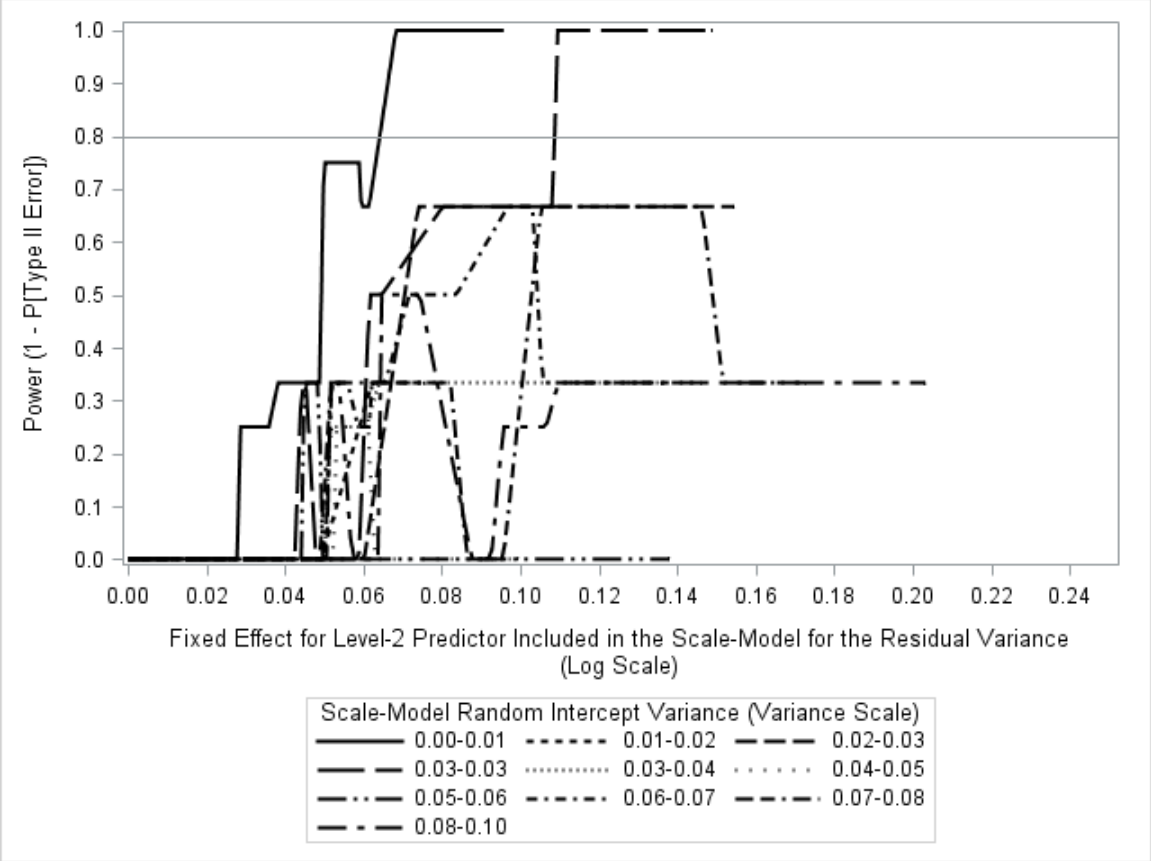


*Figure 4.12.* Power to detect the scale-model fixed effect for the level-2 predictor by the scale-model random intercept variance

Finally, an approximation of the interaction between the fixed effect of the level-2 predictor in the scale model for the residual variance, $\delta_{01}^{e}$, estimated to < 0.25 and the scale-model random intercept variance, $\sigma_{\omega_0^e}^2$, is presented in Figure 4.12. Although the

pattern of interaction is not clearly indicated, especially at smaller values of $\delta_{01}^{e}$, the power to detect $\delta_{01}^{e}$ increases more quickly for smaller values of $\sigma_{\omega_0^e}^2$. For $\sigma_{\omega_0^e}^2$ ranging from 0.00 to 0.01, 80% power to detect $\delta_{01}^{e}$ was achieved when $\delta_{01}^{e}$ was approximately 0.06. These results suggested that increasing $\sigma_{\omega_0^e}^2$ resulted in smaller parameter estimates for $\delta_{01}^{e}$ to become undetectable.

***The ability to detect the fixed effect for a level-2 predictor included in the scale model for the residual variance when the scale-model random intercept variance cannot be detected.*** Of the 4,000 replications in this study, model 3 indicated that 3,597 (89.93%) replications detected significant scale-model random intercept variance, $\sigma_{\omega_0^e}^2$. For the 403 replications in which model 3 did not detect $\sigma_{\omega_0^e}^2$, the true sampled value of the fixed effect for the level-2 predictor included in the scale model for the residual variance, $\delta_{01}^{e}$, was 0. Thus, for these 403 replications, model 1 and model 4 (i.e., the only models to include $\delta_{01}^{e}$) should never have estimated $\delta_{01}^{e}$ as statistically significant, and therefore any statistically significant estimate for $\delta_{01}^{e}$ would be a Type I error. With this in mind, if $\sigma_{\omega_0^e}^2$ was not detected but retained in the model, as in model 1, the Type I error rate for $\delta_{01}^{e}$ was 0.99%, 95% CI [0.27%,2.52%], whereas if $\sigma_{\omega_0^e}^2$ was not detected and omitted from the model, as in model 4, the Type I error rate for $\delta_{01}^{e}$ increased to 2.48%, 95% CI [1.20%,4.52%]. Taken together, these results suggest that $\sigma_{\omega_0^e}^2$ must be detected before it can be predicted, and that if $\sigma_{\omega_0^e}^2$ is not detected and either retained or omitted, continuing to model between-individual differences in residual variability with a level-2 predictor is most likely unproductive, although this does not entirely preclude continuing to evaluate for systematically-varying effects of level-2 predictors.

***The ability to detect a truly systematically-varying fixed effect for a level-2 predictor included in the scale model for the residual variance.*** Of the 3,597 replications in which scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, was detected, 177 of these replications (4.92%) indicated that the inclusion of the level-2 predictor included in the scale model for the residual variance, $\delta^e_{01}$, explained the majority of $\sigma^2_{\omega^e_0}$, such that a non-significant proportion of unexplained $\sigma^2_{\omega^e_0}$ remained (as indicated by comparing DIC between models 1 and 4). This result indicated that for these 177 replications, individual differences in residual variability were not random, but instead were truly systematically varying by $\delta^e_{01}$. For these 177 replications, the consequences of removing the remaining non-significant scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, from the model was examined. Specifically, the Type I error rate for a level-2 predictor included in the scale model for the residual variance, $\delta^e_{01}$, estimated by true model 1 that included $\sigma^2_{\omega^e_0}$ was lower than the Type I error rate for $\delta^e_{01}$ estimated by model 4 that excluded $\sigma^2_{\omega^e_0}$ (9.72%, 95% CI [4.00%,19.01%] vs. 13.89%, 95% CI [6.87%,24.06%], respectively). Although both Type I error rates were considerably higher than nominal 5% for the majority of the confidence interval, these results suggest that if residual heterogeneity is indicated to vary systematically (but not randomly otherwise), non-significant $\sigma^2_{\omega^e_0}$ should be retained in subsequent models to provide better (albeit, potentially inadequate) control of the Type I error rate.

**Discussion of the first simulation study.** The first simulation study set out to accomplish three specific aims identified to be lacking in the methodological literature regarding the mixed-effects location-scale model. First, this study calculated the

statistical power to detect the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, and further evaluated how the power function for $\sigma^2_{\omega^e_0}$ varied by several study design characteristics, as well as model parameters that would typically be estimated in a mixed-effects location-scale model. Results indicated that the power to detect $\sigma^2_{\omega^e_0}$ increased with concurrent increases in $\sigma^2_{\omega^e_0}$, which was moderated by the number of individuals as well as by the number of repeated occasions within an individual. Taken together, these results were not surprising as they align with statistical power theory for typical repeated-measures analyses. That is, more data generally increases the probability of finding a statistically significant effect. Therefore, based on the results shown in Figures 4.5 and 4.6, it is recommended that at least 100 individuals measured at a minimum of 20 occasions should result in approximately 80% power to detect $\sigma^2_{\omega^e_0}$ as small as 0.035 (i.e., the average $\sigma^2_{\omega^e_0}$ indicated in this study).

Second, the power to detect the effect of a level-2 predictor included in the scale model for the residual variance, $\delta^e_{01}$, was evaluated only for replications in which the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, was detected (although it could have been evaluated for all replications given that $\delta^e_{01} = 0$ for all replications in which $\sigma^2_{\omega^e_0}$ was not detected). Results indicated that power increased substantially when parameter estimates of $\delta^e_{01}$ were > 0.25 (or $\delta^e_{01}$ explained at least 50% of $\sigma^2_{\omega^e_0}$). A subsequent piecewise logistic regression model indicated that power was approximately 100% for replications in which $\delta^e_{01} \geq 0.25$, a result not moderated by any study design characteristic or model parameter. A more interesting result was that the power to detect $\delta^e_{01} < 0.25$ increased with increases in the number of individuals or in the number of repeated occasions within

an individual, but decreased with increases in $\sigma^2_{\omega_0^e}$ (i.e., more scale-model random intercept variance resulted in smaller $\delta_{01}^e$ parameter estimates to become undetectable). These results provided direct insight into how and when Type I errors for $\delta_{01}^e$ are likely to occur. That is, the true sampled value for $\delta_{01}^e$ was 0 for all replications in which $\delta_{01}^e$ was estimated to be less than 0.25. Thus, a Type I error for $\delta_{01}^e$ became more likely with increased power to detect $\delta_{01}^e$ resulting from drastic increases in the number of individuals or occasions and decreases in $\sigma^2_{\omega_0^e}$ (although power remained less than 80% for sampled values of individuals and occasions in this study). This statement, of course, assumes the empirical scientist is not interested in $\delta_{01}^e$ effects < 0.25, which may or may not be true given the specific field of study. In addition, given that 98.97% of replications for which $\delta_{01}^e$ was estimated to be > 0.25 had true sampled values greater than 0, the drastic increase in power observed when $\delta_{01}^e$ estimates were > 0.25 may indicate that 0.25 is the upper bound of estimation error when $\delta_{01}^e$ is actually 0.

Third, Type I error rates for the effect of a level-2 predictor included in the scale model for the residual variance, $\delta_{01}^e$, were evaluated using only replications in which the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, was not detected. The most interesting finding from this portion of the study was that all models in which $\sigma^2_{\omega_0^e}$ could not be detected had a true sampled value of $\delta_{01}^e$ = 0. This indicated that in the absence of significant $\sigma^2_{\omega_0^e}$, any statistically significant effect for $\delta_{01}^e$ was a Type I error; although Type I error rates were 0.99% for $\delta_{01}^e$ if $\sigma^2_{\omega_0^e}$ was not detected but retained in the model and only increased to 2.48% if $\sigma^2_{\omega_0^e}$ was not detected and omitted from the model. Taken together, these results suggested that if $\sigma^2_{\omega_0^e}$ cannot be detected initially, the inclusion of

level-2 predictors in the scale model for the residual variance to detect systematically-varying effects could be a potentially fruitless endeavor; however, Type I error rates were low regardless of whether $\sigma^2_{\omega^e_0}$ was retained or omitted. Therefore, if $\sigma^2_{\omega^e_0}$ was minimal (whether it was tested empirically or not), results suggested that when testing variability-related hypotheses, it is unlikely that systematically-varying effects would be detected for level-2 predictors included in the scale model for the residual variance.

Fourth, truly systematically-varying residual heterogeneity was indicated when significant scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, became non-significant upon the inclusion of the level-2 predictor in the scale-model for the residual variance, $\delta^e_{01}$. For these replications, the Type I error rate for $\delta^e_{01}$ was lower if non-significant $\sigma^2_{\omega^e_0}$ was retained in the model as opposed to being omitted from the model (9.72% vs. 13.89%, respectively, with considerable overlap in their confidence intervals). Although these results indicated control of Type I error rate was likely inadequate in the presence of systematically varying effects, it is recommended that the remaining non-significant $\sigma^2_{\omega^e_0}$ be retained in the model in the presence of truly systematically-varying effects.

The first simulation study defined the power to detect the scale-model random intercept variance and effect of a level-2 predictor included in the scale model for the residual variance for a range of study design characteristics as well as model parameters using a fixed location model. Thus, the logical next step was to inform empirical scientists about the model-building process of the mixed-effects location-scale model. Specifically, it is unknown how a misspecified location model affects the accuracy and inference of estimated scale-model fixed and random effects, or alternatively, how a misspecified scale model affects the accuracy and inferences of location-model fixed and

random effects. These inquiries are the primary purpose the second simulation study detailed next.

## Simulation Study II: The Consequences of Alternative Strategies for Estimating Location- and Scale-Model Fixed and Random Effects

In the model-building tradition, most empirical scientists attempt to identify the most inclusive yet parsimonious final model and therefore tend to remove non-statistically significant effects throughout the model-building process. To date, few methodological studies have provided insight directly into the model-building process of the mixed-effects location-scale model; however, some research has been provided by Cleveland et al. (2000), Leckie (2014), Leckie et al (2014). The literature search for this dissertation found no methodological study that explicitly evaluated the consequences of specifying the location model prior to the scale model (or vice versa), or that provided insight into whether the location and scale models should be built concurrently.

These model-building inquires essentially reduce to how misspecifying the location and/or scale model influence subsequent decisions about which of these parameters to retain during the model-building process. Therefore, the purpose of the second simulation study was to begin to address these concerns by evaluating whether a misspecified scale model affects the accuracy and inference of a fixed effect for a level-2 predictor included in the location model, and whether a misspecified location model affects the accuracy and inference of fixed and random effects included in the scale model for the residual variance.

**Description of individual parameters.** This second simulation study used the results of the first simulation study to inform the size of the parameters necessary to

ensure a conventional level of 80% statistical power to detect the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, given that random scale effects are the primary reason an empirical scientist would estimate the mixed-effects location-scale model. For this second simulation study, 100 individuals were observed for 20 repeated occasions (as recommended by the first simulation study), with the location-model random intercept variance, $\sigma^2_{u_0}$, and fixed intercept for the residual variance, $\delta^e_{00}$, both fixed at 1 (on the log scale); the correlation between the location- and scale-model random intercepts, $\rho_{u_0,\omega_0^e}$, was fixed to 0 given that the first simulation study indicated this correlation did not influence the power to detect $\sigma^2_{\omega_0^e}$.

Initially, the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, was fixed at 0.075 on the variance scale (or –2.59 on the log scale) for 1,000 replications. However, the statistical power to detect $\sigma^2_{\omega_0^e}$ was 99.80%, 95% CI [99.28%,99.98%], with drastically increased Type I error rates estimated by the true model for the fixed effect of a level-2 predictor included in both the location model and scale model for the residual variance (to be discussed in detail below). As a result, a second set of replications was simulated with $\sigma^2_{\omega_0^e}$ fixed to 0.035 on the variance scale (or –3.35 on the log scale), resulting in statistical power to detect $\sigma^2_{\omega_0^e}$ of 85.35%.

In addition to the value of the level-2 predictor $X_i$, which was sampled from a standard normal distribution defined by (4.13), the only model parameters sampled in the second simulation study were fixed effect of the level-2 predictor to be included in the location model and scale model for the residual variance, $\gamma_{01}$ and $\delta^e_{01}$, respectively. Both

parameter estimates were sampled from the zero-inflated Poisson distributions described in (4.14) and (4.16), respectively.

