Model Criticism for Growth Curve Models

via Posterior Predictive Model Checking

by

Derek M. Fay

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2015 by the
Graduate Supervisory Committee:

Roy Levy, Chair
Marilyn Thompson
Craig Enders

ARIZONA STATE UNIVERSITY

December 2015

ABSTRACT

Although models for describing longitudinal data have become increasingly sophisticated, the criticism of even foundational growth curve models remains challenging. The challenge arises from the need to disentangle data-model misfit at multiple and interrelated levels of analysis. Using posterior predictive model checking (PPMC)—a popular Bayesian framework for model criticism—the performance of several discrepancy functions was investigated in a Monte Carlo simulation study. The discrepancy functions of interest included two types of conditional concordance correlation (*CCC*) functions, two types of R^2 functions, two types of standardized generalized dimensionality discrepancy (*SGDDM*) functions, the likelihood ratio (*LR*), and the likelihood ratio difference test (*LRT*). Key outcomes included effect sizes of the design factors on the realized values of discrepancy functions, distributions of posterior predictive p-values (PPP-values), and the proportion of extreme PPP-values.

In terms of the realized values, the behavior of the *CCC* and R^2 functions were generally consistent with prior research. However, as diagnostics, these functions were extremely conservative even when some aspect of the data was unaccounted for. In contrast, the conditional *SGDDM* (*SGDDM*$_C$), *LR*, and *LRT* were generally sensitive to the underspecifications investigated in this work on all outcomes considered. Although the proportions of extreme PPP-values for these functions tended to increase in null situations for non-normal data, this behavior may have reflected the true misfit that resulted from the specification of normal prior distributions. Importantly, the *LR* and the *SGDDM*$_C$ to a greater extent exhibited some potential for untangling the sources of data-

model misfit. Owing to connections of growth curve models to the more fundamental frameworks of multilevel modeling, structural equation models with a mean structure, and Bayesian hierarchical models, the results of the current work may have broader implications that warrant further research.

DEDICATION

I dedicate this dissertation to my tremendous wife, Jaye, whose unwavering love, support, and encouragement have made this journey possible. You have shared this journey with me, which I know has come with its challenges on many levels; there is no one else who I would rather bear those challenges with. Thank you for being such a solid foundation through it all. This dissertation is also dedicated to my daughters: Ariadne and Petra. To Ariadne, we've got a lot of coloring, puzzles, and doing everything your wonderful imagination can dream up to catch up on; I can't wait to do all of it. To Petra, I have yet to meet you, but the sound of your beating heart and the opportunity to meet you in a few short months has been a remarkable source of inspiration and happiness. I'd also like to dedicate this dissertation to my parents, who have been a source of love, support, and encouragement since day one.

ACKNOWLEDGMENTS

Some say it takes a village to raise a child (it does); I believe it can also be said that it takes a village to construct a dissertation. While it at times feels like a project undertaken in isolation, nothing could be further from the truth. My success in writing this dissertation, and my time in graduate school more broadly, has been impacted by more people than I can possibly thank in this limited space. However, there are some individuals who I cannot go without mentioning.

Foremost, I'd like to extend an enormous thank you to my mentor, Dr. Roy Levy. Thank you for pushing me as a scientist and allowing me to explore my interests. There are so many other aspects of my graduate experience, ranging from coursework to the completion of various milestones that were incredible owing to such clear and patient mentoring. To my immediate committee members, Dr. Marilyn Thompson and Dr. Craig Enders, thank you for making the construction and completion of this dissertation, and other aspects of my graduate school career, a rewarding experience. Through initial experiences with growth curve modeling in applied research, coursework, and conversation, the two of you in particular have influenced aspects of this dissertation well before it was conceived of. Although not part of my immediate committee, I'd also like to thank Dr. Samuel Green for your support, mentorship, and capacity to elicit a (sometimes much needed) laugh at any given point over the years.

I'd like to thank the many friends I have made while in graduate school. Although there many, a special thanks to Katie Kunze, Nedim Yel, Dr. AAron Crawford, Dr. Lietta Scott, and Dr. Dubravka Svetina for the wonderful conversations and camaraderie. Going

back to it taking a village to construct a dissertation, I'd like to thank Cindy Williams, Jordan Williams, and the Foster family for being my family away from family. I couldn't have a better support system; I only hope I can return all of the favors you all have done for me (and Jaye and Ariadne) over the years.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

CHAPTER 1: Introduction

A common perspective among statisticians is that no statistical model should be regarded as 'correct'. Rather, the application of a statistical model represents an attempt to capture the key underlying processes that give rise to observed data through a smaller set of model parameters. With theoretical considerations as the foundation, the strength of a statistical model grows to the extent that the hypothesized process yields predictions that approximate observed data. The weaknesses of a statistical model become evident to the extent that predictions are systematically and increasingly disparate from observed data. As a result, summarizing the weaknesses and strengths of a statistical model in relation to observed data is a critical step in any statistical modeling endeavor.

Longitudinal data structures are often encountered in a variety of disciplines in the natural and social sciences. Over the course of 60 years (Baker, 1954), statistical models for summarizing the fundamental processes that underlie longitudinal data have become increasingly sophisticated (e.g., Meredith & Tisak, 1990; Raudenbush & Bryk, 2002). Modern statistical models for longitudinal data, which are broadly referred to as growth curve modeling (GCM; see Bollen & Curran, 2006; Singer & Willet, 2003), allow for the simultaneous estimation of a population trajectory and individual variation around that trajectory. Growth curve models are also remarkably parsimonious in that very few parameters are required to arrive at a model that fits with the intuition that people start in different places and change at different rates.

The popularity of GCM among applied researchers is marked by extensive use in a variety of applied research settings; the popularity of GCM among methodologists is marked by the rate at which complexity is added to the already flexible framework.

Unfortunately, the widespread use and rate by which GCM has evolved has outpaced the understanding and research for critiquing even foundational models. This state of affairs has left applied researchers with little guidance for properly evaluating the strengths and weakness of growth curve models relative to observed data.

Purpose of the Study

The criticism of growth curve models is a methodological challenge. The challenge is the result of the presence of variability at multiple and interrelated levels of analysis (Wu, West, & Taylor, 2009). At the first level of analysis, variability exists across measurement occasions within people (or some other entity, such as schools). At the second level of analysis, variability in growth exists between people. The variability may be characterized by differing rates of growth, differing functional forms of growth altogether, or perhaps some combination of both. In addition to these sources of variation, the question of whether the functional form of the average and person-specific trajectories resemble observed data remains open. These issues are more fully explored in Chapter 2.

In terms of assessing the strengths and weakness of a particular GCM to observed data, there is a critical need to identify or engineer fit functions that are suitable for disentangling the sources of data-model misfit at different levels of analysis. The purpose of this study was to explore the performance of a collection of discrepancy functions that have been purported or are hypothesized to have this capacity. Moreover, the performance of the selected discrepancy functions was assessed using posterior predictive model checking (PPMC; Gelman, Meng, & Stern, 1996), a popular Bayesian approach to model criticism. Drawing from the limited available research and theoretical

considerations, several factors hypothesized to impact the behavior of the discrepancy functions, and the success of model criticism, were manipulated in a Monte Carlo simulation study. With the details more fully explicated in Chapter 3, the underlying theme of all design decisions can be viewed as maintaining relevance to applied research that makes use of GCM.

Couching the performance of the selected discrepancy functions within the PPMC framework is an important addition to the methodological literature for GCM for at least two reasons. First, the analytical characteristics for most of the selected discrepancy functions are currently unknown, rendering it difficult to understand how the functions perform in various null and non-null conditions. As a resampling procedure, PPMC serves to construct the reference distribution for which to compare the value(s) of the discrepancy functions computed for the observed data. Although this statement holds for non-Bayesian resampling techniques, such as the parametric bootstrap, PPMC uniquely incorporates the uncertainty of model parameters into the criticism of the model (Gelman et al., 1996; Levy, 2011). Second, inasmuch as there is a general scarcity of methodological research pertaining to the criticism of growth curve models, no such research has systematically investigated the utility of a Bayesian approach for assessing the strengths and weaknesses of GCM. This need is compounded by the sharp increase in the number of Bayesian applications in recent years (e.g., Rupp, Zumbo, & Dey, 2004).

CHAPTER 2: Literature Review

Growth Curve Models

Traditional approaches to modeling longitudinal data, such as the analysis of variance (ANOVA; Maxwell & Delaney, 2003) model for repeated measures, have emphasized the estimation of the population trajectory. The population trajectory can be viewed as the regression line that summarizes the relationship between the means of a repeatedly measured outcome and the passage of time. In effect, this model assumes that individuals are the same at the beginning of the study and change in the same way with the passage of time; discrepancies from this assumption are subsumed into the model as random errors in prediction.

Although estimating an overall growth trajectory for the population is informative in its own right, there are some key disadvantages to the traditional method for modeling longitudinal data. First, the assumption that the population trajectory is sufficient to describe the growth trajectories for all members of the population often falls counter to theory. It is generally more natural to view the starting points and rate of change as varying between individuals. Second, the between-individual variation in growth may be systematically related to other variables. To the applied researcher, it is often of interest to pursue predictors that account for variations in growth or use growth as a predictor of some distal outcome (Bollen & Curran, 2006; Singer & Willet, 2003); investigating such questions is impossible when the growth trajectory is assumed constant over individuals.

As mentioned above, GCM represents the modern approach for flexibly summarizing the key processes that underlie longitudinal data. With the capacity to simultaneously estimate a population trajectory and summarize the amount of between-

person variation around that trajectory, GCM allows for models that more closely resemble theory, and perhaps more fundamentally, intuition about the processes that underlie real longitudinal data. Notably, the foundational GCM is a very flexible framework that can easily be extended to handle more complicated situations. For example, researchers can investigate the relationship between parallel growth processes (Cheong, MacKinnon, & Khoo, 2003; McArdle, 1989); model change for latent variables that are measured by multiple indicators (Duncan & Duncan, 1996; McArdle, 1988); or even explore whether there is group-level heterogeneity when group membership is unknown (Muthén & Shedden, 1999; Nagin, 1999). Although these extensions are interesting and highlight the flexibility of GCM, the matter of model criticism, which is the concern for this work, remains challenging for foundational growth curve models.

The signature characteristics of the GCM approach include the concurrent estimation of an average (i.e., population) growth trajectory and between-person variation in growth around that average trajectory. Using fictitious data, Figure 1 depicts a GCM that allows for unique linear growth trajectories across five measurement occasions for 25 individuals. The figure is structured such that scores on the outcome measure, which is shown on the vertical axis, are regressed on the passage of time, which is shown on the horizontal axis. The heavy solid line represents the average trajectory, and the dashed lines represent the trajectories that are specific to individuals.

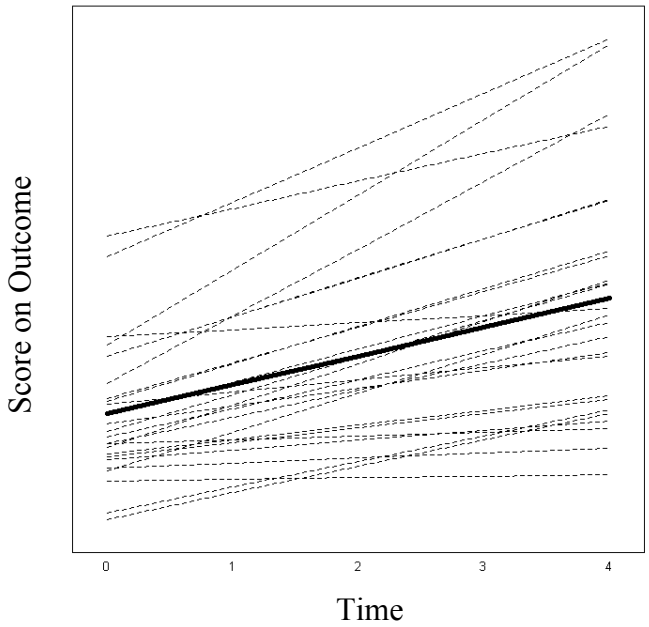

Figure 1. Linear growth curve model with between-individual variation in the intercept and slope.

With Figure 1 as a visual heuristic, it is made apparent that the aspects of growth can be viewed as being distributed around the average trajectory. Although relatively simple, the example model captures the essential features of GCM that are germane to more complicated extensions. At the intercept (i.e., Time = 0), there is clear separation in the outcome between individuals. In terms of the rate of change, some individual trajectories are essentially parallel to the average trajectory while others are flatter or exhibit a greater incline than the average trajectory. In addition, the individual trajectories with intercepts below that of the average trajectory tend to be flatter than the individual trajectories with intercepts above the average trajectory. This suggests that the aspects of change at the level of the individual are correlated such that higher intercepts tend to exhibit faster linear change.

Parameterization of GCM. Growth curve models have been framed as special cases of multilevel modeling (MLM; e.g., Raudenbush & Bryk, 2002; Singer & Willett,

2003) and structural equation modeling (SEM; e.g., Kline, 2005). It has been widely acknowledged that once stripped of surface-level notation, the MLM and SEM frameworks are mathematically identical representations of GCM (MacCallum & Kim, 2000; Preacher, Wichman, MacCallum, & Briggs, 2008; Rovine & Molenaar, 2000; Willett & Sayer, 1994). These frameworks for fitting growth curve models are described briefly here; additional details about these models and the connections between them can be reviewed in Appendix A. With respect to GCM, the fundamental difference between the MLM and SEM is a matter of philosophical orientation about the aspects of change. From the perspective of MLM, the scores on the repeatedly observed outcome are viewed as nested within individuals. In accounting for this hierarchical structure, the person level regression coefficients that model the variation in the observed data are in turn assumed to be random variables that vary between people. From the SEM perspective, the unobserved aspects of change are viewed as unmeasured variables that give rise to the observed data (Meredith & Tisak,1990). The basis of the SEM model for growth is a confirmatory factor analytic (CFA; Jöreskog, 1969) model with a mean structure (e.g., Bentler & Yuan, 2000; Jöreskog & Sörbom, 1985) among the latent variables with factor loadings (typically) fixed to integer values to specify the functional form of growth (Bollen & Curran, 2006).

The difference of philosophical orientation has bearing on approaches to estimation that in turn have practical consequences that may motivate the selection of one framework over the other. The core advantage of the MLM framework over SEM is the relative of ease of specifying models with more than two levels. Building off of the description above, a model with three levels may be needed if repeated measures are

nested within students who are in turn nested within schools. Specifying this model is likely to prove quite difficult if not impossible for most currently available software dedicated to the estimation of SEM models. The fundamental advantage of the SEM framework over MLM is a matter of flexibility. Excepting to the incorporation of additional levels of analysis, extensions to the standard GCM are easily accommodated and easily specified in the SEM framework. For example, it is relatively simple to specify models with simultaneous growth processes for different sets of repeatedly measured outcomes; estimate a growth process for latent outcomes that are not free of measurement error, as is assumed for observed outcomes; flexibly model the growth parameters as outcomes, predictors, or correlates with other observed or latent variables; and/or explore for the presence of mixtures of growth trajectories (e.g., Cheong et al., 2003; Duncan & Duncan, 1996; McArdle, 1988, 1999; MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997; MacCallum & Kim, 2000; Muthén & Shedden, 1999; Nagin, 1999). Specifying these models is impossible for most dedicated software packages for estimating multilevel models; key exceptions include software packages such as WinBUGS (Lunn, Spiegelhalter, Thomas, & Best, 2009), JAGS (Plummer, 2013), and STAN (Stan Development Team, 2015).

For the purposes of this work, the parameterization of GCM that follows is framed as a Bayesian hierarchical model (Gill, 2007; Raudenbush & Bryk, 2002). The motivation for this choice is two fold. First, the only distinction with respect to the nature of variables from the Bayesian perspective is that some variables are observed, and can be treated as random until known, while others variables are unobserved, and treated as random and unknown. All unobserved variables are assigned a prior distribution. The

implication is that all of the advantages of MLM and SEM are combined into a single framework. More specifically, the Bayesian hierarchical model, and currently available software for Bayesian estimation, can accommodate multiple levels of analysis (with no theoretical limit) without sacrificing the flexibility afforded by SEM.

The central goal of a Bayesian analysis is to blend observed data and any prior beliefs about model parameters to arrive at a posterior distribution. In terms of mathematical machinery, Bayes' theorem is a mechanism for inverting probabilities (e.g., Gill, 2007). Consider the simple case in which there is inferential interest in some parameter θ given some observed data **Y**:

$$P(\theta|\mathbf{Y}) = \frac{P(\theta) \times P(\mathbf{Y}|\theta)}{P(\mathbf{Y})}. \quad (1)$$

The components of Bayes' theorem include the *posterior distribution* of the model parameter, $P(\theta|\mathbf{Y})$; the *prior distribution* of the model parameter, $P(\theta)$; the probability of the data given the model parameter, $P(\mathbf{Y}|\theta)$; and the marginal distribution of the data, $P(\mathbf{Y})$. Notably, $P(\mathbf{Y}|\theta)$ is equally thought of as the likelihood of the model parameter given the data, $L(\theta|\mathbf{Y})$. Bayes' theorem makes it possible to direct inferences from the **Y** to θ, or alternatively from the θ to **Y**. From the classical perspective (i.e., frequentist), from which estimation routines for SEM and MLM are based in, the analytical goal is centered on finding a point estimate of the parameter that maximizes the probability of the observed data. In standard applications of Bayesian analyses, the analytical goal is to obtain the posterior distribution to support inferences about the unknown parameters given the observed data, as is presented above.

The prior distribution for θ is specified as exhibiting some distributional form with parameters set by the analyst. In the case of Bayesian hierarchical models, which serve as the foundation for GCM from a Bayesian perspective, the parameters that govern the form of the prior distribution of θ are in turn conditional on some other unknown parameter, ψ. As an unknown parameter, ψ must also be assigned a prior distribution. This is formally characterized in Bayes' theorem by specifying a prior for θ that is conditional on ψ, $P(\theta|\psi)$, and adding the prior distribution of $P(\psi)$:

$$P(\theta, \psi|\mathbf{Y}) = \frac{P(\theta|\psi) \times P(\psi) \times P(\mathbf{Y}|\theta)}{P(\mathbf{Y})}. \quad (2)$$

As noted by Gill (2007), the layering of prior distributions is *theoretically* unlimited. However, the number of layers is *practically* limited by the interpretational challenges that arise with each additional level. In addition to interpretational challenges, layers that are closer to the top of the hierarchy are more distant from, and are therefore less informed by **Y** (Goel, 1983; Titterington, Smith, & Makov, 1985).

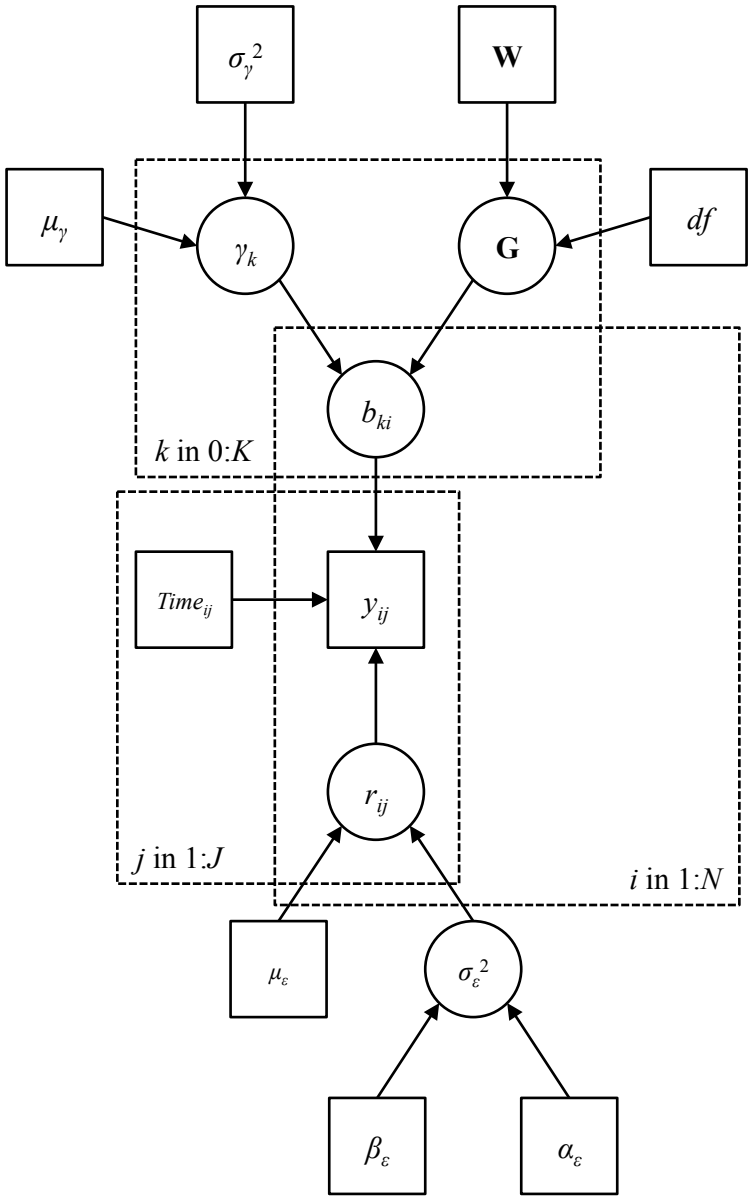

Figure 2. Directed acyclic graph (DAG) of a growth curve model.

Figure 2 depicts the Bayesian hierarchical model parameterization for GCM in the form of a directed acyclic graph (DAG). DAGs serve two broad purposes. First, DAGs serve as visual summary of the interrelationships among observed (the data and fixed parameters) and unobserved (unknown model parameters) variables. Second, DAGs structure the sources of dependence and conditional independence in the joint

distribution. The key features of DAGs include nodes, directed edges, and plates. Square nodes represent observed variables and any parameters that are directly specified by the analyst. Nodes shown as a circle represent unknown parameters for which a posterior distribution is to be obtained. The directed edges are shown as arrows that begin at one node, which is labeled a parent, and point to a different node, which is labeled a child of the node from which the edge originates. The graph is acyclic in that the flow of dependence can never trace back from children to parents, hence the unidirectional arrows. Plates indicate that the enclosed node(s) is (are) applicable to (i.e., subscripted by) some features of the data, such as the number of individuals or measurement occasions.

The underlying parametric form of the Bayesian hierarchical model for growth is a regression model in which scores on the outcome (y_{ij}) are regressed on the passage of time ($Time_{ij}$):

$$y_{ij} = b_{0i} + b_{1i}\left(Time_{ij}\right) + r_{ij}. \quad (3)$$

For the sake of simplicity, a GCM with a linear functional form is shown here. The model includes an intercept coefficient (denoted b_{0i}), a slope coefficient (denoted b_{1i}), and a residual score (denoted r_{ij}). The residual score is the difference between the observed score and the score implied by the model (denoted $\hat{y}_{ij}$, such that $\hat{y}_{ij} = b_{0i} + b_{1i}\left(Time_{ij}\right)$) for a given individual *i* at a given measurement occasion *j* (where *j* = 1,2,…, *J*). In foundational growth curve models, the vector of residuals for a given individual (denoted $\mathbf{r}_i$) are assumed to arise from a multivariate distribution as follows:

$$\mathbf{r}_i \sim MVN\left(\boldsymbol{\mu}_\varepsilon = \begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix}, \mathbf{R} = \begin{bmatrix}\sigma_\varepsilon^2 & & & & \\ 0 & \sigma_\varepsilon^2 & & & \\ 0 & 0 & \sigma_\varepsilon^2 & & \\ 0 & 0 & 0 & \sigma_\varepsilon^2 & \\ 0 & 0 & 0 & 0 & \sigma_\varepsilon^2\end{bmatrix}\right). \quad (4)$$

This structure for the distribution of residual scores reflects an assumption that the residual scores (a) have mean of zero across measurement occasions, (b) are homoscedastic across measurement occasions, and (c) are uncorrelated for any two measurement occasions. In standard applications, the vector of means for residual scores (denoted $\boldsymbol{\mu}_\varepsilon$) is fixed to zero and the variances, or in this case, a common variance for all measurement occasions is estimated. In the Bayesian framework, an inverse Wishart distribution is the typical choice of prior for a multivariate specification of (co)variances. As shown in the DAG, the residual scores are shown to arise from a univariate distribution with some fixed mean (typically $\mu_\varepsilon = 0$) and some residual variance, σ_ε^2. For the residual variance, which is an estimated parameter, the typical prior specification is an inverse gamma with shape parameters that are specified by the analyst, $\sigma_\varepsilon^2 \sim IG(\alpha_\varepsilon, \beta_\varepsilon)$.

The potential for unique growth trajectories between people is reflected by subscripting the intercept and slope coefficients by *i* (where $i = 1, 2, \ldots, N$). As unknown parameters, the person-specific intercept (b_{0i}) and slope (b_{1i}) in Equation 3 are assigned a multivariate normal prior distribution with some vector of unknown means ($\boldsymbol{\gamma}$) and a covariance matrix with unknown elements (**G**):

$$\begin{bmatrix}b_{0i}\\b_{1i}\end{bmatrix} \sim MVN\left(\boldsymbol{\gamma} = \begin{bmatrix}\gamma_{00}\\\gamma_{10}\end{bmatrix}, \mathbf{G} = \begin{bmatrix}\tau_{00} & \\ \tau_{10} & \tau_{11}\end{bmatrix}\right). \quad (5)$$

As described above, the signature characteristic of GCM is the capacity to simultaneously estimate person-specific growth trajectories and the trajectory for the population. Since inferential interest also lies in the population trajectory, which is defined by $\boldsymbol{\gamma}$, the hierarchical feature of the Bayesian GCM obtains by assigning prior distributions to the elements of $\boldsymbol{\gamma}$. In standard applications, each element in $\boldsymbol{\gamma}$ is assumed to be univariately normally distributed with fixed values of the mean (suppose 0) and variance (suppose 1):

$$\gamma_{00} \sim N(\mu_{\gamma_{00}} = 0, \sigma^2_{\gamma_{00}} = 1) \text{ and } \gamma_{10} \sim N(\mu_{\gamma_{10}} = 0, \sigma^2_{\gamma_{10}} = 1). \quad (6)$$

Notably, since the elements are unknown, the covariance matrix of the person-specific growth parameters (denoted **G**) is also assigned a prior distribution. Since the person-specific growth parameters are (typically) assumed to exhibit multivariate normality, the prior distribution for **G** is typically an inverse Wishart distribution, **G** ~ *IW*(**W**, *df*), where **W** is some positive definite matrix and *df* represent the degrees of freedom (as mentioned above, an inverse Wishart could also be specified as the prior for the **R** matrix in the case of a multivariate specification of residual scores).

A Note on Language for GCM. Owing to the development of GCM from different frameworks, there is significant variability in the language for describing components of GCM models. In the parlance of MLM, the submodel for the data (i.e., Equation 3) is typically labeled the level-1 model; the submodel for the growth coefficients (i.e., Equation 5) is typically labeled the level-2 model. Unfortunately, indexing the levels of the hierarchy in MLM does not directly sync with the levels of the prior distributions in the Bayesian hierarchical model. In the SEM framework, the growth parameters are collectively referred to as the latent growth parameters; aside from the growth parameters, the remaining components are typically referenced using

conventional language associated with factor analysis (e.g., factor loadings, factor means). Unfortunately, the language of factor analysis carries little meaning with respect to the description of GCM provided above for Bayesian hierarchical models.

Wu, West, and Taylor (2009) employ a general nomenclature for referencing the components of growth curve models that is used throughout the remainder of this work. Using their terminology, the GCM is comprised of a mean structure and the overall covariance structure. The mean structure is further separated into the conditional mean structure and the marginal mean structure. The former represents the regression of observed scores on the passage of time for each person; Equation 3 defines this component of GCM. The latter represents the regression of sample means on the passage of time, *averaged* over people within measurement occasions; the equation for this model is akin to that shown in Equation 3 with the key exception that the average regression coefficients (i.e., γ_{00}, γ_{10}) are substituted in for the person-level regression coefficients (i.e., b_{0i}, b_{1i}). The overall covariance structure is further separated into the within-subject matrix, which they denote as the **R** matrix (as was the case above), and the between-subject matrix, which they denote as the **G** matrix (as was the case above). As described above, the **R** matrix includes the variances and covariances of the residual scores among the measurement occasions. The **G** matrix includes the variances and covariances among the person-specific estimates of the growth parameters.

Posterior Predictive Model Checking

PPMC is an extremely flexible framework for evaluating the statistical strengths and weakness of a given model. As evidenced and perhaps owing to the default use of PPMC when Bayesian estimation is employed in software packages such as Mplus 6

(Muthén & Muthén, 2010), PPMC is by far the most well-known Bayesian approach for model criticism. PPMC is conceptually viewed as a comprehensive framework that directly compares the observed data at hand to data that are consistent with a model for which the posterior distribution has been obtained. The section begins with a description of the procedural aspects of PPMC and concludes with discussion of some key advantages and disadvantages of PPMC.

Obtaining the Posterior Distribution. The necessary ingredient for conducting a PPMC analysis is the posterior distribution. Let $P(\boldsymbol{\Omega}|\mathbf{Y})$ and $P(\boldsymbol{\Omega})$ respectively represent the joint posterior and prior distributions of multiple model parameters; consistent with Equation 1, $\mathbf{Y}$ represents the observed data. Bayes' theorem with multiple model parameters is given by:

$$P(\boldsymbol{\Omega}|\mathbf{Y}) = \frac{P(\boldsymbol{\Omega})P(\mathbf{Y}|\boldsymbol{\Omega})}{P(\mathbf{Y})} = \frac{P(\boldsymbol{\Omega})P(\mathbf{Y}|\boldsymbol{\Omega})}{\int_{\boldsymbol{\Omega}} P(\boldsymbol{\Omega})P(\mathbf{Y}|\boldsymbol{\Omega})d\boldsymbol{\Omega}}. \quad (7)$$

The posterior distribution can be obtained analytically for simple situations. Unfortunately, the models that are often encountered in applied research settings, such as growth curve models, typically involve complex multivariate systems of variables. This implication is that the high dimensional integral in the denominator of Equation 7 quickly becomes computationally intractable (e.g., Gill, 2007; Rupp et al., 2004). A key realization was that all of the information necessary for obtaining the posterior distribution is contained in the prior distribution and the likelihood, and the role of the denominator is to ensure that the posterior is proper (e.g., Gill, 2007; Rupp et al., 2004). That is, the posterior distribution exhibits a proportional relationship to the product of the prior distribution and the likelihood:

$$P(\mathbf{\Omega}|\mathbf{Y}) \propto P(\mathbf{\Omega})P(\mathbf{Y}|\mathbf{\Omega}). \quad (8)$$

This relationship is leveraged by computational routines to sample from the distribution of inferential interest many times over. In the limit the sample converges to the stationary distribution, which is in turn taken to be an approximation to the posterior distribution (Levy, 2006, 2009).

Markov chain Monte Carlo (MCMC; Brooks, 1998; Casella & George, 1992; Chib & Greenberg, 1995; Gelfand & Smith, 1990; Geman & Geman, 1984; Gilks, Richardson, & Spiegelhalter, 1996; Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller & Teller, 1953) methods are the primary class of algorithms that are used to sample from what eventually becomes an empirical approximation to the true posterior distribution (Levy, 2006, 2009). As a simulation environment, MCMC can be viewed as a 'divide-and-conquer' strategy in that the end goal is to sample from the full space of the posterior distribution (in a finite amount of time). Once the analyst has deemed that the algorithm has converged to the stationary distribution and that the chains have been run sufficiently long, the MCMC chain is terminated, and after discarding all draws prior to convergence, the resulting collection of draws are labeled the posterior distribution. Since the features of the common MCMC samplers, like the one used in this work, have been rigorously defined elsewhere (e.g., Cowles, 2002), the details of MCMC algorithms are not described here. Moreover, the features of MCMC are beyond the scope of the current work. It is sufficient for this work to recognize the link between the role of MCMC for constructing the posterior distribution and PPMC.

Posterior Predictive Distribution. The link between MCMC and a PPMC analysis is the construction of the *posterior predictive distribution*. With Bayes' theorem as the machinery, the posterior predictive distribution is given by:

$$P(\mathbf{Y}^{\text{rep}}|\mathbf{Y}) = \int_{\mathbf{\Omega}} P(\mathbf{Y}^{\text{rep}}|\mathbf{\Omega})P(\mathbf{\Omega}|\mathbf{Y})d\mathbf{\Omega}. \quad (9)$$

Using the individual draws that comprise the posterior distribution, *replicated datasets* with the same dimensions as the observed data are generated to be consistent with the model at hand. For example, if the posterior distribution was empirically approximated based on R = 1,000 draws from an MCMC sampler, the posterior predictive distribution will in turn be empirically approximated by 1,000 replicated datasets generated by the model (i.e., one for each draw). The full collection of replicated datasets (denoted $\mathbf{Y}^{\text{rep}}$) represents an empirical sampling distribution under the model at hand (Gelman et al., 1996; Levy, 2011; Sinharay, Johnson, & Stern, 2006).

Conducting PPMC. The PPMC framework builds naturally off of the use of MCMC to build the posterior distribution. Figure 3 schematically depicts the general process of PPMC, with two possible pathways that are determined on the basis of the characteristics of the function to be computed. The literature on PPMC distinguishes between functions that depend only on the data, which are labeled *test statistics* (denoted $T(\mathbf{Y})$), and functions that depend on the data and the model parameters, which are labeled *discrepancy measures* (denoted $D(\mathbf{Y}, \mathbf{\Omega})$) (e.g., Gelman et al., 1996; Levy, 2011; Sinharay et al., 2006). Test statistics typically include functions used to summarize distributions such as measures of central tendency (e.g., mode, median, mean), variability (e.g., range, variance, standard deviation), shape (e.g., skew, kurtosis), or association

(e.g., covariance, correlation). In the PPMC framework, a test statistic is computed for each posterior predictive dataset and the observed data. The collection of values of the test statistic constitutes a reference distribution for which to locate the value of the test statistic associated with the observed data.