**Data generation.** All replications were simulated in SAS v. 9.4, with the level-2 predictor, $X_i$, the location-model fixed effect for the level-2 predictor, $\gamma_{01}$, and the fixed effect of the level-2 predictor included in the scale model for the residual variance, $\delta_{01}^{e}$, sampled pseudo-randomly using a seed based on the system clock (i.e., seed = 0). For the first set of 1,000 replications, the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, was fixed at 0.075; the second set simulated 2,000 additional replications fixed $\sigma^2_{\omega_0^e}$ at 0.035. The number of replications was doubled in the second set simply to afford more certainty when assessing whether the increased Type I error rates for the true model calculated from the first set of analyses were simply a coincidence.

Table 4.6
*The Effects Included in the Mixed-Effects Location-Scale Models for the Second Simulation Study*

| | Location Model | | | Scale Model for Level-2 Variance Components | | | Scale Model for Level-1 Residual Variance | | |
|---|---|---|---|---|---|---|---|---|---|
| | Fixed Effects | | Random Effects | Fixed Effects | | | Fixed Effects | | Random Effects |
| | $\gamma_{00}$ | $\gamma_{01}$ | $u_{0,i}$ | $\alpha_0^{u_0}$ | $\alpha_0^{\omega_0^e}$ | $\alpha_0^{u_0;\omega_0^e}$ | $\delta_{00}^{e}$ | $\delta_{01}^{e}$ | $\omega_{0,i}^{e}$ |
| Model 5 | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| Model 6 | ● | | ● | ● | | | ● | | |
| Model 7 | ● | ● | ● | ● | | | ● | | |
| Model 8 | ● | ● | ● | ● | ● | ● | ● | | ● |
| Model 9 | ● | | ● | ● | ● | ● | ● | | ● |
| Model 10 | ● | | ● | ● | ● | ● | ● | ● | ● |

*Note.* A ● in a given column indicates that the effect is included in the model. $\gamma_{00}$ = location-model fixed intercept. $\gamma_{01}$ = location-model fixed effect for level-2 predictor $X_i$. $u_{0,i}$ = location-model random intercept for individual $i$. $\alpha_0^{u_0}$ = log of the scale-model fixed intercept for the location-model residual variance. $\alpha_0^{\omega_0^e}$ = log of the scale-model fixed intercept for the scale-model random intercept. $\alpha_0^{u_0;\omega_0^e}$ = inverse hyperbolic tangent of the scale-model fixed intercept for the correlation between the location- and scale-model random intercepts. $\delta_{00}^{e}$ = log of the scale-model fixed intercept for the residual variance. $\delta_{01}^{e}$ = log of the scale-model fixed effect for level-2 predictor $X_i$. $\omega_{0,i}^{e}$ = scale-model random intercept for individual $i$.

**Estimated model sequence.** In this second simulation study, six models were estimated for each replication as shown in Table 4.6; the models in this second simulation study are numbered 5-10 to avoid any overlap with models 1-4 from the first simulation study. Similar to the estimated models described for the first simulation study, all models included the location-model random intercept variance, $\sigma^2_{u_0}$; both $\sigma^2_{u_0}$ and the scale-model random intercept variance as well as their correlation were estimated as homogeneous between individuals (i.e., $\mathbf{G}_i = \mathbf{G}$).

Further, based on the increased Type I error rates presented by Leckie (2014), Leckie et al. (2014), and the replication of these studies presented in the first simulation study, no model was estimated that included a level-2 predictor in the scale model for the residual variance, $\delta^e_{01}$, in the absence of the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$. That is, given that $\sigma^2_{\omega^e_0}$ had adequate statistical power to be detected, omitting this variance component would have led to increased Type I error rates for $\delta^e_{01}$. A description of each estimated model is provided below.

***Model 5.*** The fifth estimated model was the true model (identical to model 1 from the first simulation study), which estimated both the correct location model, as shown in (4.1), and correct scale model for the residual variance, as shown in (4.2) and (4.3). In this true model, individual mean differences in the location-model fixed intercept, $\gamma_{00}$, were estimated via $u_{0,i}$, with the location-model random intercept variance, $\exp\left(\alpha_0^{u_0}\right) = \sigma^2_{u_0}$, explained by the fixed effect for the level-2 predictor, $\gamma_{01}$.

Further, the scale model for the residual variance estimated individual differences in residual variability via $\omega^e_{0,i}$, and also estimated the fixed effect of the level-2 predictor,

$\delta_{01}^{e}$, which allowed the residual variance to be heterogeneous across values of $X_i$, and to reduce the scale-model random intercept variance, $\exp\left(\alpha_0^{\omega_0^e}\right) = \sigma_{\omega_0^e}^2$.

***Model 6.*** The sixth estimated model misspecified both the location and scale models, in which the location model, as shown in (4.21), allowed for individual mean differences via $u_{0,i}$, but omitted the fixed effect for the level-2 predictor, $\gamma_{01}$, to predict those individual differences.

$$\begin{aligned} &\text{Level 1:} \quad Y_{t,i} = \beta_{0,i} + e_{t,i} \\ &\text{Level 2:} \quad \beta_{0,i} = \left(\gamma_{00} + u_{0,i}\right) \\ &\text{Combined:} \quad Y_{t,i} = \left(\gamma_{00} + u_{0,i}\right) + e_{t,i} \end{aligned} \tag{4.21}$$

The scale model for the residual variance for model 6 was identical to the scale model for model 2 shown in (4.17), and was misspecified by omitting the fixed effect for the level-2 predictor, $\delta_{01}^{e}$, and further assumed that residual variance was constant across individuals by omitting the scale-model random intercept, $\omega_{0,i}^{e}$. Thus, this model is a typical empty (or unconditional) linear mixed-effects model, and therefore the subscript $i$ would not be required for any effects included in the scale model for the residual variance.

***Model 7.*** The seventh model estimated the correct location model shown above in (4.1), where individual mean differences were estimated via $u_{0,i}$, and explained by the fixed effect for the level-2 predictor, $\gamma_{01}$. However, the scale model for the residual variance was misspecified by omitting the fixed effect of the level-2 predictor, $\delta_{01}^{e}$, and further assumed that residual variance was constant across individuals by omitting the scale-model random intercept, $\omega_{0,i}^{e}$, as shown for model 2 in (4.17).

***Model 8.*** The eighth model estimated the correct location model shown in (4.1), where individual mean differences were estimated via $u_{0,i}$, and explained by the fixed effect for the level-2 predictor, $\gamma_{01}$, but misspecified the scale model for the residual variance by omitting the fixed effect of the level-2 predictor, $\delta_{01}^{e}$, as shown for model 3 in (4.18). Therefore, this model allowed for individual differences in residual variance, but did not attempt to explain (or predict) why these individual differences existed.

***Model 9.*** The ninth estimated model misspecified both the location and scale models. Here, the location model allowed for individual mean differences via $u_{0,i}$, but omitted the fixed effect for the level-2 predictor, $\gamma_{01}$, to predict these individual differences, as shown for model 6 in (4.21). Further, the scale model for the residual variance included the scale-model random intercept, $\omega_{0,i}^{e}$, but omitted the fixed effect of the level-2 predictor, $\delta_{01}^{e}$, as shown above for model 3 in (4.18). Therefore, this model allowed for individual differences in residual variance, but did not attempt to explain why these individual differences existed.

***Model 10.*** The tenth, and final, model estimated the correct scale model for the residual variance shown in (4.2), where individual differences in residual variability were estimated via $\omega_{0,i}^{e}$, with a fixed effect of the level-2 predictor, $\delta_{01}^{e}$, allowing residual variance to be heterogeneous across values of $X_i$, and to reduce the scale-model random intercept variance, $\exp\left(\alpha_0^{\omega_0^e}\right) = \sigma_{\omega_0^e}^2$. However, the location model was misspecified, such that although it included individual mean differences via $u_{0,i}$, it omitted the fixed effect for the level-2 predictor, $\gamma_{01}$, to predict these individual differences, as shown for model 6 in (4.21).

**MCMC estimation, parameter convergence, recovery, and accuracy.** Similar to the first simulation study, all mixed-effects location-scale models were estimated using the MCMC estimator described in chapter 3. The procedure for estimation was identical to the first simulation study, where start values were determined by a traditional mixed-effects model, with 20 initial tuning chains for 50 iterations, a Markov chain of 5,000 iterations, a burn-in period of 1,000 iterations, and a thinning interval of 1. Convergence was determined by satisfying Geweke's and/or the Gelman and Rubin criterion based on the 4,000 iterations following burn in. Finally, parameter recovery estimates were calculated as the proportion of replications in which the estimated 95% credible interval contained the true sampled value, whereas parameter accuracy was measured by signed bias, as shown in (4.20), calculated for each parameter within each of the six models, presented alongside its 95% confidence interval.

**Estimated model comparisons.** The first set of comparisons for the second simulation study determined the effect a misspecified scale model for the residual variance had on the Type I and Type II error rates for the fixed effect for the level-2 predictor included in the location model, $\gamma_{01}$. Because $\gamma_{01}$ was sampled pseudo-randomly, the significance of $\gamma_{01}$ was first determined by the 95% credible interval excluding 0 using the true model 5. For all replications in which $\gamma_{01}$ was actually sampled to be 0, Type I error rates for $\gamma_{01}$ were calculated based on $\gamma_{01}$ estimated by model 7 and model 8 and presented alongside their Clopper-Person 95% confidence interval. Alternatively, for replications in which $\gamma_{01}$ from model 5 was statistically significant (with true values $\neq 0$), Type II error rates were calculated for $\gamma_{01}$ estimated by model 7 and model 8 and presented alongside their Clopper-Person 95% confidence

interval. Further, for all replications, signed bias in the standard deviation of the posterior distribution of $\gamma_{01}$ was evaluated by comparing the posterior standard deviation of $\gamma_{01}$ estimated by model 7 and model 8 to the posterior standard deviation of $\gamma_{01}$ estimated by model 5.

The second set of comparisons determined the effect that a misspecified location model had on the scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, and the fixed effect for the level-2 predictor included in the scale model for the residual variance, $\delta_{01}^e$. Although $\sigma^2_{\omega_0^e}$ should have adequate statistical power, the significance of $\sigma^2_{\omega_0^e}$ was evaluated by comparing the DIC from model 6 to the DIC from model 9. In addition, because the effect of $\delta_{01}^e$ was sampled pseudo-randomly, the statistical significance of $\delta_{01}^e$ was evaluated initially using the 95% credible interval from the true model 5. For replications in which $\delta_{01}^e$ was actually sampled to be 0, Type I error rates were calculated for $\delta_{01}^e$ estimated from model 10 and presented alongside the Clopper-Person 95% confidence interval, whereas for replications in which true model 5 estimated $\delta_{01}^e$ as significant (with true values $\neq 0$), Type II error rates were calculated for $\delta_{01}^e$ estimated from model 10 and presented alongside the Clopper-Person 95% confidence interval. Significance was determined by the 95% credible interval of the posterior distribution of the fixed effect for the level-2 predictor included in the scale model for the residual variance, $\delta_{01}^e$, that excluded 0. Finally, for all replications, signed bias in the standard deviation of the posterior distribution of $\delta_{01}^e$ was evaluated by comparing the posterior standard deviation of $\delta_{01}^e$ estimated by model 10 to the posterior standard deviation of $\delta_{01}^e$ estimated by true model 5.

**Results of simulation study II.** The results of the second simulation study are presented below. First, the rationale behind simulating a second set of 2,000 replications for this simulation study is discussed, which is then followed by the results with respect to convergence and signed bias. Finally, the consequences of modeling location-model fixed effects in the presence of a misspecified scale model for the residual variance are discussed, followed by the consequences of modeling scale-model fixed effects in the presence of a misspecified location model.

***Why two sets of simulated data were required.*** The first set of 1,000 replications, in which the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, was fixed to 0.075, indicated that the Type I error rate for the fixed effect of the level-2 predictor included in the location model, $\gamma_{01}$, estimated by true model 5 was 11.23%, 95% CI [8.89%,13.94%]; the Type I error rate for the level-2 fixed effect included in the scale model for the residual variance, $\delta^e_{01}$, estimated by true model 5 was 6.61%, 95% CI [5.00%,8.54%]. Although the error rate for $\delta^e_{01}$ was closer to the nominal 5%, the error rate for $\gamma_{01}$ in the location model was exceedingly high. This was unexpected given that the Type I error rates from the first simulation study estimated by true model 1 were 6.35%, 95% CI [5.39%,7.43%], and 3.19%, 95% CI [2.60%,3.86%], for $\gamma_{01}$ and $\delta^e_{01}$, respectively.

The increased Type I error rates observed in the second simulation study for the fixed effects for the level-2 predictor included in the location model and scale model for the residual variance, $\gamma_{01}$ and $\delta^e_{01}$, were hypothesized to result from an overpowered scale-model random intercept variance, $\sigma^2_{\omega^e_0}$. That is, the empirical power rate for $\sigma^2_{\omega^e_0}$ from the first set of 1,000 replications was 99.80%, 95% CI [99.28%,99.98%]. Therefore, a second set of 2,000 replications were simulated with $\sigma^2_{\omega^e_0}$ decreased from 0.075 to

0.035; the number of replications was increased to afford more certainty in determining whether error rates from the first set of simulated replications were simply a coincidence. Using this second set of replications, the empirical power rate to detect $\sigma^2_{\omega^e_0}$ decreased to 85.35%, 95% CI [83.72%,86.87%], with Type I error rates estimated by true model 5 for $\gamma_{01}$ to be 5.81%, 95% CI [4.60%,7.23%], and for $\delta^e_{01}$ to be 2.02%, 95% CI [1.40%,2.83%], both of which were much closer to Type I error rates from the first simulation study. Thus, the hypothesis regarding an overpowered $\sigma^2_{\omega^e_0}$ was, at best, partially supported given that this change in its population model value cannot be disentangled from the concomitant increase in replications in the second set.

Further, in this second set of replications, models 6 through 10 could not be estimated for replication 1,484 and it was dropped from all subsequent analyses. It was hypothesized that this replication was non-estimable because the true sampled value for the fixed effect of the level-2 predictor included in the scale model for the residual variance, $\delta^e_{01}$, was 4; thus, $\exp(4) \approx 54.60$, which would have made estimation of any level-2 random effect impossible given that $\text{ICC}_i \to 0$ as $X_i$ becomes increasingly positive.

Overall, given that the majority of research questions for the second simulation study were based primarily on the Type I and Type II error rates across models, any result from the *first set* of replications was considered untrustworthy because Type I error rates for the true model were upwardly biased. Thus, only the results of the *second set* of 1,999 replications are presented below.