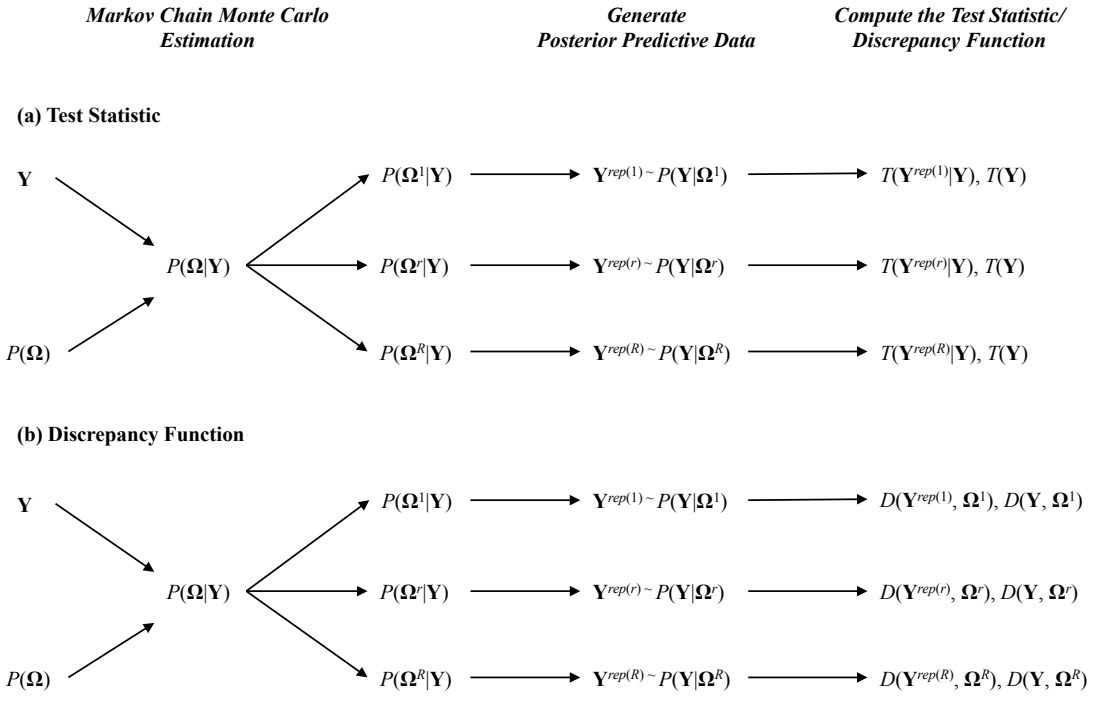

Figure 3. Schematic of PPMC for (a) test statistics and (b) discrepancy measures.

There is significantly greater variability in the form, logic, and intended use of discrepancy functions. However, the general logic of discrepancy functions is to compare the data at hand—either posterior predictive datasets or the observed data—to the model implications that are defined by the functions of model parameters. As was the case for test statistics, a discrepancy function is computed for each posterior predictive dataset, and in effect, forms an empirical reference distribution for the discrepancy function given the model. Unlike test statistics, the values of discrepancy functions rely on the model parameters. Since there exist a posterior distribution of model parameters, a distribution

of the discrepancy function is also constructed for the observed data. When critiquing a model or a particular feature of a model with a discrepancy function, the analysis involves comparing the posterior predictive values of the discrepancy function and the observed values of the discrepancy function that were computed using the same values of the model parameters.

For a given discrepancy function, the final step of a PPMC analysis is to compare the values computed using the observed data, which are labeled the *realized value(s)*, to those computed using posterior predicted datasets, which are referred to as the *posterior predicted values* of the function(s). One approach that summarizes the comparison between realized and posterior predictive values is to compute the *posterior predicted p-value* (PPP value; Gelman et al., 1996). Formally, PPP values for any function (denoted *f*), whether it is a test statistic or discrepancy measure, is described as:

$$PPP = [f(\mathbf{Y}^{\text{rep}}) \geq f(\mathbf{Y})]/R. \quad (10)$$

The PPP value is conceptually akin to (and yet not the same as) an upper-tailed *p*-value rendered from a formal statistical test in the hypothesis-testing framework. PPP values that approach zero indicate that the model underpredicts the feature in question; PPP values that approach 0.5 suggests that the model is consistent with the observed data for the feature in question; and PPP values that approach one indicate that the model over-predicts the feature in question.

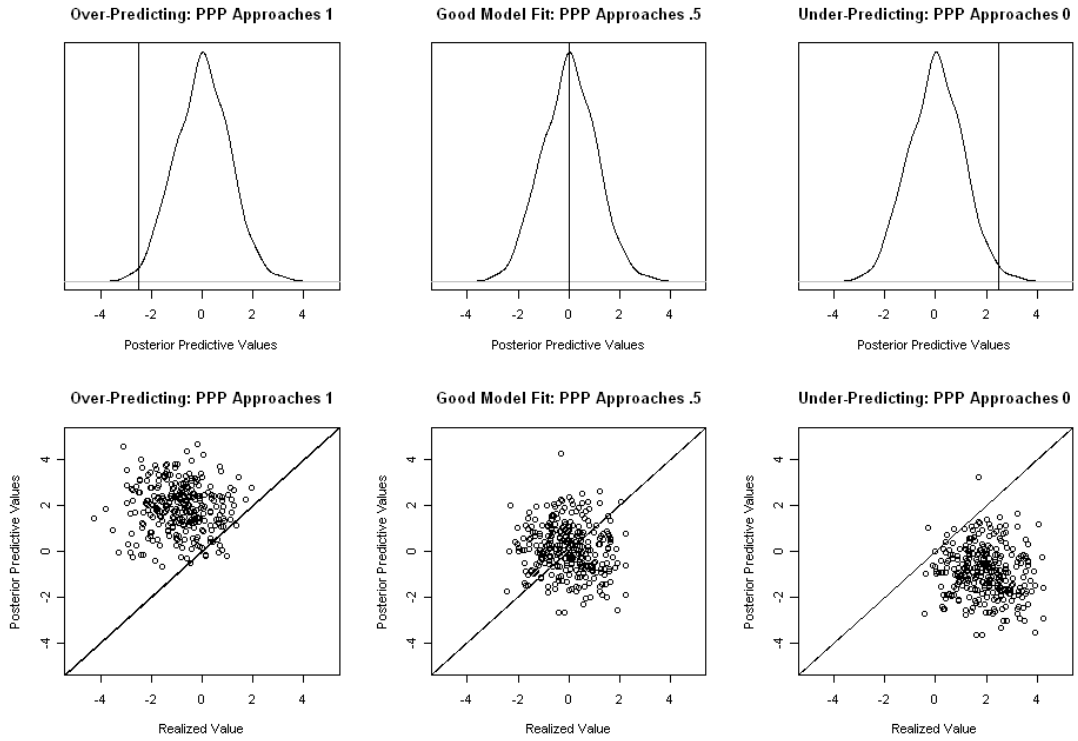

Figure 4. Example graphical representations and approximate PPP values for test statistics (the top row) and discrepancy measures (the bottom row) rendered from PPMC analysis.

Another approach, which is often viewed as a key strength of PPMC, is the ability to construct attractive graphical displays of the results (e.g., Sinharay, 2005). Figure 4 shows typical graphical displays for test statistics (top row) and discrepancy measures (bottom row). As described above, test statistics are functions that only depend on the data at hand. The realized value of the test statistics is shown as a solid vertical line; this line serves to locate the realized value of the test statistic within the reference distribution. The associated PPP value is the proportion of points to the right of the vertical line. In contrast to test statistics, the values of discrepancy measures also depend on the values of model parameters. This feature of discrepancy measures introduces variability into the values of the function computed for both the observed and posterior

predicted data; this yields a collection of values of the discrepancy measure for both the observed and posterior predicted data. Accordingly, graphical displays for discrepancy measures typically show the association between the realized values (i.e., those based on the observed data) and posterior predictive values. The diagonal line represents equality between the realized values and their posterior predictive counterpart; the PPP value is the proportion of points that fall above and to the left of the line. For both types of functions, the left column shows an over-prediction by the model (PPP values approach one), the middle column shows adequate model-fit (PPP values approach .50), and the right column shows an under-prediction by the model (PPP values approach 0).

Advantages of PPMC. The PPMC framework offers two key advantages over other approaches to model criticism. The first advantage is flexibility. Within the PPMC framework, one can employ any function that is believed to be sensitive to targeted sources of data-model misfit (e.g., global fit, person fit, bivariate fit). This advantage is the result of empirically constructing the reference distribution rather than appealing to asymptotic arguments (Levy, 2006, 2011). The payoff is that one's choice of functions is restricted to those that are believed to be informative about sources of misfit rather than to those with known asymptotic behavior (Jannsen, Tuerlinckx, Meulders, & De Boeck, 2000). Moreover, among functions with known asymptotic behavior, situating these functions within the PPMC framework may prove useful when the regularity conditions for the asymptotic behavior to hold are violated.

Notably, the advantages just described also apply to parametric bootstrapping, a frequentist approach to model checking that involves the use of model-based resampling techniques to construct an empirical reference distribution (Kline, 2005). The underlying

process of the parametric bootstrap is similar to PPMC in that both are resampling techniques that involve comparing the values of a measure (whether it be a test statistic or discrepancy function) to a reference distribution that is constructed from data that are consistent with the model. A key difference between parametric boostrapping and PPMC comes to the choice of values that are used for generating data that are consistent with the model. The former relies on the use of point estimates of model parameters—which are usually obtained with the use of least squares or maximum likelihood approaches to estimation—to generate predictive data. In contrast, the latter relies on the complete posterior distribution to generate predictive data. The implication is that parametric bootstrapping underestimates the degree of uncertainty in the model parameters (Levy, 2006; Meng, 1994). As recognized by Levy (2006, 2011), in using the full posterior distribution, PPMC seamlessly incorporates the uncertainty in the model parameters into the criticism of the model.

Disadvantages of PPMC. Although PPMC has an intuitive appeal and significant flexibility, the potential disadvantage of PPMC is the matter of implementation. To take full advantage of the flexibility of PPMC, it is usually necessary for the analyst to (a) obtain the posterior distribution using software or an independently written MCMC algorithm, (b) save the draws that comprise the posterior distribution, and then (c) import the draws into a software platform that allows the analyst to write a program to implement PPMC, which includes the manual coding of the test statistics and/or discrepancy functions. Although commercial software packages such as Mplus (Muthén & Muthén, 1998-2010) have recently implemented Bayesian estimation (i.e., MCMC)

and PPMC, the user is still required to use an external program to employ any function that is not produced by the software.

Summary of PPMC. With the use of Bayesian analysis methods, the analyst is afforded remarkable flexibility to support the application of existing and innovative models to observed data (e.g., Crawford, 2014; Levy, 2006). With the use of PPMC, this flexibility is carried over into the process of model checking. The advantages of flexibility and the modeling of uncertainty overcome the limitations of alternative approaches to model criticism, namely null hypothesis significance testing and the parametric bootstrap. Notably, the matter of implementation difficulty is becoming less salient with improvements in software and to the extent that interest and knowledge in Bayesian analysis methods continues to grow. The advantages of PPMC described above are particularly salient for GCM owing to rate at which methods for model-based analysis of longitudinal data are evolving (Bollen & Curran, 2006; Preacher et al., 2008).

Review of Data-Model Fit for Growth Curve Models

This section reviews the available methodological literature pertaining to model checking in the context of GCM. To mimic good practice with respect to model checking—which is most successful when discrepancy functions are selected or engineered to target particular sources of misfit—key sources of data-model misfit are described prior to reviewing the literature on the performance of discrepancy functions.

Sources of Data-Model Misfit. As briefly mentioned in the first chapter, the challenges of critiquing growth curve models is the result of disentangling sources of data-model misfit that are inherently intertwined. The key sources of misfit include the

marginal mean structure, conditional mean structure, the **G** matrix, and the **R** matrix (Wu et al., 2009). These sources of data-model misfit are described in turn below.

Marginal Mean Structure. The marginal mean structure is defined as the vector of model-implied means across people. The vector includes a model-implied mean for each measurement occasion, $\hat{\bar{\mathbf{y}}} = \left(\hat{\bar{y}}_1, \hat{\bar{y}}_j, \ldots, \hat{\bar{y}}_J\right)$. For a GCM with a linear functional form, the model-implied mean at a given measurement occasion is determined by the average coefficients, $\boldsymbol{\gamma} = (\gamma_{00}, \gamma_{10})$, and the passage of time for a given individual, $Time_{ij}$:

$$\hat{\bar{\mathrm{y}}}_{ij} = \gamma_{00} + \gamma_{10}\left(Time_{ij}\right). \quad (11)$$

The subscript *i* in $Time_{ij}$ can be dropped if the measurement schedule is identical across people. Discrepancies in the marginal mean structure become larger to the extent that the predicted means depart from the observed sample means.

Conditional Mean Structure. The conditional mean structure is the vector of model-implied scores for each person. The vector of model-implied scores includes the predicted scores (i.e., conditional means) for each measurement occasion, $\hat{\mathbf{y}}_i = \left(\hat{\mathrm{y}}_{i1}, \hat{\mathrm{y}}_{ij}, \ldots, \hat{\mathrm{y}}_J\right)$. For a GCM with a linear functional form, the model-implied score at a given measurement occasion for a given person is determined by the person-specific coefficients, $\boldsymbol{b}_i = (b_{0i}, b_{1i})$, and the passage of time for that person, $Time_{ij}$:

$$\hat{\mathrm{y}}_{ij} = b_{0i} + b_{1i}\left(Time_{ij}\right). \quad (12)$$

The subscripting of time by *i* can be dropped if the measurement schedule is identical across people. Discrepancies in the conditional mean structure for a given person becomes larger to the extent that the predicted scores depart from the observed scores for that person.

***G Matrix* (*Between-Person Covariance Matrix*).** The elements of the **G** matrix include the variances and covariances among the person-specific growth parameters. Stated generally, the source of discrepancies is the typically the result of assuming an element is zero, whether the element is a variance or covariance, when it shouldn't be.

***R Matrix* (*Within-Person Covariance Matrix*).** The elements of the **R** matrix include the variances and covariances of the residual scores (see Equation 4). In standard applications of GCM, the residual variances (the diagonal elements of **R**) are assumed to be homogeneous across measurement occasions and the covariances (the off-diagonal elements of **R**) are assumed to be zero. If the residual variances are assumed homogeneous, data-model misfit may occur if the variance of residuals scores differs across measurement occasions. In terms of the covariances, failing to include a predictor—whether the predictor pertains to the passage of time, a covariate of the scores at each measurement occasion, or a characteristic of the person—has the potential to manifest as non-zero covariance values. This latter situation represents a failure to meet the assumption that the scores are *conditionally independent* given the model (described in greater detail below).

***Dependencies among the Covariance Matrices*.** Given estimates of between-person (denoted $\widehat{\mathbf{G}}$) covariance matrix, within-person (denoted $\widehat{\mathbf{R}}$) covariance matrix, and the vector of scores that represent the passage of time for one person (denoted $\mathbf{T}_i$), the model-implied overall covariance matrix $\left(\hat{\Sigma}_i\right)$ is given by:

$$\left(\hat{\Sigma}_i\right) = \mathbf{T}_i\widehat{\mathbf{G}}\mathbf{T}_i' + \widehat{\mathbf{R}} \tag{13}$$

such that:

$$\mathbf{T}_i = \begin{bmatrix} 1 & Time_{i1} \\ 1 & Time_{ij} \\ \vdots & \vdots \\ 1 & Time_{iJ} \end{bmatrix}. \quad (14)$$

When measurement schedules are identical across people, the subscript *i* can be dropped. Given that $\mathbf{T}_i$ typically remains unchanged, the interrelationship between **G** and **R** are of central interest here. The key to understanding the relationship is that **G** and **R** are part of the same system; imposing or removing restrictions on one has implications for the amount of variance in the other. Wu et al. (2009) consider a case in which the functional form is underspecified (e.g., fitting a linear GCM to data that are better described by a quadratic GCM). In this scenario, removing restrictions in the **R** matrix would serve to improve the fit of the model at the risk of underestimating variability in the **G** matrix and overestimating variability in the **R** matrix. Similarly, removing restrictions in the **G** matrix would serve to improve the fit of the model, but aside from introducing greater complexity to the growth process, this may be considered reasonable since it is generally better to reduce the amount of variability in the **R** matrix. This of course begs the question as to whether the improvement in fit is worth the added complexity. Verbeke and Molenbergh (2000) argued that more information is gained by removing restrictions on the **G** matrix before removing restrictions on the **R** matrix. Under these circumstances, if the **G** matrix is optimized (i.e., there are no restrictions on any element) for the functional form at hand and there still exists substantial variability in the **R** matrix, particularly among the covariances, then evidence of an underspecified functional form is obtained (Wu et al., 2009).

Data-Model Fit for the Marginal Mean and Covariance Structure and the Impact on the Conditional Mean Structure. The agreement between observed scores and the conditional mean structure depends on the fit of the marginal mean structure and the covariance structure. That is, achieving solid fit for either one is not sufficient when taken alone in terms of ensuring fit of the conditional mean structure. Although unlikely, it is possible for the marginal mean structure to agree with the sample means while the conditional mean structure exhibits greater complexity than the marginal mean structure. For example, this can occur if the trajectory for each person exhibits a unique quadratic form that yields a linear trajectory when marginalized over (Wu et al., 2009; Wu & West, 2013). A GCM with a linear functional form would yield adequate fit for marginal mean structure but fail to account for the variability of the quadratic effect between people.

Since poor fit in the conditional mean structure can be remedied by removing restrictions in the **R** matrix (holding the functional form as constant across people), close agreement between the observed and model implied covariance matrix is also not sufficient by itself. For the reasons given above in the description of the relationships among the covariance matrices, the need to restrict elements in the **R** matrix, particularly the elements off the diagonal, may be taken as evidence that the functional form is underspecified.

Other Factors That May Impact the Analysis of Data-Model Misfit. Wu et al. (2009) identify two additional factors that may have bearing on the success of the criticism of growth curve models. The first factor is the structure of time, which can broadly be characterized as balanced or unbalanced. The structure of time is balanced if all people have identical measurement schedules. The structure of time is *un*balanced if

measurement schedules are (potentially) unique to each person. As discussed in greater detail below, the structure of time has less to do with the *performance* of discrepancy functions and more to do with *whether* a single value of the function can be computed.

The second factor identified by Wu et al. (2009) is the matter of distributional assumptions. A popular assumption of GCM is that the observed scores for the collection of repeated measures, growth parameters, and the residual scores exhibit multivariate normality (though not necessarily jointly). Even in the case of categorical outcomes for the observed repeated measures, it is typically assumed that observed scores are the result of discretizing an underlying latent response distribution (see Wirth & Edwards, 2007 for a general description of latent response distributions in the context of factor analysis with categorical outcomes), which is typically assumed to be normally distributed. A substantial number of the common fit indices used in the SEM framework incorporate the maximum likelihood (mis)fit function, the use of which is asymptotically justified under multivariate normality. In the presence of non-normal data, functions that incorporate the maximum likelihood (mis)fit function have been shown to indicate poor fit often when the data generation and analysis models are consistent (Curran, West, & Finch, 1996; Satorra, 1992; Yu, 2002). Although the regression-type diagnostics developed for the criticism of MLM (described below) do not appeal to distributional assumptions, no research has investigated the impact of non-normality on these measures.

A third factor, which was not pertinent to the review given by Wu et al. (2009), is the impact of prior distribution specifications in the Bayesian analysis context. On the one hand, prior distributions may be specified to contribute little to no information for the construction of the posterior distribution. The specification of *diffuse* prior distributions is

common, particularly for complex multivariate systems for which the analyst is unlikely to have prior knowledge about the variables. On the other hand, prior distributions may be specified to meaningfully contribute to the construction of the posterior distribution. The specification of *informative* priors can be useful if they are centered near the "true values" of model parameters but may produce inaccurate results that overwhelm the information in the data, particularly when the sample size is small (Depaoli, 2010, 2012).

Review of Literature on Data-Model Fit Indices. This section reviews the existing literature on the performance of select discrepancy functions as applied to growth curve models. Attention was primarily (but not exclusively) given to discrepancy functions that make minimal assumptions about the data and can be applied regardless of the structure for time. This section is organized as follows. First, a rationale is provided for the exclusion of a class of measures that are collectively referred to as "fit indexes", all of which are associated with (but not specific to usage within) the SEM framework at large. This is followed by a description of the discrepancy functions of interest for the current work. Consistent with the general theme of maintaining relevance to applied research, all of the discrepancy functions described in what immediately follows can be used for all types of growth curve models.

On the Exclusion of Fit Indexes. Although appearing in much of the existing methodological research pertaining to the criticism for growth curve models, a class of functions that are broadly referred to as fit indices are neither considered in this literature review nor the study (e.g., Liete & Stapleton, 2011; Liu, Rovine, & Molenaar, 2012; McMurray, 2010; West & Wu, 2010). These functions have a number of features in common. Many of them incorporate the maximum likelihood fit function, typically with

the use of some penalty factor that reduces the impact of sample size and/or penalizes model complexity. As a short list, common examples of these functions include the comparative fit index (CFI; Bentler, 1990), goodness-of-fit index (GFI; Jöreskog & Sörbom, 1984), root-mean-square error of approximation (RMSEA; Steiger & Lind, 1980); standardized root-mean-square residual (SRMR; Jöreskog & Sörbom, 1981); and the Tucker-Lewis index (TLI; Tucker & Lewis, 1973). Given reliance on the maximum likelihood fit function, many of the fit functions invoke assumptions of multivariate normality. In addition, the interpretation of fit functions often involves some appeal to cut off values that indicate some categorical decision about adequacy of the model (e.g., good fit, adequate fit, poor fit). These features are often criticisms associated with the use fit functions. However, owing to the construction of an empirical reference distribution, situating fit functions in the PPMC framework obviates the need to appeal to asymptotic behavior or cut off values.

The issue with these functions that cannot be overcome with the use of PPMC is the inability to use these functions with data in which time is unstructured (e.g., Wu et al., 2009; Wu & West, 2013). As seen in Equation 13, allowing for unique measurement schedules between people yields the potential for the model-implied covariance matrix to vary across individuals. The fit indices listed above, and many others like them, rely on the existence of a single covariance matrix that applies to the entire sample of data at hand, thereby limiting use to longitudinal designs in which the structure of time is balanced. Due to this limitation, fit indexes such as those listed above were not considered in the following literature review or the study.

Conditional Concordance Correlations and R^2 Measures. Recall that discrepancies in the conditional mean structure are captured by the difference between the observed and model-implied scores for individual *i* at a given measurement occasion *j* (i.e., the residual score, r_{ij}). Two measures that have shown promise include the conditional concordance correlation for the conditional mean structure (CCC_C; Lin, 1989; Vonesh, 1992; Vonesh, Chinchilli, & Pu, 1996) and the conditional R^2 (R_C^2; Singer & Willett, 2003). The components of these measures include the vectors of observed (denoted $\mathbf{y}_i$) and model-implied scores (denoted $\hat{\mathbf{y}}_i$, see Equation 12 for the definition of a particular predicted conditional mean score for a given *i* at measurement occasion *j*) for a given person *i*; the respective observed (denoted $\bar{y}_{GM}$) and model-implied (denoted $\hat{\bar{y}}_{GM}$) grand means averaged over individuals and measurement occasions; and for the *CCC* measures only, the total number of observations ($O = N \times J$ assuming no data are missing). The CCC_C is a measure of the agreement between the observed and model implied scores:

$$CCC_C = 1 - \left(\frac{\sum_{i=1}^{N} (\mathbf{y}_i - \hat{\mathbf{y}}_i)'(\mathbf{y}_i - \hat{\mathbf{y}}_i)}{\left[\sum_{i=1}^{N} (\mathbf{y}_i - \bar{y}_{GM})'(\mathbf{y}_i - \bar{y}_{GM})\right] + \left[\sum_{i=1}^{N} (\hat{\mathbf{y}}_i - \hat{\bar{y}}_{GM})'(\hat{\mathbf{y}}_i - \hat{\bar{y}}_{GM})\right] + O(\bar{y}_{GM} - \hat{\bar{y}}_{GM})^2} \right), \tag{15}$$

and R_C^2 is the squared correlation between the observed and model-implied scores:

$$R_C^2 = \frac{\left[\sum_{i=1}^{N} (\mathbf{y}_i - \bar{y}_{GM})'(\hat{\mathbf{y}}_i - \hat{\bar{y}}_{GM})\right]^2}{\left[\sum_{i=1}^{N} (\mathbf{y}_i - \bar{y}_{GM})'(\mathbf{y}_i - \bar{y}_{GM})\right] \times \left[\sum_{i=1}^{N} (\hat{\mathbf{y}}_i - \hat{\bar{y}}_{GM})'(\hat{\mathbf{y}}_i - \hat{\bar{y}}_{GM})\right]}. \tag{16}$$

As noted by Wu and West (2013), the difference in metric between CCC_C and R_C^2 usually results in the values of the former being larger. The values become closer to the extent that there is greater agreement between the observed and model-implied scores; given

perfect agreement between the observed and model-implied scores, the values will both be one.

Analogous functions of the *CCC* and R^2 measures have also been constructed to target the marginal mean structure. The structure of the marginal versions for these functions is obtained by replacing the vector of model-implied scores (denoted $\hat{\mathbf{y}}_i$) with the vector of model-implied means (denoted $\hat{\bar{\mathbf{y}}}_i$; see Equation 11 for the definition of a particular predicted marginal mean score for a given *i* at measurement occasion *j*). After substituting the terms, the resulting conditional concordance correlation for the marginal means (CCC_M) is in turn given by:

$$CCC_M = 1 - \left(\frac{\sum_{i=1}^{N} (\mathbf{y}_i - \hat{\bar{\mathbf{y}}}_i)'(\mathbf{y}_i - \hat{\bar{\mathbf{y}}}_i)}{\left[\sum_{i=1}^{N} (\mathbf{y}_i - \bar{y}_{GM})'(\mathbf{y}_i - \bar{y}_{GM})\right] + \left[\sum_{i=1}^{N} (\hat{\bar{\mathbf{y}}}_i - \hat{\bar{y}}_{GM})'(\hat{\bar{\mathbf{y}}}_i - \hat{\bar{y}}_{GM})\right] + O(\bar{y}_{GM} - \hat{\bar{y}}_{GM})^2} \right), \quad (17)$$

and the marginal R^2 (R_M^2) by:

$$R_M^2 = \frac{\left[\sum_{i=1}^{N} (\mathbf{y}_i - \bar{y}_{GM})'(\hat{\bar{\mathbf{y}}}_i - \hat{\bar{y}}_{GM})\right]^2}{\left[\sum_{i=1}^{N} (\mathbf{y}_i - \bar{y}_{GM})'(\mathbf{y}_i - \bar{y}_{GM})\right] \times \left[\sum_{i=1}^{N} (\hat{\bar{\mathbf{y}}}_i - \hat{\bar{y}}_{GM})'(\hat{\bar{\mathbf{y}}}_i - \hat{\bar{y}}_{GM})\right]}. \quad (18)$$

The CCC_M is the agreement between the observed scores and the marginal means; the R_M^2 is the squared correlation between the observed scores and the marginal means. As was the case of the conditional versions of these measures, the CCC_M will typically be larger than the R_M^2. Since the observed scores will be more closely related to the model-implied *scores* than to the model-implied *means*, the CCC_M and R_M^2 will usually be smaller than the corresponding conditional versions.

A simulation study by Wu and West (2013) is the only simulation study to date that has assessed the performance of the *CCC* and R^2 measures. In a factorial design, the

authors manipulated the strength of the average quadratic coefficient, the magnitude of the residual variance, and sample size. Since the reference distributions are not known, the authors submitted the values of the *CCC* and R^2 measures to an ANOVA; the reported outcomes included the average values of the *CCC* and R^2 measures and the strength of the effect size as measured by η^2. In null conditions (the data analysis model matched the data generation model), the means of the *CCC* and R^2 were affected by the magnitude of the residual variance but not other design factors. The effect was much stronger for the conditional measures (η^2 = .98 for CCC_C; η^2 = .97 for R_C^2) than for the marginal measures (η^2 = .38 for CCC_M; η^2 = .38 for R_M^2).

The conditional versions of the measures performed as expected. When the functional form of the trajectory was underspecified, the CCC_C and R_C^2 were sensitive to the strength of the average quadratic effect. As expected by the authors, sensitivity to the underspecification of the functional form declined with increases in the magnitude of the residual variance. A surprising result was the minimal impact of the strength of the average quadratic effect on the means of the CCC_M and R_M^2 when the only misspecification was in the marginal mean structure. The reported means of the measures were in fact identical across the levels of corresponding to the strength of the quadratic effect. Interestingly, when the data analysis model failed to account for the presence of a quadratic mean and a quadratic variance, the impact of the quadratic mean strength was evidenced for the CCC_M and R_M^2. For both measures, it is the *observed scores* that are compared to the *predicted means*; this finding may reflect this construction of the measures.

Likelihood Ratio Test. One of the byproducts of estimation is the raw value of the likelihood function (Singer & Willett, 2003). Under the assumption that the data from individual cases are independently and identically distributed, the raw value of the likelihood function given the parameters, $L(\boldsymbol{\Omega})$, is given by:

$$L(\boldsymbol{\Omega}) = \prod_{i=1}^{N} P(\mathbf{y}_i|\boldsymbol{\Omega}). \tag{19}$$

To ease the computation of the overall model likelihood, the sum of the individual *log* likelihoods, $LL_i(\boldsymbol{\Omega})$, is typically computed to form the overall model log likelihood, $LL_M(\boldsymbol{\Omega})$:

$$LL_M(\boldsymbol{\Omega}) = \sum_{i=1}^{N} LL_i(\boldsymbol{\Omega}), \tag{20}$$

such that:

$$LL_i(\boldsymbol{\Omega}) = p_i - \frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(\mathbf{y}_i - \hat{\mathbf{y}}_i)'\Sigma_i^{-1}(\mathbf{y}_i - \hat{\mathbf{y}}_i), \tag{21}$$

where p_i represents the number of observations provided by person *i* and the remaining terms retain the same meaning as described above. Notably, the version of the log likelihood shown in Equation 21 uses all of the information available in the observed data to estimate the model parameters (Bollen & Curran, 2006; Preacher et al., 2008). This version of the *LL* is appropriate for use in the presence of missing data, which the situation of varying measurement schedules across people is a special case of. For the case in which all data are observed and measurement schedules are invariant across people, the $LL_i(\boldsymbol{\Omega})$ simplifies to (see Coffman & Millsap, 2006; Preacher et al., 2008):

$$LL_i(\boldsymbol{\Omega}) = p - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(\mathbf{y}_i - \hat{\mathbf{y}})'\Sigma^{-1}(\mathbf{y}_i - \hat{\mathbf{y}}). \tag{22}$$

Multiplying the individual log likelihoods by -2 (Equations 21 or 22, depending on the situation at hand) and summing across people yields the *deviance* statistic (e.g., Coffman & Millsap, 2006; Singer & Willett, 2003). The deviance statistic will be referenced in the current work as the -2LL_M to reflect the foundation from which it is determined and to be consistent with how it is referenced throughout applied and methodological research. The diagnostic value of the -2LL_M usually comes in the form of evaluating whether the removal of some restriction on a given model results in a meaningful improvement (in the statistical sense) in data-model fit:

$$LRT = {}^{-}2\,LL_{M_A} - {}^{-}2\,LL_{M_B} = {}^{-}\,2\,LL_{M_A} + \,2\,LL_{M_B}. \quad (23)$$

The difference between the -2 *LL* value for the more restricted model (M_A) and the corresponding value with some restriction removed (M_B) yields the likelihood ratio difference test (*LRT*). The *LRT* is a general tool for evaluating whether it's statistically worthwhile to add complexity to the model. In the null hypothesis framework, the *LRT* is assumed distributed as a χ^2 with degrees of freedom equal to the difference in the number of parameters between M_B and M_A.

Notably, when testing variance components in the null hypothesis-testing framework, the *LRT* evaluates whether the variance(s) are significantly larger than zero, which is at the boundary of the parameter space. The implication is that the assumed χ^2 distribution does not hold. Rather, the correct distribution is a mixture of $b + 1$ χ^2 distributions where *b* is the number parameters on the boundary of the parameter space (Stoel, Galindo, & van den Wittenboer, 2006). When the distribution is not corrected, the *LRT* becomes more conservative to the extent that *b* increases (see Figures 2A through 2D in Stoel et al., 2006). Having said that, there is very little loss in power when the

incorrect reference distribution is used when $b = 1$ (Stoel et al., 2006; Verbeke & Molenberghs, 2003). Moreover, these issues are obviated in the context of resampling approaches in which the reference distribution is constructed.

A recent Monte Carlo study by Leite and Stapleton (2011) investigated the performance of the *LRT* for finding evidence of misspecifications in the shape of growth trajectories. In particular, the performance of the *LRT* was investigated when a linear growth model was incorrectly applied to data generated under various non-linear growth models (quadratic, plateau, piecewise) across different levels of sample size (100; 200; 500; 1,000; 2,000) and severity of misspecification (low, moderate, high). Importantly, the severity of misspecification was manipulated for the marginal mean structure and the **G** matrix. When manipulating the severity of misspecification of the marginal mean structure, the **G** matrix was held constant. When manipulating the severity of misspecification of the **G** matrix, the marginal mean structure was held constant.

The authors did not pursue estimates of Type I error or power as an outcome. Instead, the actual values of fit indices for were submitted to a between-condition ANOVA; given excessive power due to the large number of replications, they reported their results as effect sizes (i.e., partial η^2) for each main effect and interaction in the design of the ANOVA. They found that the *LRT* was impacted by sample size, the severity of the misspecification, and the interaction between these two factors. The nature of the interaction was such that the increase in the *LRT* associated with the change in misspecification severity was greater for larger sample sizes. Although this is actually an encouraging finding when the data analysis model actually does underspecify the underlying functional form, performance was not evaluated in null conditions. As a

result, it is difficult to gauge whether inferences about the performance of the *LRT* should be tempered by a tendency to increase with larger sample sizes.

Liu et al. (2012) evaluated the performance of the *LRT* to select covariance structures generated from a GCM. The performance of the measures was investigated across four different covariance structures (compound symmetry, first-order autoregressive, first-order moving average, random-coefficients), two levels of sample size (20, 100), and magnitude of covariances (small, medium, large). The outcome in their study was the proportion of times the correct covariance structure was selected. The *LRT* was found to be superior to all other measures considered irrespective of sample size, magnitude of the covariance, and type of covariance structure. Although the covariance structures investigated by the authors are not considered here, the results of the simulation study highlight the strength of the *LRT* as a diagnostic tool for evaluating highly specific hypotheses despite its general formulation.