***Convergence evaluation.*** The number and proportion of parameters across replications satisfying Geweke's and/or the Gelman and Rubin convergence criterion are

Table 4.7

*Number and Proportion of Parameters Satisfying Geweke's and/or Gelman and Rubin's Convergence Criterion (N = 1,999)*

| | Model 5 | Model 6 | Model 7 | Model 8 | Model 9 | Model 10 |
|---|---|---|---|---|---|---|
| Location Model | | | | | | |
| $\gamma_{00}$ | 1966 (98.30) | 1979 (98.95) | 1986 (99.30) | 1981 (99.05) | 1984 (99.20) | 1969 (99.45) |
| $\gamma_{01}$ | 1955 (97.75) | - | 1984 (99.20) | 1962 (98.10) | - | - |
| Scale Model | | | | | | |
| $\delta_{00}^{e}$ | 1992 (99.60) | 1997 (99.85) | 1999 (99.95) | 1991 (99.55) | 1989 (99.45) | 1991 (99.55) |
| $\delta_{01}^{e}$ | 1993 (99.65) | - | - | - | - | 1995 (99.75) |
| $\alpha_{0}^{u_0}$ | 1990 (99.50) | 1994 (99.70) | 1991 (99.55) | 1991 (99.55) | 1994 (99.70) | 1991 (99.55) |
| $\alpha_{0}^{u_0;\omega_0^e}$ | 1982 (99.10) | - | - | 1880 (94.00) | 1985 (99.25) | 1987 (99.35) |
| $\alpha_{0}^{\omega_0^e}$ | 1984 (99.20) | - | - | 1988 (99.40) | 1978 (98.90) | 1975 (98.75) |

*Note.* Data presented as frequency (%). $\gamma_{00}$ = location-model fixed intercept. $\gamma_{01}$ = location-model fixed effect for level-2 predictor $X_i$. $u_{0,i}$ = location-model random intercept for individual $i$. $\alpha_{0}^{u_0}$ = scale-model fixed intercept for the location-model residual variance. $\alpha_{0}^{\omega_0^e}$ = scale-model fixed intercept for the scale-model random intercept. $\alpha_{0}^{u_0;\omega_0^e}$ = scale-model fixed intercept for the correlation between the location- and scale-model random intercepts. $\delta_{00}^{e}$ = scale-model fixed intercept for the residual variance. $\delta_{01}^{e}$ = scale-model fixed effect for level-2 predictor $X_i$. $\omega_{0,i}^{e}$ = scale-model random intercept for individual $i$.

Table 4.8

*Parameter Recovery for Location- and Scale-Model Effects (N = 1,999)*

| | Model 5 | Model 6 | Model 7 | Model 8 | Model 9 | Model 10 |
|---|---|---|---|---|---|---|
| Location Model | | | | | | |
| $\gamma_{00}$ | 1809 (90.45) | 1860 (93.00) | 1832 (91.60) | 1815 (90.75) | 1828 (91.40) | 1840 (92.00) |
| $\gamma_{01}$ | 1795 (89.75) | - | 1830 (91.50) | 1795 (89.75) | - | - |
| Scale Model | | | | | | |
| $\delta_{00}^{e}$ | 1886 (94.30) | 1455 (72.75) | 1454 (72.70) | 1890 (94.50) | 1886 (94.30) | 1889 (94.45) |
| $\delta_{01}^{e}$ | 1889 (94.45) | - | - | - | - | 1894 (94.70) |
| $\alpha_{0}^{u_0}$ | 1917 (95.85) | 1443 (72.15) | 1902 (95.10) | 1911 (95.55) | 1433 (71.65) | 1430 (71.50) |
| $\alpha_{0}^{u_0;\omega_0^e}$ | 1999 (99.95) | - | - | 1999 (99.95) | 1872 (93.60) | 1999 (99.95) |
| $\alpha_{0}^{\omega_0^e}$ | 1895 (94.75) | - | - | 1544 (77.20) | 1546 (77.30) | 1888 (94.40) |

*Note.* Data presented as frequency (%). $\gamma_{00}$ = location-model fixed intercept. $\gamma_{01}$ = location-model fixed effect for level-2 predictor $X_i$. $u_{0,i}$ = location-model random intercept for individual $i$. $\alpha_{0}^{u_0}$ = scale-model fixed intercept for the location-model residual variance. $\alpha_{0}^{\omega_0^e}$ = scale-model fixed intercept for the scale-model random intercept. $\alpha_{0}^{u_0;\omega_0^e}$ = scale-model fixed intercept for the correlation between the location- and scale-model random intercepts. $\delta_{00}^{e}$ = scale-model fixed intercept for the residual variance. $\delta_{01}^{e}$ = scale-model fixed effect for level-2 predictor $X_i$. $\omega_{0,i}^{e}$ = scale-model random intercept for individual $i$.

Table 4.9

*Signed Bias for Location- and Scale-Model Effects (N = 1,999)*

| | Model 5 | Model 6 | Model 7 | Model 8 | Model 9 | Model 10 |
|---|---|---|---|---|---|---|
| Location Model | | | | | | |
| $\gamma_{00}$ | 0.00 [–0.01,0.01] | 0.01 [0.00,0.02] | 0.00 [–0.01,0.01] | 0.00 [0.00,0.01] | 0.00 [–0.01,0.01] | 0.00 [–0.01,0.01] |
| $\gamma_{01}$ | 0.01 [0.00,0.01] | - | 0.01 [0.00,0.01] | 0.00 [0.00,0.01] | - | - |
| Scale Model | | | | | | |
| $\delta_{00}^{e}$ | 0.00 [–0.01,0.00] | 0.13 [0.12,0.15] | 0.13 [0.12,0.15] | 0.00 [–0.01,0.00] | 0.00 [–0.01,0.00] | 0.00 [–0.01,0.00] |
| $\delta_{01}^{e}$ | 0.00 [0.00,0.00] | - | - | - | - | 0.00 [0.00,0.00] |
| $\alpha_{0}^{u_0}$ | 0.03 [0.02,0.03] | 0.18 [0.16,0.19] | 0.00 [–0.01,0.00] | 0.03 [0.02,0.03] | 0.21 [0.19,0.22] | 0.21 [0.19,0.22] |
| $\alpha_{0}^{u_0;\omega_0^e}$ | 0.00 [0.00,0.00] | - | - | 0.00 [0.00,0.00] | 0.08 [0.06,0.09] | 0.00 [0.00,0.00] |
| $\alpha_{0}^{\omega_0^e}$ | 0.06 [0.03,0.08] | - | - | 0.73 [0.66,0.80] | 0.72 [0.65,0.79] | 0.06 [0.04,0.08] |

*Note.* Data presented as mean [95% CI]. $\gamma_{00}$ = location-model fixed intercept. $\gamma_{01}$ = location-model fixed effect for level-2 predictor $X_i$. $u_{0,i}$ = location-model random intercept for individual $i$. $\alpha_{0}^{u_0}$ = scale-model fixed intercept for the location-model residual variance. $\alpha_{0}^{\omega_0^e}$ = scale-model fixed intercept for the scale-model random intercept. $\alpha_{0}^{u_0;\omega_0^e}$ = scale-model fixed intercept for the correlation between the location- and scale-model random intercepts. $\delta_{00}^{e}$ = scale-model fixed intercept for the residual variance. $\delta_{01}^{e}$ = scale-model fixed effect for level-2 predictor $X_i$. $\omega_{0,i}^{e}$ = scale-model random intercept for individual $i$.

presented in Table 4.7. The convergence criterion was met for a high proportion of replications across all six models. As in the first simulation study, the parameter with the lowest convergence rate was the correlation between the location- and scale-model random intercepts, $\alpha_0^{u_0;\omega_0^e}$, for model 8. Of the 120 replications that failed to converge, a pseudo-random sample of 12 trace plots (10%) were examined, with each trace plot having a subtle peak or valley across iterations that was nearly identical to the trace plot shown in Figure 4.1 presented previously for the first simulation study. Similar to the first simulation study, this decrease in convergence was hypothesized to result from the omission of the effect of the level-2 predictor in the scale model for the residual variance, $\delta_{01}^e$, which necessarily increased estimated scale-model random intercept variance, $\sigma_{\omega_0^e}^2$ (as $\delta_{01}^e$ explicitly explains, or reduces, $\sigma_{\omega_0^e}^2$). This hypothesis was supported given that the convergence rate for $\alpha_0^{u_0;\omega_0^e}$ from model 5, which included $\delta_{01}^e$, was 99.10%, and the convergence rate from model 9, which excluded the level-2 fixed effects in both the location model and scale model for the residual variance (resulting in a similar increase to location-model random intercept variance), was 99.25%.

***Parameter recovery and accuracy.*** The number and proportion of replications for which the estimated 95% credible interval for each parameter contained the true sampled value are presented in Table 4.8, whereas signed bias for each parameter is presented in Table 4.9.

Model 5 (i.e., the true model) recovered all scale-model fixed and random effects, as well as the location-model random intercept variance, with expected error rates for a 95% credible interval. Recovery for both location model fixed effects, $\gamma_{00}$ and $\gamma_{01}$, was only around 90%, which (similar to the first simulation study) was unexpected given that

the starting values for these parameters estimated by the traditional linear mixed effect model were hypothesized to be nearly identical to the final model estimates from the mixed-effects location-scale model—a hypothesis that was supported by signed bias estimates of at 0.00 for both effects. Therefore, given the ideal starting values and absence of signed bias for both $\gamma_{00}$ and $\gamma_{01}$, it was hypothesized that the MCMC algorithm may have sampled from highly leptokurtic posterior distributions (i.e., small posterior standard deviations), which may have resulted in extremely narrow 95% credible intervals that only just excluded the true sampled parameter. Further, the (log of the) scale-model random intercept variance, $\alpha_0^{\omega_0^e}$, was overestimated by an average of 0.06 on the log scale (or 6.18% on the variance scale; $[\exp(0.06) - 1] * 100$). Although overestimation by 6.18% appeared significant, it only represented an average difference between the true value and the model-estimated value of 0.002 on the variance scale (true value of $\sigma^2_{\omega_0^e}$ = 0.035 vs. mean $\sigma^2_{\omega_0^e}$ estimated by model 1 = 0.037). Therefore, the overestimation of $\alpha_0^{\omega_0^e}$ was not considered to bias any results presented below.

When considering misspecified models 6 through 10, their recovery and bias estimates followed expected patterns. That is, if the location-model fixed effect of the level-2 predictor, $\gamma_{01}$, was omitted (as in models 6, 9, and 10), the (log of the) location-model random intercept variance, $\alpha_0^{u_0}$, had poor recovery and was overestimated by an average of 0.18 to 0.21 on the log scale (or 19.72% to 23.37% on the variance scale; e.g., $[\exp(0.18) - 1] * 100 = 19.72\%$), expected given that $\gamma_{01}$ would have reduced location-model random intercept variance, $\exp\left(\alpha_0^{u_0}\right) = \sigma^2_{u_0}$. Similarly, if the fixed effect for the level-2 predictor, $\delta_{01}^e$, was omitted from the scale model for the residual variance

(as in models 6 through 9), the (log of the) scale-model random intercept variance, $\alpha_0^{\omega_0^e}$, had poor recovery and was overestimated by 0.72 to 0.73 on the log scale (or 105.44% to 107.51% on the variance scale; e.g., $[\exp(0.72) - 1] * 100 = 105.44\%$), expected given that $\delta_{01}^e$ would have reduced scale-model random intercept variance, $\exp\left(\alpha_0^{\omega_0^e}\right) = \sigma_{\omega_0^e}^2$. Finally, if both the fixed and random effects were omitted from the scale model for the residual variance (as in models 6 and 7), the fixed intercept for the residual variance, $\delta_{00}^e$, had poor recovery and was overestimated by 0.13 on the log scale (or 13.88% on the variance scale; e.g., $[\exp(0.13) - 1] * 100 = 13.88\%$), expected given that that partitioning $\sigma_{\omega_0^e}^2$ out of $\delta_{00}^e$ would have necessarily decreased $\delta_{00}^e$.

***The effect of misspecifying the scale model for the residual variance on the fixed effect for a level-2 predictor included in the location model.*** Of the 1,999 total replications, 1,290 (64.53%) replications had true values for the fixed effect of the level-2 predictor included in the location model, $\gamma_{01}$, sampled to be 0. Using these 1,290 replications, the Type I error rate for $\gamma_{01}$ estimated from model 7 was 4.50%, 95% CI [3.43%,5.77%], whereas the Type I error rate for $\gamma_{01}$ estimated from model 8 was 4.81%, 95% CI [3.70%,6.12%]. Both of these error rates were near the expected 5%, and provide evidence that misspecifying the scale model for the residual variance should not bias the inference of the fixed effect for a level-2 predictor included in the location model.

In addition, based on true model 5, the fixed effect for the level-2 predictor included in the location model, $\gamma_{01}$, was significant for 784 replications, of which 75 replications had true values of $\gamma_{01}$ sampled to be 0. Thus, considering only the remaining 709 replications that had a true sampled value of $\gamma_{01} > 0$, the Type II error rates for $\gamma_{01}$ estimated from model 7 (which omitted the fixed effect for the level-2 predictor in the

scale model for the residual variance, $\delta_{01}^{e}$, as well as the scale-model random intercept variance, $\sigma_{\omega_0^e}^2$) and from model 8 (which omitted $\delta_{01}^{e}$, but included $\sigma_{\omega_0^e}^2$) were identical at 0.14%, 95% CI [0.00%,0.78%]. These results suggest that misspecifying the scale model for the residual variance has little effect on the ability to detect the fixed effect for a level-2 predictor included in the location model, given that this effect actually exists.

Finally, the posterior standard deviation for $\gamma_{01}$ was identical for models 5, 7, and 8, and was estimated to be 0.16, 95% CI [0.16,0.16]. Thus, when comparing the posterior standard deviations from true model 5 to misspecified models 7 and 8, no signed bias was indicated, as the difference in posterior standard deviations for $\gamma_{01}$ was 0.00, 95% CI [0.00,0.00]. When considered alongside the Type II and Type I error rates reported above, these results suggest rather convincingly that misspecifying the scale model for the residual variance has very minimal effect on the correct inference and accuracy of the fixed effect of a level-2 predictor included in the location model.

***The effect of misspecifying the location model on the fixed effect for a level-2 predictor included in the scale model for the residual variance.*** Based on a comparison of model 6 and model 9, the statistical power to detect the scale-model random intercept variance was 85.35%, 95% CI [83.72%,86.87%], $n_{significant}$ = 1,707.

When considering the 1,707 replications in which scale-model random intercept variance was detected, 1,340 (78.50%) replications had the true value of $\delta_{01}^{e}$ sampled to be 0. For these replications, the Type I error rate for $\delta_{01}^{e}$ estimated from model 10 was 2.09%, 95% CI [1.39%,3.01%]. In addition, true model 5 indicated that $\delta_{01}^{e}$ was significant for 400 replications, of which 33 of these replications had true values of $\delta_{01}^{e}$ sampled to be 0. Thus, considering only the remaining 367 replications that had a true

sampled values of $\delta_{01}^{e} > 0$ (values of $\delta_{01}^{e}$ which will always have 100% power to be detected, as observed in the first simulation study), the Type II error rate for $\delta_{01}^{e}$ estimated from model 10 (which omitted location-model fixed effect for the level-2 predictor, $\gamma_{01}$) was 0.00%, 95% CI [0.00%,0.00%]. Further, the posterior standard deviation for $\delta_{01}^{e}$ was identical for models 5 and 10, and was estimated to be 0.04, 95% CI [0.04,0.04]. Thus, when comparing the posterior standard deviations from model 5 to model 10, no signed bias was indicated, as the difference in posterior standard deviations for $\delta_{01}^{e}$ was 0.00, 95% CI [0.00, 0.00]. Taken together, these results indicate rather convincingly that misspecifying the location model by excluding a level-2 fixed effect has no effect on the correct detection or accuracy of the fixed effect for a level-2 predictor included in the scale model for the residual variance.

In addition, the consequences of misspecifying the location model have on the effect of a level-2 predictor included in the scale-model for the residual variance, $\delta_{01}^{e}$, were considered using only the 292 (14.61%) replications in which scale-model random intercept variance, $\sigma_{\omega_0^e}^2$, was not detected. For all 292 replications, the true sampled value of $\delta_{01}^{e} = 0$ (results consistent with the first simulation study), with the Type I error rate for $\delta_{01}^{e}$ estimated by model 10 to be 2.05%, 95% CI [0.76%,4.42%]. Further, the posterior standard deviations for $\delta_{01}^{e}$ were identical for models 5 and 10 (and identical to signed biases reported previously), estimated to be 0.04, 95% CI [0.04,0.04]. Thus, when comparing the posterior standard deviations from model 5 to model 10, no signed bias was indicated, as the difference in posterior standard deviations for $\delta_{01}^{e}$ was 0.00, 95% CI [0.00, 0.00]. Taken together, these results provide further support that misspecifying the location model by excluding a level-2 fixed effect has no effect on the correct detection

and accuracy of the fixed effect of a level-2 predictor included in the scale model for the residual variance when $\sigma^2_{\omega^e_0}$ was not detected. Further, in the presence of a misspecified location model, these results corroborate those presented in the first simulation study suggesting that $\sigma^2_{\omega^e_0}$ must exist before it can be predicted and that it is acceptable to continue to model residual heterogeneity as systematically varying using a level-2 predictor although this endeavor would most likely be unproductive.

**Discussion of the second simulation study.** The results of the second simulation study suggest that misspecifying the scale model for the residual variance by omitting fixed and/or random effects ($\delta^e_{01}$ and/or $\sigma^2_{\omega^e_0}$) had no effect on recovery, accuracy, or Type I and Type II error rates of the fixed effect for the level-2 predictor included in the location model. Similarly, misspecifying the location model by omitting a fixed effect for a level-2 predictor ($\gamma_{01}$) had no effect on Type I or Type II error rates for the fixed effect of the level-2 predictor included in the scale model for the residual variance ($\delta^e_{01}$).

Given that misspecifying the scale model for the residual variance did not bias the fixed effect for the level-2 predictor included in the location model, these results suggest that failing to model individual differences in residual variance only limits potential research questions available to the empirical scientist. That is, if an empirical scientist has only been concerned with the prediction of individual differences in mean level, their inferences for the fixed effects for level-2 predictors included in the location model to explain these differences remained unbiased regardless of whether significant fixed or random effects were erroneously omitted in the scale model. However, estimating individual differences in mean level only accounts for half of the potential research questions available to the empirical scientist who collect repeated-measures data. This is

because failing to estimate fixed and random effects in the scale model for the residual variance precludes the examination of specific research questions pertaining specifically to individual differences in outcome variability.

The primary purpose of including fixed effects of level-2 predictors in a mixed-effect location-scale model is to reduce (or explain) their specific random intercept variance component. Not surprisingly then, the model substantially overestimated the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, if the scale model for the residual variance was misspecified by omitting the fixed effect of the level-2 predictor, $\delta^e_{01}$; a similar result was indicated for the location-model random intercept variance, $\sigma^2_{u_0}$, if the location model omitted the fixed effect for the level-2 predictor, $\gamma_{01}$, albeit to a lesser extent. Although signed bias indicated that overestimation was larger for $\sigma^2_{\omega^e_0}$ compared to $\sigma^2_{u_0}$ (105.44% to 107.51% vs. 19.72% to 23.37% on the variance scale, respectively), the scale of each variance component must be considered—$\sigma^2_{\omega^e_0}$ was fixed to 0.035, whereas $\sigma^2_{u_0}$ was fixed to 2.72. Because both level-2 predictor effects, $\gamma_{01}$ and $\delta^e_{01}$, were sampled from a zero-inflated Poisson distribution, an identical sampled non-zero value (say, a value of 1) would have explained a greater proportion of $\sigma^2_{\omega^e_0}$ relative to $\sigma^2_{u_0}$ simply because there was considerably more variance available to be predicted in $\sigma^2_{u_0}$ within the location model. Thus, omitting the level-2 fixed effect resulted in $\sigma^2_{\omega^e_0}$ in the scale model appearing to be overestimated to a greater extent compared to $\sigma^2_{u_0}$ in the location model.

In addition, if both the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, and the fixed effect for the level-2 predictor in the scale model for the residual variance, $\delta^e_{01}$, were omitted (akin to modeling homogeneous residual variance in a linear mixed-effects

model), the model overestimated the fixed intercept for the residual variance, $\delta_{00}^{e}$, by approximately 13.88%. What is important to note is that the overestimation of residual variance did not affect bias of the fixed effect for the level-2 predictor included in the location model, $\gamma_{01}$; however, bias for $\gamma_{01}$ would have been considerable if the location-model random intercept variance, $\sigma_{u_0}^2$, had not been initially partitioned out of residual variance. This result was hypothesized to occur because $\sigma_{\omega_0^e}^2$, that was initially partitioned out of residual variance, was re-aggregated into total residual variance during analysis, as described by (4.6), which is then used in the calculation of $\mathbf{V}_i$ (in the presence of location-model random effects) or $\mathbf{R}_i$ (in the absence of location-model random effects).

Finally, two sets of replications were simulated for the second simulation study, with Type I error rates from the true model for both the location- and scale-model fixed effects of a level-2 predictor, $\gamma_{01}$ and $\delta_{01}^{e}$, varying greatly between sets of replications. It was initially hypothesized that the substantial Type I error rate for $\gamma_{01}$ (i.e., 11.23%) observed using the first set of replications resulted from an overpowered scale-model random intercept variance, $\sigma_{\omega_0^e}^2$. Based on the results of the first simulation study, this hypothesis appeared to be partially supported given that the second set of replications had more reasonable Type I error rates for $\gamma_{01}$ (i.e., 5.81%). However, the second simulation study failed to provide further support for this hypothesis given that the estimated value of $\sigma_{\omega_0^e}^2$ had no effect on the Type I error rate for $\gamma_{01}$. Therefore, when considering this additional evidence, the unusually large Type I error rates from the first set of replications was considered to result primarily from random sampling, as Type I error rates decreased to more appropriate levels upon sampling a larger set of replications.

## Limitations of the Simulation Studies and Directions for Future Methodological Research

The purpose of the simulation studies provided in this chapter was to 1) provide empirical scientists with the information necessary to conduct *a priori* power analyses for hypotheses specifically regarding constant individual differences in outcome variability, and 2) inform empirical scientists about the consequences misspecifying the location model or scale model for the residual variance have on fixed effects for a level-2 predictor. Because model building is a multi-step process, the synthesis of both studies provides useful information for model-building in practice, as will be described in the general discussion provided in chapter 6. With that said, both simulation studies had limitations that could be addressed by future research.

First, both simulation studies were conducted under highly controlled (i.e., ideal/simulated) conditions, and although study design characteristics and model parameter values were sampled from ranges informed by a small existing literature, complete and consistent recommendations cannot be made unless observed data collected by the empirical scientist satisfy the specific conditions for each parameter studied. More specifically, extrapolation is not recommended beyond the range of values sampled for parameter, only complete data was used, and all models assumed the location-model random intercept variance was estimated before scale-model random intercept variance.

The results of both simulation studies were based on analyses of complete data, which is unrealistic in real-world data collection, especially for repeated-measures data from longitudinal designs. Missing data has been shown to deteriorate statistical power and have potentially disastrous effects on the accuracy and recovery of fixed and random

effects, especially given that traditional methods of handling missing data such as listwise deletion, mean/median substitution, or last observation carried forward are still being used routinely by empirical scientists (Enders, 2010; Saha & Jones, 2009; Shao & Zhong, 2003). Although the effects that missing data have on location-model fixed and random effects from repeated-measures data have been well documented (assuming the missing data mechanism is known; e.g., missing at random; see Yang & Maxwell, 2014), future research should evaluate the effects missing data have on the recovery, accuracy, and error rates for level-1 and level-2 effects included in the scale model.

In addition, the location-model random intercept variance, $\sigma^2_{u_0}$, was included in every model under the assumption that individual mean differences must be accounted for prior to evaluating individual differences in outcome variability. This assumption has not been tested empirically; thus, future research should evaluate the recovery, accuracy, and Type I error rates resulting from modeling scale-model random intercept variance, $\sigma^2_{\omega_0^e}$, in the absence of significant $\sigma^2_{u_0}$.

Fourth, the results of the first simulation study indicated that the correlation between location- and scale-model random intercepts, $\rho_{u_0;\omega_0^e}$, did not influence power to detect $\sigma^2_{\omega_0^e}$. However, this result could not explicitly indicate that $\rho_{u_0;\omega_0^e}$ had no effect on the recovery or accuracy of estimated $\sigma^2_{u_0}$ when erroneously omitting $\sigma^2_{\omega_0^e}$ (and, as a result, $\rho_{u_0;\omega_0^e}$). In the second simulation study, $\rho_{u_0;\omega_0^e}$ was fixed to 0, but future research should evaluate whether a true non-zero value of $\rho_{u_0;\omega_0^e}$ influences estimates of $\sigma^2_{u_0}$.

Further, the first set of replications from the second simulation study indicated the true model had large Type I error rates for the fixed effect of a level-2 predictor included in the location model, $\gamma_{01}$, which were mitigated in the second simulation study by

decreasing the scale-model random intercept variance, $\sigma^2_{\omega^e_0}$, and doubling the number of replications. Based on the results of the second simulation study, it was hypothesized that the decrease in Type I error rates for $\gamma_{01}$ was primarily a function of resampling and increased replications, not from the decrease in $\sigma^2_{\omega^e_0}$. However, this hypothesis needs to be tested explicitly; thus, future research should determine the effect of a drastically overpowered $\sigma^2_{\omega^e_0}$ on the Type I error rates of location-model fixed effects.

Additionally, both simulation studies only considered the effect of a level-2, individual-level predictor. This was intentional, as random effects in repeated-measures data represent between-individual differences to be predicted by individual-level variables. However, this provided only partial coverage of the effects available to be tested in the mixed-effects location-scale model. Thus, future research should evaluate how fixed and random effects of level-1, occasion-specific predictors included in (or omitted from) the scale model for the residual variance influence recovery, accuracy, and error rates of location-model fixed and random effects, and vice versa, as explaining residual variance by level-1 fixed effects in the location model results in less residual variance to be partitioned into random scale variance. This same suggestion can also be made for level-2 predictors included in the scale model for the random effects in $\mathbf{G}_i$, which were assumed homogeneous between individuals for both simulation studies.

Finally, regarding the actual estimation of the mixed-effects location-scale model, the MCMC estimates from this study were based on a single Markov chain for each parameter and estimates may have been more accurate and less biased (where bias existed) had starting values been chosen based on the true values. Although it was noted that starting values for specific scale-model effects were set to 0 to more realistically

simulate the uncertainty of true values typically encountered in real-world applications, additional research should be conducted to using multiple chains and diverse starting values to provide empirical scientists with information regarding best practices to ensure true values are accurately recovered with minimal bias when estimating the mixed-effects location-scale model using MCMC estimation. In addition, given the use of uninformative priors for all parameters, all posterior means provided by the MCMC estimator approximated the ML estimates from a frequentist framework (e.g., by either PROC NLMIXED or MIXREGLS as described in chapter 3). As a result, the posterior means for the variance components provided in both simulation studies may have been downwardly biased to an unknown extent; thus, developing a REML estimator to be used within the MCMC estimation procedure is another area where future research is needed.

**Chapter Summary**

The purpose of chapter 4 was to present the results of two simulation studies. The first simulation study provided power curves to detect and predict scale-model random intercept variance, with the primary purpose of providing empirical scientists with necessary information regarding study design characteristics and model parameters that will allow them to conduct *a priori* power analyses for variability-related hypotheses using the mixed-effects location-scale model. The second simulation study provided empirical scientists with information regarding the consequences of misspecifying the location and/or scale model for the residual variance, results of which were used in the empirical analysis, presented next in chapter 5, to inform the model-building procedure for testing fixed and random effects in both the location model and scale model for the residual variance.

# CHAPTER 5: MOVEMENT VARIABILITY IN OLDER ADULTS WITH AND WITHOUT PROBABLE MILD ALZHEIMER'S DISEASE

Until recently, evaluating variability-based research questions had required empirical scientists to use non-model-based statistical techniques such as intra-individual standard deviation ($ISD_i$) or the coefficient of variation ($CV_i$); the limitations of both approaches were detailed in chapter 1. With that said, recent instructional papers (e.g., Cleveland et al., 2000; Hedeker et al., 2008; Hedeker & Mermelstein, 2012; Leckie et al., 2014; Rast, Hofer, & Sparks, 2012) and software tutorials (Hedeker & Nordgren, 2013; Leckie, 2014; Rast et al., 2012) have allowed variability-based questions to be answered directly using a model-based technique known as the mixed-effects location-scale model. Examples of empirical studies using this model were presented in chapter 1; briefly, the mixed-effects location-scale model has been applied to positive and negative affect (Hedeker et al., 2008, Hedeker et al., 2009, Hedeker et al., 2012; Pugach et al., 2014; Rast et al., 2012) as well as to eye-tracking data (Lee & Noh, 2012).

In this chapter, an empirical data analysis is presented showing the capability of the mixed-effects location-scale model to estimate and predict individual differences in unstructured physical activities between older adults with and without probable mild Alzheimer's disease (AD), using a design and data similar to the illustrative example provided in chapter 2.

## Method

**Participants and data collection.** The data were collected from 92 older adults, 39 with probable mild AD and 53 healthy individuals, using ecological momentary assessments. Note that individuals with probable AD may have had slight variations in

severity, as their clinical dementia rating could range from 0.5 (i.e., very mild) to 1.0 (mild). Data collection was complete as of February 14, 2015. Physical activity was measured using the Actigraph GT3X+ tri-axial accelerometer worn on the hip 24 hours a day for 5 to 13 consecutive days. The GT3X+ eliminates noise outside of an individual's movement based on a sampling frequency between 30Hz and 100Hz (ActiGraph, 2012). These raw sampling frequencies were then converted to acceleration units known as *activity counts* (measured in units of gravity, $g$), which represent the aggregated physical movement in the *medio-lateral* (ML, front-to-back, sagittal), *antero-posterior* (AP, side-to-side, frontal), and *vertical* planes (VT, rotational, transverse) for a given length of occasion, known as an *epoch*. Movement in each orthogonal axis could have been evaluated separately, but for this study, movement was aggregated into a tri-axial composite metric known as *average vector magnitude*, calculated as $\mathrm{VM} = \sqrt{\mathrm{ML}^2 + \mathrm{AP}^2 + \mathrm{VT}^2}$ (ActiGraph, 2012).

For the purposes of data analysis, only physical activity during waking hours were evaluated, as determined by self-report diary data from each individual. Real-time accelerometer data was binned into 60-minute epochs for two reasons. First, the length of this epoch is consistent with time between occasions in the individual's self-report diary. Second, a shorter epoch is viewed as unnecessary because individuals with mild AD generally spend 60-75% of their day in sedentary activities, which highlights the purpose of the present study to predict location- and scale-model differences between individuals with and without mild AD.

**Measures.** The primary outcome was the observed average vector magnitude, $VM_{t,i}$, aggregated every 60 minutes only during epochs in which the

individual was awake. Vector magnitude was chosen (versus the other three uni-axial metrics) to evaluate whether this tri-axial metric provided insight into the *total movement* of traditionally sedentary individuals and to begin to provide support for evaluating tri-axial movement.

The primary independent variable was mild AD status, $mAD_i$, which was a binary, level-2 predictor for individual $i$, where 0 = no AD and 1 = mild AD. Probable mild AD was determined by comprehensive clinical assessment by a licensed practitioner as detailed by Burns, Cronk, Anderson, Donnely, Thomas, Harsha, and Swerdlow (2008).