Standardized Generalized Dimensionality Discrepancy Measure. As alluded to above, one way to detect whether a key process or variable has been omitted from a multivariate system is to evaluate the off-diagonal elements in the **R** matrix. This approach is popular in the context of measurement models such as confirmatory factor analysis (CFA; Kline, 2005), item response theory (IRT; Embretson & Reise; 2000), and latent class models (LCA; Collins & Lanza, 2010). Although the different measurement models vary with respect to assumptions about the nature of observed and latent variables, the common thread among them is to adequately specify the latent structure to render the off-diagonal elements of **R** to (essentially) zero (e.g., Levy, 2006). If this is achieved, the specified latent structure is deemed to have successfully achieved

conditional independence among the observables at hand[1]. As acknowledged by Crawford (2014), functions for evaluating the tenability of conditional independence can be constructed to be a measure of global fit, which involves the full collection of observables, or a measure of local fit, which involves a subset of the observables (Levy & Svetina, 2011). In the case of local fit, the subset can be viewed as collection of observables that measure something unique from the observables that are not in the subset; the subset must include at least two observables.

One function that has been shown to be successful is the standardized generalized dimensionality discrepancy measure (*SGDDM*; Levy, Xu, Yel, & Svetina, 2015; see Levy & Svetina, 2011 for the unstandardized version, *GDDM*). The $SGDDM_C$ is an aggregation of conditional associations, specifically those elements in the off-diagonal of the **R** matrix. In the case of GCM, recall that variability exists (a) across measurement occasions within people and (b) between people in the magnitude and/or functional form of growth. The original form of the *SGDDM* (subscripted by *C* for the purposes of this

[1] Technically, the evaluation of conditional independence as described represents an assessment of *weak* conditional independence (WCI). This form of conditional independence requires that all bivariate associations among residual scores be equal to zero; since higher order relationships may still be present, WCI is not sufficient for meeting the strong form of conditional independence (e.g., Ip, 2000; Levy, 2006; Stout, 1987; Zhang & Stout, 1999a, 1999b). WCI is generally deemed *empirically* sufficient in that the higher-order relationships required for SCI are unlikely if bivariate associations are not present (McDonald, 1994).

work) can be used to determine the portion of the **R** matrix that is due to underspecification across measurement occasions within people:

$$SGDDM_C = \frac{\sum_{j \neq j'} \left| \frac{\frac{\sum_{i=1}^{N}(y_{ij} - \hat{y}_{ij})(y_{ij'} - \hat{y}_{ij'})}{N}}{\sqrt{\frac{\sum_{i=1}^{N}(y_{ij} - \hat{y}_{ij})}{N}} \sqrt{\frac{\sum_{i=1}^{N}(y_{ij'} - \hat{y}_{ij'})}{N}}} \right|}{J(J-1)/2}. \quad (24)$$

Each of the components retains the same meaning as described above. The marginal version of the *SGDDM* (labeled $SGDDM_M$) obtains by replacing the model-implied score (denoted $\hat{y}_{ij}$; see Equation 12 for the definition of a particular predicted conditional mean score for a given *i* at measurement occasion *j*) by the model-implied mean (denoted $\hat{\bar{y}}_{ij}$; see Equation 12 for the definition of a particular predicted marginal mean score for a given *i* at measurement occasion *j*)

$$SGDDM_M = \frac{\sum_{j \neq j'} \left| \frac{\frac{\sum_{i=1}^{N}(y_{ij} - \hat{\bar{y}}_{ij})(y_{ij'} - \hat{\bar{y}}_{ij'})}{N}}{\sqrt{\frac{\sum_{i=1}^{N}(y_{ij} - \hat{\bar{y}}_{ij})}{N}} \sqrt{\frac{\sum_{i=1}^{N}(y_{ij'} - \hat{\bar{y}}_{ij'})}{N}}} \right|}{J(J-1)/2}. \quad (25)$$

The *SGDDM* discrepancy measures range from zero to one. Values closer to zero are indicative of close agreement between the observed associations and the associations implied by the model. The evidence in favor of data-model misfit grows to the extent that values approach one. To date, the $SGDDM_M$ has neither been used in applied research nor has it been systematically investigated in any methodological work, but it is included here as a method of to evaluate the marginal mean structure. In contrast, the $SGDDM_C$ has been used with success in applied psychometric research (e.g., Levy, Crawford, Fay, &

Poole, 2011; Rupp et al., 2012) and investigated methodologically (e.g., Crawford, 2014; Levy et al., 2015) in the context of psychometric models. Although the $SGGDM_C$ has exhibited a strong performance for different types of highly complex models, it remains an open question how it will perform in the context of GCM.

What Impacts the Success of Data-Model Fit Analyses in GCM?

Having reviewed the available methodological literature, several factors emerge as salient to the success of data-model fit analyses for GCM. In the case of underspecifying the function form of the growth trajectory, it is clearly the case that the degree of underspecification matters. For example, if a GCM with linear function form is applied to data that exhibit a quadratic functional form, identifying the underspecification of the functional form becomes easier to the extent that the quadratic effect is large (Leite & Stapleton, 2011; Wu & West, 2010; Wu & West, 2013). A similar argument holds in the case of underspecifying the variability in the conditional mean structure. For example, if a GCM with a quadratic functional form but no quadratic variance is applied to data in which the quadratic relationship with time varies over people, identifying the underspecification of the functional form becomes easier to the extent that the quadratic variance is large (Leite & Stapleton, 2011; Wu & West, 2013).

Another important feature is sample size. In the case of the correlation-based functions described above (i.e., *CCC* and R^2 measures), sample size does *not* appear to affect the value of these measures holding all else constant (Wu & West, 2013). Although not investigated in the context of GCM, it is expected that the performance of the *SGDDM* functions would similarly be unaffected by sample size. In contrast to the correlation-based functions, likelihood-based functions are generally known to detect

trivial data-model misfit to the extent that sample size is increased across modeling contexts (e.g., Bollen & Curran, 2006; Wu et al., 2009). In a hypothesis-testing framework, the impact is that likelihood-based functions become more likely to reject even trivial discrepancies between the model and the data (e.g., Millsap & Coffman, 2006). When situated within a resampling framework in which the reference distribution is constructed, some research suggests that the impact of sample size is diminished and may even be irrelevant. As evidence of this claim, Nyland, Muthén, and Asparouhov (2007) showed that the performance of a parametrically bootstrapped version of the *LRT* was insensitive to sample size and consistently favored the data-generation model across a variety of models, including a growth mixture model (a model-based extension of GCM that extracts unobserved groups that may exhibit varying growth process; see Bauer & Curran, 2003; Muthén & Shedden, 1999; and Nagin, 1999 for a general description of these models).

The magnitude of the residual variances has also been identified as a key factor that ubiquitously affects the performance of the *CCC* functions, R^2 functions, and likelihood-based functions (Wu & West, 2013). This beckons the question of whether the magnitude of the residual variance would similarly have bearing on the values of the *LRT* and *SGDDM* functions. As it involves a *comparison* of model log likelihoods, it is difficult to predict how the magnitude of residual variances might impact the *LRT*. Although the individual log likelihood might be shifted upwards (indicating greater data-model misfit), the behavior of the *difference* in the log likelihood (i.e., the *LRT*) may be quite different. As described above, the *SGDDM* functions target the off-diagonal elements of the **R** matrix (i.e., $j \neq j'$). In contrast, the *CCC* and R^2 functions target the

diagonal elements of the **R** matrix. In cases of underspecification, residual variances (i.e., the diagonal elements of **R**) will be large since model-implied scores are discrepant from observed scores (or means in the case of the functions that target the fit marginal mean structure). Similarly, associations among residual scores will also be large owing to the presence of systematic discrepancies between the observed scores and data implied by the model. However, when the data analysis model matches the data generation model, the *CCC* and R^2 functions may be more susceptible to suggesting underspecification of the functional form than the *SGDDM* measures to the extent that the repeatedly observed measures are unreliable (this was implied in the form of lower values with increasing values of the residual variance in Wu & West, 2013). Holding all else constant, unreliable measures will yield large residual variances (due to large errors in prediction) but lower covariation among residual scores (due to less overlap among the observed measures), even if the data analysis and generation models are matched.

One feature that has been identified as potentially important is the shape of the distribution for observed variables (Wu et al., 2009). All of the methodological studies described above involved the generation and analysis of growth curve models under the assumption that the data arise from a multivariate normal distribution. Notably, the *CCC*, R^2, and *SGDDM* measures do not make the assumption that the observed data exhibit multivariate normality. Although this assumption is not made, the question of how these functions behave when the observed data are non-normal is an empirical one that has not yet been evaluated. Unlike the measures listed above, the likelihood-based *LRT* does rely on multivariate normality. However, it has been shown that it is the standard errors of model parameters rather than the estimates of model parameters that are impacted by

non-normal data (e.g., Min, 2008). Since the *LRT* is constructed in part from the estimates of model parameters, the question of how the *LRT* performs, particularly in a parametric resampling framework, is an empirical one that has not been evaluated.

Summary

This chapter has summarized the core concepts and relevant literature that inform the design decisions of the current study. To date, the limited amount of methodological work pertaining to data-model fit for GCM is situated within the classical frequentist paradigm; no study to the author's knowledge has pursued a fully Bayesian approach for conducting data-model fit analyses for growth curve models. As described in greater detail in the next chapter, the goal of the current work was to investigate the utility of the discrepancy functions described above to critique the fit of growth curve models using the popular Bayesian approach of PPMC. Notably, many of the functions described above do not have known reference distributions; this is a non-issue in the PPMC context, which serves as a mechanism for empirically constructing the appropriate reference distribution while also acknowledging the uncertainty in the model parameters.

CHAPTER 3: Methods

Data Generation Model

The general GCM used to generate data allowed for person-specific intercept (b_{0i}), slope (b_{1i}), and quadratic (b_{2i}) coefficients as follows:

$$y_{ij} = b_{0i} + b_{1i}(Time_j) + b_{2i}(Time_j)^2 + r_{ij}. \quad (26)$$

The person-specific regression coefficients were in turn simulated from a multivariate normal distribution with the mean vector ($\boldsymbol{\gamma}$) defined by the mean regression coefficients (γ_{00} = intercept mean, γ_{10} = slope mean, γ_{20} = quadratic mean) and the **G** matrix as follows:

$$\begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim MVN\left(\boldsymbol{\gamma} = \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \end{bmatrix}, \mathbf{G} = \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ 0 & 0 & \tau_{22} \end{bmatrix}\right). \quad (27)$$

The measurement schedule with $J = 5$ occasions was assumed identical across people ($Time_j$) such that $Time = 0, 1, 2, 3, 4$. The variance of the residual scores were assumed homogeneous and uncorrelated across measurement occasions; in effect, the residuals scores were generated as follows:

$$\mathbf{r}_i \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} \sigma_\varepsilon^2 & & & & \\ 0 & \sigma_\varepsilon^2 & & & \\ 0 & 0 & \sigma_\varepsilon^2 & & \\ 0 & 0 & 0 & \sigma_\varepsilon^2 & \\ 0 & 0 & 0 & 0 & \sigma_\varepsilon^2 \end{bmatrix}\right). \quad (4, \text{repeated})$$

As dependent on the particular condition (which are described below), appropriate restrictions were placed on the generating model. Some conditions were generated to follow GCM with a quadratic functional form but no between-person variation in the strength of the quadratic effect. These conditions were obtained by fixing the quadratic

variance (τ_{22} in the **G** matrix) to zero. As a further restriction, some conditions were generated to follow GCM with a linear functional form. These conditions were obtained by fixing the quadratic mean (γ_{20} in the $\boldsymbol{\gamma}$ vector) and quadratic variance (τ_{22} in the **G** matrix) to zero.

A common practice in the design of Monte Carlo simulation studies is to specify particular values for model parameters and other features that facilitate the generation of data. This approach has the benefit of producing highly comparable independent trials. The drawback to this approach is that the results capitalize on the chance features of the particular values that are chosen. To minimize the impact of capitalizing on the whims of particular values, the values that governed the creation of data were drawn from random uniform distributions with minimum and maximum values. Although this choice reduces the comparability of independent trials within a condition somewhat, the advantage that is gained is a more general representation of the condition to applied research settings. Moreover, provided the boundaries of random uniform distributions that distinguish levels of the manipulated variables are sufficiently separated and yield qualitatively distinct levels of the manipulated variable, the principal effects of the manipulated variables on the outcome will still be observed.

Some features of the simulated model were consistent across conditions. The mean intercept (γ_{00}) was drawn from a random uniform distribution with a minimum of nine and a maximum of 11. In applied research, the values of the intercept are dependent on the scale of the repeatedly measured outcome. Owing to the dependence of the intercept on the particular outcome, the choices above were arbitrarily selected. The mean slope (γ_{10}) was specified as a percentage (minimum = 30%, maximum = 50%) of the

simulated value of the intercept. Values of the variances for the intercept and slope parameters were first specified on a standard deviation metric, such that $\sqrt{\tau_{00}} \sim U(3, 4)$, and then converted to a variance for the purposes of data generation; the standard deviation of the slope was in turn 30% to 50% of $\sqrt{\tau_{00}}$. The correlation between the intercept and slope coefficients assumed a value between $.25 \leq \rho_{10} \leq .35$; the covariance necessary for simulating the person-specific regression parameters was in turn determined by $\tau_{10} = \rho_{10} \times \left(\sqrt{\tau_{00}} \times \sqrt{\tau_{10}}\right)$. The remaining characteristics of the generated data were determined by the manipulated variables described below. Importantly, these values and those that follow were selected after reviewing examples in both applied research (e.g., Chou, Bentler, & Pentz, 1998; Shevlin & Millar, 2006; Mäkikangas, Bakkar, Aunola, & Demeraouti, 2010; You & Sharkey, 2009; Shaeffer, Petras, Ialongo, Poduska, & Kellam, 2003) and methodological research (e.g., Coffman & Millsap, 2006; Wu & West, 2010, 2013).

Manipulated Variables

The design of the Monte Carlo study consisted of manipulating the following factors: sample size (3 levels); the strength of the quadratic mean (3 levels); magnitude of the quadratic variance (3 levels); magnitude of the residual variance (2 levels); and distribution shape (2 levels). These factors are described in turn below.

Sample Size. The three levels of sample size (Small: $N \sim U(225, 275)$; Moderate: $N \sim U(475, 525)$; Large: $N \sim U(975, 1{,}025)$) were in part selected to support comparisons to other methodological research pertaining to data-model fit for growth curve models (e.g., Liete & Stapleton, 2011; Wu & West, 2013). Although there is significant

variability in the range of sample sizes observed in applied research settings, many applications of GCM fall within the range of values used in this work.

Quadratic Mean Strength. As was the case for the mean linear effect (γ_{10}), the values of the mean quadratic effect (γ_{20}) were defined as some percentage of the drawn value of the intercept mean (γ_{00}). For data generated to follow GCM with a linear functional form, the quadratic mean was fixed to zero (i.e., $\gamma_{20} = 0$). The remaining two levels were defined by random uniform distributions with some minimum and maximum percentage of the mean intercept (i.e., γ_{00}). In conditions generated with a small quadratic mean, $\gamma_{20} = U(-0.03 \times \gamma_{00}, -0.01 \times \gamma_{00})$. For conditions with a large quadratic mean, $\gamma_{20} = U(-0.13 \times \gamma_{00}, -0.10 \times \gamma_{00})$. For both levels of the quadratic mean strength, the negative sign indicates that the rate of linear growth decelerates with the passage of time. For the current study, this means that the *average rate of linear change* decreases by the amount of the average quadratic effect with each unit increase in time.

Quadratic Variance. Figure 5 shows the lower and upper boundaries for each level of the quadratic variance manipulation. For each panel, the horizontal axis represents the passage of time, and the observed score on some outcome is shown on the left vertical axis. For reference, the leftmost column shows a GCM with a quadratic functional form with no variation between people in the functional form of growth. The panels to the right of reference figure are configured as a 2 × 2 matrix in which the rows represent the two levels of quadratic variance magnitude (small and large) and the columns represent the boundaries of the corresponding random uniform distributions.

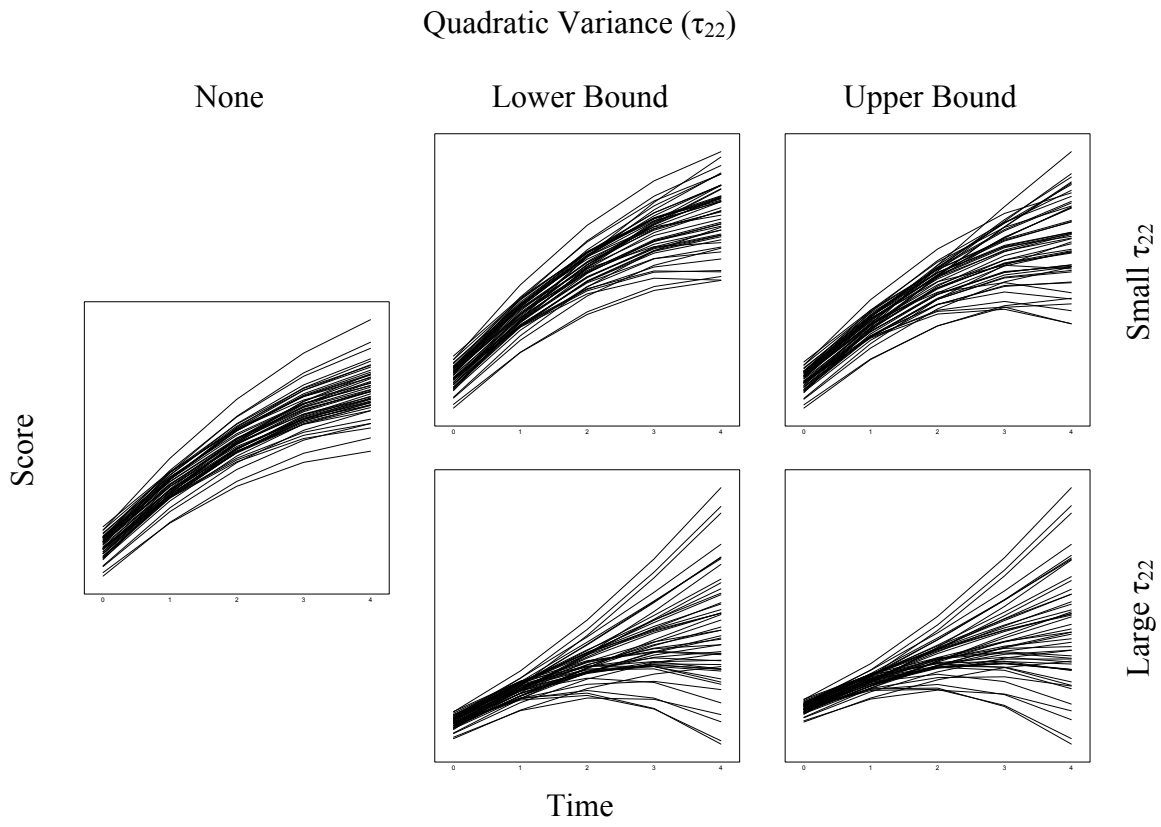

Figure 5. Visual representation of the lower and upper bounds of the random uniform distributions that distinguish the levels of the small and large quadratic variance factor.

Looking within a level of the quadratic variance, the goal was to specify values that produce similar patterns in the individual-level trajectories; this was to ensure some consistency among trials within conditions. Looking across levels (regardless of which boundary), the goal was to specify values of τ_{22} that clearly produce different patterns in the person-level trajectories. By setting a seed value, each of the figures exhibit exactly the same form excepting to the magnitude of τ_{22}. In conditions with a small τ_{22}, the values (in the standard deviation metric) ranged from 0.1 to 0.2. For conditions generated with a large τ_{22}, the values (in the standard deviation metric) ranged from 0.5 to 0.6.

Residual Variance. As was the case for determining the levels of the quadratic variance, graphical representations were pursued to determine reasonable intervals for the levels of the residual variance. Using the same general structure shown in Figure 5, Figure 6 shows the lower and upper boundaries for each level of the residual variance manipulation. In this case, the rows of the 2 × 2 matrix correspond to the levels of the σ_ε^2. The trajectories within panels were generated to exhibit the exact same structure (a linear GCM model) with the only difference between panels being the magnitude of σ_ε^2. In conditions with a small σ_ε^2, the values (in the standard deviation metric) ranged from 0.75 to 1.25. For conditions generated with a large τ_{22}, the values (in the standard deviation metric) ranged from 2.25 to 2.75.

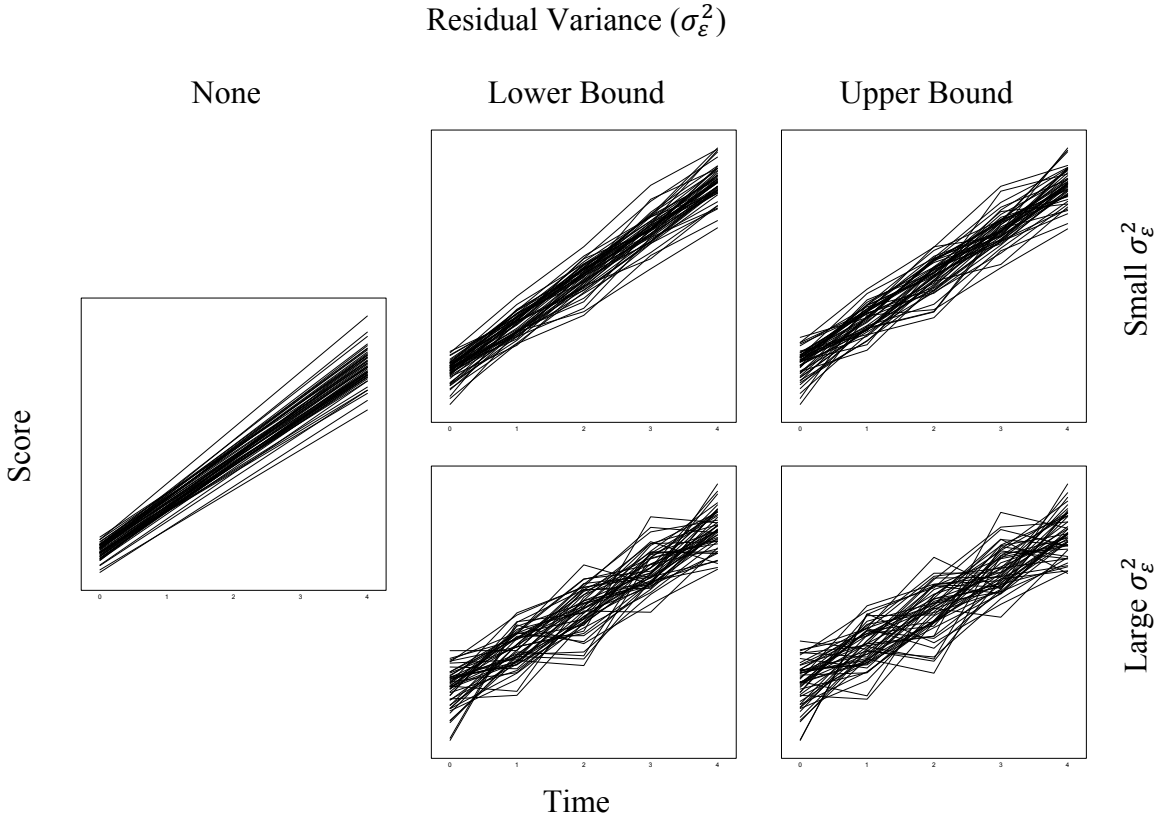

Figure 6. Visual representation of the lower and upper bounds of the random uniform distributions that distinguish the small and large levels of the residual variance factor.

Distribution Shape. For the univariate case, Fleishman (1978) developed a procedure to generate non-normal *univariate* data with desired levels of skewness and kurtosis. Fleishman's method involves applying a polynomial transformation to convert a normally distributed variable (denote it X) to a new variable (denote it U) with the desired skewness and kurtosis as follows:

$$U = a + bX + cX^2 + dX^3, \tag{28}$$

such that the coefficients (a, b, c, d) represent the unknown values that are necessary to transform X to U. Vale and Maurelli (1983) extended Fleishman's method to the bivariate case. Although the details of their method is beyond the intentions and scope of this work, the key idea is to apply Fleishman's method for the univariate case shown in Equation 28 to obtain the coefficients to transform both variables in question. Then, given the original correlation between the variables, an intermediate correlation is computed for each pairing of applicable variables. The $J \times J$ matrix of intermediate correlations are in turn used to generate multivariate data with the desired skewness and kurtosis.

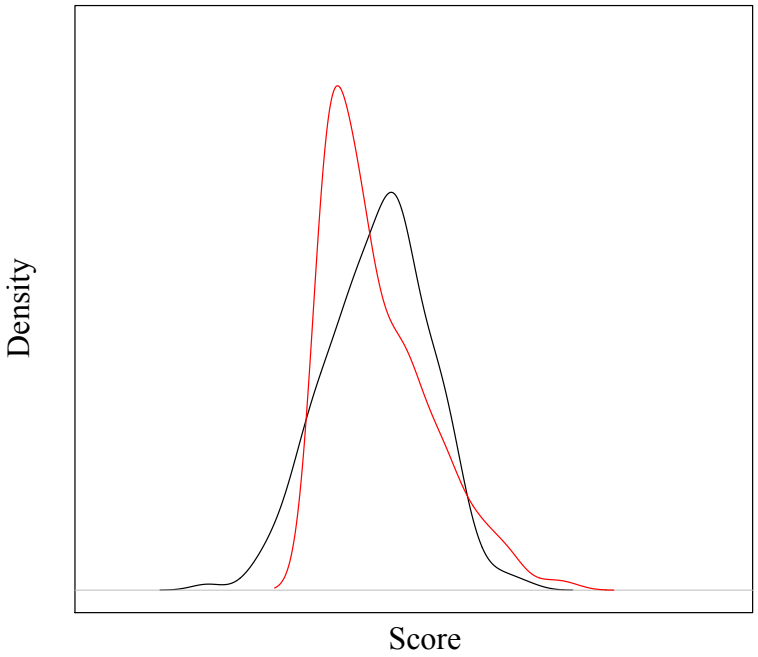

Figure 7. Comparison of density distributions that reflect the application of Vale and Maurelli's (1983) method for generating non-normal data. The black line shows a variable prior to applying Vale and Maurelli's method to generate non-normal data; the red line shows the same variable following the application of their method.

The levels for the distribution shape manipulation consisted of two levels: normal and non-normal. Conditions characterized by non-normally distributed data were moderately skewed in the positive direction (right-tailed) and moderately leptokurtic (peaked). The values of skew and kurtosis were independently drawn from a random uniform distribution with a minimum of 0.8 and a maximum of 1.2. After generating data to exhibit multivariate normality, the drawn values of skew and kurtosis were entered into an optimization routine to find the unknown values of the polynomial coefficients and convert the original data to be non-normally distributed. Figure 7 (shown above) reflects the success of the optimization routine for implementing Vale and Maurelli's method (1983). The density distribution shown in black represents one variable prior to be transformed, and the distribution shown in red represents the same variable after applying the method. Excepting to the skew, kurtosis, and median (which was shifted towards the

peak of the distribution), other features such as the means, standard deviations, and correlations remained identical after applying the routine.

Summary of Data Generation. Table 1 summarizes the boundaries of the random uniform distributions that distinguish among the levels for each of the manipulated variables. If the levels of manipulated variables were fully crossed, the study would consist of 108 design cells. However, design cells representing combination of a null quadratic mean (i.e., $\gamma_{20} = 0$) and non-null quadratic variance (i.e., $\tau_{22} > 0$) were omitted since this situation is not likely to be encountered in applied research settings. Excluding these conditions, the study consisted of 84 design cells that represent unique longitudinal data structures that approximate situations that may be encountered applied research settings (e.g., Chou et al., 1998; Shevlin & Mullar, 2006; Mäkikangas et al., 2010; You & Sharkey, 2009; Shaeffer et al., 2003). Using R 3.1.1 (R Core Team, 2014), 100 independent trials were simulated for each design cell. For each trial within a condition, the parameters that governed the generation of data were unique from all other trials but were based on the same combination of random uniform distributions corresponding to the levels of the manipulated variables that define that condition.

Table 1

Minimum and Maximum Values of the Random Uniform Distributions for Each Level of the Manipulated Variables

		Random Uniform Distribution Boundaries	
Manipulated Variable	Level	Minimum	Maximum
Sample Size	Small	225	275
	Moderate	475	525
	Large	975	1,025
Quadratic Mean Magnitude	None	0	0
	Small	$-(.03 \times \gamma_{00})$	$-(.01 \times \gamma_{00})$
	Large	$-(.13 \times \gamma_{00})$	$-(.10 \times \gamma_{00})$
Quadratic Standard Deviation (Variance) Magnitude	None	0	0
	Small	.10 (.01)	.20 (.04)
	Large	.50 (.25)	.60 (.36)
Residual Standard Deviation (Variance) Magnitude	Small	.75 (.56)	1.25 (1.56)
	Large	2.25 (5.06)	2.75 (7.56)
Distribution Shape	Normal		
	Skew	0	0
	Kurtosis	0	0
	Non-Normal		
	Skew	.8	1.2
	Kurtosis	.8	1.2

Data Analysis Models

For each of the 8,400 trials (i.e., 100 trials within each of the 84 conditions), the posterior distributions were constructed for three data analysis models. The likelihood of

the data at a given measurement occasion was characterized by a normal distribution with a common residual variance[2] across measurement occasions:

$$y_{ij} \sim N(\hat{y}_{ij}, \sigma_{\varepsilon}^2). \quad (29)$$

such that:

$$\hat{y}_{ij} = b_{0i} + b_{1i}(Time_{ij}) + b_{2i}(Time_{ij}^2). \quad (30)$$

As captured by integer values in which the first occasions represents the intercept (i.e., *Time* = 0, 1, 2, 3, 4), measurement schedules were assumed equivalent across $J = 5$ measurement occasions. A common residual variance (σ_{ε}^2) was assumed across measurement occasions, and was assigned the following prior:

$$\sigma_{\varepsilon}^2 \sim IG(1,1). \quad (31)$$

With the mean of the intercept, slope, and quadratic terms as the mean vector and the **G** matrix as the covariance matrix, the prior distribution of the person-specific regression coefficients was specified by a multivariate normal distribution:

$$\begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim MVN\left(\boldsymbol{\gamma} = \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \end{bmatrix}, \ \mathbf{G} = \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ 0 & 0 & \tau_{22} \end{bmatrix}\right). \quad (32)$$

Each of the elements in $\boldsymbol{\gamma}$ were independently assigned a univariate normal prior:

[2] The software package used for model-fitting parameterizes variability in the precision metric, which is simply the inverse of the variance. The choice to express the data analysis models in the variance metric represents an assumption that the precision is relatively unfamiliar metric. In the precision metric, the analogous prior distribution in the univariate (multivariate) case is a gamma (Wishart) distribution (Gill, 2007).

$$\gamma_{00} \sim N(0,1)$$

$$\gamma_{10} \sim N(0,1) \tag{33}$$

$$\gamma_{20} \sim N(0,1)$$

with the prior for the **G** matrix specified as an inverse Wishart distribution

$$\mathbf{G} \sim IW\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, v = K - 1\right) \tag{34}$$

where *K* is the number of elements in the marginal mean structure.

As mentioned, three data analysis models were applied to each independent trial in all conditions. The first model (M1) reduces to a linear GCM by fixing the quadratic mean, $\gamma_{20} = 0$, and quadratic variance to zero, $\tau_{22} = 0$. The second model (M2) included a non-zero quadratic mean but no quadratic variance, $\tau_{22} = 0$. The third model (M3) included both a quadratic mean, and a quadratic variance term, $\tau_{22} > 0$. When taken in sequence, M1, M2, and M3 can be viewed as natural steps of the model-building process often used when fitting growth curve models (assuming growth is in fact present and there are no covariates of interest). The posterior distribution of model parameters from all three data analysis models were obtained using JAGS 3.4.0 (Plummer, 2013) as interfaced through the `rjags` package in R 3.1.1 (R Core Team, 2014).

Estimation. A trial study was conducted to determine (a) the requisite number of burn-in iterations to consistently achieve convergence and (b) a sufficient thinning interval to consistently yield approximately independent draws from the posterior distribution. For each of the 84 conditions considered in this work, the posterior distribution was obtained for two independent trials under each of the three data analysis models. To determine (a) and (b) above, the following graphical representations were

examined for the marginal posterior distributions of all model parameters: the time series showing successive draws across MCMC iterations; the lagged autocorrelation; and the degree of overlap between the marginal posterior densities rendered from two chains. The results of the trial study suggested that 500 burn-in iterations were sufficient to achieve convergence with two chains for all model parameters associated with each of the data analysis models in all data generation conditions.

Separate thinning intervals were required for different combinations of the data analysis and data generation models. The thinning interval was largely dependent on the magnitude of the residual variance such that conditions generated with a small residual variance (Minimum = 15, Median = 25, Maximum = 30) required a smaller interval than conditions generated with large residual variances (Minimum = 20, Median = 40, Maximum = 55). In all models, the posterior distributions were constructed from 300 draws with 150 draws coming from each of two chains. Owing to differences in thinning intervals, the total number of iterations following burn-in ranged from 4,500 (= 300 × 15) to 16,500 (= 300 × 55) iterations. All posterior draws were written to a file for post-processing.

The estimation of all trials was performed on an iMAC with a 3.2 quad-core Intel i5 processor with 8 GB of memory. The amount of time required for estimation was largely determined by sample size; the smaller sample size conditions required about 40 minutes, and the larger sample size conditions required about 4 hours and 30 minutes. Although minimal, model complexity also contributed to estimation time. The total amount of time required to estimate all models for all trials was approximately 8.5 days.

Posterior Predictive Model Checking

Discrepancy Functions. Eight discrepancy functions that target different features of the growth curve models were investigated in the current work. The following indicates which discrepancy functions were used in the current work to target different sources of data-model misfit.