Six covariates were also included: day of study as well as the individual's age, years of formal education, biologic sex, cardiorespiratory capacity, and body composition, as described by Watts et al. (2013). As stated above, individuals wore the accelerometer for 5 to 13 consecutive days. However, because 99.07% of observations occurred within the first seven days (only 84 total observations were recorded after day 7; 0.93%), any observation after the seventh study day were excluded from analysis. In addition, technically the data presented a three-level model; that is, observations were nested within days nested within individuals. However, because no systematic or random change in average vector magnitude was expected (or of specific research interest), to account for all variability and to remove its level of nesting, day of study was modeled in both the location model and scale model for the residual variance as a categorical, level-1 predictor, $Day1_{t,i}$ through $Day7_{t,i}$, with the first study day serving as reference.

Regarding the individual-specific covariates, both an individual's age, $Age_i$, and years of formal education, $Ed_i$, were continuous, level-2 predictors for individual $i$, where higher values indicated older and more educated individuals, respectively. The

biologic sex covariate indicated whether the individual was a woman, $Woman_i$, and was therefore a binary, level-2 predictor for individual $i$, where 0 = man and 1 = woman. Cardiorespiratory capacity was evaluated using a treadmill test and measured by peak oxygen volume (VO2), $VO2_i$, which was a continuous, level-2 predictor for individual $i$, where greater $VO2_i$ indicated greater cardiorespiratory capacity. Finally, body composition was quantified by body mass index (BMI), $BMI_i$, calculated by the individual's height and weight (after a morning bowel voiding attempt). Therefore, $BMI_i$ was a continuous, level-2 predictor for individual $i$, where greater $BMI_i$ indicated a greater ratio of weight to height.

**Research questions.** Given the novelty of evaluating individual differences in variability for older adults with and without probable mild AD, four research questions were proposed in lieu of specific hypotheses.

The first research question set to determine whether individual differences in mean movement in average vector magnitude existed, and whether the average amount of movement adequately describe everyone in the sample, or whether some individuals move more than others, on average. If movement showed significant individual differences, the second research question set to determine whether these individual differences could be predicted by probable mild AD status after controlling for day in study, age, years of formal education, biologic sex, cardiorespiratory capacity, and body composition.

Similarly, the third research question set to determine whether individual differences in the residual variance for average vector magnitude existed. That is, whether the fixed residual variance estimate assumed by traditional linear models (e.g.,

linear regression, linear mixed-effects model) adequately described everyone in the sample, or whether movement was actually more variable for some individuals compared to others, on average. If individual differences existed in movement variability, the fourth research question set to determine whether these individual differences could be predicted by probable mild AD status after controlling for day in study, age, years of formal education, biologic sex, cardiorespiratory capacity, and body composition.

**Data considerations.** Because average vector magnitude could not assume negative values, the truncation of the data at zero created semi-continuous data. Briefly, semi-continuous data is considered to result from two separate processes—one process for the zero values (termed the *binary part*) and the other process for the non-zero values (termed the *continuous part*). With application to repeated-measures data in which occasions are nested within individuals, the binary part would be estimated using a mixed-effects logistic regression model, where average vector magnitude would be dichotomized into zero and non-zero values (i.e., no movement, $VM_{t,i} = 0$, and some movement, $VM_{t,i} > 0$), with the probability of $VM_{t,i} = 0$ being modeled. By contrast, the continuous part would exclude zero values and only model data where $VM_{t,i} > 0$. Thus, the model for the continuous part would be estimated using a linear (or generalized linear) mixed-effects model using an appropriate conditional distribution. Note that both parts of the model could include predictors and additional random effects as appropriate. Simultaneous estimation of the continuous and binary part would require a two-part model (see Olsen & Schafer, 2001); however, because the mixed-effects location-scale model applies only to the continuous part, any observation where average vector

magnitude equaled zero was excluded from the current analyses (note this exclusion does not preclude future estimation of the binary part in future analyses).

In addition, examination of residual values from an unconditional traditional linear mixed-effects model indicated a right-skewed distribution of residuals. Therefore, average vector magnitude was natural log transformed for all models estimated below similar in nature to using an identity link with a log-normal distribution of residuals. Following transformation, residual values were normally distributed.

**Estimated models.** The second simulation study described in chapter 4 found that the fixed effects for level-2 predictors included in the location model were unbiased by effects omitted in the scale model for the residual variance (and vice versa); thus, a traditional model-building approach was retained where the location-model fixed and random effects were estimated prior to the fixed and random effects within the scale model for the residual variance. Note that all model equations presented below use the multi-level, scalar notation described in chapter 2.

***Research question 1.*** The first research question of whether individual differences in (the log of) average vector magnitude existed was addressed directly by comparing the DIC between an unconditional location-model random intercept model to a single-level linear model without any random effects; a smaller DIC indicated a better fitting model. The estimated single-level linear model had a location model that assumed no individual mean differences (i.e., no random intercept) in (the log of) average vector magnitude as shown in (5.1). As stated above, day of study was included in all location models to account for all systematic and random variability due to day of study.

$$
\begin{aligned}
\text{Level 1:} \quad \log(VM_{t,i}) &= \beta_{0,i} + \beta_{1,i}(Day2_{t,i}) + \beta_{2,i}(Day3_{t,i}) + \\
&\quad \beta_{3,i}(Day4_{t,i}) + \beta_{4,i}(Day5_{t,i}) + \\
&\quad \beta_{5,i}(Day6_{t,i}) + \beta_{6,i}(Day7_{t,i}) + e_{t,i} \\
\text{Level 2:} \quad \beta_{0,i} &= \gamma_{00} \\
\beta_{1,i} &= \gamma_{10} \\
\beta_{2,i} &= \gamma_{20} \\
\beta_{3,i} &= \gamma_{30} \\
\beta_{4,i} &= \gamma_{40} \\
\beta_{5,i} &= \gamma_{50} \\
\beta_{6,i} &= \gamma_{60} \\
\text{Combined:} \quad \log(VM_{t,i}) &= \gamma_{00} + \gamma_{10}(Day2_{t,i}) + \gamma_{20}(Day3_{t,i}) + \\
&\quad \gamma_{30}(Day4_{t,i}) + \gamma_{40}(Day5_{t,i}) + \\
&\quad \gamma_{50}(Day6_{t,i}) + \gamma_{60}(Day7_{t,i}) + e_{t,i}
\end{aligned}
\tag{5.1}
$$

Here, $\gamma_{00}$ is the location-model fixed intercept representing (the log of) average vector magnitude across all observations and individuals specifically on the first study day, and $\gamma_{10}$ through $\gamma_{60}$ are location-model fixed effects representing the difference in (the log of) average vector magnitude between the first study day and the second through seventh study days, respectively. Further, $e_{t,i}$ is the residual value representing the difference between (the log of) observed average vector magnitude at day $t$ for individual $i$ and the mean of (the log of) average vector magnitude on a specific day $t$.

The scale model for the residual variance estimated for this single-level model is shown in (5.2), where residual variances were allowed to be heterogeneous across study

days. Note that although residual variances were modeled as heterogeneous across study days given the inclusion of the day of study indicator variables, residual values were still assumed to be independent within individuals, such that within-individual correlations in **R** were assumed to be 0 (a current limitation of the MH algorithm discussed in chapter 3). With that said, the mixed-effects location-scale model defined by (5.1) and (5.2) does account for individual mean differences via the location-model random intercept; thus, the model does not completely ignore within-individual correlation.

$$
\begin{aligned}
\text{Level 1:} \quad & \log\left(\sigma_{e_t}^2\right) = \tau_0^e + \tau_1^e\left(Day2_{t,i}\right) + \tau_2^e\left(Day3_{t,i}\right) + \\
& \tau_3^e\left(Day4_{t,i}\right) + \tau_4^e\left(Day5_{t,i}\right) + \\
& \tau_5^e\left(Day6_{t,i}\right) + \tau_6^e\left(Day7_{t,i}\right) \\
\text{Level 2:} \quad & \tau_0^e = \delta_{00}^e \\
& \tau_1^e = \delta_{10}^e \\
& \tau_2^e = \delta_{20}^e \\
& \tau_3^e = \delta_{30}^e \\
& \tau_4^e = \delta_{40}^e \\
& \tau_5^e = \delta_{50}^e \\
& \tau_6^e = \delta_{60}^e \\
\text{Combined:} \quad & \log\left(\sigma_{e_t}^2\right) = \delta_{00}^e + \delta_{10}^e\left(Day2_{t,i}\right) + \delta_{20}^e\left(Day3_{t,i}\right) + \\
& \delta_{30}^e\left(Day4_{t,i}\right) + \delta_{40}^e\left(Day5_{t,i}\right) + \\
& \delta_{50}^e\left(Day6_{t,i}\right) + \delta_{60}^e\left(Day7_{t,i}\right)
\end{aligned}
\tag{5.2}
$$

Here, $\delta_{00}^e$ is the (log of the) fixed intercept for the residual variance specifically on the first study day, and $\delta_{10}^e$ through $\delta_{60}^e$ are scale-model fixed effects for the residual variance

representing the difference in residual variance between the first study day and the second through seventh study days, respectively.

To evaluate individual differences in mean level of (the log of) average vector magnitude, an unconditional location-model random intercept model was estimated as shown in (5.3); the scale model for the residual variance was as shown in (5.2).

$$
\begin{aligned}
&\text{Level 1:} \quad \log(VM_{t,i}) = \beta_{0,i} + \beta_{1,i}(Day2_{t,i}) + \beta_{2,i}(Day3_{t,i}) + \\
&\qquad\qquad \beta_{3,i}(Day4_{t,i}) + \beta_{4,i}(Day5_{t,i}) + \\
&\qquad\qquad \beta_{5,i}(Day6_{t,i}) + \beta_{6,i}(Day7_{t,i}) + e_{t,i} \\
&\text{Level 2:} \quad \beta_{0,i} = (\gamma_{00} + u_{0,i}) \\
&\qquad\qquad \beta_{1,i} = \gamma_{10} \\
&\qquad\qquad \beta_{2,i} = \gamma_{20} \\
&\qquad\qquad \beta_{3,i} = \gamma_{30} \\
&\qquad\qquad \beta_{4,i} = \gamma_{40} \\
&\qquad\qquad \beta_{5,i} = \gamma_{50} \\
&\qquad\qquad \beta_{6,i} = \gamma_{60} \\
&\text{Combined:} \quad \log(VM_{t,i}) = (\gamma_{00} + u_{0,i}) + \gamma_{10}(Day2_{t,i}) + \\
&\qquad\qquad \gamma_{20}(Day3_{t,i}) + \gamma_{30}(Day4_{t,i}) + \\
&\qquad\qquad \gamma_{40}(Day5_{t,i}) + \gamma_{50}(Day6_{t,i}) + \\
&\qquad\qquad \gamma_{60}(Day7_{t,i}) + e_{t,i}
\end{aligned}
\tag{5.3}
$$

Here, $u_{0,i}$ is the location-model random intercept value for each individual $i$ representing the individual-specific deviation (or difference) from the location-model fixed intercept, $\gamma_{00}$, thus allowing for constant individual mean differences across all study days. Note

that the variance of the $u_{0,i}$ values was quantified by the location-model random intercept variance, $\sigma^2_{u_0}$, which was assumed constant between individuals (i.e., no subscript $i$; $\mathbf{G}_i = \mathbf{G}$), as shown in (5.4).

$$\log\left(\sigma^2_{u_0}\right) = \alpha_0^{u_0} \tag{5.4}$$

***Research question 2.*** Whether individual differences in (the log of) average vector magnitude could be predicted by probable mild AD status after controlling for day in study, an individual's age, years of formal education, biologic sex, cardiorespiratory capacity, and body composition were addressed by including these level-2, individual-level predictors in the location model.

The inclusion of a specific covariate in the final location model was determined based on a series of preliminary, covariate-only location models where the significance of the specific covariate's location-model fixed effect was evaluated. Each preliminary, covariate-only location model included only one covariate in addition to day of study; thus, five separate location-model random intercept models were estimated, one for each covariate. An example of the preliminary location model for day in study as well as for an individual's age, $Age_i$, is presented in (5.5).

$$\begin{aligned}
\text{Level 1:} \quad \log\left(VM_{t,i}\right) &= \beta_{0,i} + \beta_{1,i}\left(Day2_{t,i}\right) + \beta_{2,i}\left(Day3_{t,i}\right) + \\
&\quad \beta_{3,i}\left(Day4_{t,i}\right) + \beta_{4,i}\left(Day5_{t,i}\right) + \\
&\quad \beta_{5,i}\left(Day6_{t,i}\right) + \beta_{6,i}\left(Day7_{t,i}\right) + e_{t,i} \\
\text{Level 2:} \quad \beta_{0,i} &= \left(\gamma_{00} + u_{0,i}\right) + \gamma_{01}(Age_i) \\
\beta_{1,i} &= \gamma_{10} \\
\beta_{2,i} &= \gamma_{20}
\end{aligned} \tag{5.5}$$

$$
\begin{aligned}
\beta_{3,i} &= \gamma_{30} \\
\beta_{4,i} &= \gamma_{40} \\
\beta_{5,i} &= \gamma_{50} \\
\beta_{6,i} &= \gamma_{60} \\
\text{Combined: } \log(VM_{t,i}) &= \gamma_{00} + \gamma_{10}(Day2_{t,i}) + \gamma_{20}(Day3_{t,i}) + \\
&\quad \gamma_{30}(Day4_{t,i}) + \gamma_{40}(Day5_{t,i}) + \\
&\quad \gamma_{30}(Day6_{t,i}) + \gamma_{40}(Day7_{t,i}) + \\
&\quad \gamma_{01}(Age_i) + e_{t,i}
\end{aligned}
\tag{5.5}
$$

Here, $\gamma_{10}$ through $\gamma_{60}$ are location-model fixed effects representing the difference in (the log of) average vector magnitude between the first study day and the second through seventh study days, respectively, whereas $\gamma_{01}$ represents the difference in (the log of) average vector magnitude for each additional year of age. Models for additional covariates were estimated by retaining all day of study fixed effects, but substituting the other covariates in place of for $Age_i$ alongside its unique fixed effect $\gamma$.

***Research question 3.*** Whether individual differences in the residual variability of (the log of) average vector magnitude existed was addressed directly by modifying the scale model for the residual variance to include the scale-model random intercept for each individual $i$, as shown in (5.6). Note that this model used the final location model determined by research question 2.

$$
\begin{aligned}
\text{Level 1:} \quad \log(\sigma^2_{e_{t,i}}) &= \tau^e_{0,i} + \tau^e_1(Day2_{t,i}) + \tau^e_2(Day3_{t,i}) + \\
&\quad \tau^e_3(Day4_{t,i}) + \tau^e_4(Day5_{t,i}) + \\
&\quad \tau^e_5(Day6_{t,i}) + \tau^e_6(Day7_{t,i})
\end{aligned}
\tag{5.6}
$$

$$
\begin{aligned}
\text{Level 2:} \qquad \tau_{0,i}^{e} &= \left(\delta_{00}^{e} + \omega_{0,i}^{e}\right) \\
\tau_{1}^{e} &= \delta_{10}^{e} \\
\tau_{2}^{e} &= \delta_{20}^{e} \\
\tau_{3}^{e} &= \delta_{30}^{e} \\
\tau_{4}^{e} &= \delta_{40}^{e} \\
\tau_{5}^{e} &= \delta_{50}^{e} \\
\tau_{6}^{e} &= \delta_{60}^{e} \\
\text{Combined:} \quad \log\left(\sigma_{e_{t.i}}^{2}\right) &= \left(\delta_{00}^{e} + \omega_{0,i}^{e}\right) + \delta_{10}^{e}\left(Day2_{t,i}\right) + \\
&\quad \delta_{20}^{e}\left(Day3_{t,i}\right) + \delta_{30}^{e}\left(Day4_{t,i}\right) + \\
&\quad \delta_{40}^{e}\left(Day5_{t,i}\right) + \delta_{50}^{e}\left(Day6_{t,i}\right) + \\
&\quad \delta_{60}^{e}\left(Day7_{t,i}\right)
\end{aligned}
\tag{5.6}
$$

Here, $\delta_{00}^{e}$ is the fixed intercept for the (log of the) residual variance, and $\omega_{0,i}^{e}$ is the scale-model random intercept representing the deviation from the fixed intercept for the (log of the) residual variance specifically for individual $i$, thus allowing for constant between-individual differences in residual variance. The variance of the $\omega_{0,i}^{e}$ values is quantified by the scale-model random intercept variance, $\sigma_{\omega_0^e}^2$. Further, it was assumed that the location- and scale-model random intercept variances and correlation were constant across individuals (i.e., no subscript $i$; $\mathbf{G}_i = \mathbf{G}$). This assumption was shown for the location-model random intercept variance in (5.4), and is shown below in (5.7) and (5.8) for the scale-model random intercept variance and the correlation between the location- and scale-model random intercepts, respectively.

$$
\log\left(\sigma_{\omega_0^e}^{2}\right) = \alpha_0^{\omega_0^e} \tag{5.7}
$$

$$\tanh^{-1}\left(\rho_{u_0;\omega_0^e}\right) = \alpha_0^{u_0;\omega_0^e} \tag{5.8}$$

***Research question 4.*** Whether individual differences in the residual variability of (the log of) average vector magnitude were predicted by probable mild AD status after controlling for day of study, an individual's age, years of formal education, biologic sex, cardiorespiratory capacity, and body composition were addressed by including these predictors in the scale model for the residual variance.

The model-building approach used for the scale model for the residual variance shadowed the approach detailed above for the location model, where day of study indicator variables were included to account for all systematic and random variability due to day of study, and the inclusion of a specific covariate in the final scale model for the residual variance was determined based on the preliminary significance of a specific covariate's fixed effect. Note that although residual variances were modeled as heterogeneous across study days given the inclusion of the day of study variables, residual values were still assumed to be independent within individuals (i.e., all residual correlations in $\mathbf{R}_i$ were still assumed to be 0).

Each scale model for the residual variance included only one covariate in addition to day of study; thus, five separate mixed-effects location-scale models were estimated, one for each covariate. An example covariate-only scale model for the residual variance which includes day in study as well as an individual's age, $Age_i$, is presented in (5.9).

$$\begin{aligned} \text{Level 1:} \quad \log\left(\sigma^2_{e_{t,i}}\right) = {} & \tau^e_{0,i} + \tau^e_1\left(Day2_{t,i}\right) + \tau^e_2\left(Day3_{t,i}\right) + \\ & \tau^e_3\left(Day4_{t,i}\right) + \tau^e_4\left(Day5_{t,i}\right) + \\ & \tau^e_5\left(Day6_{t,i}\right) + \tau^e_6\left(Day7_{t,i}\right) \end{aligned} \tag{5.9}$$

$$
\begin{aligned}
\text{Level 2:} \qquad \tau_{0,i}^{e} &= \left(\delta_{00}^{e} + \omega_{0,i}^{e}\right) + \delta_{01}^{e}(Age_i) \\
\tau_{1}^{e} &= \delta_{10}^{e} \\
\tau_{2}^{e} &= \delta_{20}^{e} \\
\tau_{3}^{e} &= \delta_{30}^{e} \\
\tau_{4}^{e} &= \delta_{40}^{e} \\
\tau_{5}^{e} &= \delta_{50}^{e} \\
\tau_{6}^{e} &= \delta_{60}^{e}
\end{aligned} \tag{5.9}
$$

$$
\begin{aligned}
\text{Combined:} \quad \log\left(\sigma_{e_{t.i}}^{2}\right) = {} & \left(\delta_{00}^{e} + \omega_{0,i}^{e}\right) + \delta_{10}^{e}\left(Day2_{t,i}\right) + \\
& \delta_{20}^{e}\left(Day3_{t,i}\right) + \delta_{30}^{e}\left(Day4_{t,i}\right) + \\
& \delta_{40}^{e}\left(Day5_{t,i}\right) + \delta_{50}^{e}\left(Day6_{t,i}\right) + \\
& \delta_{60}^{e}\left(Day7_{t,i}\right) + \delta_{01}^{e}(Age_i)
\end{aligned}
$$

Here, $\delta_{10}^{e}$ through $\delta_{60}^{e}$ are fixed effects representing the difference in residual variance between the first study day and the second through seventh study days, respectively, whereas $\delta_{01}^{e}$ represents the difference in residual variance for each additional year of age. The scale model for the residual variance that included additional covariates was estimated by retaining all day of study fixed effects, but substituting the other covariates in place of $Age_i$ alongside its unique fixed effect $\delta^{e}$.

**Analytic strategy.** All available data (excluding zeros) collected during the first seven study days were included in analysis; no imputation methods were employed for missing data. As detailed in chapter 3, the inclusion of any additional random effect was evaluated by model comparison using DIC, where smaller DIC indicated improved model fit. To describe random variation around a given location- or scale-model fixed effect, 95% random effect confidence intervals were calculated as: fixed effect ±

$1.96\sqrt{\text{random effect variance}}$. Further, effect sizes are reported as proportion of variance explained using pseudo-$R^2$ calculated as: ($\text{variance}_{\text{larger}} - \text{variance}_{\text{smaller}}$) / $\text{variance}_{\text{larger}}$, with the specific variance component explained by the level-2 predictor dependent on whether the predictor was included in the location or scale model. In addition, the decision to include a given covariate in the final location or scale model was determined by the 80% credible interval (i.e., the $10^{\text{th}}$ to $90^{\text{th}}$ percentile of the posterior distribution) which excluded zero. For the final location and scale models, all continuous covariates were centered near their mean to ensure meaningful interpretation of the intercept. After including all relevant covariates, the significance of probable mild AD status in the location and scale models was determined by a fixed effect for which its 95% credible interval excluded zero.

**Model estimation via MCMC.** All models were estimated using the MCMC estimator detailed in chapter 3. Start values for all location-model fixed effects, the location-model random intercept variance, and the residual variance (which served as a proxy for the scale-model fixed intercept), were based on preliminary estimation of a traditional linear mixed-effects model using the lme4 package in R developed by Bates et al. (2014). Start values for additional fixed and random effects included in the scale model for the residual variance as well as the correlation between the location- and scale-model random intercepts were set to zero. Prior to initiation of the Markov chain, the candidate-generating distribution of all parameters were tuned to achieve an acceptance rate of 45% using 20 tuning chains of 100 iterations. Following tuning, the Markov chain was specified to sample 100,000 iterations (due to increased model complexity compared to the models estimated in the simulation studies), with a burn-in period set to the first

25,000 iterations and a thinning interval of 1 (i.e., no thinning). Convergence was assessed empirically by Geweke's diagnostic test (Geweke, 1992) and the Gelman and Rubin criterion (Gelman & Rubin, 1992), both calculated using the CODA package in R (Plummer et al., 2006). Convergence was defined as satisfying at least one criterion; all parameters met convergence criteria.

**Results**

Average vector magnitude was measured between 5 and 13 days, with 92 individuals providing 9,251 total observations ($M$ = 124.76, $SD$ = 129.10, range = 0.00 to 1734.20), with the average of 99.16 observations per individual ($SD$ = 12.26, range = 55 to 120). Of these observations, 291 (3.12%) were excluded due to zero values (a much lower proportion than expected) and an additional 84 (0.94%) observations were excluded because they occurred after the seventh study day. In addition, within the probable mild AD group, five individuals were missing their baseline cardiorespiratory capacity (i.e., VO2) measurements with one additional individual missing years of formal education; no missing data was observed for healthy individuals. As a result, the final location and scale models were based on 8,442 total observations from 86 individuals (data scale: $M$ = 131.80, $SD$ = 130.68, range = 0.10 to 1734.20; log scale: $M$ = 4.37, $SD$ = 1.27, range = –2.30 to 7.46).

Table 5.1
*Group-Specific Descriptive Statistics*

| | Healthy Individuals ($n$ = 53) | | Individuals with Probable Mild AD ($n$ = 33) | |
|---|---|---|---|---|
| | $M$ ($SD$) | Range | $M$ ($SD$) | Range |
| Age | 73.19 (6.53) | 62 to 92 | 72.73 (7.47) | 60 to 86 |
| Years of Education | 17.32 (3.38) | 12 to 25 | 15.61 (2.94) | 10 to 20 |
| VO2 max | 1.60 (0.45) | 0.79 to 2.74 | 1.73 (0.56) | 0.62 to 3.10 |
| Body Mass Index | 26.42 (4.42) | 19.69 to 36.68 | 27.18 (5.03) | 19.31 to 38.38 |

Group-specific descriptive statistics are provided in Table 5.1. Of the 86 individuals included in the final models, 47 (54.65%) were women, of which 37 (69.81%) were healthy individuals and 10 (30.30%) were individuals with probable mild AD, $X^2$ (1, $N$ = 86) = 12.81, $p < .001$. A statistically significant between-group difference was also indicated for years of education, $t(84) = 2.40$, $p < .05$, with healthy individuals being slightly more educated, on average (although both groups averaged an equivalent of a bachelor's degree); the groups were similar in terms of age, VO2 max, and BMI.

**Building the location model.** An initial location-model random intercept model which included day of study as location model fixed effects with an unconditional scale model for the residual variance estimated a homogeneous ICC of 0.19, indicating that 19% of the variability in (the log of) average vector magnitude was due to level-2, between-individual differences. That is, up to 19% of *total variability* in (the log of) average vector magnitude was available to be explained by probable mild AD status.

***Research question 1.*** Individual differences in (the log of) average vector magnitude was evaluated by comparing an unconditional location-model random intercept model defined in (5.3) to the single-level linear model defined in (5.1). Results indicated that the location-model random intercept model fit better than the single-level linear model, DIC = 28,051.44 vs. 29,723.48, respectively, indicating significant individual mean differences in (the log of) average vector magnitude (fixed intercept, $\gamma_{00}$ = 4.38 on log scale; location-model random intercept variance = 0.31 on the variance scale). A 95% random effects confidence interval indicated that approximately 95% of the sample had predicted individual (log of) average vector magnitude intercepts ranging between 3.77 and 4.98 (data scale: 43.54 and 145.78), indicating most activity was

sedentary (i.e., sitting, reading, or lying down; Crouter, Horton, & Bassett, 2012). These results indicated that significant constant individual differences in (the log of) average vector magnitude existed and could be potentially predicted by mild AD status.

***Research question 2.*** Whether individual differences in (the log of) average vector magnitude were predicted by mild AD status after controlling for day in study, an individual's age, years of formal education, biologic sex, cardiorespiratory capacity, and body composition was addressed by including these level-2 predictors in the location model. Preliminary, covariate-only location models indicated that body composition was not statistically significant and was therefore excluded from the final location model.

Results of the final location-model are provided in Table 5.2. Note that age was centered at 70, years of education was centered at 16 (i.e., a college graduate), and cardiorespiratory capacity (VO2) was centered at 1.50 mL/(kg·min). Results indicated that the location-model level-2 predictors as a set explained approximately 16.85% of location-model random intercept variance. Compared to healthy individuals, individuals with probable mild AD averaged nonsignificantly lower average vector magnitude by 18.94% (i.e., $[1 - \exp(-0.21)] * 100$), $B$ = –0.21, 95% CI [–0.47,0.02], after controlling for day of study, age, years of formal education, biologic sex, and cardiorespiratory capacity. Probable mild AD status explained approximately 1.89% of location-model random intercept variance and was not statistically significant as the 95% credible interval included zero. Therefore, results indicated individual differences in (the log of) average vector magnitude were not predicted by mild AD status after controlling for the age, years of formal education, biologic sex, and cardiorespiratory capacity; however, a difference of 18.94% in total unstructured movements may indicate some clinical utility.

Table 5.2
*Results of the Final Location Model*

| Location Model | Posterior Mean | Posterior SD | 95% Credible Interval Lower | 95% Credible Interval Upper |
|---|---|---|---|---|
| Fixed Intercept, $\gamma_{00}$ | 4.36 | 0.15 | 4.07 | 4.63 |
| Day of Study[a] | | | | |
| Day 2, $\gamma_{10}$ | –0.06 | 0.04 | –0.14 | 0.02 |
| Day 3, $\gamma_{20}$ | –0.06 | 0.04 | –0.13 | 0.01 |
| Day 4, $\gamma_{30}$ | –0.05 | 0.04 | –0.12 | 0.03 |
| Day 5, $\gamma_{40}$ | –0.01 | 0.04 | –0.09 | 0.06 |
| Day 6, $\gamma_{50}$ | –0.03 | 0.04 | –0.11 | 0.04 |
| Day 7, $\gamma_{60}$ | **–0.20** | 0.06 | –0.31 | –0.09 |
| Age (0 = 70), $\gamma_{02}$ | –0.01 | 0.01 | –0.03 | 0.00 |
| Years of Education (0 = 16), $\gamma_{03}$ | 0.02 | 0.02 | –0.01 | 0.04 |
| Woman, $\gamma_{04}$ | 0.18 | 0.14 | –0.05 | 0.44 |
| VO2 (0 = 1.50), $\gamma_{05}$ | **0.28** | 0.12 | 0.07 | 0.50 |
| Mild Alzheimer's Status, $\gamma_{01}$ | –0.21 | 0.12 | –0.47 | 0.02 |
| Scale Model for the Residual Variance | | | | |
| Residual Variance Fixed Intercept, $\delta_{00}^{e_t}$ | 0.18 | 0.04 | 0.11 | 0.26 |
| Day of Study[a] | | | | |
| Day 2, $\delta_{10}^{e_t}$ | **0.16** | 0.05 | 0.06 | 0.26 |
| Day 3, $\delta_{20}^{e_t}$ | 0.08 | 0.05 | –0.02 | 0.18 |
| Day 4, $\delta_{30}^{e_t}$ | 0.06 | 0.05 | –0.04 | 0.16 |
| Day 5, $\delta_{40}^{e_t}$ | **0.12** | 0.05 | 0.02 | 0.23 |
| Day 6, $\delta_{50}^{e_t}$ | 0.03 | 0.05 | –0.07 | 0.13 |
| Day 7, $\delta_{60}^{e_t}$ | **0.33** | 0.06 | 0.21 | 0.46 |
| Variance Components | | | | |
| Location-Model Random Intercept, $\alpha_0^{u_0}$ | –1.36 | 0.17 | –1.68 | –1.02 |

*Note.* Statistically significant effects are presented in boldface. DIC = 26,392.28.
[a]Reference day was study day 1.

**Building the scale model for the residual variance.** All models estimated below used the final location model that included the location-model random intercept variance and location-model fixed effects for mild AD status, age, years of education, biologic sex, and cardiorespiratory capacity.

***Research question 3.*** When compared to the final location model, the addition of the scale-model random intercept variance improved model fit, DIC = 26,392.28 vs.

25,399.39, respectively, indicating significant individual differences in the residual variability of (the log of) average vector magnitude, with (the log of) the residual variance fixed intercept being 0.06 and scale-model random intercept variance of 0.29. The 95% random effects confidence interval indicated approximately 95% of the sample had predicted individual (log of the) residual variance intercepts ranging between –0.50 and 0.62 (variance scale: 0.60 and 1.86), indicating that significant constant individual differences in residual variability were available to be predicted by mild AD status.

***Research question 4.*** Whether individual differences in the residual variability of (the log of) average vector magnitude were predicted by mild AD status after controlling for day of study, an individual's age, years of formal education, biologic sex, cardiorespiratory capacity, and body composition was addressed by including these level-2 predictors in the scale model for the residual variance. Preliminary, covariate-only models indicated that cardiorespiratory capacity and body composition did not predict individual differences in residual variability and were excluded from the final scale model for the residual variance.