Overall Covariance Structure and Marginal Mean Structure. One function that is often used to simultaneously assess the data-model fit of the overall covariance and marginal mean structures is the maximum likelihood discrepancy function (F_{ML}; e.g., Bollen & Curran, 2006):

$$F_{ML} = \ln|\boldsymbol{\Sigma}| - \ln|\mathbf{S}| + tr(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - J - [(\bar{y} - \hat{\bar{y}})' \boldsymbol{\Sigma}^{-1} (\bar{y} - \hat{\bar{y}})], \quad (35)$$

such that **S** is the overall covariance matrix for the sample and the other terms retain the same meaning as described above. The F_{ML} discrepancy function is in turn multiplied by N - 1 to yield the likelihood ratio (LR) statistic:

$$LR = (N - 1)F_{ML}. \quad (36)$$

Due to its common use to critique many types of models, the LR was included owing to its use as a general-purpose function for assessing data-model fit (e.g., Schienes, Hoijtink, & Boomsma, 1999). Wu and West found (2013) that the values of the LR (represented as T_{ML} in their study) was sensitive to data-model misfit in the mean structure, and in particular, the marginal mean structure. Notably, the LR assumes a common covariance for the entire sample; in the case of varying measurement schedules, the LR cannot be computed (Bollen & Curran, 2006; Preacher et al., 2008; Wu et al., 2009). Although this dependency on the time structure limits the utility of the LR statistic,

it warrants investigation in the current study as a balanced structure was used for the current work.

Absolute Fit of the Conditional Mean Structure. The conditional versions of the *CCC* and R^2 measures were employed to assess the congruence between observed and model-implied scores (i.e., the conditional mean structure). These measures can be viewed as being reflective of the portion of variance along the diagonals of the **R** matrix that is due to differences between the observed and model-implied scores.

Absolute Fit of the Marginal Mean Structure. The marginal versions of the *CCC* and R^2 measures were employed to assess the fit of the marginal mean structure. These measures can be viewed as being reflective of the portion of variance along the diagonals of the **R** matrix that is due to differences between the observed scores and the model-implied means (i.e., the marginal mean structure).

Discrepancies in the Off-Diagonal Elements of the R Matrix. The conditional and marginal versions of the *SGDDM* measures are separated from the corresponding *CCC* and R^2 discrepancy functions as they differ in logic. As described above, the *SGDDM* functions measure the amount of variance present in the off-diagonals of the **R** matrix. The conditional version of the *SGDDM* captures the portion of that variance that arises due to the unaccounted associations between the observed and model-implied scores (i.e., the conditional mean structure). The marginal version of the *SGDDM* captures the portion of the off-diagonal variance that arises due to the unaccounted for associations between the observed scores and model-implied means (i.e., the marginal mean structure).

Likelihood Ratio Difference Test (LRT) for Relative Fit. The *LRT* is designed to evaluate the relative fit of two models (see Equations 19 – 23). Owing to the use of the entire posterior distribution for constructing the reference distribution, the evaluation of relative fit is not as straightforward as in frequentist-based frameworks such as null hypothesis significance testing (NHST) and parametric bootstrapping. In the NHST framework, two models are applied to the observed data; the likelihood for each model is in turn based on the same data and one set of model parameters (usually obtained via ordinary least squares or maximum likelihood estimation). The *LRT* is then computed and the observed value is compared to a critical χ^2 with degrees of freedom equal to the difference in the number of parameters estimated between the models. The key difference for the parametric bootstrap is that reference distribution of the *LRT* is constructed. This is achieved by simulating some large number of datasets using the model parameters from the more restricted model. For each simulated dataset, the models being compared are fitted to each simulated dataset, and ultimately, yields likelihoods for the two models. This step in turn facilitates the computation of the *LRT* for each simulated dataset given data that are consistent with the restricted model. The observed value of the *LRT* is then located within the reference distribution to arrive at a decision about the suitability of the restricted model for the observed data.

Like the parametric bootstrap method, PPMC is a resampling technique for constructing reference distributions of test statistics and discrepancy functions. However, owing to the availability of the entire posterior distribution for the model parameters, arriving at an appropriate null distribution is more complicated than for the parametric bootstrap approach. Although there are number of approaches to constructing the

reference distribution, the following steps—which were used for the current work—describe one way to do so within the PPMC framework.

1. For a given observed dataset, construct the posterior distribution of model parameters under the more restricted model (M_A) and the less restricted model (M_B). For the purposes of this work, the posterior distributions were obtained via MCMC estimation.
2. Using the posterior distribution for M_A, generate R posterior predictive datasets that are consistent with M_A.
3. Using the observed data, and the respective posterior distributions of model parameters, compute the realized values of the -2 log likelihood (see Equation 22) for M_A and M_B. Compute the *LRT* as the difference between the -2 log likelihoods (see Equation 23).
4. For each of R posterior predictive datasets (which were generated to follow M_A in Step 2), compute the -2 log likelihoods using the posterior distribution of model parameters associated with M_A and M_B (see Equation 22). Compute the *LRT* as the difference between the -2 log likelihoods between M_A and M_B (see Equation 23).
5. Compare the realized values of the *LRT* to the posterior predictive values. One way to compare these values is to compute the PPP-value of the *LRT* (see Equation 10).
6. Given an a priori selected threshold (akin to the selection of α), determine whether the PPP-value is extreme. If the PPP-value is greater than the threshold, the more restricted (i.e., M_A) model is favored; if the PPP-value is less than or

equal to the threshold, the added complexity of the less restricted model (i.e., M_B) is favored.

This series of steps is akin to the parametric bootstrap in that data are simulated from M_A, the more restricted model. The key difference is that models M_A and M_B are not fit to the simulated datasets. Rather, in conjunction with data generated under the more restricted model, the posterior distributions of model parameters are used to compute the -2 log likelihood for each model.

Data Analyses

There were three outcomes of interest for the current work: the realized values, the distribution of PPP-values, and the proportion of extreme PPP-values. To date, simulation studies on the performance of discrepancy functions in the GCM context have focused more on the behavior of the actual values of discrepancy functions rather than on performance from a NHST (i.e., empirical Type I error rates and power) perspective (e.g., Liete & Stapleton, 2011; Wu & West, 2010; Wu & West, 2013). Accordingly, by evaluating the realized values, it is made possible to connect the results of the current work to prior research. The analysis of realized values involved fitting a series of factorial analysis of variance (ANOVA) models. The ANOVA models were applied separately to discrepancy functions for null and non-null situations. The null conditions consisted of fitting the model that matches the generating process; non-null conditions involved fitting a model that underspecifies a key feature of the generating process. The design of the ANOVA models consisted of all factors that were relevant to the data generation process. Irrespective of the relationship between the data generation and data

analysis models, each cell in the ANOVA design consisted of 30,000 observations (= 100 trials × 300 realized values).

The second outcome of interest was the marginal distribution of PPP-values. The shape of the PPP-value distribution is hypothesized to depend on the relationship between the data and the model. Given alignment between the data generation and analysis models, PPP-values are expected to be uniform throughout the range of PPP-values (i.e., 0 to 1), as this pattern reflects the inability to distinguish between the observed data and data that are consistent with model (Hjort, Dahl, & Steinbakk, 2006). In contrast, when the model fails to capture a key process underlying the observed data, PPP-values are expected to be concentrated near the appropriate boundary that reflects underspecification of the model. Accordingly, in cases of underspecification, the PPP-values for the *CCC* and R^2 measures were expected to be close to one while PPP-values for the *SGDDM* functions and the *LR* were expected to cluster close to zero. These hypotheses were evaluated graphically. From the applied researcher's perspective, the distribution of PPP-values provides a very broad sense of the types of decisions that are likely to be made about a model in the absence of knowledge about underlying but unknown features about the data (e.g., quadratic strength, residual variance size).

The proportion of extreme PPP-values, which is closely related to the marginal posterior distributions of PPP-values, was the third outcome investigated in the current work. The inclusion of this outcome has two key advantages that derive from constructing the reference distribution. First, PPP-values involve comparing the realized values of discrepancy functions to values that could have been observed given the model (or model comparison) at hand (i.e., the posterior predictive values). Second, in a

simulation context, the proportion of extreme PPP-values are the Bayesian analogue of estimates of empirical Type I error rates (in null conditions) and power (in non-null conditions). The relationship between the manipulated variables and the proportion of extreme PPP-values were explored graphically.

Defining Extreme PPP-Values. In the case of the *SGDDM* functions (both conditional and marginal), the *LR*, and the *LRT*, larger values are indicative of greater data-model misfit, and accordingly, were deemed extreme if 5% or fewer of the posterior predictive values were *smaller* than the corresponding realized values. In contrast, smaller values of the *CCC* and R^2 measures (both conditional and marginal) are indicative of a weak association between data and the model-implied values. Accordingly, the *CCC* and R^2 measures were deemed extreme if 95% or more of the posterior predictive values were *larger* than the corresponding realized values.

CHAPTER 4: Results

ANOVA Results for Realized Values of Measures of Absolute Fit

This section presents the results from the series of factorial ANOVA models for identifying relationships between the manipulated variables and the realized values of the seven discrepancy functions for assessing the model-fit in the absolute sense. In terms of organization, the various subsections represent some relationship between the data generation and data analysis models. The first subsection describes the results for the three situations in which the data generation and data analysis models were aligned. Then, the three subsections that follow present the results for cases in which the data analysis model fails to capture one or more features of the data generation model. Features of the ANOVA designs are also described in greater detail in the respective subsections. Owing to the large number of trials in each design cell, key attention was given to the effect sizes (as measured by partial η^2) for each factor included in the design; complex relationships are presented graphically to facilitate interpretation.

Match Between Data Analysis and Data Generation Model. The collection of results presented in this subsection represent a match between the data generation and analysis models. Recall that the design of the simulation study resulted in three broad types of data generation models: a GCM with a linear functional form (M1), a GCM with a quadratic functional form that was equal in strength for all individuals (M2), and a GCM with a quadratic functional form that varied across individuals (M3). In what follows, the results for each of these models are described in turn. Owing to the similarity of the results, the description of patterns in the results follows the description of each design.

Table 2 presents the effect sizes of each design factor on the realized values for each discrepancy function of absolute fit for data generated and analyzed as a GCM with a linear functional form. With the realized values associated with M1 as the outcome, the factors in the design included sample size (denoted N, 3 levels), the shape of the distribution (denoted DS, 2 levels), and reliability of the observed measures as captured by the size of the residual variance (denoted RV, 2 levels). These three factors were fully crossed to yield 12 design cells.

Table 2

Effect Sizes of Each Effect on the Realized Values for Discrepancy Functions of Absolute Fit for Data Generated and Analyzed as a Linear Growth Curve Model

	Conditional			Marginal			
Effect	*CCC*	R^2	*SGDDM*	*CCC*	R^2	*SGDDM*	*LR*
DS	< .001	< .001	.044	< .001	< .001	.010	.057
N	< .001	.012	.194	< .001	< .001	< .001	.007
RV	.451	.456	< .001	.004	.005	.486	< .001
N × DS	< .001	< .001	.003	< .001	< .001	.003	.003
N × RV	.001	.001	< .001	< .001	< .001	.002	< .001
RV × DS	.007	.007	< .001	< .001	< .001	< .001	< .001
N × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	.001

Note. DS = distribution shape, N = sample size, RV = residual variance.

Table 3 shows the effect sizes of each design factor on the realized values for each discrepancy function of absolute fit for data generated and analyzed as a GCM that exhibits a quadratic functional form that holds across all individuals. In addition to all of the factors in the design relevant to the linear GCM, the strength of the quadratic effect (as operationalized by a more negative quadratic mean) was also included (denoted QM, 2 levels). The outcome in the design was the realized values associated with M2. All factors were fully crossed to yield 24 design cells.

Table 3

Effect Sizes of Each Effect on the Realized Values for Discrepancy Functions of Absolute Fit for Data Generated and Analyzed as a Growth Curve Model with a Quadratic Functional Form that Does Not Vary Over Individuals

	Conditional			Marginal			
Effect	*CCC*	R^2	*SGDDM*	*CCC*	R^2	*SGDDM*	*LR*
DS	< .001	< .001	.036	< .001	< .001	.004	.046
N	< .001	.003	.096	< .001	.001	< .001	.002
QM	.002	.007	< .001	.184	.185	< .001	< .001
RV	.233	.255	< .001	.006	.010	.338	.001
N × DS	< .001	< .001	.005	< .001	< .001	< .001	.001
N × QM	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × RV	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × RV	.021	.019	< .001	< .001	.002	< .001	< .001
RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	.004
N × QM × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × RV	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	.002
QM × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001

Note. DS = distribution shape, N = sample size, QE = quadratic mean, RV = residual variance.

Table 4

Effect Sizes of Each Effect on the Realized Values for Discrepancy Functions of Absolute Fit for Data Generated and Analyzed as a Growth Curve Model with a Quadratic Functional Form that Does Vary Over Individuals

	Conditional			Marginal			
Effect	*CCC*	R^2	*SGDDM*	*CCC*	R^2	*SGDDM*	*LR*
DS	< .001	< .001	.005	< .001	< .001	.002	.007
N	< .001	.003	.062	< .001	< .001	< .001	.002
QM	.001	.004	< .001	.105	.110	< .001	< .001
QV	< .001	.001	.001	.014	.018	.057	.015
RV	.165	.175	< .001	.002	.003	.158	.001
N × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QV	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × RV	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × QV	< .001	< .001	< .001	.001	.003	< .001	< .001
QM × RV	.014	.011	< .001	.001	.002	< .001	< .001
QV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QV × RV	.007	.006	< .001	< .001	< .001	.007	< .001
RV × DS	.003	.003	< .001	< .001	< .001	< .001	< .001
N × QM × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × QV	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × RV	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QV × RV	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × QV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × QV × RV	.002	.001	< .001	< .001	< .001	< .001	< .001
QM × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QV × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × QV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × QV × RV	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QV × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × QV × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × QV × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001

Note. DS = distribution shape, N = sample size, QM = quadratic mean, QV = quadratic variance, RV = residual variance.

Table 4 shows the effect sizes of design factor in the ANOVA on the realized values for each discrepancy function of absolute fit for data generated and analyzed as a GCM with a quadratic functional form that varies across all individuals. In addition to the factors in the design included all of those relevant to the design of the ANOVAs models for the linear and non-varying quadratic models, the size of the quadratic variance was also included (denoted QV, 2 levels). With the realized values associated with M3 as the outcome, all design factors were fully crossed to yield 48 design cells.

In general, when the data analysis model captured all aspects germane to the data generation model, the patterns of results for most of the discrepancy functions were similar irrespective of the model that was used to generate data. For instance, although the effect diminished with increased complexity of the data generation model, the average realized values for several discrepancy functions tended towards values that were indicative of greater discrepancies between the data and the model with increased residual variation. In the case of the CCC_C and R_C^2, larger residual variances led to smaller realized values on average.

The realized values of the $SGDDM_C$ also exhibited a similar pattern of results across the types of growth curve models generated in this work. The shape of distribution for the observed outcomes and sample size exhibited main effects on the realized values of the $SGDDM_C$. On average, the realized values of the $SGDDM_C$ (a) were larger when outcomes were positively skewed rather than normally distributed and (b) in that values became smaller with increased sample size.

The manipulated factors were generally not found to impact the average realized values for the LR, CCC_M, and R_M^2 functions, particularly for the data generated to follow

a linear GCM. For the latter two functions, the key exception was the presence of the main effect associated with the strength of the quadratic relationship with time. Irrespective of whether the quadratic effect varied across individuals, the realized values of the CCC_M and R_M^2 functions became larger with increased strength of the quadratic effect, on average.

For the three types of data-generation models, the realized values of the $SGDDM_M$ were most strongly impacted by the main effect of the residual variance; in the case of data generated with quadratic form that varies over individuals, the realized values were also impacted the size of the quadratic variance. The direction of both effects was such that such that larger values of the residual or quadratic variance were associated with smaller values of the realized values for the $SGDDM_M$, on average. This pattern of results suggests that the data analysis model, which matches the data generation model in this case, tends to fit the data better, as represented by smaller realized values, with increasing values of the residual variance or quadratic variance, on average.

Underspecified Marginal Mean Structure. The results presented in this section pertain to the situation in which a GCM with a linear functional form was applied to data that exhibited a quadratic relationship with the passage of time. The quadratic effect was generated as equal across individuals. Although it is the case that failing to capture the quadratic effect would yield an underspecification for both the marginal and conditional mean structures, including a non-varying quadratic effect would adequately capture the generating process for both mean structures. Accordingly, the relationship between the data analysis and data generation models is foundationally an underspecification of the marginal mean structure.

The design of the ANOVA included all features relevant to the generation of data that follow a GCM with a quadratic functional form but no quadratic variance component (i.e., M2). With the realized values associated with the linear GCM (i.e., M1) as the outcome, the factors in the model included sample size (3 levels), the strength of the quadratic effect (2 levels), the reliability of the measures as captured by the size of the residual variance (2 levels), and the shape of the distribution for the observed variables (2 levels). All factors were fully crossed to yield 24 design cells.

Table 5 shows the effect sizes (as measured by partial η^2) for each design factor on the realized values for each discrepancy functions designed for evaluating model fit in absolute terms. The *CCC* (both conditional and marginal), R^2 (both conditional and marginal), $SGDDM_C$, and *LR* were impacted by the main effect associated with the strength of the quadratic mean. For each of these functions, increasing the strength of the quadratic mean resulted in (a) smaller realized values of the *CCC* and R^2 functions and (b) larger realized values of the $SGDDM_C$ and *LR*, on average. Substantively, the direction of this effect suggests that it becomes easier to detect underspecification of the marginal mean structure to the extent that the strength of the quadratic effect is increased. Notably, the impact of the quadratic strength on the realized values of the $SGDDM_C$ and *LR* was moderated by other factors. These interactive relationships are presented graphically below to facilitate interpretation.

The residual variance main effect impacted the CCC_C, R_C^2, and $SGDDM_M$. Increased values of the residual variance were associated with smaller values for these functions, on average. In the case of the CCC_C and R_C^2, the decreased values would be interpreted as a weaker correspondence between the data and model expectations. For the

$SGDDM_M$, decreased realized values would be interpreted as less unaccounted for conditional associations given the model.

Table 5

Effect Sizes of Each Effect on the Realized Values for Discrepancy Functions of Absolute Fit for Data Generated to Follow a Growth Curve Model with a Quadratic Functional Form that was Equal Across Individuals but Analyzed as a Linear Growth Curve Model

	Conditional			Marginal			
Effect	*CCC*	R^2	*SGDDM*	*CCC*	R^2	*SGDDM*	*LR*
DS	< .001	< .001	.004	< .001	< .001	.004	< .001
N	< .001	.001	.009	< .001	.002	< .001	.030
QM	.246	.322	.722	.346	.313	< .001	.261
RV	.081	.117	.003	.006	.010	.338	.002
N × DS	< .001	< .001	.001	< .001	< .001	< .001	< .001
N × QM	< .001	< .001	.003	< .001	< .001	< .001	.621
N × RV	< .001	< .001	.001	< .001	< .001	< .001	.010
QM × DS	< .001	< .001	.003	< .001	< .001	< .001	< .001
QM × RV	.004	< .001	.398	.002	.004	< .001	.070
RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × DS	< .001	< .001	.002	< .001	< .001	< .001	.002
N × QM × RV	< .001	< .001	.002	< .001	< .001	< .001	.245
N × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001

Note. DS = distribution shape, N = sample size, QM = quadratic mean, RV = residual variance.

Figure 8 shows the interaction between the strength of the quadratic effect and the residual variance size on the realized values of the $SGDDM_C$. The average realized values of the $SGDDM_C$ (shown along the vertical axis) plotted against the levels associated with the size of the residual variance (shown along the horizontal axis). Within the levels associated with size of the residual variance, the results for small and large quadratic

effects (i.e., the value of quadratic mean) are represented by gray and black bars, respectively.

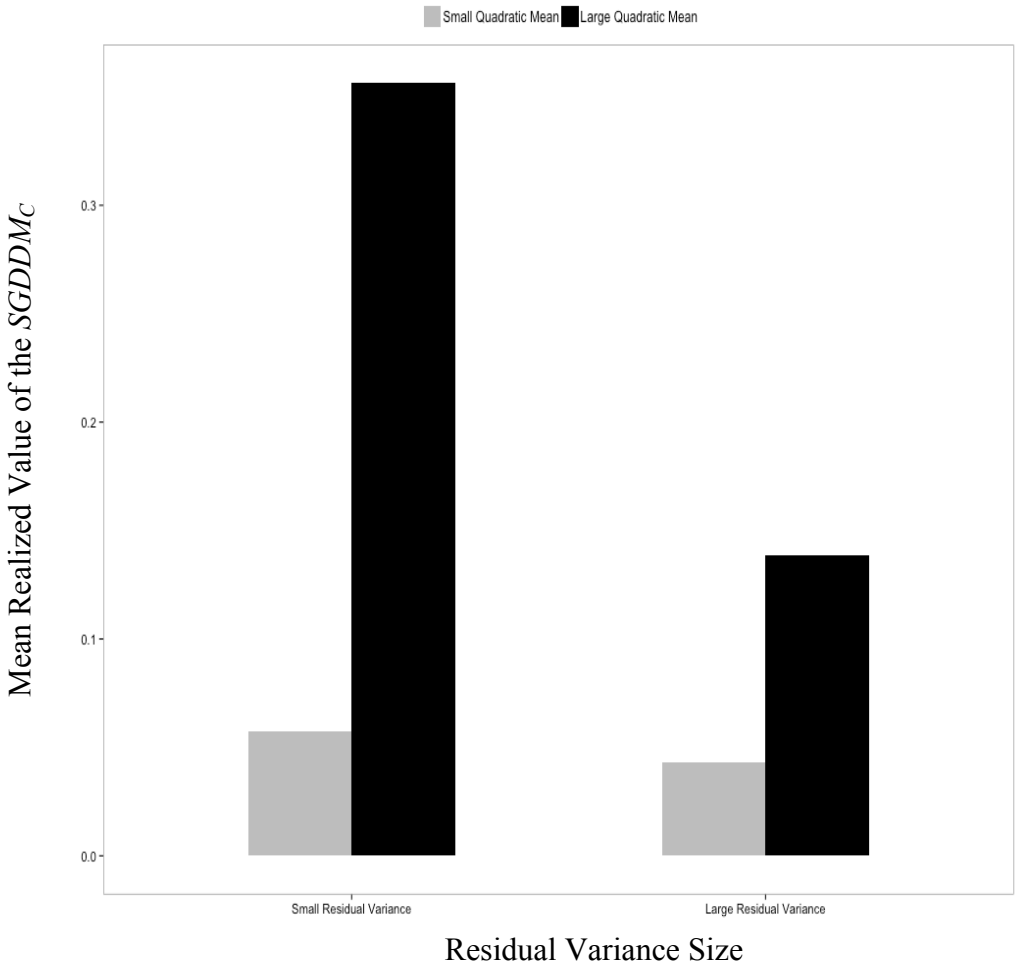

Figure 8. Interaction effect between the quadratic mean strength and the size of the residual variance on the realized values of the $SGDDM_C$ when the data analysis model underspecifies the functional form of the marginal mean structure.

Looking within any one level of the residual variance size, the positive relationship between the strength of the quadratic effect and the realized values is seen. That is, the realized values of the $SGDDM_C$ associated with the application of a linear GCM to data that follow GCM with non-varying quadratic effect were more indicative of data-model misfit when the quadratic effect was large rather than small. Although this

pattern held across the levels of the residual variance, the effect associated with the quadratic strength was stronger for data with a small residual variance than for data generated with a large residual variance.

The realized values of the *LR* were impacted by an interaction between sample size, the strength of the quadratic effect, and the amount of residual variation. Figure 9 was constructed to visualize the nature of the three-way interaction on the realized values of the *LR*. The figure includes two panels that represent the strength of the quadratic effect. Looking within panels, the average realized value of the *LR* (shown along the vertical axis) for each level of the residual variance size (gray and black bars represent the small and large residual variance conditions, respectively) is plotted against sample size (shown along the horizontal axis). For simplicity, the midpoints of the random uniform distributions for drawing values of sample size serve as the labels of sample size.

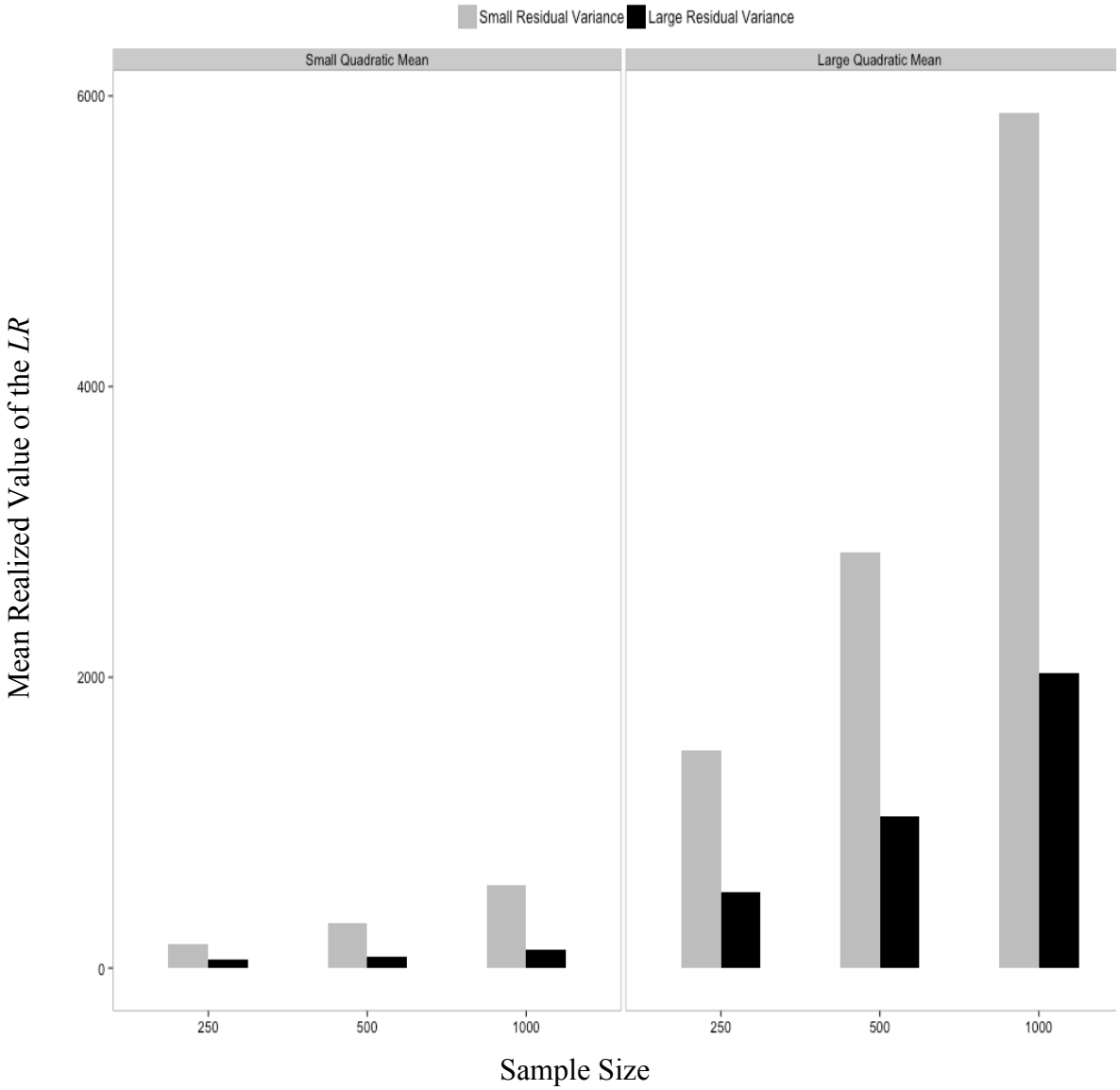

Figure 9. Interaction effect between sample size, the strength of the quadratic effect, and the size of the residual variance on the average realized values of the *LR* when the data analysis model underspecifies the functional form of the marginal mean structure.

With the sample size held constant, increasing the strength of the quadratic effect (looking across panels) resulted in larger values of the *LR*, on average. Moreover, holding the sample size constant, increasing the strength of the quadratic effect had a larger impact on the realized values of the *LR* for data generated with a small residual variance

than for data generated with a large residual variance. This two-way interaction between the strength of the quadratic mean and the size of the residual variance was intensified with increasing sample size, hence the three-way interactive effect of these variables on the realized values of the *LR*.

Underspecified Conditional Mean Structure. The results presented in this section are based on the situation in which a GCM with quadratic form that is equal across individuals was applied to data that follow a GCM with a quadratic form that varies across individuals. Accordingly, the results are based on the situation in which the data analysis model underspecifies the conditional mean structure of the data generation model. With the realized values associated with the non-varying quadratic GCM (i.e., M2) as the outcome, the ANOVA model included all factors relevant to the generation of data that follow GCM with quadratic functional form that varies across individuals. The factors in ANOVA design included sample size (3 levels), the strength of the quadratic effect (2 levels), the size of the quadratic variance (2 levels), the size of the residual variance (2 levels), and the shape of the distribution of the observed variables (2 levels). All factors were fully crossed to yield 48 design cells. Table 6 presents the effect size (as measured by partial η^2) for each effect on the realized values.

Table 6

Values of Partial η^2 for each Effect on the Realized Values for Discrepancy Functions of Absolute Fit when the Conditional Mean Structure was Underspecified

	Conditional			Marginal			
Effect	*CCC*	R^2	*SGDDM*	*CCC*	R^2	*SGDDM*	*LR*
DS	< .001	< .001	.004	< .001	< .001	.002	< .001
N	< .001	.003	.007	< .001	< .001	< .001	.001
QM	.002	.004	< .001	.106	.110	< .001	< .001
QV	.003	.002	.405	.014	.018	.057	.093
RV	.151	.165	.005	.002	.003	.158	< .001
N × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QV	< .001	< .001	.005	< .001	< .001	< .001	.369
N × RV	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × QV	< .001	< .001	< .001	.001	.003	< .001	< .001
QM × RV	.013	.011	< .001	.001	.002	< .001	< .001
QV × DS	< .001	< .001	.002	< .001	< .001	< .001	< .001
QV × RV	.007	.007	.166	< .001	< .001	.007	.031
RV × DS	.003	.003	< .001	< .001	< .001	< .001	< .001
N × QM × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × QV	< .001	< .001	.002	< .001	< .001	< .001	.005
N × QM × RV	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QV × DS	< .001	< .001	< .001	< .001	< .001	< .001	.008
N × QV × RV	< .001	< .001	< .001	< .001	< .001	< .001	.160
N × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × QV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × QV × RV	.002	.001	< .001	< .001	< .001	< .001	< .001
QM × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QV × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × QV × DS	< .001	< .001	< .001	< .001	< .001	< .001	.003
N × QM × QV × RV	< .001	< .001	.001	< .001	< .001	< .001	.003
N × QM × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QV × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	.004
QM × QV × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × QV × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	.002

Note. DS = distribution shape, N = sample size, QM = quadratic mean strength, QV = quadratic variance, RV = residual variance.

Excepting to the $SGDDM_C$ and the *LR*, interactive effects had little bearing on the variability of the realized values for the absolute discrepancy functions. The average realized values of the CCC_C, R_C^2, and $SGDDM_M$ functions were impacted by the residual variance main effect. For each of these functions, decreasing the reliability of the observed measures (i.e., increasing the residual variance) resulted in smaller realized values, on average. To a lesser extent, the realized values of the $SGDDM_M$ were also influenced by the quadratic variance main effect such that higher realized values were observed with an increased quadratic variance, on average.

Although the realized values of the $SGDDM_C$ were impacted by the main effect associated with the quadratic variance, the influence was moderated by the amount of residual variation. Figure 10 displays the means of the $SGDDM_C$ for all combinations of these two factors. The structure of the graphic is generally identical to that described for Figure 8. The caveat is that the gray and black bars represent conditions generated with small and large between-person variation in the strength of the quadratic effect rather than representing the strength of the marginal quadratic effect.

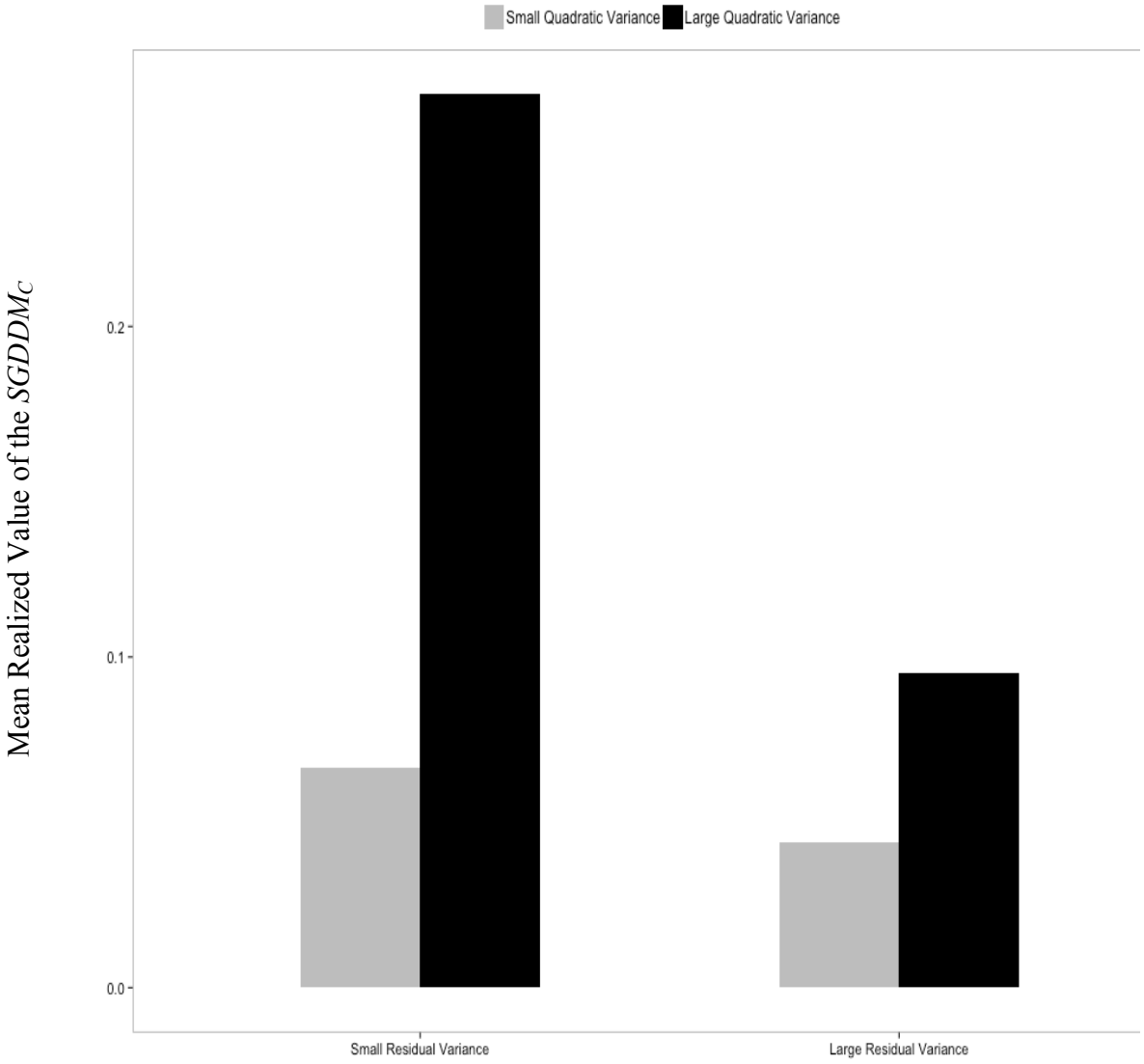

Figure 10. Interaction effect between the size of the quadratic variance and the size of the residual variance on the realized values of the $SGDDM_C$ when the data analysis model underspecifies the functional form of the conditional mean structure.