Results indicated that the level-2 predictors included in the scale model for the residual variance as a set explained approximately 8.69% of the scale-model random intercept variance. Individuals with probable mild AD averaged 1.01% (i.e., $[\exp(0.01) - 1] * 100$) more residual variability compared to healthy controls, $B = 0.01$, 95% CI [–0.21,0.22], after controlling for day of study, age, years of formal education, and biologic sex. The fixed effect of probable mild AD status explained 0.00% of scale-model random intercept variance and was not statistically significant as the 95% credible interval included zero. This result indicated that individual differences in the residual

Table 5.3
*Results of the Final Location and Scale Model for the Residual Variance*

| | | | 95% Credible Interval | |
|---|---|---|---|---|
| Location Model | Posterior Mean | Posterior SD | Lower | Upper |
| Fixed Intercept, $\gamma_{00}$ | 4.37 | 0.14 | 4.11 | 4.62 |
| Day of Study[a] | | | | |
| Day 2, $\gamma_{10}$ | –0.03 | 0.04 | –0.11 | 0.04 |
| Day 3, $\gamma_{20}$ | –0.05 | 0.04 | –0.12 | 0.02 |
| Day 4, $\gamma_{30}$ | –0.06 | 0.04 | –0.13 | 0.01 |
| Day 5, $\gamma_{40}$ | 0.01 | 0.04 | –0.07 | 0.08 |
| Day 6, $\gamma_{50}$ | –0.03 | 0.04 | –0.10 | 0.04 |
| Day 7, $\gamma_{60}$ | **–0.19** | 0.05 | –0.29 | –0.09 |
| Age (0 = 70), $\gamma_{02}$ | –0.01 | 0.01 | –0.03 | 0.01 |
| Years of Education (0 = 16), $\gamma_{03}$ | 0.01 | 0.02 | –0.02 | 0.05 |
| Woman, $\gamma_{04}$ | 0.18 | 0.14 | –0.09 | 0.44 |
| VO2 (0 = 1.50), $\gamma_{05}$ | **0.26** | 0.10 | 0.07 | 0.46 |
| Mild Alzheimer's Status, $\gamma_{01}$ | **–0.23** | 0.11 | –0.45 | –0.02 |
| Scale Model for the Residual Variance | | | | |
| Residual Variance Fixed Intercept, $\delta_{00}^{e_t}$ | 0.19 | 0.12 | –0.02 | 0.44 |
| Day of Study[a] | | | | |
| Day 2, $\delta_{10}^{e_t}$ | **0.16** | 0.05 | 0.06 | 0.26 |
| Day 3, $\delta_{20}^{e_t}$ | **0.14** | 0.05 | 0.04 | 0.23 |
| Day 4, $\delta_{30}^{e_t}$ | 0.08 | 0.05 | –0.02 | 0.19 |
| Day 5, $\delta_{40}^{e_t}$ | **0.12** | 0.05 | 0.01 | 0.23 |
| Day 6, $\delta_{50}^{e_t}$ | –0.01 | 0.05 | –0.10 | 0.09 |
| Day 7, $\delta_{60}^{e_t}$ | **0.39** | 0.07 | 0.25 | 0.54 |
| Age (0 = 70) , $\delta_{02}^{e_t}$ | 0.01 | 0.01 | –0.01 | 0.02 |
| Years of Education (0 = 16) , $\delta_{03}^{e_t}$ | –0.02 | 0.02 | –0.06 | 0.01 |
| Woman, $\delta_{04}^{e_t}$ | **–0.28** | 0.12 | –0.54 | –0.10 |
| Mild Alzheimer's Status, $\delta_{01}^{e_t}$ | 0.01 | 0.11 | –0.21 | 0.22 |
| Variance Components | | | | |
| Location-Model Random Intercept, $\alpha_0^{u_0}$ | –1.36 | 0.18 | –1.70 | –1.00 |
| Location-Scale Random Intercept Correlation, $\alpha_0^{u_0;\omega_0^e}$ | –0.15 | 0.04 | –0.23 | –0.09 |
| Scale-Model Random Intercept, $\alpha_0^{\omega_0^e}$ | –1.33 | 0.18 | –1.67 | –0.97 |

*Note.* Statistically significant effects are presented in boldface. DIC = 25,396.46.
[a]Reference day was study day 1.

variability of (the log of) average vector magnitude were not predicted by mild AD status after controlling for age, years of formal education, and biologic sex.

Finally, all location-model posterior means and standard deviations in Table 5.3 were very similar to those estimated by the final location model presented in Table 5.2. However, in the final location and scale model for the residual variance, the fixed effect for mild AD status was statistically significant, $B = –0.23$, 95% CI [–0.45,–0.02], with individuals with probable mild AD averaging 20.65% (i.e., $[1 - \exp(-0.23)] * 100$) lower average vector magnitude compared to healthy individuals. This result updated the previous result reported for the second research question and indicated that probable mild AD status does predict individual differences in (the log of) average vector magnitude.

**Discussion**

This chapter explicitly detailed the applicability of the mixed-effects location-scale model to an empirical dataset examining total movement (measured in three axes of motion) in a sample of older adults with and without probable mild AD. A traditional model-building approach was employed throughout, in which the location model was specified prior to specifying the scale model for the residual variance. Although significant individual differences in the residual variability of (the log of) average vector magnitude were observed, as indicated by the scale-model random intercept variance, results indicated probable mild AD status did not explain a significant proportion of these individual differences after adjusting for an individual's age, years of formal education, and biologic sex.

With that said, significant individual differences in the mean level of (the log of) average vector magnitude were observed, as indicated by the location-model random

intercept variance. Prior to including the individual-level predictors in the final scale model for the residual variance, no statistically significant differences were observed in (the log of) average vector magnitude between individuals with and without probable mild AD after controlling for an individual's age, years of formal education, biologic sex, and cardiorespiratory capacity. However, this difference became statistically significant upon inclusion of level-2 predictors the final scale model for the residual variance, as individuals with probable mild AD status had significantly lower (log of) average vector magnitude compared to healthy individuals after controlling for the covariates. On the surface, this appeared to contradict the results of the second simulation study. However, in the final location model presented in Table 5.2, in which the scale model for the residual variance did not include any level-2 predictors nor scale-model random intercept variance, the 95% credible interval for the location-model fixed effect of probable mild AD status only just included zero, 95% CI [–0.47,0.02], whereas when including the level-2 predictors and random intercept variance in the scale model for the residual variance, the 95% credible interval for the location-model fixed effect of probable mild AD status just excluded zero, 95% CI [–0.45,–0.02]. The difference in parameter estimates of mild AD status between models that omitted and included the final scale model for the residual variance was only 0.02 on the log scale (i.e., –0.21 vs. –0.23) or 1.71% on the data scale (i.e., 18.94% vs. 20.65%). From a strictly statistical perspective, it is difficult to determine which result is the least wrong, but it is a safe assumption that this minimal difference does not indicate significant bias. Thus, it is the task of the empirical scientist to determine whether the consistent observed difference of approximately 20% is clinically meaningful, irrespective of statistical significance.

In the second simulation study reported in chapter 4, the location-model fixed effects for level-2 predictors and location-model random intercept variance, as well as their associated precision (i.e., posterior standard deviations), remained unbiased regardless of whether the level-2 predictors and/or random intercept were included in the scale model for the residual variance, with no modification of statistical significance across models. With that said, the current location model and scale model for the residual variance included day of study as a set of level-1 fixed effects, although this empirical analysis cannot quantify potential bias in location-model fixed effects of level-2 predictors resulting from the exclusion of relevant level-1 predictors in the scale model for the residual variance.

Finally, the primary statistical limitation of this study was that although residual variances were specified to be heterogeneous across study days, the residual values were assumed to be independent. As stated throughout, the assumption of independence was untestable using the MCMC estimator developed for this dissertation, and may have introduced bias into inferences of location- and scale-model fixed effects to an unknown extent.

**Chapter Summary**

The purpose of this chapter was to provide an explicit example showing the capability of the mixed-effects location-scale model to evaluate mean- and variability-related hypotheses using empirical data collected from 86 older adults with and without mild probable Alzheimer's disease. The chapter began with a brief description of the study design, data collection procedure, and description of the variables collected, and concluded with a description of the mixed-effects location-scale model building

procedure employed to determine whether individual differences in both mean level and variability of unstructured movement was predicted by a clinical diagnosis of probable mild Alzheimer's disease after controlling for relevant covariates.

In chapter 6, an overarching summary the simulation studies and empirical analysis conducted in this dissertation is presented followed by a discussion of specific empirical research areas that could benefit from evaluating and predicting individual differences in outcome variability through estimation of the mixed-effects location-scale model.

# CHAPTER 6: GENERAL DISCUSSION

This final chapter presents an overarching summary of the simulation studies and empirical analysis conducted in this dissertation as well as specific empirical research areas that could benefit from evaluating and predicting individual differences in outcome variability through estimation of the mixed-effects location-scale model.

## Brief Summary of Results

Conducting methodological studies without application is an exercise in futility. Thus, following the two simulation studies presented in chapter 4, a direct application of the mixed-effects location-scale model was presented in chapter 5 to show a novel application of this model when evaluating individual differences in mean level and variability of unstructured movement in individuals with and without probable mild Alzheimer's disease. Given its limited precedent in the literature, application of the mixed-effects location-scale model will most likely be relatively novel to empirical scientists across many fields of study. It is this novelty that is most exciting, as it potentially engenders additional variability-based research questions, especially to those empirical scientists who have focused solely on questions related to mean levels of an outcome. Indeed, what is noteworthy is the fact that mean- and variability-based questions can be evaluated in a single model.

To begin to address some of the nuances of studying individual differences in response variability, two simulations studies were conducted. The first simulation study provided power curves for detecting scale-model random intercept variance (i.e., individual differences in response variability) across several study design characteristics and model parameters. This study provided initial insight for empirical scientists to

design studies to evaluate individual differences in variability—and to ensure these differences could be detected if they actually exist. It was recommended that a minimum of 20 repeated occasions from at least 100 individuals be collected to ensure at least 80% power to detect the scale-model random intercept variance of 0.035. After ensuring individual differences in residual variability can be detected, power curves were then calculated for the prediction of these individual differences by an individual-level variable. Results indicated approximately 100% power to detect the fixed effect of an individual-level predictor that was ≥ 0.25 (or that explained ≥ 50% of the scale-model random intercept variance), which was not moderated by any study design characteristic or model parameter. Further, the power to detect the fixed effect of an individual-level predictor that was < 0.25 (or that explained < 50% of the scale-model random intercept variance) increased with increases in the number of individuals and the number of repeated occasions within an individual (although the number of individuals and occasions used in this study failed to achieve 80% power for these effects), but decreased with increases in $\sigma^2_{\omega_0^e}$ (i.e., more scale-model random intercept variance resulted in smaller parameter estimates, or effect sizes, for the individual-level predictor to become undetectable). Further, if individual differences in variability were not detected, it was unlikely that systematic differences in variability could be predicted by level-2 fixed effects. With that said, Type I error rates for these systematically-varying effects were lower when retaining non-significant scale-model random intercept variance in the model, often near or below the nominal 5%. Thus, results suggest that although it may be appropriate to model systematically-varying effects, the detection of the scale-model random intercept variance should occur prior to predicting individual-level differences in

residual variability to increase the likelihood that these predictor effects would be detected.

The second simulation study was conducted to inform empirical scientists about the consequences of predicting individual differences in the presence of a misspecified location and/or scale model. Traditionally, the location model has been specified prior to the scale model; however, there is no consensus among methodologists (e.g., Bryk & Raudenbush, 1988) and this model-building procedure has not been studied extensively. Given that model-building is a multi-step process, results of this simulation study informed model-building practice by indicating that it was unnecessary to explicitly specify the location model before the scale model for the residual variance (or vice versa) when estimating individual differences in both the location and scale model, and subsequently predicting them with individual-level predictors, given that Type I error rates remained low and bias was non-existent.

Finally, when considering the theoretical framework presented for the mixed-effects location-scale model alongside the two simulation studies and empirical analysis, this dissertation provides empirical scientists with incredibly useful information as they progress from study design through analysis, interpretation, and reporting for publication. With this in mind, what follows below is a discussion of specific empirical research areas that could benefit from evaluating and predicting individual differences in residual variability through the use of mixed-effects location-scale model.

**Directions for Future Empirical Research**

Given the novelty of the mixed-effects location-scale model, empirical scientists from a wide array of fields would benefit from resources (e.g., pedagogical articles)

pertaining to the use of this model, specifically informing them how individual differences in response variability can be identified and predicted. Presenting the mixed-effects location-scale model in this format, although requiring considerable cognitive effort, would not be overwhelmingly daunting as many empirical scientists have some existing familiarity with the linear mixed-effects model used to evaluate individual differences in mean levels and in change over time. With this information in mind, the remainder of this chapter describes potential empirical research areas for which the mixed-effects location-scale model could be directly applicable.

Regarding the practical application of the mixed-effects location-scale model, the overwhelming majority of publications that have actually applied this model have been found within nicotine and tobacco research (e.g., Hedeker & Mermelstein, 2007; Hedeker et al., 2008; Hedeker et al., 2009). Each of these studies used ecological momentary assessments to evaluate positive and negative affect immediately following the use of a combustible cigarette. This same approach could be applied directly when studying individuals who use electronic cigarettes (or e-cigarettes). For example, although it is important to determine whether the lung function (as measured by, say, methacholine challenge) of e-cigarette users improves, on average, compared to traditional cigarette smokers, it would also be important to determine whether the lung function of e-cigarette users is more variable given that individuals are essentially self-titrating their nicotine, as well as the characteristics of an individual or e-cigarette that may predict increases (or decreases) in variability.

In addition, pediatric obesity research has embraced the technological side of data collection by routinely using accelerometers to continuously collect physical activity

data, in the presence or absence of an intervention, over a given study period. The sheer quantity of data per child produced by an accelerometer lends itself nicely to the application of a mixed-effects location-scale model. To date, pediatric obesity research has focused primarily on a child's mean level of physical activity, under the assumption that all children respond to the intervention in a consistent manner (i.e., homoscedastic residuals; see Cushing, Walters, & Hoffman, 2014). Individual differences in the variability of physical activity could (and should) also be evaluated, especially with a non-homogenous sample of children (i.e., not a randomized controlled trial). It may be that a novel intervention is more effective for some children compared to others, where a differential treatment effect necessarily increases response variability. Therefore, although mean levels of physical activity may increase for the intervention condition compared to a control condition, the intervention may increase variability to the point that a sizeable proportion of children have physical activity levels in the presence of the intervention at or below their baseline/control level. An awareness of this information would be critically important! Thus, the ability to identify individual differences in the variability in physical activity, and predict why these individual differences exist, would afford knowledge of the child-level characteristics related to increased variability in the presence of the intervention. Given that physical activity *should be* a lifestyle or habit, the goal of any intervention should be to create a consistent response. Thus, it would be important to identify the characteristics of the children who responded positively *and consistently* to the intervention, and continue to modify the intervention for the children who displayed increased variability regardless of whether their mean levels increased or decreased.

Although this dissertation has spoken specifically to repeated-measures data, it does not preclude the mixed-effects location-scale model from being applied to cross-sectional data (see Leckie et al., 2014). For example, within-school variability in standardized test scores could be identified and subsequently predicted by the proportion of students who meet some fitness standard, or, perhaps more likely, be predicted by the proportion of students within a school that receive free/reduced lunch as a proxy for socio-economic status. This information could be used subsequently to inform potential policy changes.

In conclusion, estimation of the mixed-effects location-scale model allows empirical scientists to pose and answer a plethora of research questions that they may have never considered in their specific area of research. Indeed, given that the use of this model is relatively rare in most fields of study, the purpose of this dissertation was to provide empirical scientists with practical information to ensure they have the tools necessary to design an appropriate study to answer variability-related questions or hypotheses in addition to presenting them with explicit examples of model estimation and interpretation. Although this dissertation explores the tip of the iceberg when considering the methodological studies that still need to be conducted regarding the mixed-effects location-scale model, it provides important initial groundwork for empirical scientists to begin confidently and competently using this model in their own area of research.

## REFERENCES

ActiGraph (2012). *ActiLife 6 user's manual*. Pensacola, FL: ActiGraph Software Department. Retrieved from http://www.actigraphcorp.com/support/downloads/#manuals

Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician, 52*, 119-126.

Ahlskog, J. E., Geda, Y. E., Graff-Radford, N. R., & Petersen, R. C. (2011). Physical exercise as a preventative or disease-modifying treatment of dementia and brain aging. *Mayo Clinic Proceedings, 86*, 876-884.

Allison, P. D. (1999). *Multiple regression: A primer*. Thousand Oaks, CA: Sage Publications, Inc.

Almeida, D. M., Piazza, J. R., & Stawski, R. S. (2009). Interindividual differences and intraindividual variability in the cortisol awakening response: An examination of age and gender. *Psychology and Aging, 24*, 819-827.

Barnard, J., McCulloch, R., & Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica, 10*, 1281-1311.

Bates, D., Mächler, M., Bolker, B. B., & Walker, S. C. (2014). Lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. Retrieved from http://CRAN.R-project.org/package=lme4

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions, 53*, 370-418.

Berry, W. D., & Feldman, S. (1985). *Multiple regression in practice*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-050. Newbury Park, CA: Sage Publications, Inc.

Blair, R. C., Higgins, J. J., Topping, M. E. H., & Mortimer, A. L. (1983). An investigation of the robustness of the t test to unit of analysis violations. *Educational and Psychological Measurement, 43*, 69-80.

Bolstad, W. M. (2007). *Introduction to Bayesian Statistics* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Box, G. E. P., & Tidwell, P. W. (1962). Transformation of the independent variables. *Technometrics*, 4, 531-550.

Breusch, T., & Pagan, A. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47, 1287-1294.

Browne, W. J., Charlton, C. M. J., Michaelides, D. T., Parker, R. M. A., Cameron, B., Szmaragd, C., Yang, H., Zhang, Z, Goldstein, H., Jones, K., Leckie, G., & Moreau, L. (2013). A beginner's guide to Stat-JR's TREE interface (version 1.0.0) [Computer software manual]. Bristol, UK: Center for Multilevel Modelling. Available from http://www.bristol.ac.uk/cmm/software/statjr/manuals/

Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin, 104*, 396-404.

Burns, J. M., Cronk, B. B., Anderson, H. S., Donnelly, J. E., Thomas, G. P., Harsha, A., & Swerdlow, R. H. (2008). Cardiorespiratory fitness and brain atrophy in early Alzheimer disease. *Neurology, 71*, 210-216.

Charlton, C. M. J., Michaelides, D. T., Parker, R. M. A., Cameron, B., Szmaragd, C., Yang, H., Zhang, Z., Frazer, A. J., Goldstein, H., Jones, K., Leckie, G., Moreau, L., & Browne, W. J. (2013). Stat-JR (Version 2.1) [software]. Available from http://www.bristol.ac.uk/cmm/software/statjr/

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician, 49*, 327-335.

Cleveland, W., Denby, L., & Liu, C. (2000). Random location and scale effects: Model building methods for a general class of models. *Computing Science and Statistics*, *32*, 3-10.

Cleveland, W., Denby, L., & Liu, C. (2002). Random scale effects. Technical report, Bell Labs, http://cm.bell-labs.com/stat/doc/randomscale.pdf.

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limited illustrated in the case of the binomial. *Biometrika, 26*, 404-413.

Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association, 91*, 883-904.

Cribari-Neto, F. (2007). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis*, *45*, 215-233.

Crouter, S. E., Horton, M., & Bassett, D. R. (2012). Use of a 2-regression model for estimating energy expenditure in children. *Medicine & Science in Sports & Exercise, 44*, 1177-1185.

Cushing, C. C., Walters, R. W., & Hoffman, L. (2014). Aggregated N-of-1 randomized controlled trials: Modern data analytics applied to a clinically valid method of intervention effectiveness. *Journal of Pediatric Psychology, 39,* 138-150.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression I. *Biometrika, 37*, 409-428.

Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression II. *Biometrika, 38*, 159-177.

Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121-138.

Ertekin E., & Grossman, J. C. (2008). Introduction to Monte Carlo methods. [PDF document]. Retrieved from Lecture Notes Online Website: http://nanohub.org/resources/4034/download/02-12-08-forweb.pdf

Estabrook, R., Grimm, K. J., & Bowles, R. P. (2012). A Monte Carlo simulation of the reliability of intraindividual variability. *Psychology and Aging, 27*, 560-576.

Evans, M., & Moshonov, H. (2006). Checking for prior-data conflict. *Bayesian Analysis, 1*, 893-914.

Fox, J. (2008). *Applied regression analysis and generalized linear models* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Gallager, R. G. (1996). *Discrete stochastic processes*. Norwell, MA: Kluwer Academic Publishers.

Gelfand, A. E., Dey, D. K., & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.). *Bayesian Statistics 4*. Oxford, UK: Oxford University Press.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* New York: Cambridge University Press.

Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 6,* 733-807.

Gelman, A., & Rubin, D. B. (1992). Inferences from iterative simulation using multiple sequences (with discussion). *Statistical Science, 7,* 457-511.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distribution and Bayesian restoration of images. *IEE Transactions on Patter Analysis and Machine Intelligence, 6,* 721-741.

Greene, W. H. (2002). *Econometric Analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.). *Bayesian statistics 4*. Oxford, UK: Oxford University Press.

Gill, J. (2007). Bayesian methods: A social and behavioral sciences approach (2nd ed.). New York: CRC Press.

Haldane, J. B. S. (1955). The measurement of variation. *Evolution, 9,* 484.

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

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica, 44*, 461-466.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association, 72*, 320-338.

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science, 1*, 297-318.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika, 57*, 97-109.

Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods, 39*, 709-722.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ: John Wiley & Sons, Inc.

Hedeker, D., & Mermelstein, R. J. (2012). Mood changes associated with smoking in adolescents: An application of a mixed-effects location scale model for longitudinal Ecological Momentary Assessment (EMA) data (pp. 59-79). In G. R. Hancock & J. Harring (Eds.), *Advances in Longitudinal Methods in the Social and Behavioral Sciences*. Charlotte, NC: Information Age Publishing.

Hedeker, D., Mermelstein, R. J., Berbaum, M. L., & Campbell, R. T. (2009). Modeling mood variation associated with smoking: An application of heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. *Addiction, 104*, 297-307.

Hedeker, D., Mermelstein, R. J., & Demitras, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics, 64*, 627-634.

Hedeker, D., Mermelstein, R. J., & Demitras, H. (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in Medicine, 31*, 3328-3336.

Hedeker, D., & Nordgren, R. (2013). MIXREGLS: A program for mixed-effects location scale analysis. *Journal of Statistical Software, 52*, 1-38.

Hilden-Minton, J. A. (1995). *Multilevel diagnostics for mixed and hierarchical linear models* (Unpublished doctoral dissertation). University of California, Los Angeles.

Hoffman, L. (2014). *Longitudinal analysis: Modeling within-person fluctuation and change*. New York: Routledge.

Houseman, E. A., Ryan, L. M., & Coull, B. A. (2004). Cholesky residuals for assessing normal errors in a linear model with correlated outcomes. *Journal of the American Statistical Association, 99*, 383-394.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*. Berkeley, CA: University of California Press.

Hultsch, D. F., MacDonald, S. W. S., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults:

Comparisons of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology, 14*, 588-598.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998). Markov chain Monte Carlo in Practice: A roundtable discussion. *The American Statistician, 52*, 93-100.

Kauermann, G., & Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association, 96,* 1387-1396.

Kim, J. S., & Bolt, D. M. (2007). Estimating item response theory models using Markov chain Monte Carlo Methods. *Educational Measurement: Issues and Practice, 26*, 38-51.

Koenker, R. (1981). A Note on studentizing a test for heteroscedasticity. *Journal of Econometrics, 29*, 305-326.

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione (On the empirical determination of a distribution law). *Giornale dell'Istituto Italiano degli Attuari, 4*, 83-91.

Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence form project STAR. *The Elementary School Journal, 108*, 275-291.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*, 963-974.

Leckie, G. (2014). runmixregles: A program to run the MIXREGLS mixed-effects location scale software from within Stata. *Journal of Statistical Software, 59*, 1-41.

Leckie, G., French, R., Charlton, C., & Browne, W. (2014). Modeling heterogeneous variance-covariance components in two-level models. *Journal of Educational and Behavioral Statistics, 39*, 307-332.

Lee, Y., & Nelder, J. A. (2006). Double hierarchical generalized linear models. *Applied Statistics, 55*, 139-185.

Lee, Y., & Noh, M. (2012). Modelling random effect variance with double hierarchical generalized linear models. *Statistical Modelling, 12*, 487-502.

Li, B., Bruyneel, L., & Lesaffre, E. (2014). A multivariate multilevel Gaussian model with a mixed effects structure in the mean and covariance part. *Statistics in Medicine, 33*, 1877-1899.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute, Inc.

Littell, R. C., Pendergast, J., & Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine, 19*, 1793-1819.

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician, 54*, 217-224.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10*, 325-337.

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. New York: Springer.

Manning, W. G. (1998). The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics, 1998*, 283-295.

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticty-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics, 29,* 305-325.

Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association, 44,* 335-341.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics, 21*, 1087-1092.

Nesselroade, J. R., & McCollam, K. M. S. (2000). Putting the process in developmental process. *International Journal of Behavioral Development, 24*, 295-300.

Nesselroade, J. R., & Ram, N. (2004). Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research in Human Development, 1*, 9-29.

Nesselroade, J. R., & Salthouse, T. A. (2004). Methodological and theoretical implications of intraindividual variability in perceptual-motor performance. *Journal of Gerontology, 59B*, 49-55.

Neuhaus, J. M., & Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics, 54,* 638-645.

Olsen, M. K., & Schafer, J. L. (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association, 96*, 730-745.

Pagach, O., Hedeker, D., & Mermelstein, R. (2014). A bivariate mixed-effects location-scale model with application to ecological momentary assessment. *Health*

*Services and Outcomes Research Methodology*. Advance online publication. Retrieved from http://link.springer.com/content/pdf/10.1007%2Fs10742-014-0126-9.pdf.

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika, 58,* 545-554.

Patterson, H. D., & Thompson, R. (1974). Maximum likelihood estimation of components of variance. *Proceedings of the 8th International Biometric Conference*, 197-207.

Patz, R., & Junker, B. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics, 24,* 342-366.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York: Harcourt Brace.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the third international workshop on distributed statistical computing* (pp. 20-22). Retrieved from http://www.r-project.org/conferences/DSC-2003/Proceedings/

Plummer, M. (2012). JAGS version 3.3.0 user manual [Computer software manual]. Lyon, France: Author. Retrieved from http://sourceforge.net/projects/mcmc-jags/files/

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnostics and output analysis. *R News, 6,* 7-11.

Pugach, O., Hedeker, D., & Mermelstein, R. (2014). A bivariate mixed-effects location-scale model with application to ecological momentary assessment. *Nicotine and Tobacco Research, 16*, S151-S158.

Pullenayegum, E. M., & Thabane, L. (2009). Teaching Bayesian statistics in a health research methodology program. *Journal of Statistics Education, 17*, 1-16.

R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Raftery, A. E., & Lewis, S. (1992). How many iterations in the Gibbs sampler? In J. M. Bernardo, J. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics* (Vol. 4, pp. 763-773). Oxford, UK: Oxford University Press.

Ram, N., Rabbitt, P., Stollery, B., & Nesselroade, J. R. (2005). Cognitive performance inconsistency: Intraindividual change and variability. *Psychology and Aging, 20*, 623-633.

Rast, P., Hofer, S. M., & Sparks, C. (2012). Modeling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using BUGS/JAGS. *Multivariate Behavioral Research, 47*, 177-200.

Rast, P., & Zimprich, D. (2011). Modeling within-person variance in reaction time data of older adults. *The Journal of Gerontopsychology and Geriatric Psychiatry, 24*, 169-176.

Raudenbush, S. W., & Bryk, A. S. (1987). Examining correlates of diversity. *Journal of Educational Statistics, 12*, 241-269.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Ed.). Thousand Oaks, CA: Sage.

Roberts, G. O., Gelman, A., & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability, 7,* 110-120.

Roberts, G. O., & Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science, 16,* 351-367.

Rupp, A. A., Dey, D. K., & Zumbo, B. D. (2004). To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to modeling. *Structural Equation Modeling, 11,* 424-451.

Saha, C., & Jones, M. P. (2009). Bias in the last observation carried forward method under informative dropout. *Journal of Statistical Planning and Inference, 139,* 246-255.

Shao, J., & Zhong, B. (2003). Last observation carry-forward and last observation analysis. *Statistics in Medicine, 22,* 2429-2441.

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2009). On the relation of mean reaction time and intraindividual reaction time variability. *Psychology and Aging, 24,* 841-857.

Schneider, S., Junghaenel, D. U., Keefe, F. J., Schwartz, J. E., Stone, A. A., & Broderick, J. E. (2012). Individual differences in the day-to-day variability of pain, fatigue, and well-being in patients with rheumatic disease: Associations of psychological variables. *Pain, 153,* 813-822.

Singer, J. D., & Willett, J. B. (2003*). Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

Sliwinski, M. J., Hoffman, L., & Hofer, S. (2010). Evaluating convergence of within-person change and between-person age differences in age-heterogeneous longitudinal studies. *Research in Human Development, 7,* 45-60. DOI:10.1080/15427600903578169.

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24,* 236-243.

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics, 19,* 279-281.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage Publications Inc.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2003) WinBUGS user manual version 1.4. MRC Biostatistics Unit, Institute for Public Health, Cambridge, England. Retrieved from: http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine, 24,* 236-243.

Tractenberg, R. E., & Pietrzak, R. H. (2011). Intra-individual variability in Alzheimer's disease and cognitive aging: Definitions, context, and effect sizes. *PLOS ONE, 6,* 1-9.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer-Verlag.

Walsh, B. (2004). Markov chain Monte Carlo and Gibbs sampling. [PDF document]. Retrieved from Lecture Notes Online Website: http://web.mit.edu/~wingated/www/introductions/mcmc-gibbs-intro.pdf.

Watts, A. S., Vidoni, E. D., Loskutove, N., Johnson, D. K., & Burns, J. M. (2013). Measuring physical activity in older adults with and without early stage Alzheimer's disease. *Clinical Gerontologist, 36,* 356-374.

White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica,* 48, 817-838.

Williams, M. N., Grajales, C. A. G., & Kurkiewicz, D. (2013). Assumptions of multiple regression: Correcting two misconceptions. *Practical Assessment, Research & Evaluation, 18,* 1-14.

Yang, M., & Maxwell, S. E. (2014). Treatment effects in randomized longitudinal trials with different types of nonignorable dropout. *Psychological Methods, 19,* 188-210.