Looking within the levels of the residual variance factor, the difference in the heights of the bars indicates realized values of the $SGDDM_C$ were larger for data generated with greater between-person variation in the strength of the quadratic effect. Holding the degree of between-person variation constant (i.e., focusing on either the gray or black bar) and looking across the levels of the residual factor, it is clear that less

reliable outcomes (i.e., increasing residual variation) yield smaller realized average values of the $SGDDM_C$. The source of the two-way interaction between the quadratic variance and the residual variance factors on the realized of the $SGDDM_C$ was the result of large quadratic variance conditions exhibiting a greater decline with increasing residual variation compared to data generated with a small quadratic variance.

Although a strong two-way interaction between sample size and the size of the quadratic variance influenced the realized values of the *LR*, it was in turn moderated by the size of the residual variance. Figure 11 displays the average realized values for each combination of these factors. Specifically, the average realized value (shown along the vertical axis) is plotted against sample size (shown as increasing along the horizontal axis) for each level of the residual variance size (small and large residual variance conditions represented by gray and black bars, respectively). The impact of increasing the size of the quadratic variance is seen looking across panels.

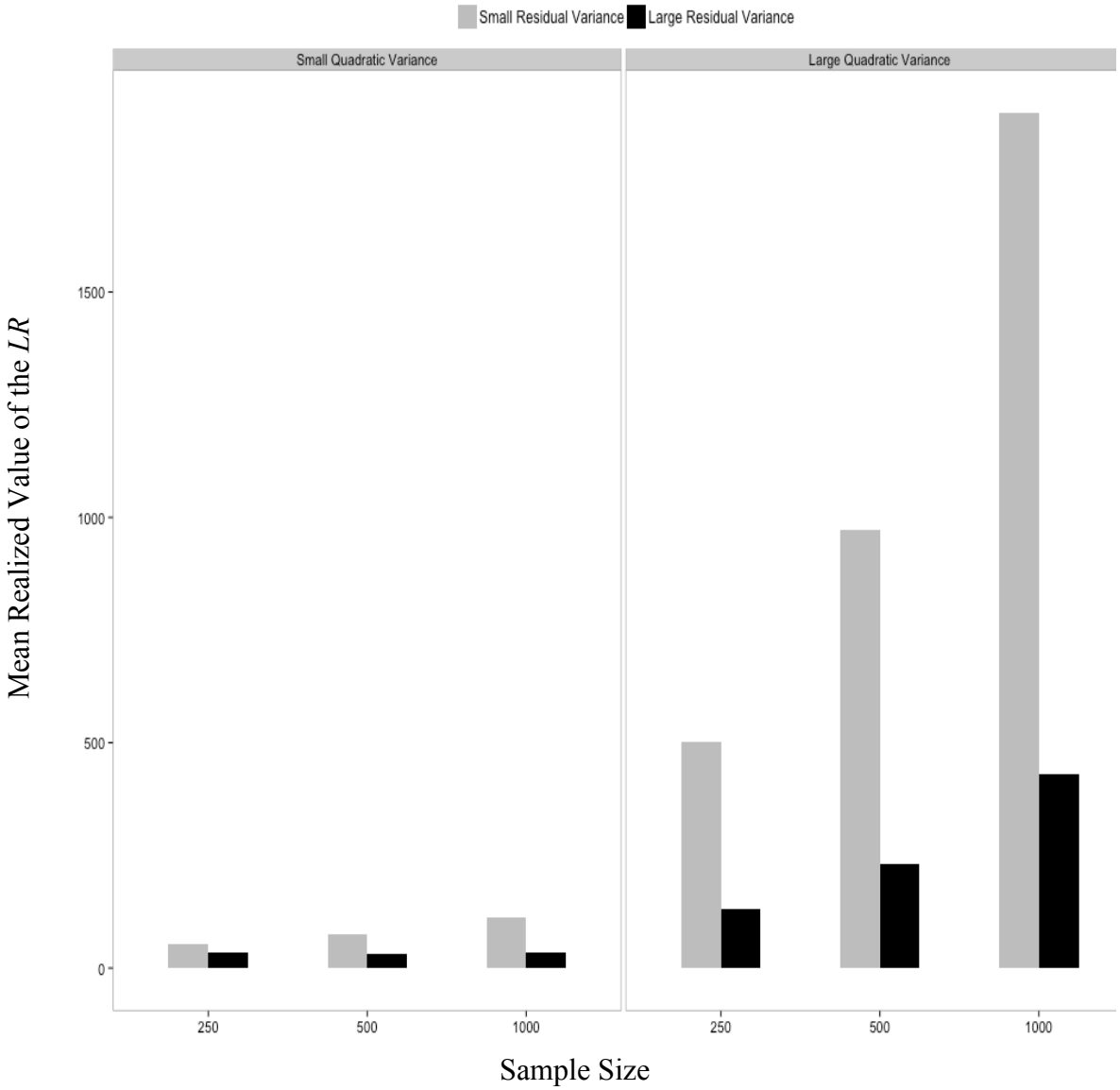

Figure 11. Interaction effect between sample size, the degree of between-person variation (i.e., quadratic variance), and the size of the residual variance on the average realized values of the *LR* when the data analysis model underspecifies the functional form of the conditional mean structure.

Irrespective of the size of the quadratic variance or sample size, the realized values of the *LR* became smaller with increased residual variation. Larger average values of the *LR* were observed with increasing sample size irrespective of the amount of residual variation or the amount of quadratic variation. Taken in combination, the impact

of increasing the amount of residual variation was accentuated with increased sample size. This interplay between sample size and the size of the residual variance was in turn made larger with increased variation in the functional form of quadratic effect between individuals.

Underspecified Marginal and Conditional Mean Structures. The results presented in this section are based on the situation in which a GCM with linear function form was applied to data that follow a GCM with a quadratic form that varies across individuals. Accordingly, the results are based on the situation in which the data analysis model underspecifies both the marginal and conditional mean structures of the data generation model. With the realized values associated with the linear GCM (i.e., M1) as the outcome, the ANOVA model included all factors relevant to the generation of data that follow GCM with between-person variation in the strength of the quadratic effect (i.e., M3). The factors in ANOVA design included sample size (3 levels), the strength of the quadratic effect (2 levels), the size of the quadratic variance (2 levels), the size of the residual variance (2 levels), and the shape of the distribution of the observed variables (2 levels). All factors were fully crossed to yield 48 design cells. Table 7 presents the effect size (as measured by partial η^2) for each effect on the realized values.

Table 7

Values of Partial η^2 for each Effect on the Realized Values for Discrepancy Functions of Absolute Fit when the Marginal and Conditional Mean Structures were Underspecified

	Conditional			Marginal			
Effect	*CCC*	R^2	*SGDDM*	*CCC*	R^2	*SGDDM*	*LR*
DS	< .001	< .001	.002	< .001	< .001	.002	< .001
N	< .001	.001	.004	< .001	< .001	< .001	.019
QM	.165	.218	.497	.218	.195	< .001	.139
QV	< .001	< .001	.305	.013	.018	.057	.015
RV	.049	.070	.007	.002	.003	.158	.001
N × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM	< .001	< .001	.003	< .001	< .001	< .001	.443
N × QV	< .001	< .001	.003	< .001	< .001	< .001	.076
N × RV	< .001	< .001	< .001	< .001	< .001	< .001	.007
QM × DS	< .001	< .001	.002	< .001	< .001	< .001	< .001
QM × QV	.009	.014	.231	.005	.008	< .001	.007
QM × RV	.007	.002	.169	.002	.002	< .001	.028
QV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QV × RV	.002	.003	.110	< .001	< .001	.007	.004
RV × DS	< .001	.001	< .001	< .001	< .001	< .001	< .001
N × QM × DS	< .001	< .001	< .001	< .001	< .001	< .001	.003
N × QM × QV	< .001	< .001	.003	< .001	< .001	< .001	.053
N × QM × RV	.002	.002	< .001	< .001	< .001	< .001	.140
N × QV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QV × RV	< .001	< .001	< .001	< .001	< .001	< .001	.025
N × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × QV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × QV × RV	.002	< .001	.093	< .001	< .001	< .001	.002
QM × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QV × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × QV × DS	< .001	< .001	< .001	.001	< .001	< .001	.002
N × QM × QV × RV	.001	.001	.001	< .001	< .001	< .001	.024
N × QM × RV × DS	.002	.002	< .001	< .001	< .001	< .001	.001
N × QV × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
QM × QV × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	< .001
N × QM × QV × RV × DS	< .001	< .001	< .001	< .001	< .001	< .001	.002

Note. DS = distribution shape, N = sample size, QM = quadratic mean strength, RV = residual variance.

Excepting to the $SGDDM_C$ and *LR*, interactive effects had little bearing on the behavior of the discrepancy functions when the marginal and conditional mean structures were both underspecified. Despite the presence of interactive effects, the patterns in the

realized values of the *LR* were similar to the situation in which only the marginal mean structure was underspecified, and accordingly, are not further discussed here. The *CCC* and R^2 functions (both marginal and conditional) were impacted by the strength of the quadratic mean such that realized values became smaller (i.e., reflecting greater underspecification of the functional form), on average, with increasing strength of the quadratic effect. The average realized values of the CCC_C, R_C^2, and $SGDDM_M$ functions were impacted and became smaller with increased residual variation. However, unlike the CCC_C and R_C^2, decreasing realized values of the $SGDDM_M$ are indicative of better fit between the data and the model.

A three-way interaction involving the strength of the quadratic mean, size of the quadratic variance, and size of the residual variance influenced the realized values of the $SGDDM_C$. Figure 12 displays the mean realized values of the $SGDDM_C$ for each combination of these three factors. Within panels, the mean realized values of the $SGDDM_C$ (shown along the vertical axis) for each level of the residual variance factor (gray and black bars represent the small and large residual variance conditions, respectively) are plotted against the strength of the quadratic mean (shown as increasing from left to right along the horizontal axis). The impact of increasing the size of the quadratic variance is seen moving from left to right across panels.

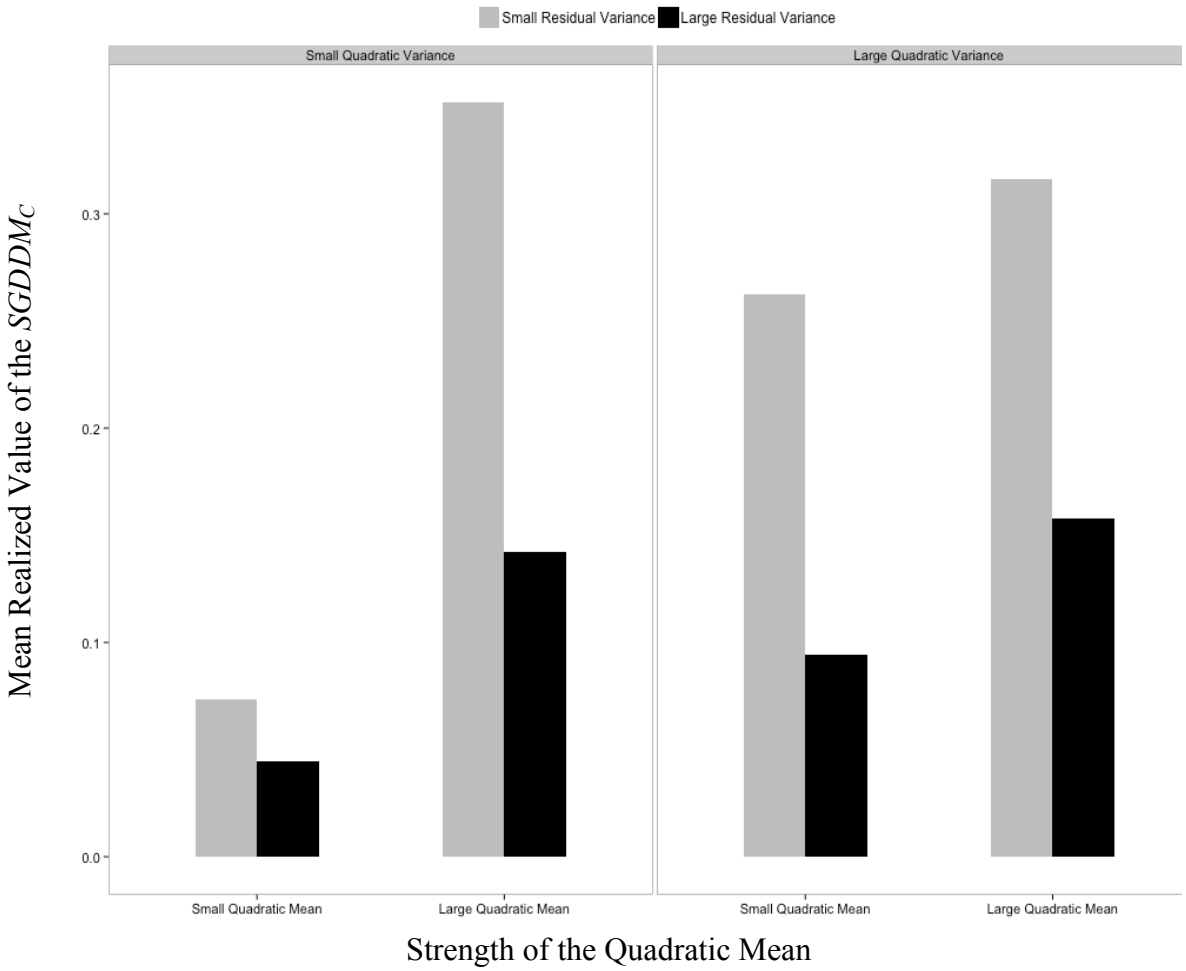

Figure 12. Interaction effect between strength of the quadratic mean, the size of the quadratic variance, and the size of the residual variance on the realized values of the $SGDDM_C$ when the data analysis model underspecifies the functional form of the marginal and conditional mean structures.

Looking within panels, it is clear that increasing the strength of the quadratic mean resulted in larger realized values of the $SGDDM_C$, on average, irrespective of the amount of residual variation in the observed measures or the size of the quadratic variance. Looking across panels, increasing the size of the quadratic variance resulted in an increased realized values of the $SGDDM_C$, on average. In combining these two factors, the realized values of the $SGDDM_C$ were (a) more strongly impacted by the strength of the quadratic mean when the quadratic variance was small and (b) more

strongly impacted by the size of the quadratic variance when the quadratic mean was small. Although these patterns generally held across the levels of the residual variance, increasing the size of the quadratic variance had a smaller impact in the presence of a large quadratic effect when the residual variance was large.

Distribution of PPP-Values for Absolute Measures

For the purposes of the current work, evaluating the realized values allows for comparing and contrasting the results to prior research. However, the results of a PPMC analysis are often summarized by PPP-values, which are determined by comparing the realized and posterior predictive values of a discrepancy function. Using graphical representations, this section presents the marginal distributions of PPP-values for each of the discrepancy functions designed to evaluate absolute model fit. Key attention is given to the relationship between the behavior of the PPP-values across the different relationships between the data analysis and generation models.

Figure 13 displays the marginal density of PPP-values for the conditional and marginal versions of the *CCC*, R^2, and the $SGDDM_C$ for different relationships between the data generation and analysis models. Owing to the presence of particularly concentrated densities that rendered the shape of the PPP-values distributions for other functions unclear, the densities of PPP-values for the $SGDDM_M$ and *LR* are plotted separately in in Figures 14 and 15, respectively. Since it was necessary to separate the results for the functions, the description of results follows the presentation of each of these figures.

Figure 13 consists of 20 panels organized into a 4 × 5 matrix. Each row represents one of the four possible relationships between the data generation and analysis models:

matched, underspecified functional form of the marginal mean structure, underspecified functional form of the conditional mean structure, and underspecified form of the marginal and conditional mean structures. The columns represent the six discrepancy functions of interest. The first three columns show the results for the conditional versions of the *CCC*, R^2, and *SGDDM*; the last two columns show the results for the CCC_M and R^2_M. Looking within panels, the relative likelihood (i.e., density) of each PPP-value is shown, and accordingly, reflects the distribution of PPP-values.

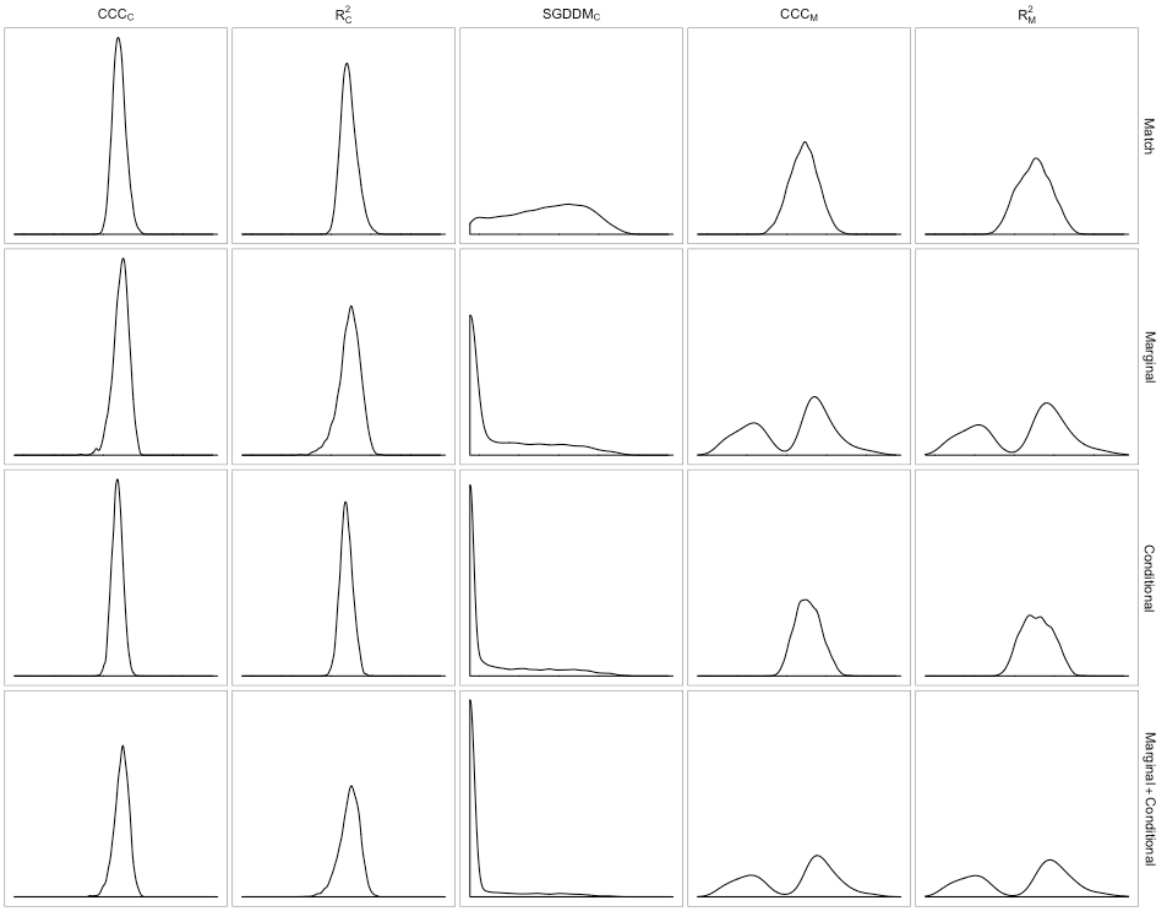

Figure 13. Plots of the marginal (i.e., over all manipulated factors) densities for the conditional and marginal versions of the *CCC*, R^2, and $SGDDM_C$ discrepancy functions for different associations between the data generation and analysis models.

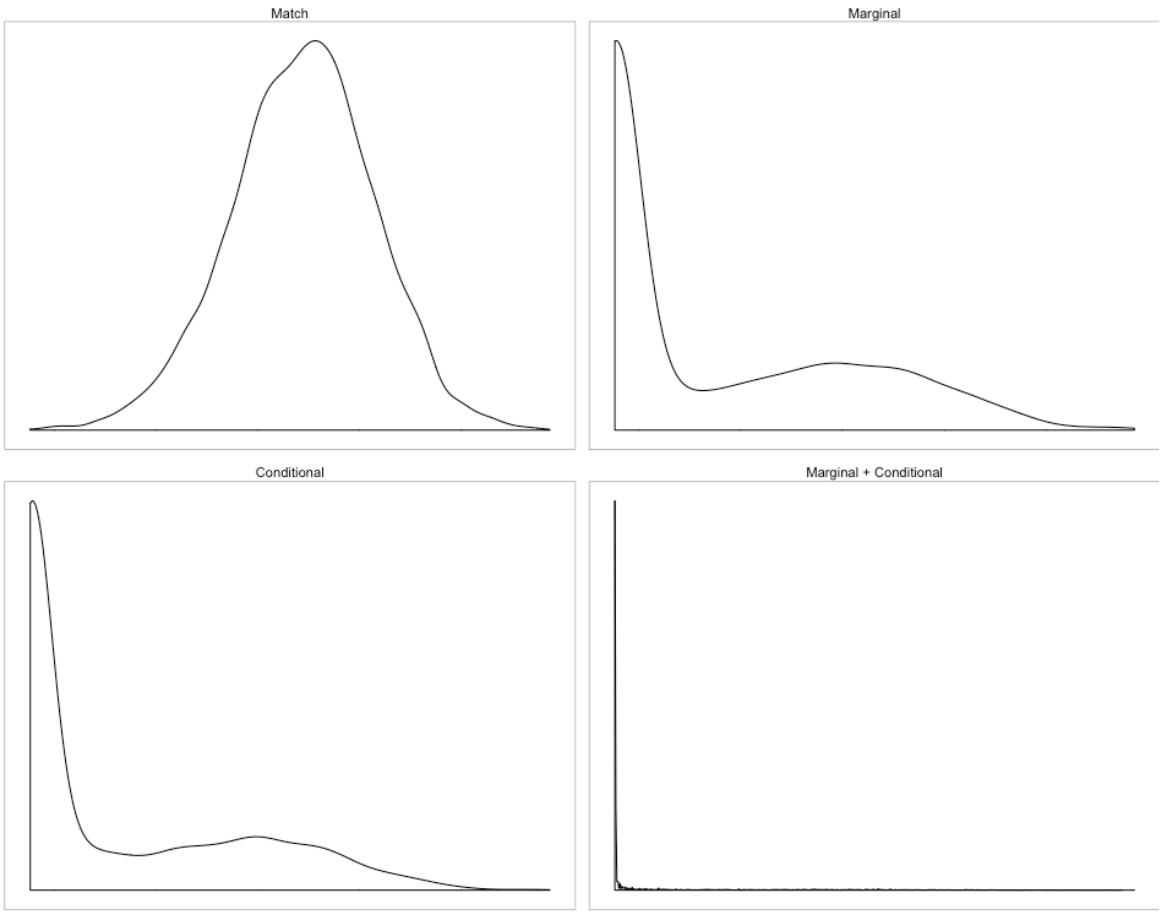

Figure 14. Plots of the marginal (i.e., over all manipulated factors) densities for the $SGDDM_M$ discrepancy functions for different associations between the data generation and analysis models.

Figures 14 and 15 display the densities of PPP-values for the $SGDDM_M$ and LR, respectively. These figures are organized into a 2 × 2 matrix such that each panel represents one of the four relationships between the data generation and analysis models. The top left panel displays the results for all situations in which the data generation and analysis models were aligned. All other panels show the results given some underspecification of the data analysis model in relationship to the process used to generate data.

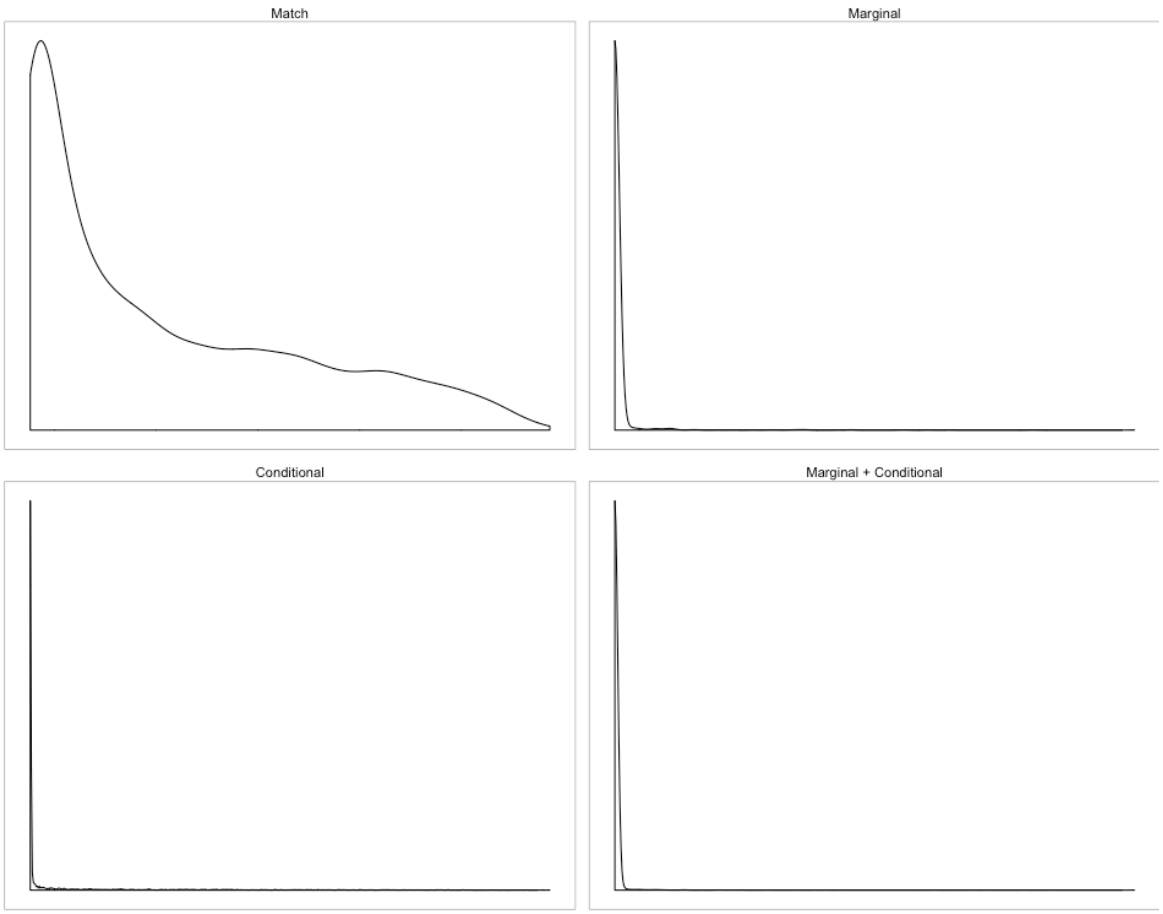

Figure 15. Plots of the marginal (i.e., over all manipulated factors) densities for the *LR* discrepancy functions for different associations between the data generation and analysis models.

Match Between Data Generation and Analysis Models. Focusing on the first row in Figure 13 and the upper left panels in Figures 14 and 15, which shows results for the situation in which the data generation and analysis models were aligned, it is desired for PPP-values to be uniformly distributed (Hjort & Dahl, 2006). Although the PPP values for the $SGDDM_C$ were unlikely to be higher than .75 in null situations, the distribution was quite uniform for smaller PPP-values.

In null situations, the mass of the distributions of PPP-values for the conditional *CCC* and R^2 functions (first two columns in the top row of Figure 13), marginal *CCC* and

R^2 functions (last two columns top row of Figure 13), and $SGDDM_M$ (upper left panel in Figure 14) were centered around 0.50. Although the distributions of PPP-values for these functions were not uniform, the location of the distributional mass indicates that these functions are quite unlikely to detect misspecification when the data generation and analysis models are in fact aligned. In contrast to the other measures of absolute fit, the mass of the PPP-values for the *LR* (upper left panel in Figure 15) is more concentrated at lower PPP-values. This finding suggests that the *LR* may exhibit a tendency to indicate model inadequacy when the analysis model is actually aligned with the process used to generate the data.

Underspecification of the Marginal Mean Structure. When the marginal mean structure was underspecified, the distributions of PPP-values of the CCC_C and R_C^2 (second row in the first two columns of Figure 13) were tightly centered around 0.50; this finding suggests that these functions are quite unlikely to detect underspecification of the marginal mean structure. In contrast, the masses of the distributions for the $SGDDM_C$ (second row in the third column of Figure 13), $SGDDM_M$ (upper right quadrant in Figure 14), and *LR* (upper right quadrant in Figure 15) were clustered at the lower end of the PPP-scale. However, the tail of the distributions of PPP-values for the *SGDDM* functions was thicker than that of the *LR*, particularly the $SGDDM_M$, which had a secondary mode. Further analyses revealed that the additional thickness in the tails of the *SGDDM* functions corresponded to the manipulation of the quadratic mean strength; the higher peak for the $SGDDM_M$ is associated with data generated with a stronger quadratic mean, and the tails are largely comprised of the PPP-values associated the conditions generated with a weaker quadratic mean. As evidenced by very distinctive peaks, a similar and

much stronger effect of increasing the quadratic mean was observed for the CCC_M and R_M^2 (second row in the last two columns of Figure 13) when the marginal mean structure was underspecified.

Underspecified Conditional Mean Structure. Given the situation of an underspecified conditional mean structure, the mass of the distribution of PPP-values for the CCC_C and R_C^2 (third row in the first two columns of Figure 13) were centered around 0.50. Although the distribution was less peaked than their respective conditional counterparts, the distributions of PPP-values of the CCC_M and R_M^2 (third row in the last two columns of Figure 13) were similarly located. In stark contrast to the *CCC* and R^2 functions, the distributions of PPP-values for the $SGDDM_C$ (third column in the third row of Figure 13), $SGDDM_M$ (bottom left quadrant of Figure 14), and the *LR* (bottom left quadrant of Figure 15) were clustered at lower PPP-values, particularly those consistent with underspecification of the data analysis model.

Underspecified Marginal and Conditional Mean Structures. Excepting to the $SGDDM_M$, the distributions of PPP-values when the conditional and marginal mean structures were underspecified are consistent with those described when only the latter mean structure was underspecified. In the case of the $SGDDM_M$, the distribution of PPP-values when either one of the mean structures was underspecified exhibited multimodality such that (a) the mass of the highest peak was closer to zero and (b) the mass of the smaller peak was closer to 0.50. This multimodality was no longer present in the distribution of PPP-values for the $SGDDM_M$ when both mean structures were underspecified. Rather, the distribution was more tightly clustered closer to PPP-values

of zero, and accordingly was quite sensitive to the joint underspecification of both mean structures.

Proportion of Extreme PPP-Values for Measures of Absolute Fit

As mentioned above, PPP-values summarize the results of a PPMC analysis for a given discrepancy function. In applied settings, a decision about the adequacy of the model is sometimes made on the basis of the PPP-value, which often involves appealing to some criterion that results in labeling the PPP-value as extreme or not extreme. Using the criterion described above for each discrepancy function, this section presents the proportion of extreme values across relevant design factors in both null and non-null situations. For null situations, the proportion of extreme PPP-values can be thought of as an empirical Type I error rate. For non-null situations, the proportion of extreme PPP-values can be thought as an empirical estimate of power.

Figures 16 to 21 display the proportions of extreme PPP-values for each of the discrepancy functions of absolute fit across the manipulated variables in the simulation design. The key difference across the figures is the relationship between the data generation and analysis models. Although some structural differences among the figures are dependent on the features that were used to generate data (explained in greater detail below), the within-panel structure is shared across the figures. For each figure, the within-panel structure is such that the proportion of extreme PPP-values is plotted against sample size. The three levels of sample size are represented by the midpoints of the respective random uniform distributions that were used to draw values.

Each line within panels corresponds to one of the seven discrepancy functions of absolute fit. Blue, red, and green lines represent the *CCC*, R^2, and *SGDDM* functions,

respectively. Solid and empty markers respectively represent the conditional and marginal versions of these functions. Finally, the black line represents the *LR* with empty markers. Owing to highly similar performances among discrepancy functions in some cases, a value was drawn from a random uniform distribution with a range of -0.035 to 0.035 and added to the observed result. The pattern of results as they appear a below are consistent with the results without jittering and allow the lines for similarly performing functions to be distinguished.

Match Between Data Analysis and Data Generation Model. Figures 16 through 18 show the proportion of extreme PPP-values for the three situations in which the data generation and analysis models were aligned. From the NHST perspective, the results shown in this section are analogous to Type I error rates. Owing to the similarity of the results across the three types of data generation models for null situations, a summary of the key patterns follow the presentation of all results for the proportion of extreme PPP-values. Figure 16 displays the results for data generated to follow a linear GCM. The effect of increasing the amount of residual variation is seen moving from left to right. The results for normally distributed outcomes are shown in the first row; and the analogous results for non-normally distributed outcomes are shown in the second row.

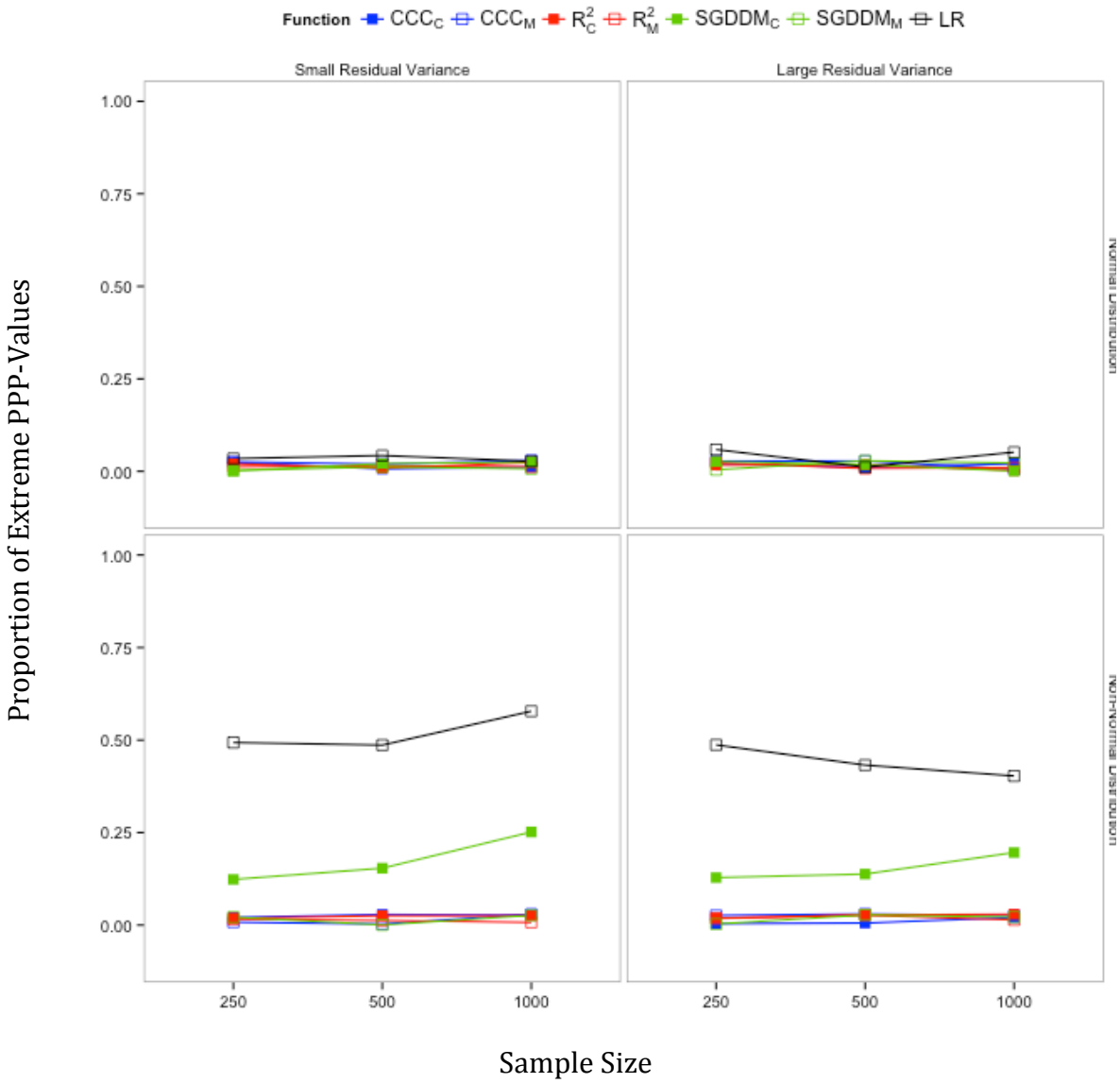

Figure 16. Proportion of extreme PPP-values for data generated and analyzed as a growth curve model with a linear functional form.

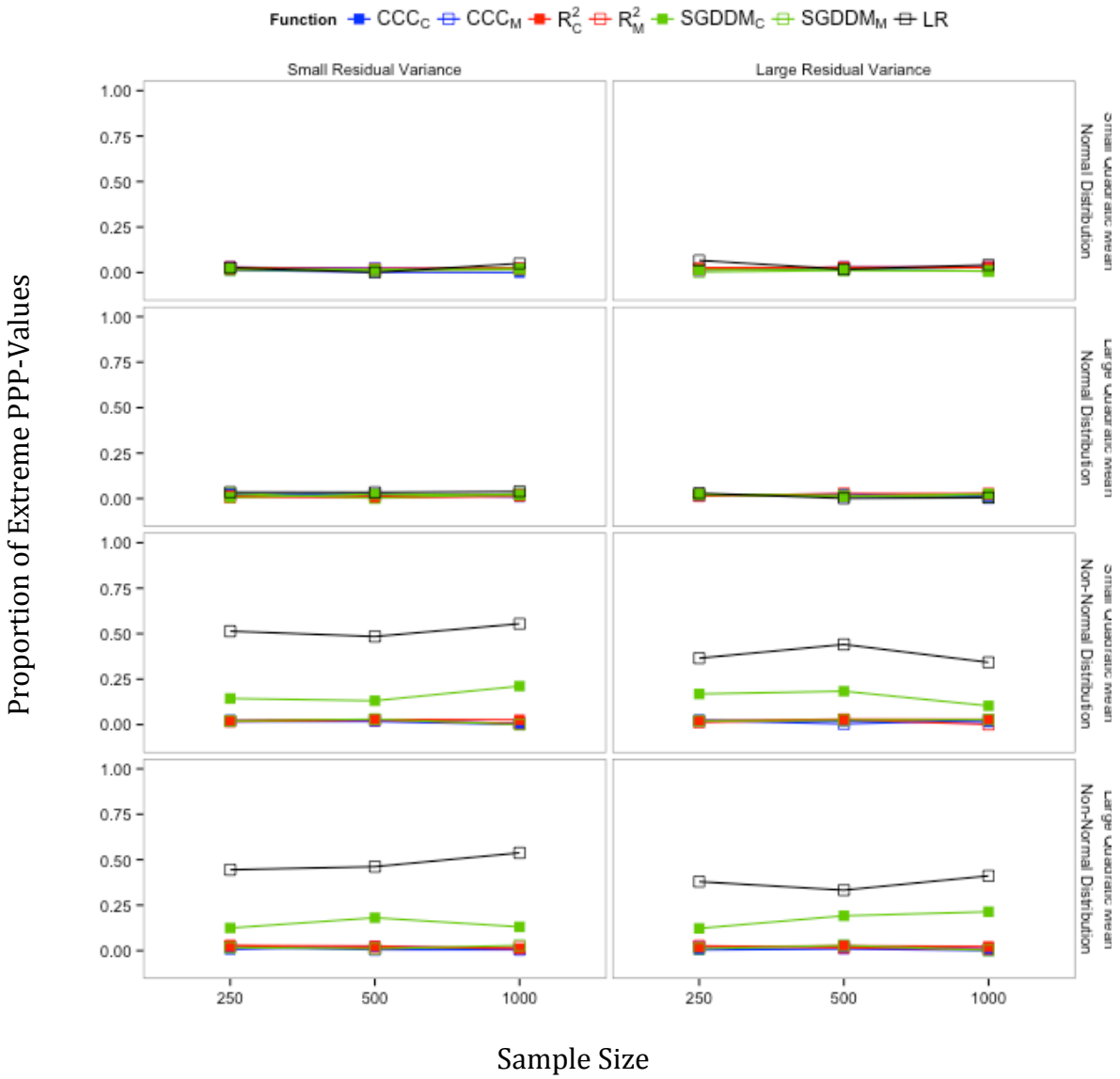

Figure 17. Proportion of extreme PPP-values for data generated and analyzed as a growth curve model with a quadratic functional form that was equal across individuals.

Figure 17 displays the proportion of PPP-values for all absolute measures of data-model fit for data generated and analyzed by a GCM with quadratic functional form that did not vary across individuals. The effect of increasing residual variation is observed looking from left to right. The top (bottom) two rows show the results for data generated with normally distributed (non-normal) outcomes. Holding the shape of the distribution

constant (i.e., normal, non-normal), the impact of increasing the strength of the quadratic effect is observed looking down a column.

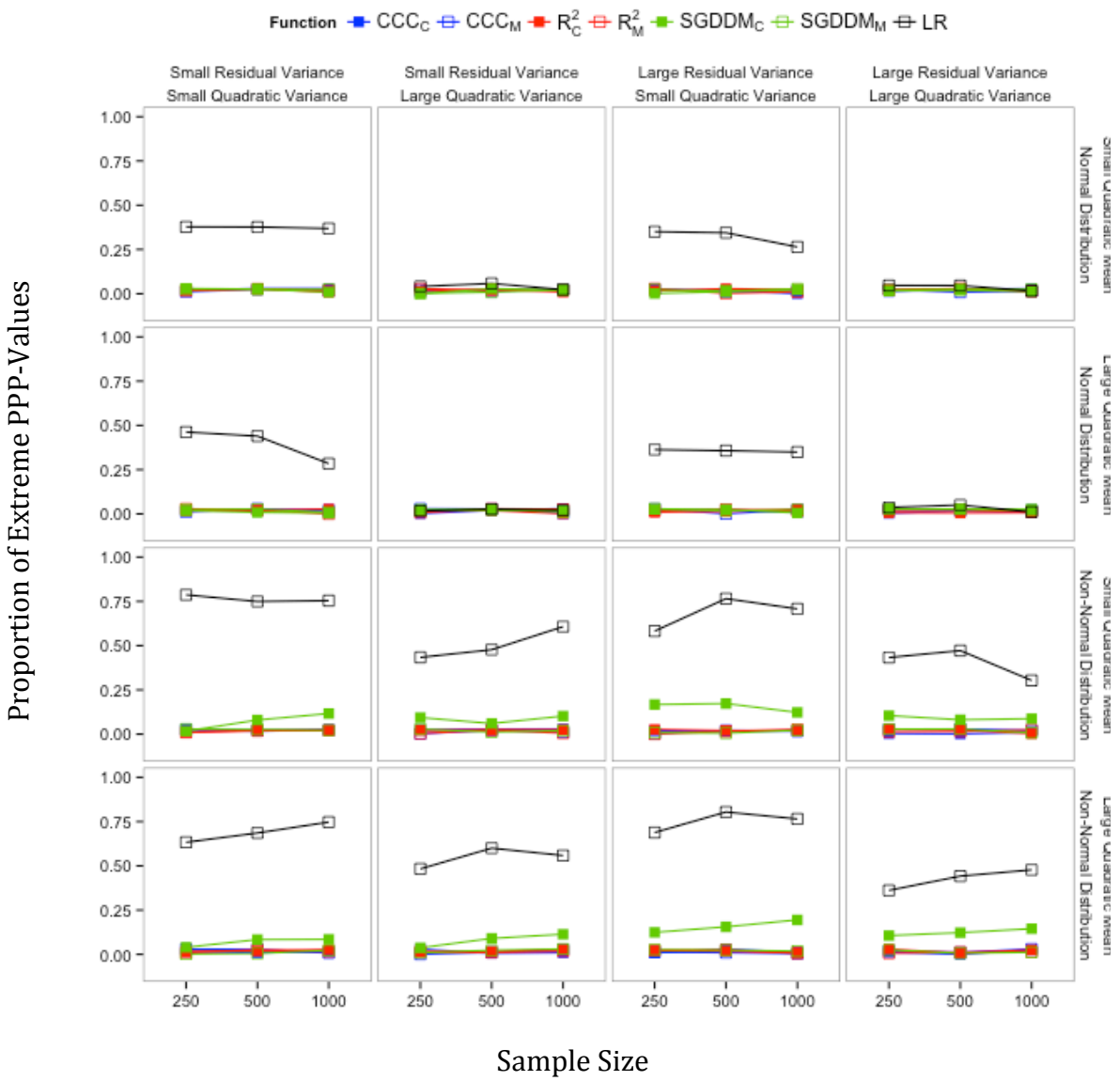

Figure 18. Proportion of extreme PPP-values for data generated and analyzed as a growth curve model with a quadratic functional form that varied across individuals.

Figure 18 show the proportion of extreme PPP-values for all absolute measures of data-model fit for data generated and analyzed by a GCM in which the strength of quadratic effect varied across individuals. Each column represents one of the four

combinations that result from crossing the levels of the residual and quadratic variance sizes. The results for data generated with a small (large) residual variance are shown in the two columns shown on the left (right). Looking within either level associated with the size of the residual variance, the effect of increasing the size of the quadratic variance is observed moving from left to right. The rows represent one of the four combinations that result from crossing the strength of the quadratic mean and the shape of the distribution. The top (bottom) two rows show the results for data generated with normally distributed (non-normal) outcomes. Holding the shape of the distribution constant (i.e., normal, non-normal), the impact of increasing the strength of the quadratic effect is observed looking down a column.

Drawing from the NHST perspective, the discrepancy functions would ideally reject the model about 5% of the time given a cutoff of .05 for declaring a PPP-value as extreme when the data generation and analysis models are aligned. Irrespective of the type of data generation model, the proportions of extreme PPP-values were consistently low and even conservative for the *CCC* (both conditional and marginal), R^2 (both conditional and marginal), and the $SGDDM_M$. The same was also true for the $SGDDM_C$ and *LR* for data generated from a normal distribution. However, for these latter two functions, the proportions of extreme PPP-values were larger than the ideal of 5% for data generated from a non-normal distribution. With proportions of extreme PPP-values generally hovering close to 0.50, this was particularly true for the *LR*. Interestingly, the proportions of extreme PPP-values for the *LR* were close 0.50 for data generated with a small quadratic variance even for normally distributed data.

Underspecified Marginal Mean Structure. Figure 19 displays the proportion of extreme PPP-values when the marginal mean structure was underspecified. Although the structure of the figure matches that of Figure 17, the proportions are based on fitting a linear GCM to data that followed a GCM with a quadratic effect but no quadratic variance. Given this relationship between the data generation and analysis models, the proportion of extreme PPP-values would ideally by close to one. From NHST perspective, such a result is reflective of high power to detect model inadequacy when it is present.

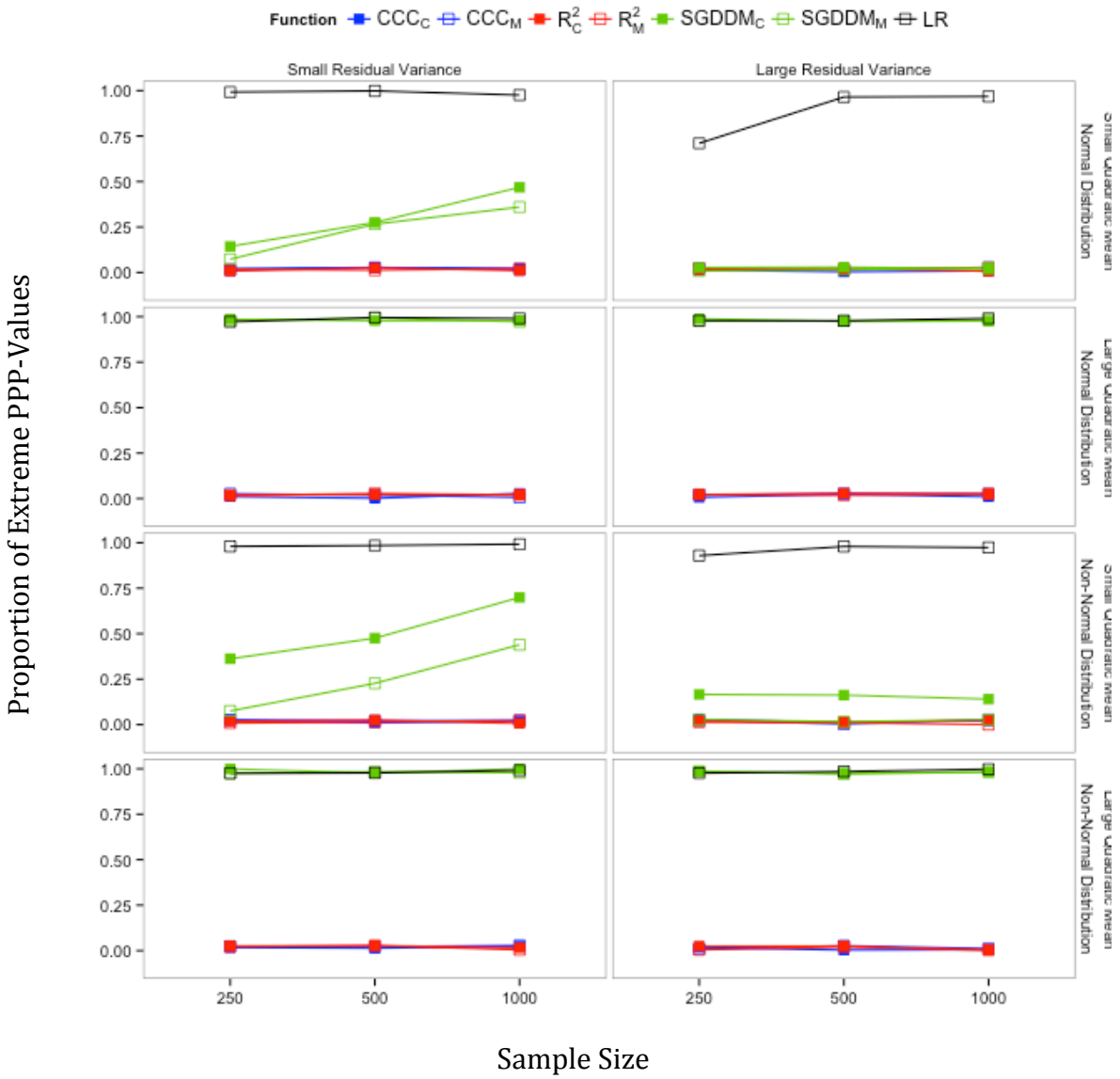

Figure 19. Proportion of extreme PPP-values for data generated by a growth curve model with a quadratic of equal strength across individuals but analyzed as a growth curve model with a linear functional form.

With proportions of extreme PPP-values consistently close to zero, neither the conditional nor marginal versions of the *CCC* and R^2 (blue and red lines, respectively) functions exhibited sensitivity to the underspecification of the marginal mean structure. With one exception, the proportions of extreme PPP-values for the *LR* (black line) were consistently at or close to one. Although still quiet reflective of being sensitive to the

underspecified marginal mean structure, the proportion of extreme PPP-values for the *LR* decreased somewhat with (a) the smallest sample size, (b) a small quadratic mean, and (c) a large residual variance for data generated with a normal distribution.

With proportions of extreme PPP-values at or close to one, the *SGDDM* functions (both conditional and marginal) were consistently sensitive to the underspecified marginal mean structure with a strong quadratic mean (see the green lines in the second and fourth rows). However, the performance of these functions deteriorated when the quadratic mean was small (green lines in the first and third rows), particularly when the residual variance was large (second column). Given a small residual variance, the proportions of extreme PPP-values for these functions increased with sample size. Notably, given a small quadratic mean irrespective of the size of the amount of residual variation, the proportion of extreme PPP-values for the $SGDDM_C$ were also larger for non-normal rather than normally distributed data.

Underspecified Conditional Mean Structure. Figure 20 (see below) displays the proportion of extreme PPP-values when the marginal mean structure was adequately specified but the conditional mean structure was underspecified. In the context of the current work, this involved applying a GCM with a quadratic effect that does not exhibit between-person variation to data from a GCM in which the quadratic effect that *does* exhibit between-person variation. Given the relationship between the data generation and analysis models, ideal performance depends on the characteristic of the model targeted by the discrepancy function. Among functions that target the marginal mean structure, low proportions of extreme PPP-values would be ideal since the marginal mean structure of the data analysis model is aligned with that of the data generation model. However, since

high proportions may reflect indirect manifestations of the underspecified conditional mean structure or even some other as yet unaccounted for source of misfit, there is hesitation to refer to the proportion of extreme PPP-values for marginal functions as Type I error rates. In contrast, the proportions of extreme PPP-values would ideally be high for functions that target the conditional mean structure when the data analysis model fails to capture between-personal variation in the function form of growth. Accordingly, the proportion of extreme PPP-values for the conditional functions can be viewed as estimates of power. Notably, misfit of the conditional mean structure hypothetically should have no bearing on the fit of the marginal mean structure (Wu et al., 2009; Wu & West, 2013). As the intent of having conditional and marginal versions of discrepancy is to separate sources of misfit that arise from the two mean structures, it is of interest whether the ideals described above are realized as this would reflect a capacity to disentangle misfit in the conditional mean structure from that of the marginal mean structure.

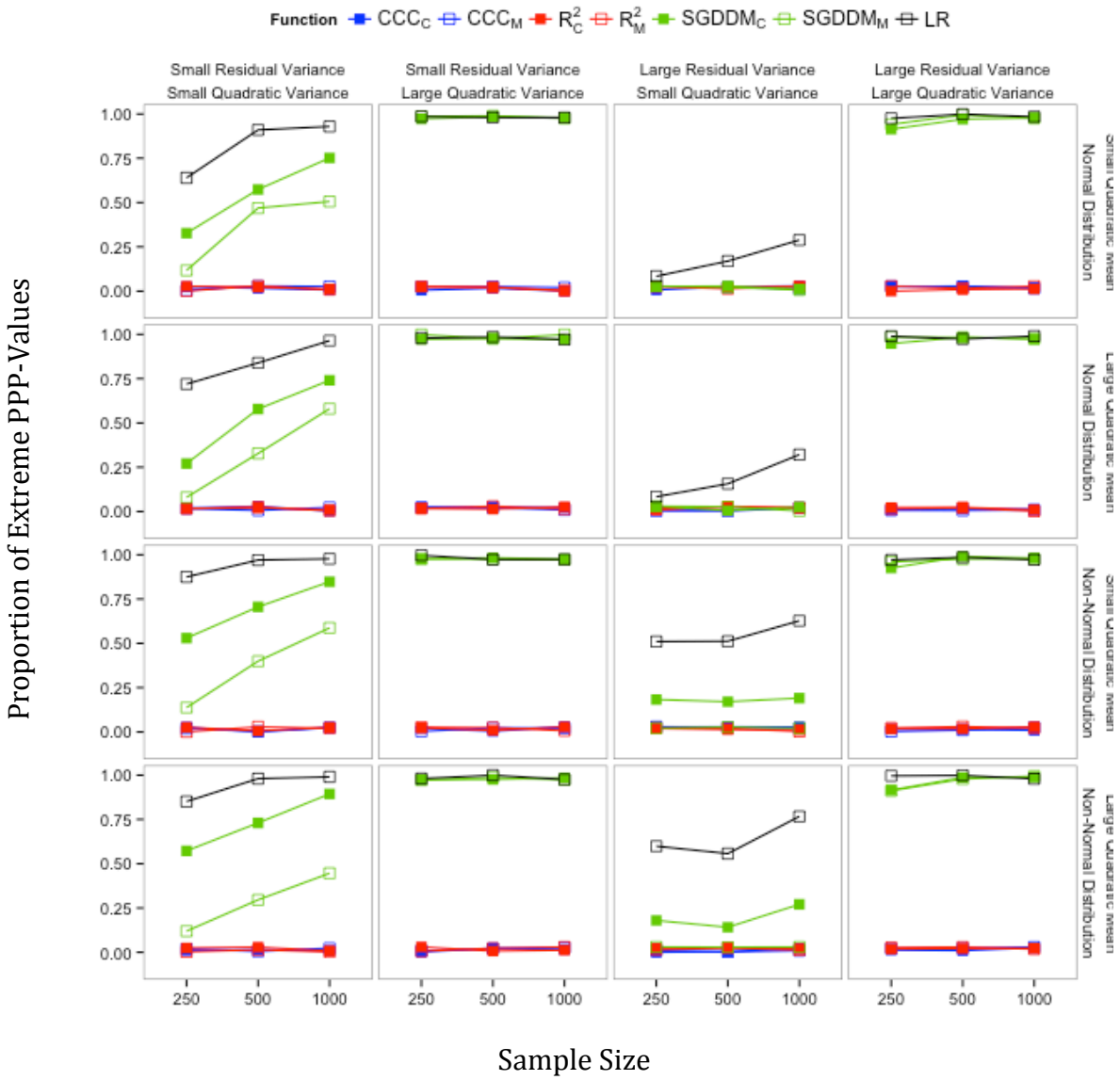

Figure 20. Proportion of extreme PPP-values for data generated as a growth curve model with a quadratic functional form that varied across individuals but analyzed by a growth curve model with a quadratic functional form assumed equal across individuals.

Consistent with the ideal described above for the CCC_M and R^2_M, the proportions of extreme PPP-values were low and even conservative across all conditions in which the conditional mean structure was underspecified. Unfortunately, and inconsistent with ideal performance for separating sources of misfit between mean structures, the same was also true for the CCC_C and R^2_C. The performances of the other discrepancy functions of

absolute fit was largely driven by (a) the size of the quadratic variance, (b) sample size, and (c) the shape of the distribution of observed variables. Given a large quadratic variance, the proportion of PPP-values for the *LR* and the *SGDDM* functions (both conditional and marginal) were at or close to one. For each of these functions, reducing the size of the quadratic variance resulted in lower proportions of extreme PPP-values. Decreasing the amount of quadratic variation had a smaller impact on the proportion of extreme PPP-values when the residual variance was small rather than large. Given a small residual variance, the loss in the proportion of extreme PPP-values was offset with increased sample size, particularly for the *SGDDM* functions, as they were not subject to ceiling effects. Given a large residual variance, decreasing the amount of quadratic variance had larger effect on the proportion of extreme PPP-values. The effect was stronger on the performance of the *SGDDM* functions than for the *LR*. Notably, the decrease in the proportion of extreme PPP-values for the $SGDDM_C$ and *LR* was smaller for non-normal data than for normally distributed data.

Underspecified Marginal and Conditional Mean Structure. Figure 21 displays the proportion of extreme PPP-values when the marginal and conditional mean structures were both underspecified. In the context of the current work, this involved applying a linear GCM to a GCM in which the quadratic effect that exhibits between-person variation. For this relationship between the data analysis and generation models, the proportions of extreme PPP-values would ideally approach one for all discrepancy functions.

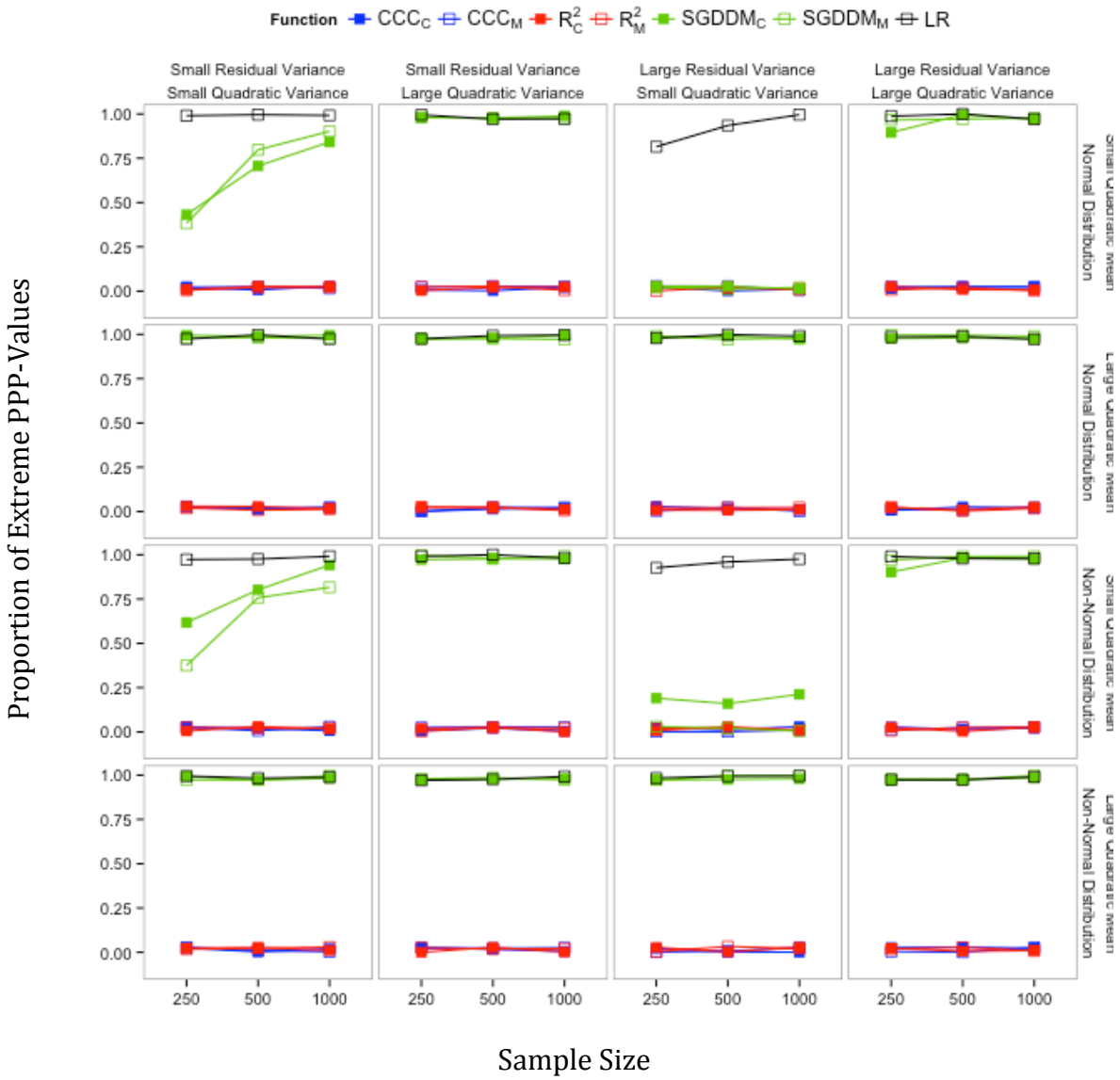

Figure 21. Proportion of extreme PPP-values for data generated as a growth curve model with a quadratic functional form that varied across individuals but analyzed by a growth curve model with a linear functional form.

With proportions of extreme PPP-values consistently close or equal to zero, the *CCC* and R^2 functions (both conditional and marginal) were not sensitive to the joint underspecification of the conditional and marginal mean structures. In stark contrast, the *LR* was quite sensitive with proportions of extreme PPP-values at or close to one in almost all conditions. Although the performance of the *LR* was still strong, performance

was found to deteriorate slightly in conditions generated with (a) the smallest sample size, (b) a small quadratic mean, (c) a small quadratic variance, and (d) a large residual variance.

Given a large quadratic mean and/or quadratic variance, the proportions of extreme PPP-values for the *SGDDM* (both conditional and marginal) were at or close to one. Decreasing the size of the quadratic mean and/or quadratic variance resulted in smaller proportions of extreme PPP-values, particularly for data generated with a large residual variance. Notably, increased sample size served to offset the loss in the proportion of extreme PPP-values when the residual variance was small.

ANOVA Results for Realized Values of the *LRT*

As was done for measures of absolute fit, a series of ANOVA models were conducted to evaluate the impact of factors in the simulation design on the realized values of the *LRT*, the only measure of relative fit. Owing to the number of observations within cells, the outcome of interest was the effect sizes (as measured by partial η^2) of factors in the ANOVA design. The effect sizes served to identify the strongest effects on the realized values of the *LRT*, which were in turn used to identify effects to be explored graphically. The presentation of results for the *LRT* differs from the presentation of results for absolute measures. The focus of the latter was on the particular relationships between data generation and analysis models. For the *LRT*, the results pertaining to each model comparison (i.e., M1 *v*. M2 and M2 *v*. M3) are presented for each of the three types of data analysis models. As a reminder, M1 represents a linear GCM; M2 represents a GCM with a quadratic effect of equal strength across individuals; and M3 represents a GCM with quadratic effect that varies in strength across individuals.

Linear GCM Model. Table 8 displays the effect sizes for factor in the ANOVA design for data generated to follow a GCM with a linear functional form. The factors in the ANOVA model consisted of sample size (denoted N, 3 levels), the size of the residual variance (denoted RV, 2 levels), and the shape of the distribution of outcomes (denoted DS, 2 levels). All factors were full crossed to yield 12 cells in the ANOVA design. Separate analyses were conducted for the realized values of the *LRT* associated with the two model comparisons.

Table 8

Effect Sizes of Each Effect on the Realized Values for the LRT for Data Generated to Follow a Linear Growth Curve Model (GCM)

	Model Comparison	
	Linear (M1) *v.*	Equal Quadratic (M2) *v.*
Effect	Equal Quadratic (M2)	Varying Quadratic (M3)
DS	< .001	.005
N	< .001	.074
RV	< .001	.093
N × DS	< .001	.002
RV × DS	< .001	.010
N × RV	< .001	.001
N × RV × DS	< .001	< .001

Note. DS = distribution shape, N = sample size, RV = residual variance.

For data generated to follow a linear GCM model, none of the factors had meaningful bearing on the realized values of the *LRT* when comparing M1 to M2. In contrast, the realized values of the *LRT* when comparing M2 to M3 were impacted by (a) the main effect of sample size and (b) the main effect of the residual variance size. The realized values for both comparisons (M1 *v* M2 is represented by gray bars, M2 *v*. M3 is represented by black bars) are plotted against sample size (shown along the horizontal

axis) for each level of the residual variance (shown as increasing from left to right across panels).

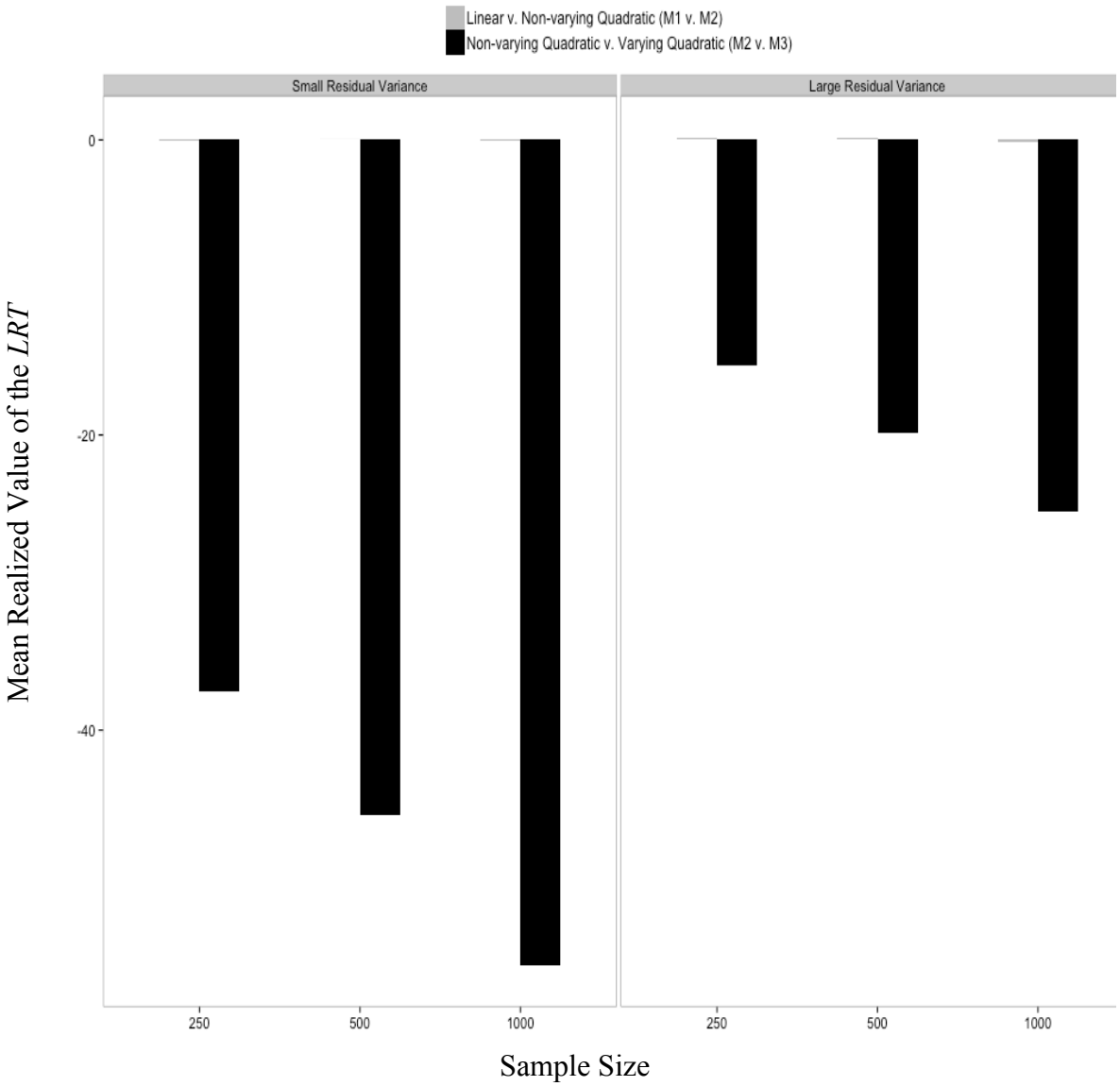

Figure 22. Plot of the interaction between sample size and the size of the residual variance on the mean realized values of the *LRT* for data generated to follow a linear GCM.

For the comparison between M1 and M2, the average realized values of the LRT were not impacted by sample size, the size of the residual variance, or the interplay between these two factors. Despite the relatively small size, the striking result was the

nature of the effects of sample size and the size of the residual variance for the comparison between M2 and M3, particularly for data generated with a small residual variance; as seen in Figure 22 the average realized values were negative. Although not shown below, the *LRT* was found to exhibit similar behavior for null comparisons involving the more complicated models considered in this work. As discussed more in the subsequent chapter, this finding was particularly surprising and clearly has implications for the use of the *LRT* within the PPMC framework.

GCM with Equal Strength of the Quadratic Mean Across Cases. Table 9 shows the effect sizes for each effect in the ANOVA design for data generated to follow a GCM with a quadratic functional form that was equivalent in strength for all individuals. The factors in the ANOVA model included sample size (3 levels), the strength of the quadratic mean (denoted QM, 2 levels), the size of the residual variance (2 levels), and the shape of the distribution for the observed variables (2 levels). All factors were full crossed to yield 24 cells in the ANOVA design. Separate analyses were conducted for the two model comparisons.

When comparing the realized values of the *LRT* for M1 and M2, a three-way interaction was found between the strength of the quadratic mean, sample size, and the size of the residual variance. This effect trumped a two-way interaction between the strength of the quadratic mean and sample size and the main effect of the quadratic strength. Excepting to small main effects of sample size and the size of the residual variance, the factors manipulated in the simulation had virtually no impact on the realized of the *LRT* when comparing the fit of M2 to that of M3. However, consistent with the null findings presented for the linear model, negative values were evidenced.

Table 9

Effect Sizes of Each Effect on the Realized Values for the LRT for Data Generated to Follow a Growth Curve Model (GCM) with an Equal Quadratic From Across Cases

	Model Comparison	
	Linear (M1) *v.*	Equal Quadratic (M2) *v.*
Effect	Equal Quadratic (M2)	Varying Quadratic (M3)
DS	< .001	.003
N	.030	.047
QM	.262	< .001
RV	.002	.042
N × DS	< .001	< .001
N × QM	.620	< .001
N × RV	.010	.009
QM × DS	< .001	< .001
QM × RV	.070	< .001
RV × DS	< .001	< .001
N × QM × DS	.002	< .001
N × QM × RV	.245	< .001
N × RV × DS	< .001	< .001
QM × RV × DS	< .001	< .001
N × QM × RV × DS	< .001	< .001

Note. DS = distribution shape, N = sample size, QM = quadratic mean, RV = residual variance.

Figure 23 displays the average realized values of the *LRT* involving the comparison between M1 and M2 for each combination of the levels associated with sample size (moving from left to right within panels), the strength of the quadratic effect (small and large quadratic effects are represented by gray and black bars, respectively), and the size of the residual variance (shown as increasing from left to right across panels). For simplicity, the midpoints of the random uniform distributions for drawing values of sample size serve as the labels of sample size. The realized values of the *LRT* were found to increase with increasing sample size, increasing strength of the quadratic mean, and smaller residual variance. The nature of the interaction between these three

factors was such that increasing the strength of the quadratic mean enhanced the impact of sample size but increasing the size of the residual variance hindered this combined effect.

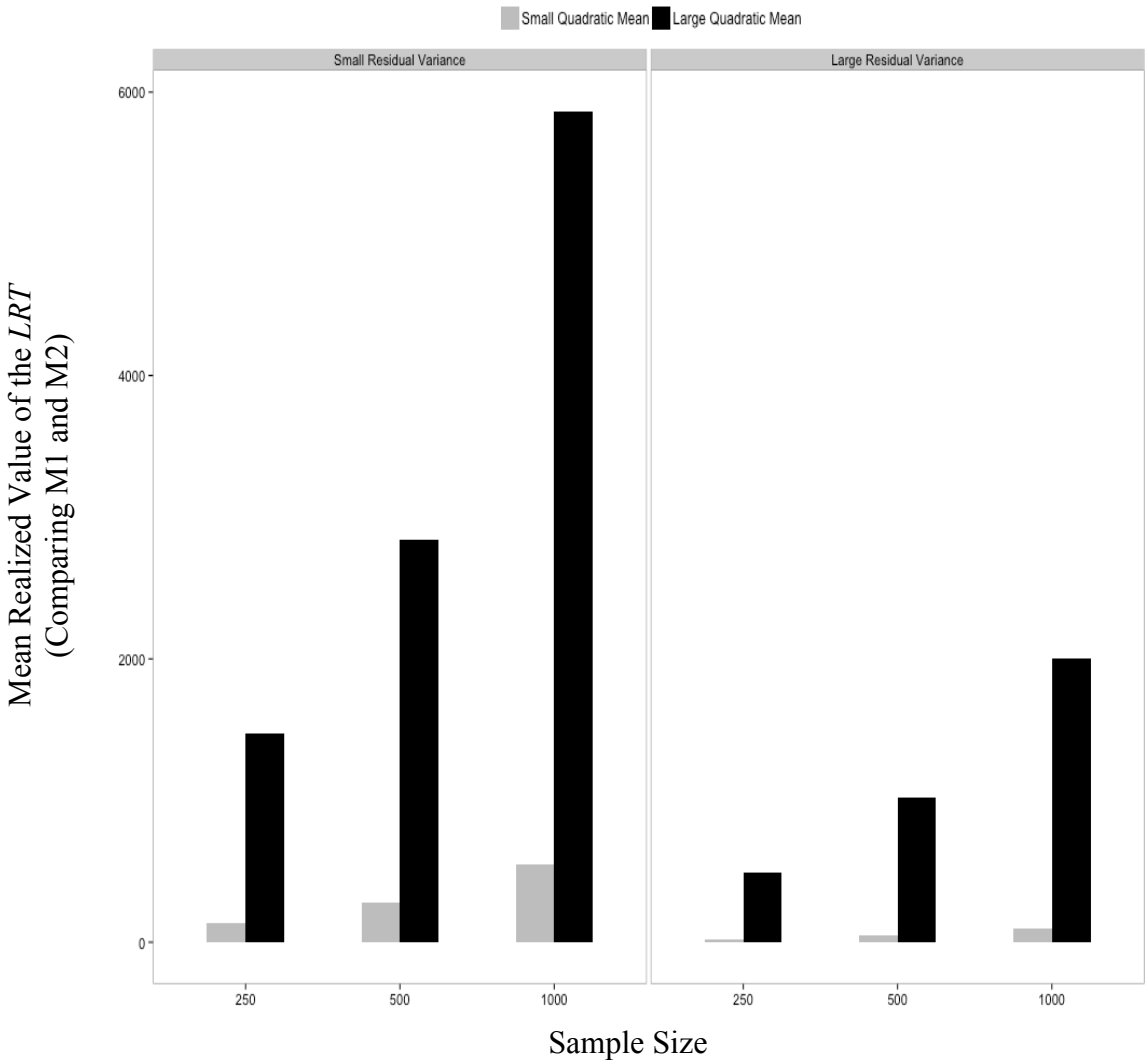

Figure 23. Plot of the interaction between sample size, the strength of the quadratic mean, and the size of the residual variance on the mean realized values of the *LRT* when comparing M1 and M2 for data generated to follow a GCM with a quadratic effect that is identical across individuals.

GCM with Varying Strength of the Quadratic Mean Across Cases. Table 10 displays the effect sizes for each effect in the ANOVA design for data generated to follow GCM generated with a quadratic functional form that varied in strength across

individuals. The ANOVA design consisted of five factors including sample size (3 levels), the strength of the quadratic mean (2 levels), the size of the quadratic variance (denoted QV, 2 levels), the size of the residual variance (2 levels), and the shape of the distribution (2 levels). All factors were fully crossed to yield 48 cells in the ANOVA design. Separate analyses were conducted for the realized values associated with each model comparison of interest (i.e., M1 *v*. M2 and M2 *v*. M3).

Focusing first on the comparison between M1 and M2, the key effects on the realized values of the *LRT* included the strength of quadratic mean, a two-way interaction between sample size and the strength of the quadratic mean, and a three-way interaction involving these two factors and the size of the residual variance. Turning to the comparison between M2 and M3, a similar pattern of effects on the *LRT* was found involving the size of the quadratic variance rather than the strength of the quadratic mean. Unlike the prior two types of generation models, the realized values of the *LRT* were not tied to a null comparison. Notably, the behavior of the *LRT* mirrors that displayed in Figure 23, which shows the three-way interaction between the same factors for data that follow a GCM with a quadratic effect that did not vary across people. The pattern of results held for the three-way interaction between sample size, the strength of the quadratic mean (variance), and the realized values of the *LRT* for the comparison between M1 (M2) and M2 (M3). Given the similarity of the results, these effects were not plotted below.

Table 10

Effect Sizes of Each Effect on the Realized Values for the LRT for Data Generated to Follow a Growth Curve Model (GCM) with a Quadratic From that Varied Across Cases

	Model Comparison	
Effect	Linear (M1) *v.* Equal Quadratic (M2)	Equal Quadratic (M2) *v.* Varying Quadratic (M3)
DS	< .001	< .001
N	.022	.001
QM	.187	< .001
QV	< .001	.099
RV	.001	< .001
N × DS	< .001	< .001
N × QM	.525	< .001
N × QV	.003	.368
N × RV	.008	< .001
QM × DS	< .001	< .001
QM × QV	.011	< .001
QM × RV	.040	< .001
QV × DS	< .001	< .001
QV × RV	< .001	.032
RV × DS	< .001	< .001
N × QM × DS	.003	< .001
N × QM × QV	.057	.005
N × QM × RV	.185	< .001
N × QV × DS	< .001	.008
N × QV × RV	.001	.159
N × RV × DS	< .001	< .001
QM × QV × DS	< .001	< .001
QM × QV × RV	.003	< .001
QM × RV × DS	< .001	< .001
QV × RV × DS	< .001	< .001
N × QM × QV × DS	< .001	.003
N × QM × QV × RV	.025	.003
N × QM × RV × DS	.001	< .001
N × QV × RV × DS	< .001	.004
QM × QV × RV × DS	< .001	< .001
N × QM × QV × RV × DS	< .001	.002

Note. DS = distribution shape, N = sample size, QM = quadratic mean strength, QV = quadratic variance, RV = residual variance.

Distribution of PPP-Values for the *LRT*

Figure 24 displays the marginal density (shown along the vertical axis) distribution of PPP-values (shown along the horizontal axis) for the *LRT*. The two columns represent the model comparisons. As a reminder, for a given trial within a condition, the -2 log likelihoods for a given model comparison were based on using the posterior distributions of model parameters for the two models being compared and the collection of posterior predictive datasets associated with the more restrictive model. The rows correspond to the three data generation models. The desired result is for PPP-values to be uniformly distributed when either (a) the more restricted model sufficiently captures growth process or (b) two models being compared both represent a case of overfitting. When the less restricted model provides better fit to the observed data than the more restricted model, PPP-values would ideally be concentrated towards the lower extreme of the PPP-value scale.

In the case of data generated to follow a linear GCM (the top row), the two model comparisons both represent null comparisons. The PPP-values for the comparison between M1 (i.e., linear GCM) and M2 (i.e., equal quadratic) were concentrated at the middle of the PPP-value scale. This finding indicates that the *LRT* is unlikely to point the analyst to selecting a model with a non-varying quadratic relationship with the passage time over a linear model when the actually follow the latter. A different picture emerges for the comparison between M2 and M3 (i.e., varying quadratic). For this model comparison, the distribution of PPP-values was peaked at low values and then gradually tapered off throughout the PPP-value range; this finding suggests that the *LRT* may favor

a model with a varying quadratic effect over a model with a non-varying quadratic effect for data that follow a linear GCM.

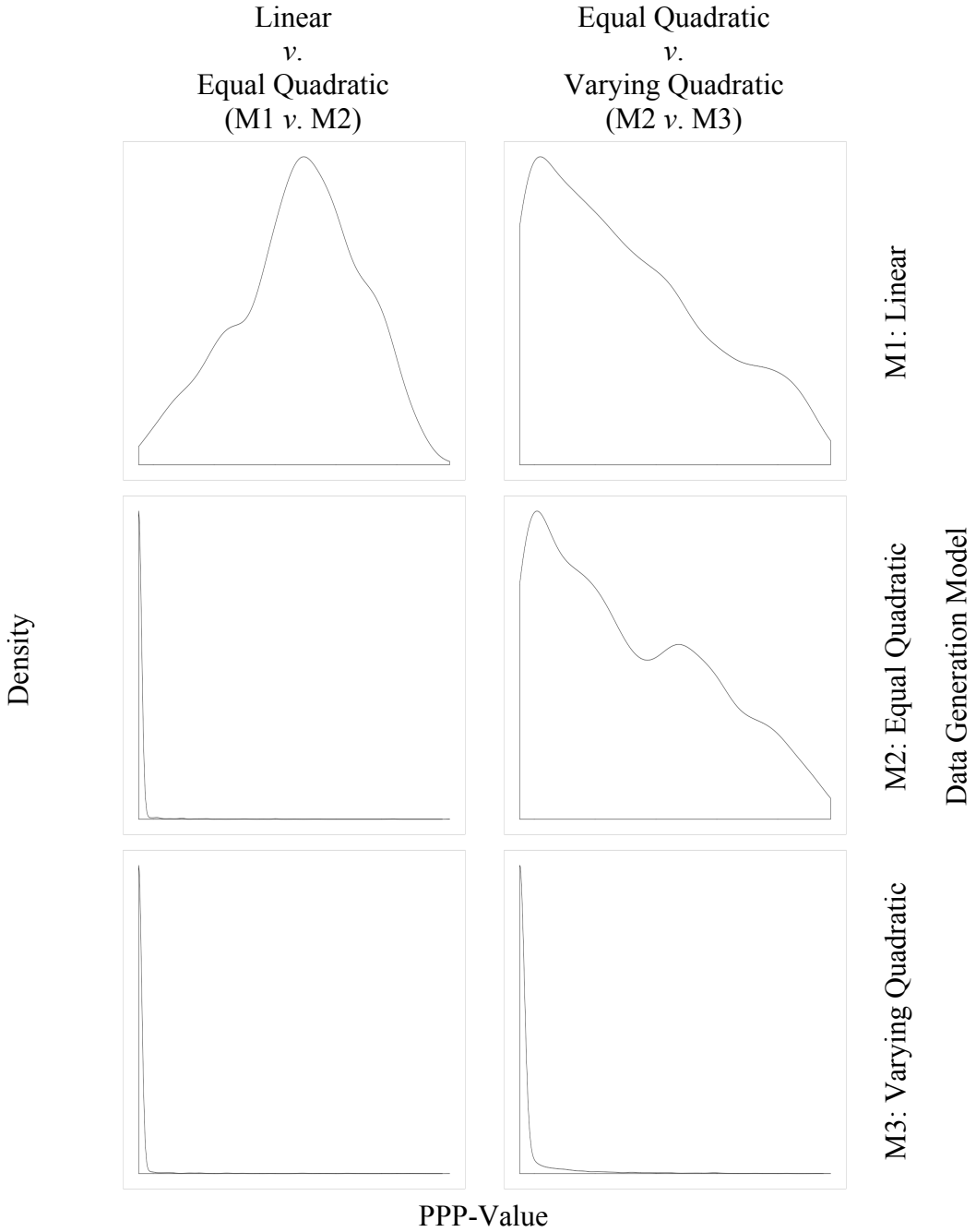

Figure 24. Plots of the marginal (i.e., over all manipulated factors) densities for the PPP-values of the *LRT* for different associations between the data generation and analysis models.

The second row displays the results for data generated to follow a GCM in which the quadratic effect was equal across individuals. Consistent with the desired result, the distribution of PPP-values was heavily concentrated at the lower boundary of the PPP-value scale. This finding suggests that *LRT* is quite likely to favor the quadratic GCM (M2) over the linear GCM (M1); this is a desirable result since the former GCM represents the data generation model. The peak of the distribution of PPP-values for the comparison between a quadratic growth curve models with (M2) and without (M3) was located at the lower boundary of the PPP-value scale. As was the case for data generated under linear GCM, the tail of the distribution gradually tapered off throughout the remaining range of the PPP-value.

The third row displays the results for data generated to follow GCM in which the quadratic effect varied across individuals. As evidenced by distributions that were heavily concentrated at the lower bound of the PPP-value scale, (a) a GCM with a quadratic effect equivalent in strength across individuals was likely to be favored over a linear GCM and (b) a GCM with a varying quadratic effect was likely to be favored over a GCM in which the quadratic effect was assumed equal across individuals.

Proportion of Extreme PPP-Values for the *LRT*

In this section, the proportions of extreme PPP-values of the *LRT* are graphically summarized across all conditions in the simulation design. Each figure presented in this section corresponds to one of the three models used to generate data. Although there are some differences in the general organization of the figures (described in greater detail below), the within-panel structure is consistent. For each panel, the proportions of extreme PPP-values (defined as a PPP-value $\leq .05$) are plotted against the three levels of

sample size. Each level of sample size is represented by the midpoint between the minimum and maximum values of the random uniform distribution used to draw values of sample size. Two bars are displayed within each level of sample size. The red bar represents the comparison between M1 (linear GCM) and M2 (GCM with equal quadratic effect); the reference distribution was based on computing the *LRT* using posterior predictive data consistent M1. The blue bar represents the comparison between M2 and M3; the reference distribution was based on computing the *LRT* using posterior predictive data consistent with M2. The remaining features of the figures were dependent on the characteristics governed the generation of data.

Linear GCM. Figure 25 displays the proportion of extreme PPP-values for data that follow a linear GCM. The figure is organized as a 2 × 2 matrix such that (a) the effect of departing from normality is seen looking down a column and (b) the effect of increasing the amount of residual variation is within rows. Looking across all conditions, the proportions of extreme PPP-values were generally conservative when comparing M1 to M2 (red bars). In contrast, the proportions of extreme PPP-values were higher when comparing M2 and M3 (blue bars). Given a small residual variance (first column), the proportions were higher for non-normal data (bottom row) than those observed for normally distributed (top row) data.

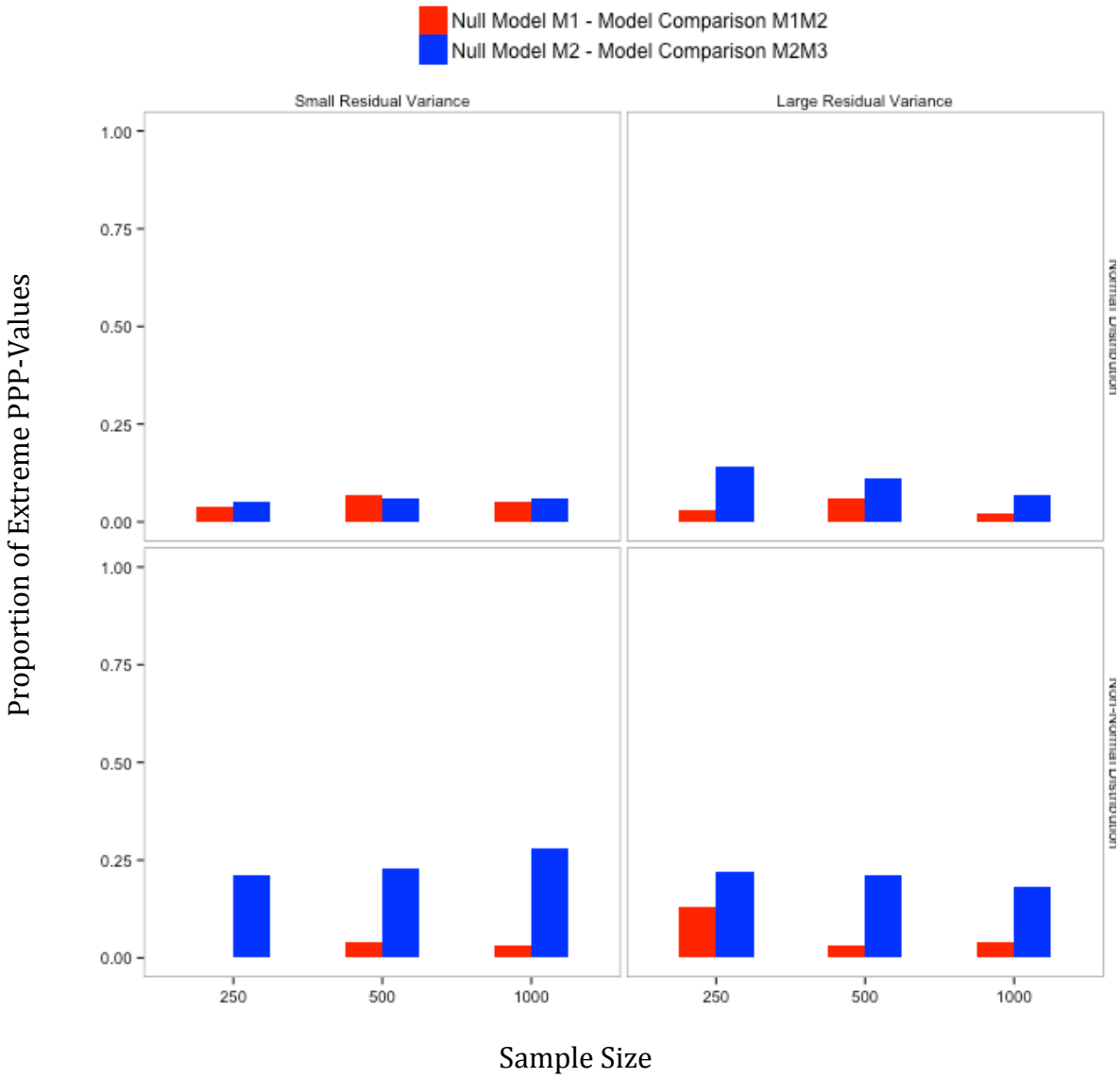

Figure 25. Proportion of extreme PPP-values for data generated as a linear growth curve model.

Non-Varying Quadratic GCM. Figure 26 displays the proportion of extreme PPP-values for data generated to follow GCM with quadratic effect that was equal across individuals. The figure is organized as a 2 × 4 matrix in which (a) rows represent the levels of distribution shape and (b) columns represent one of four combinations that result from crossing the size of the residual variance and the strength of the quadratic

mean. Conditions generated with small (large) residual variance are shown in the first and second (third and fourth) columns. The levels of the quadratic strength are shown as increasing within the levels of the residual variance.

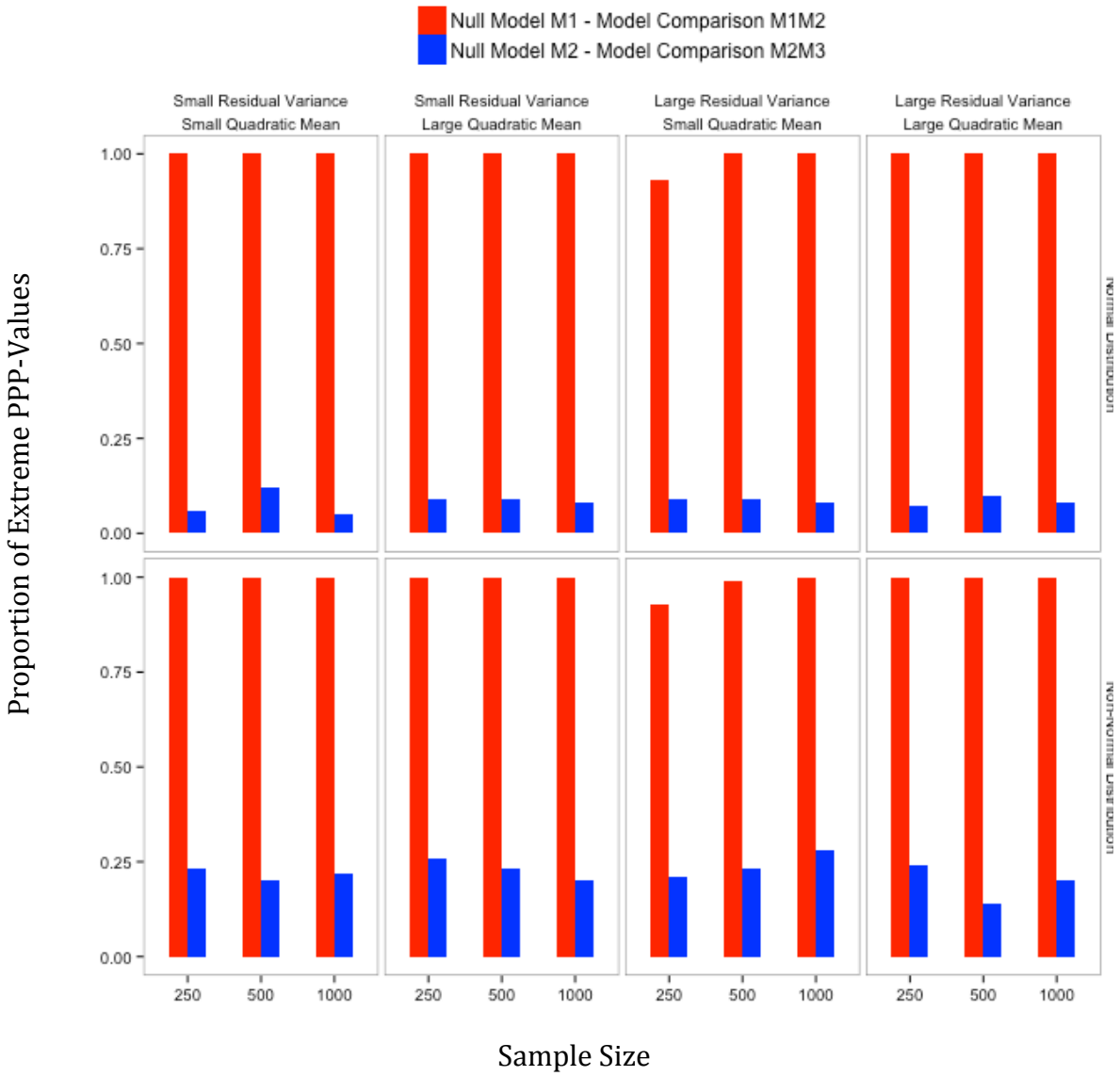

Figure 26. Proportion of extreme PPP-values for data generated as a growth curve model with a quadratic functional form that was identical across individuals.

In general, the proportions of extreme PPP-values for the *LRT* were at (or close to) one for data generated when comparing M1 and M2 (red bars). Although the

proportion of extreme PPP-values was still quite high, the proportions declined somewhat for data generated with (a) about 250 individuals, (b) a small quadratic mean, and (c) large residual variance. When comparing M2 and M3 (blue bars), the proportion of PPP-values were relatively low compared to those observed for the comparisons between M1 and M2. However, for the comparisons between M2 and M3, PPP-values were observed to increase for data generated with non-normal outcomes compared to data generated with normally distributed outcomes.

Varying Quadratic GCM. Figure 27 displays the proportion of extreme PPP-values for data generated to follow GCM with quadratic effect that varied across individuals. The figure is organized as a 4 × 4 matrix. The rows represent the four combinations that result from crossing the shape of the distribution and the strength of the quadratic mean. The top (bottom) two rows represent conditions generated with a small (large) quadratic mean. The impact of departing from normality is seen looking down a column within the levels associated with the strength of the quadratic mean. The columns represent the four combinations that result from crossing the size of the quadratic variance and the size of the residual variance. The first and second (third and fourth) two rows represent conditions generated with a small (large) quadratic variance. The levels of the residual variance are shown as increasing within the levels of the quadratic variance.

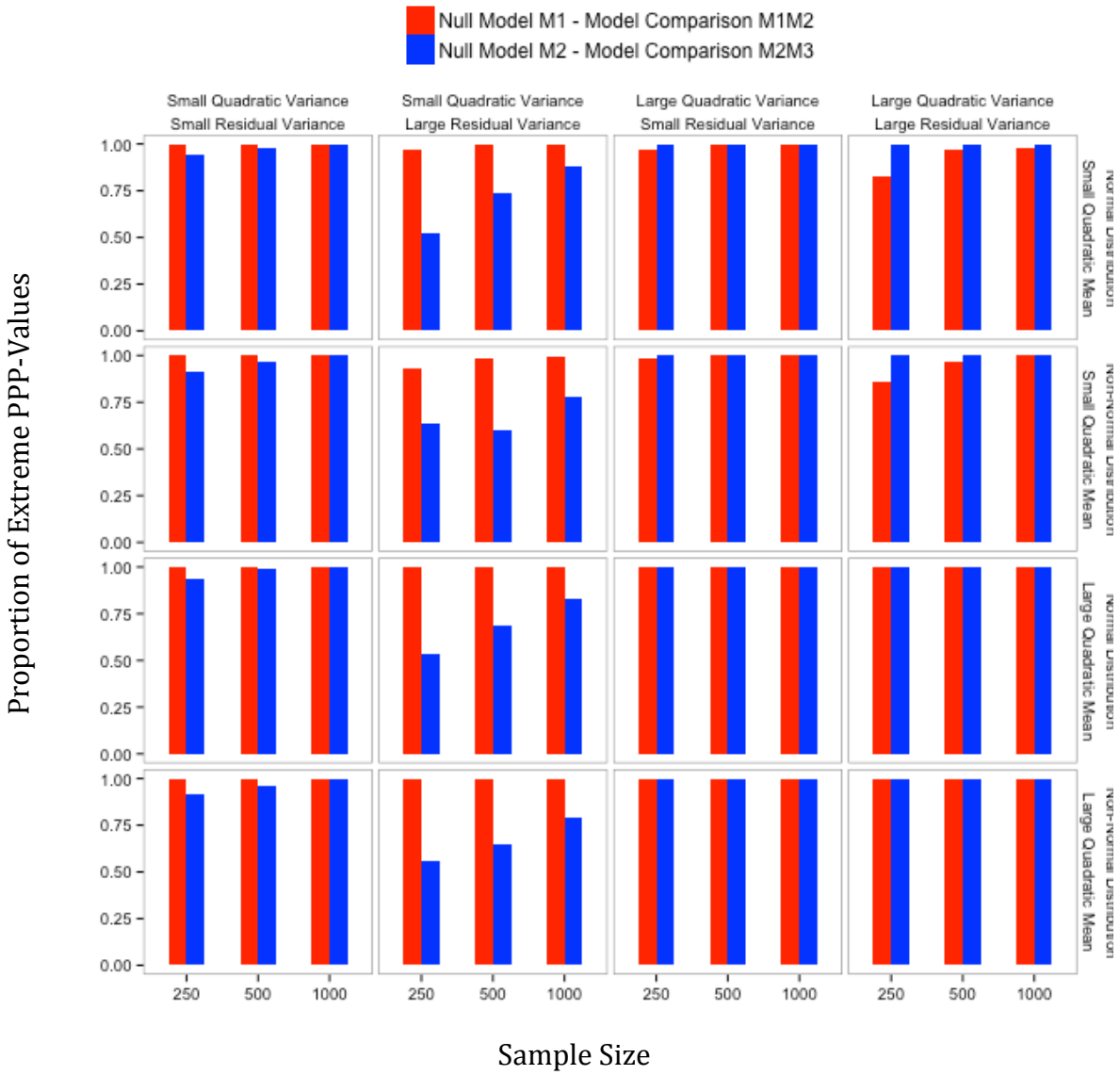

Figure 27. Proportion of extreme PPP-values for data generated as a growth curve model with a quadratic functional form that varied across individuals.

In general, the proportions of extreme PPP-values were high for both model comparisons. As observed when the quadratic effect was equal across individuals, the proportions of extreme PPP-values associated with the comparison between M1 and M2 (red bars) were lowest for data generated with a small quadratic mean and a large residual variance (top two rows in the second and fourth columns). Given these same

characteristics, increasing the size of the quadratic variance also reduced the proportions (moving from the first to the second row). Although proportions were consistently high when comparing M2 and M3 (blue bars), the loss in performance for conditions with a small quadratic variance and a large residual variance (second column) were offset with increased sample size.

Chapter 5: Discussion

Although modern approaches to GCM were developed more than 30 years ago (e.g., Baker, 1954; Meredith & Tisak, 1990), methodological research pertaining to model criticism for GCM has only recently emerged (Coffman & Millsap, 2006; Leite & Stapleton, 2011; Wu et al., 2009; West & Wu, 2010; Wu & West, 2013). The current work builds upon this growing body of research by investigating the performance of several discrepancy functions with PPMC as the foundation for model criticism. Moreover, this research has implications for applications beyond GCM that extend to the broader modeling frameworks of MLM and factor analytic models that include a mean structure.

The purpose of this study was to investigate the performance of several discrepancy functions for critiquing the fit of growth curve models with PPMC as the framework for model criticism. Many of the discrepancy functions considered in the study—namely the *CCC* (both conditional and marginal), R^2 (both conditional and marginal), *LR*, and *LRT*—have received some attention in methodological research. However, the breadth of the existing literature is limited and none of which has pursued a Bayesian approach to model criticism. This study also pursued the $SGDDM_C$—an existing function that has shown promise for critiquing psychometric models—and the $SGDDM_M$, which was constructed to target underspecification of the marginal mean structure. Taken together, the results carry implications for use in applied research and serve as a basis for future methodological research. These implications are discussed following a summary and discussion about the performance and behavior of the discrepancy functions investigated in this study.

Performance of Discrepancy Functions

The existing research pertaining to the criticism of growth curve models has only focused on patterns in the realized values of discrepancy functions. In order to connect the results to prior research, and establish a baseline for new functions, the behavior of the realized values of discrepancy functions was also considered in the current study. Unlike prior research, the PPMC framework used in the current work involves empirically constructing a reference distribution for any discrepancy function, making it possible to investigate the performance of discrepancy functions in null and non-null settings. In what follows, the goal is to synthesize the existing literature, the behavior of realized values for the discrepancy functions in the current study, and performance in null and non-null settings.

Performance in Null Conditions. Among the discrepancy functions that have been investigated in previous research, some patterns in the realized values in null conditions (see Tables 2 through 4) were consistent with past research while others were not. Wu and West (2013) performed the only study to date that has investigated the behavior of discrepancy functions in null conditions for GCM models. There were three key findings in the current work that were aligned with those reported by Wu and West when the data analysis and generation models were aligned. First, the realized values of the CCC_C and R_C^2 were found to decrease to the extent that the amount of residual variation was increased. Second, when relevant, the realized values of the CCC_M and R_M^2 exhibited a positive relationship with increasing strength of the quadratic mean. Third, with one exception that has not been considered in prior work, the realized values of the *LR* were generally not impacted by any of the design factors when the data analysis

model was aligned with the data generation model. The key exception was a small main effect in which departures from multivariate normality in the observed scores gave rise to larger realized values of the *LR*.

No research to date has pursued the $SGDDM_C$ for critiquing the fit of growth curve models. As a result, connections cannot be made with prior research. In null situations, the realized values of the $SGDDM_C$ were observed to (a) decrease with increasing sample size and (b) increase for data generated with non-normal rather than normally distributed repeated measures.

The $SGDDM_M$ was constructed for the purposes of the current work, and accordingly, there is no existing research to connect the results of the current work to. In null situations, the realized values of the $SGDDM_M$ were observed to decrease with (a) increased residual variation and (b) increased quadratic variation (seen only in Table 4). Notably, the implication of the relationship between the realized values of the marginal $SGDDM_M$ and the amount of residual variation is at odds with that observed for the CCC_C and R_C^2. Whereas these functions were found to become more indicative of increasingly greater misspecification of a null model with increasing residual variation, the $SGDDM_M$ became increasingly more indicative of better data-model fit with increasing residual variation.

Another peculiar result was observed for the $SGDDM_M$. For a given data generation model, the effect sizes were identical across the different data analysis models. This peculiar result was observed looking across tables associated with data generated with an equal quadratic mean across individuals (Tables 3 and 5) and for data generated with a varying quadratic mean across individuals (Tables 4, 6, and 7). Upon further

review, it was discovered that the realized values of the $SGDDM_M$ for a particular 'observed' dataset were constant. This was the result of the (marginal) **R** matrix being *identical* irrespective of the particular set of model-implied vector of expected means. By extension, the observed covariance matrix, which is also computed by the deviation between observed scores and some vector of means, was equivalent to the **R** matrix used to compute the $SGDDM_M$. As discussed in greater detail below, this result highlights the importance of carefully selecting/engineering discrepancy functions, particularly for data that exhibit a multilevel structure with model expectations present at multiple levels.

As for the *LRT*, there were two situations in the current work that represented a null comparison. One of which included comparing M1 (linear GCM) and M2 (equal quadratic strength GCM) for data that followed a linear GCM (see Table 8), and the other involved comparing M2 and M3 (varying quadratic strength GCM) for data that followed a GCM in which the quadratic strength was equal across individuals (see Table 9). In the former case, none of the design factors impacted the realized values of the *LRT*. Excepting to a small main effects of sample size and the size of the residual variance, a similar statement holds for the latter case. Notably, with the use of the *LRT*, one set of realized values were based on the comparison between two models that were both overparameterized relative to the data generation model. In this case, the realized values were negative on average. This finding poses serious methodological challenges to the

use of the *LRT* not just in null situations but also more broadly within the PPMC framework[3].

Before describing the impact of factors in the simulation design on the proportions of extreme PPP-values, it is worthwhile to first review patterns in the marginal distributions of PPP-values. The marginal distributions of PPP-values are aligned with the perspective of the applied analyst in that characteristics of the process that gave rise to the observed data (i.e., quadratic strength, size of residual variance, etcetera) are unknown; to the applied analyst, it only matters what decision is likely to be made given the alignment or lack thereof between the observed data and the model (or model comparison) at hand. The top row in Figure 13 displays the marginal distributions of PPP-values for the *CCC* (both conditional and marginal), R^2 (both conditional and marginal), and the $SGDDM_C$; the corresponding results for the $SGDDM_M$ and *LR* are seen

[3] Given that both models were overparameterized relative to the data generation model, additional analyses were undertaken to assess if the result (a) was a manifestation of poor convergence and/or (b) a mistake in the calculation of model log likelihoods. These additional analyses involved (a) fitting the overparameterized models to linear GCM data in commercial software via ML estimation (i.e., Mplus; Muthén & Muthén, 2010), (b) computing the log likelihoods using both sets of model parameters using the function in R, and (c) estimating the overparameterized models using fixed values of for one set of MCMC draws and comparing the log likelihoods to those obtained in R. The results of these analyses indicated that (a) the models do converge and (b) the computation of model log likelihoods was correct.

in the top left panels of Figures 14 and 15, respectively. For the *LRT*, the PPP-values that represent null comparisons are shown in (a) the first column within the top row and (b) the second column in the second row of Figure 24.

On the basis of the marginal distributions of PPP-values, most of the discrepancy functions were unlikely to indicate poor fit of a data analysis model when it was aligned with the data generation model. Specifically, the distributions of PPP-values for the *CCC* (both marginal and conditional), R^2 functions (both marginal and conditional), and $SGDDM_M$ were centered around 0.5, a result that signifies solid data-model fit. The PPP-values for the $SGDDM_C$ were generally uniform but rarely exceeded 0.75. In contrast to the performance of the other absolute measures of data-model fit, the distribution of PPP-values for the *LR* in null situations was skewed in the positive direction. This result reflects a general tendency of the *LR* to indicate a model inadequate even when all features pertinent to the data generation model have been captured in the data analysis model.

Many of the patterns and relationships described thus far carried over into the proportions of extreme PPP-values, which in null situations, represents the Bayesian proxy for Type I error rates. Figures 16 through 18 display the proportions of extreme PPP-values in null situations for all absolute functions. For each type of data generation model, the proportion of PPP-values of the *CCC* (blue lines, both marginal and conditional), R^2 (red lines, both marginal and conditional), and $SGDDM_M$ (green lines with empty markers) were consistently at or close to zero. Aside from a small, but non-trivial increase for data generated with non-normal outcomes, a similar result held for the conditional $SGDDM_C$ (green line with empty markers). Although the effect was larger

than that for the $SGDDM_C$, the proportion of extreme PPP-values for the *LR* (black line with empty markers) also increased for data non-normally distributed data. Unlike all other measures, the proportion of extreme PPP-values were strongly impacted by the size of the quadratic variance irrespective of the levels of other factors in the design. Specifically, the proportions of PPP-values for the *LR* were about 0.50 when the quadratic variance was small and close (or equal) to zero when the quadratic variance was large.

The red bars in Figure 25 (linear GCM data) and the blue bars in Figure 26 (GCM with equal strength for all individuals) represent the relevant null comparisons for the *LRT*. Apart from the methodological challenges of evaluating relative fit via the *LRT* in the PPMC context and the presence of negative values, the *LRT* was quite unlikely to result in overspecifying the complexity of the functional form for the marginal mean structure. As seen in Figure 25, the PPP-values for the comparison between M1 (linear GCM) and M2 (GCM with equal strength of the quadratic effect) were very low across all conditions. However, if an analyst were to take the unlikely route of then comparing M2 and M3 (the blue bars) for data that follow a linear GCM, the decision may result in the selection of a substantially overparameterized model, particularly for non-normal data. One possible explanation for this result is that a GCM that allows the strength of the quadratic effect to vary might actually result in better approximation to the linear conditional mean structure. In comparing a M2 to M3 (GCM with varying strength of the quadratic effect), the *LRT* (the blue bars in Figure 26) reflected a tendency to unnecessarily favor a more complex conditional mean structure. As was the case for the

other null comparison, this effect was exacerbated for data generated with non-normal outcomes.

As just described, the shape of the distribution for observed measures impacted the *LR*, $SGDDM_C$, and the *LRT* such that proportions of extreme PPP-values were larger for data generated with non-normally distributed data than for normally distributed data. On the one hand, this effect may reflect a tendency to deem a data-analysis model inadequate even when it captures the key features of the data generation process. On the other hand, irrespective of the features used to generate and analyze the data, the likelihood of the score for a particular individual at a particular measurement occasion was assumed to have arisen from a normal distribution. As a result, the increase in the proportion of extreme PPP-values may reflect a true discrepancy between the data and the specification of the likelihood.

Performance in Non-Null Conditions. Excepting to the *LRT* (see Liu et al., 2012), all of the existing literature reporting on the performance of discrepancy functions for growth curve models in non-null situations have targeted the behavior of realized values. Among the discrepancy functions that have been investigated in other work, some patterns were consistent with previous research while others were not. As was the case for the discussion of null results, the initial part of this discussion for non-null situations serves to compare and contrast the results of the current study to the existing literature. The novel aspects of the current study will follow from there.

Wu and West (2013) conducted the only existing research on the performance of the *LR*, *CCC* (both conditional and marginal), and R^2 (both conditional and marginal) in non-null situations. Starting with the *LR*, the realized values were impacted by a three-

way interaction between sample size, the strength of the quadratic mean, and the size of the residual variance when the marginal mean structure was underspecified (see Figure 9 in this work and Table 5 in Wu & West, 2013). The nature of the interaction was such that increasing the strength of the quadratic mean resulted in a larger difference in the realized of the *LR* between the levels of the residual variance (though larger residual variances always produced larger realized values of the *LR*, on average). Within the levels of the quadratic mean strength, increasing sample size served to widen the gap between the levels of the residual variance.

In the current work, the realized values of the *LR* were also impacted by a three-way interaction between sample size, the size of the quadratic variance, and the size of the residual variance (see Figure 11). The nature of the interaction was such that increasing the size of quadratic variance resulted in a smaller difference in *LR*s between the levels of the residual variance (though larger residual variances always produced larger realized values of the *LR*, on average). Within the levels of the quadratic variance size, increasing sample size served to widen the gap between the levels of the residual variance. Since none of the manipulated variables had any bearing on the realized values of the *LR* when only the conditional mean structure was underspecified, this falls in contrast to the results reported by Wu and West (2013).

Wu and West (2013) found that the realized values of the CCC_C and R_C^2 were sensitive to misspecification in the conditional mean structure, but the degree in sensitivity was attenuated with increasing residual variation. As seen in Table 6, the realized values of the CCC_C and R_C^2 were not impacted by the degree of misspecification, which was captured by the size of the quadratic variance (denoted QV). The only factor

that impacted the realized values of these functions was the size of the residual variance, which served to decrease the realized values; this finding was consistent with the results reported by Wu and West (see Table 7 in their work). The lacking effect of the quadratic variance on the realized values of the CCC_C and R_C^2 may be due to differences in how an underspecified conditional mean structure was defined. In Wu and West, the underspecified conditional mean structure was defined by fixing the quadratic variance to zero in the data analysis model for data in which (a) the form of the true marginal mean structure was linear but (b) the quadratic variance was some positive value. A similar approach for creating an underspecified conditional mean structure was used for the current work, but the true marginal mean structure exhibited a quadratic relationship with time. Notably, the strength of the quadratic mean did impact the realized values of the CCC_C and R_C^2, such that lower values were observed in an all situations in which the marginal mean structure was underspecified (see Tables 5 and 7 above). Since the impact of misspecifications in only the marginal mean structure were not reported for the CCC_C and R_C^2, it remains an open question what unique role the strength of the quadratic mean had in the study conducted by Wu and West.

The CCC_M and R_M^2 were impacted by the strength of the quadratic mean in all cases in which the data analysis model underspecified some aspect of the functional form (see Tables 5 – 7). Excepting to the situation in which both the marginal and conditional mean structures were underspecified (see Table 8 in Wu & West, 2013), this finding

varies from the results reported on the raw realized values[4] for these functions in the study conducted by Wu and West (2013), which were generally not found to be impacted by the degree of misspecification.

Despite some key differences in the simulation design, the behavior of the realized values of the *LRT* in non-null situations found in the current work align with the results reported by Leite and Stapleton (2011). Regardless of whether the misspecification came from an underspecified marginal mean structure (see Table 1 in Leite & Stapleton, 2011) or an underspecified conditional mean structure (see Table 2 in Leite & Stapleton, 2011), the realized values of the *LRT* were impacted by an interaction between sample size and the degree of underspecification. Notably, the amount of residual variation, which was not part of Leite and Stapleton's design, was found to moderate this interaction. The relevant results in the current work can be found for the comparison between M1 and M2 in Tables 9 and for the comparison between M2 and M3 in Table 10. The nature of the interactions were such that (a) the difference across the levels of varying the degree misspecification became larger with sample size and (b) the impact of increasing the degree of misspecification was lessened with increased residual variation.

[4] When only the marginal mean structure was underspecified, the impact of misspecification severity was only realized for the CCC_M and R^2_M functions when compared to the respective values when the correct data analysis model was applied (see Table 6 in Wu & West, 2013).

Among the collection of discrepancy functions, only the *SGDDM* functions (both conditional and marginal) have not been pursued in previous methodological work pertaining to GCM. Notably, once conditioned on the data generation model, the realized values of the $SGDDM_M$ were identical irrespective of the data analysis model. This matter is briefly described above in the section pertaining to null conditions and is described in greater detail below. To avoid redundancy, the patterns in the realized values for the $SGDDM_M$ are not described in non-null situations.

The realized values of the $SGDDM_C$ were sensitive to underspecification, and when the relationship between data generation and analysis models made it possible, and specific to the key sources of underspecification. The general behavior of the realized values of the $SGDDM_C$ was such that the factor(s) representing the source(s) of underspecification interacted with the amount of residual variation. For the situation in which only the marginal mean structure was underspecified (see Table 5 and Figure 8), the realized values of the $SGDDM_C$ were impacted by the interaction between the strength of the quadratic mean and the residual variance. For the situation in which only the conditional mean structure was underspecified (see Table 6 and Figure 10), the realized values of the $SGDDM_C$ were impacted by the interaction between the amount of quadratic variation and the residual variance. In both of these cases, increasing the degree of underspecification resulted in higher realized values; however, increasing the amount of residual variation resulted in bigger declines for the levels associated with a larger degree of underspecification.

As seen in Figure 12, the pattern of results for the $SGDDM_C$ when both mean structures were underspecified highlight the capacity to be specific to the sources of

misspecification, particularly when one source was small and the other was large. That is, given a small quadratic mean (variance), the $SGDDM_C$ became larger with an increase in the amount (strength) of quadratic variation (the quadratic mean). The $SGDDM_C$ was the only absolute function that exhibited this behavior.

As was the case for null conditions, no existing research has investigated the performance of the discrepancy functions within the GCM context in terms of making a decision about model adequacy (such has NHST) in non-null situations. In terms of the marginal distributions of PPP-values, the $SGDDM_C$ (the second through fourth rows in third column of Figure 14), $SGDDM_M$ (panels in the second row and second column of Figure 15), *LR* (panels in the second row and second column of Figure 16), and *LRT* (second row in the first column and bottom row in Figure 24) were the most promising for detecting misspecifications for either mean structure. For each of these functions, the masses of the PPP-value distributions were heavily concentrated close to zero given an underspecified data analysis model (in the case of the absolute measures listed above) or non-null comparison (in the case of the *LRT*). With distributions of PPP-values centered around 0.50 in non-null situations, the worst functions were the CCC_C and R_C^2 (second through fourth rows in the first and second columns of Figure 14). Although the levels of the quadratic strength impacted the distribution of PPP-values for the CCC_C and R_M^2 (second through fourth rows in the fourth and fifth columns of Figure 14), the PPP-values were not sufficiently high to be deemed extreme even when the quadratic mean was large.

Given a data analysis model that underspecifies some aspect of the data generation model, the proportion of extreme PPP-values are analogous to the NHST

concept of statistical power. Among the measures of absolute data-model fit (see Figures 19 – 21), the *CCC* and R^2 functions (both marginal and conditional) were consistently the worst performing discrepancy functions (the blue and red lines, respectively). Regardless of the relationship between the data generation and analysis model, the proportion of extreme PPP-values were consistently close or equal to zero. In contrast, the proportion of extreme PPP-values for the *LR* (black line with empty markers) were typically quite high and were often close to one. As seen in Figure 21, the only exception to this performance for the *LR* occurred with a small quadratic variance and large residual variance—particularly unfavorable conditions—when the condition mean structure was underspecified.

With rare exception, both *SGDDM* functions (green lines) almost always detected model underspecification provided that at least one source of underspecification was large. When only the marginal (conditional, see Figure 20) mean structure (see Figure 19) was underspecified, the proportions of extreme PPP-values were consistently close to one when the quadratic mean (variance) was large (see the second and fourth columns in the figures cited above). However, the sensitivity of these functions diminished with decreased underspecification (see the first and third columns in Figures 19 and 20), particularly when the residual variance was large (see the third column in Figures 19 and 20). A similar result obtained when both mean structures were underspecified (see Figure 21). That is, if the quadratic mean was small but the quadratic variance was large (second and fourth columns in the first and third rows of Figure 21), or vice versa (first and third columns in the second and fourth rows of Figure 21), the proportions of extreme PPP-values were at or close to one. With both mean structures underspecified, sensitivity for

both *SGDDM* functions only deteriorated when the degree of underspecification for both mean structures was small (first and third columns in the first and third rows of Figure 21), particularly when the residual variance was large (third column in the first and third rows of Figure 21). In cases of underspecification, the $SGDDM_C$ (green line with empty markers) generally outperformed the $SGDDM_M$ (green line with solid markers) when the degree of underspecification was small irrespective of the relationship between the data generation and analysis models.

Implications for Applied Researchers

On the Importance of Selecting/Engineering Discrepancy Functions. As recognized by Levy (2011), and exemplified in the current work by the $SGDDM_M$, the selection/engineering of discrepancy functions requires careful thought. With the balanced structure considered in this work, the intent of the $SGDDM_M$ was not realized. By exchanging the model-implied *means* in place of the model-implied *scores*, the logic for constructing the $SGDDM_M$ mirrored that of the CCC_M and R^2_M functions. Unlike the latter two functions, which target the diagonal elements of the **R** matrix, the intent of the $SGDDM_M$ was to measure the portion of the off-diagonal elements in the **R** matrix that was due to misspecifications of the marginal mean structure.

At the extreme of a completely balanced structure for time, such as that considered for the current work, the realized values of the $SGDDM_M$ will be identical regardless of the means from which observed scores were subtracted. As mentioned above, the implication is that the observed sample means would also produce the same realized value of the $SGDDM_M$; this amounts to unintentionally computing the $SGDDM_M$ using the observed covariance matrix. At the other extreme, if conditions with completely

unbalanced time structure had been pursued, the $SGDDM_M$ would be equal to the $SGDDM_C$. Thus, the intention and utility of the $SGDDM_M$ may be realized in situations in that lie somewhere in between a completely *balanced* structure for time and a completely *unbalanced* structure for time. One example of such a time structure would obtain by centering age (assuming integer values) at a particular value. This has the potential to yield "groups" that are distinguished by a particular vector of time scores such that cases within groups share the same vector of time scores and cases between groups have some other vector of time scores. In this case, the $SGDDM_M$ would be the same across cases within groups but differ between groups. More importantly, the $SGDDM_M$ would yield a perspective about data-model misfit that is not characterized by the observed covariance matrix (as with a completely balanced design for time) or the $SGDDM_C$ (as with completely unbalanced design for time).

Recommendations for Practice. A central goal of this work was to provide recommendations for practice. To this end, one of the motivating factors that guided the selection of discrepancy functions was the need to identify functions that could disentangle misspecification in the marginal mean structure from misspecification in the conditional mean structure. Drawing from the existing literature, the CCC_C and R_C^2 were selected to target the fit of the conditional mean structure, and the CCC_M and R_M^2 were selected to target the fit of the marginal mean structure. With the same intent, the current work pursued the $SGDDM_C$, which has shown promise for other types of multivariate models, and the $SGDDM_M$, which was newly constructed for investigation in the current work. Finally, the sources of misfit could also be targeted by the *LRT*, which derives meaning from the relationship between two models that are being compared.

Unfortunately, on the basis of the results for this work, these ideals that would allow the sources of misspecification to be separated were not realized as cleanly as desired. On the basis of the results, the performances of the discrepancy functions are better suited for a broader set of recommendations.

Owing to particularly poor performance, the *CCC* and R^2 (both marginal and conditional) are not recommended for use as a diagnostic for the criticism of growth curve models. On the basis of the proportion of PPP-values, these functions are just as likely to deem a data analysis model as adequate regardless of whether a key process of the data has been captured. Moreover, although the conditional and marginal versions of these functions target different sources of data-model fit, there was minimal evidence that these functions exhibit the capacity to disentangle underspecification between the two mean structures. The key exception was the distribution of PPP-values of the CCC_M and R^2_M, which consisted of two modes corresponding to the strength of quadratic mean. However, none of the PPP-values were extreme enough to indicate meaningful underspecification.

In contrast to the *CCC* and R^2 functions, the *LR*, *SGDDM* functions (particularly the conditional version), and the *LRT* exhibited some promise for critiquing the fit of growth curve models. Generally speaking, these functions were sensitive and specific to the sources of underspecification in terms patterns in the realized values and PPP-values. This is especially true if one grants the possibility that the increased PPP-values for non-normal data in null conditions observed for each of these functions reflects a true misspecification between the distributional specification of the likelihood and the data at

hand. In light of these benefits, it is important to be mindful of the unique *practical* disadvantages associated with each of these functions.

Unlike any of the other discrepancy functions considered in this work, the key disadvantage of *LR* is that it cannot be computed with unbalanced time structures owing to the absence of unique model-implied covariance matrix. The *LRT* has two key disadvantages. First, the realized values in the PPMC framework may be negative even if the models being compared reach convergence. Second, in comparison to the other discrepancy functions considered in this work, computing the *LRT* is procedurally complicated in the PPMC framework. Since there is no meaningful improvement in performance over measures of absolute fit, the additional complexity of computing the *LRT* may not be worth doing. The key disadvantage of the *SGDDM* functions, which also happens to be true of the *LR*, is the matter of isolating sources of misfit. Given that these functions are applied in the absolute sense, evidence of model inadequacy does not indicate the source of data-model misfit that gives rise to the unaccounted for conditional associations. This critique, however, can be overcome with the use of a model building strategy that typifies standard applications of growth curve modeling.

Study Limitations

There were a number of features of the current work that explored new frontiers in the space of model criticism for GCM. This work is the first to methodologically explore a Bayesian approach for the criticism of growth curve models; moreover, many of the discrepancy functions have received very limited attention in the existing research while others have not yet been suggested for use. In part due to the relative novelty of the current work, some design choices were limited to conditions that have been considered

in other methodological work to support comparisons across studies; this may limit the generalizability of the work. Some reasonable extensions to the current work—in both Bayesian and frequentist frameworks—would involve investigating the impact of missing data and different structures for the passage of time. Both of these issues have thus far received no attention with respect to matters of data-model fit for growth curve models, as far as the author knows. For the current work, the manipulation of distributional shape simultaneously altered the skew and kurtosis; the particular values used to generate data resulted in the generation of positively skewed and leptokurtic data. Reasonable extensions may involve varying the type and strengths of skewness and kurtosis, particularly such that the unique impact of these two features of distributions can be investigated. Finally, a reasonable Bayesian extension may involve (a) evaluating the impact of different prior distribution specifications and (b) pursuing Bayes Factors (see Gill, 2007) as a tool within the Bayesian analysis framework for evaluating the relative fit of models outside of the PPMC framework.

Concluding Remarks

On the basis of the results for the current work, and drawing from the extant methodological literature, separating misfit in the conditional mean structure from the marginal mean structure remains challenging. The task of disentangling between these sources of data-model misfit is particularly challenging when the sources of underspecification (e.g., strength of the mean for an underspecified effect of the marginal mean structure, amount of variation for an underspecified effect for the conditional mean structure) are small, particularly to the extent that the data unreliable (i.e., large residual variance). The applied analyst is best served by sequentially adding terms, and at each

step, evaluating the results for multiple diagnostic measures. The results of this research suggest that *SGDDM* functions (particularly the $SGDDM_C$) and the *LR* would serve as useful diagnostics at each step for critiquing the model on its own merit. The *LRT* is also a useful diagnostic for evaluating the relative improvement in fit between two models adjacent to one another in the model-building process. It is important to bear in mind, however, that a model that offers substantially better fit to the observed data does not necessarily mean that the superior model fits the data well in the absolute sense.

A reasonable question, particularly for applied analysts who may be less familiar with Bayesian analysis methods, is whether the additional time and effort required to conduct a PPMC analysis is worth the effort. One way to overcome this concern is the availability of code to conduct a PPMC analysis using the functions investigated in this work. To this end, all of the requisite components for estimating the Bayesian hierarchical model and conducting PPMC are appended below (see Appendixes C and D for estimation; see Appendixes E and F for conducting PPMC). The code is also available from the author (or the dissertation chair) upon request. In some cases, such as with the *LRT* (and other relative fit measures more broadly), the advantages unique to PPMC—such as the propagation of uncertainty in the model parameters—may not be worth the added complexity of implementation. However, in the absence of known reference distributions—such as with the $SGDDM_C$—the capacity to empirically construct the reference distribution extends the analyst's toolkit for critiquing the fit of growth curve models. Indeed, it is this flexibility of PPMC for model criticism that represents a natural complement to the increasingly generalized versions of foundational growth curve models often encountered in applied research settings.

References

Baker, G.A. (1954). Factor analysis of relative growth. *Growth*, *18*, 137-143.

Bauer, D.J., & Curran, P.J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338-363.

Bentler, P.M. (1990). Comparative fit index in structural models. *Psychological Bulletin*, *107*, 238-246.

Bentler, P.M., & Yuan, K.-H. (2000). On adding a mean structure to a covariance structure model. *Educational and Psychological Measurement*, *60*(3), 326-339.

Bollen, K.A., & Curran, P.J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley.

Brooks, S.P. (1998). Markov Chain Monte Carlo and its applications. *The Statistician*, *47*, 69-100.

Casalla, G., & George, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician*, *46*, 167-174.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, *90*, 1313-1321.

Cheong, J., MacKinnon, D.P., & Khoo, S.T. (2003). Investigation of meditational processes using parallel process latent growth curve modeling. *Structural Equation Modeling*, *10*, 238-262.

Chou, C.-P., Bentler, P.M., & M.A. Pentz, M.A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis. *Structural Equation Modeling*, *5*(3), 247–66.

Coffman, D.L., & Millsap, R.E. (2009). Evaluating latent growth curve models using individual fit statistics. *Structural Equation Modeling*, *13*(1), 1-27.

Collins, L.M., & Lanza, S.T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Hoboken, NJ: Wiley.

Crawford, A. (2014). *Posterior Predictive Model Checking in Bayesian Networks* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confir- matory factor analysis. *Psychological Methods, 1,* 16–29.

Depaoli, S. (2010). *Specification Issues in Bayesian Growth Mixture Modeling* (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses.

Depaoli, S. (2012). The ability of posterior predictive checking to identify model misspecification in Bayesian growth mixture modeling. *Structural Equation Modeling, 19,* 534-560.

Duncan, S.C., & Duncan, T.E. (1996). A multivariate latent growth curve analysis of adolescent substance use. *Structural Equation Modeling*, *3*, 323-347.

Embretson, S.E., &Reise, S.P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates, Inc.

Gelfand, A.E., & Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398-409.

Gelman, A., Meng, X., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistical Sinica*, *6*, 733-807.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721-741.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice*. London, England: Chapman & Hall.

Gill, J. (2007). *Bayesian Methods: A social and behavioral sciences approach* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Goel, P.K. (1983). Information measures and Bayesian hierarchical models. *Journal of the American Statistical Association*, *78*, 408-410.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*, 97-109.

Hjort, N., Dahl, F. A., & Steinbakk, G. H. (2006). Post-processing posterior predictive p values. *Journal of the American Statistical Association*, *101*, 1157–1174.

Ip, E. H. (2000). Adjust for information inflation due to local dependency in moderately large item clusters. *Psychometrika*, *65*¸79-91.

Jannsen, R., Tuerlinckx, F., Meulders, M., & De Boeck, P. (2000). A hierarchical IRT model for criterion-referenced measurement. *Journal of Educational and Behavioral Statistics, 25,* 285-306.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 192-202.

Jöreskog, K.G., & Sörbom, D. (1981). *LISREL VI: Analysis of linear structural relationship by maximum likelihood and least squares method*. Chicago: National Educational Resources.

Jöreskog, K.G., & Sörbom, D. (1984). *LISREL VI user's guide*. Mooresville, IN: Scientific Software.

Jöreskog, K.G., & Sörbom, D. (1985). *LISREL VI User's Guide*. Mooresville, IN: Scientific Software Inc.

Kline, R.B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guildford Press.

Leite, W.L., & Stapleton, L.M. (2011). Detecting growth shape misspecifications in latent growth models: An evaluation of fit indexes. *The Journal of Experimental Education, 79*, 361-381.

Levy, R. (2006). *Posterior predictive model checking for multidimensionality in Item Response Theory and Bayesian networks* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses.

Levy, R. (2009). The rise of Markov Chain Monte Carlo estimation for psychometric modeling. *Journal of Probability and Statistics*, 1-18.

Levy, R. (2011). Bayesian data-model fit assessment for structural equation modeling. *Structural Equation Modeling: A multidisciplinary journal, 18*, 363-685.

Levy, R., Crawford, A. V., Fay, D. M., & Poole, K. L. (2011, April). Data-model fit assessment for Bayesian networks for simulation-based assessment. In R. J. Mislevy (Chair), *Modeling strategies in a complex automated performance assessment environment*. Symposium conducted at the annual meeting of the American Educational Research Association, New Orleans, LA.

Levy, R., & Svetina, D. (2011). A generalized dimensionality discrepancy measure for dimensionality assessment in multidimensional item response theory. *British Journal of Mathematical and Statistical Psychology, 64*, 208-232.

Levy, R., Xu, Y., Yel, N., & Svetina, D. (2015). A standardized generalized dimensionality discrepancy measure and a standardized model-based covariance for dimensionality assessment for multidimensional models. *Journal of Educational Measurement, 52,* 144-158.

Lin, L. A. (1989). A concordance correlation coefficient to evaluate reproducibility. Biometrics, 45, 255–268.

Liu, S., Rovine, M.J., & Molenaar, P.C.M. (2012). Using fit indexes to select a covariance model for longitudinal data. *Structural Equation Modeling, 19,* 633-650

MacCallum, R.C., & Kim, C. (2000). Modeling multivariate change. In T.D. Little, K. U. Schnabel, & J. Baumert (Eds), *Modeling longitudinal and multilevel data: Practical Issues, applied approaches and specific examples* (pp. 51-68). Mahwah, NJ: Lawrence Erlbaum.

MacCallum, R.C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J.K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, *32*, 215-253.

Mäkikangas, A., Bakker, A.B., Aunola, K., & Demerouti, E. (2010). Job resources and flow at work: Modelling the relationship via latent growth curve and mixture model methodology. *Journal of Occupational and Organizational Psychology*, *83*, 715-814.

Maxwell, S. E., & Delaney, H. D. (2003). Designing experiments and analyzing data: A model comparison perspective. Mahwah, N.J: Lawrence Erlbaum.

McArdle, J.J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nessroade & R. B. Cattell (Eds.), *The Handbook of Multivariate Experimental Psychology*, 2nd ed. (pp. 561-614). New York: Plenum Press.

McArdle, J.J. (1989). A Structural modeling experiment with multiple growth functions. In R. Kanfer, P.L. Ackerman, & R. Cudeck (Eds.), *Abilities, motivation, and methodology: The Minneapolis symposium on learning and individual differences* (pp. 71-117). Hillsdale, NJ: Lawrence Erlbaum.

McDonald, R. P. (1994). Testing for approximate dimensionality. In D. Laveault, B. Zumbo, M. Gessaroli, & M. Boss (Eds.), *Modern theories of measurement: Problems and issues* (pp. 63-85). Ottawa, Canada:University of Ottawa Press.

McMurray, K. (2010). *A Monte Carlo study of fit indices in hierarchical linear models* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses.

Meng, X. L. (1994). Posterior predictive p-values. *The Annals of Statistics*, *22,* 1142-1160.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107-122.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machine. *Journal of Chemical Physics*, *21*, 1097-1091.

Muthén, L.K., & Muthén, B.O. (1998-2010). *M*plus *user's guide* (6th ed.) Los Angeles, CA: Muthén & Muthén.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*, 463-469.

Nagin, D. (1999). Analyzing development trajectories: A semi-parametric, group-based approach. *Psychological Methods*, *4*, 139-177.

Nyland, K.L., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monthe Carlo simulation study. *Structural Equation Modeling*, *14*, 535-569.

Plummer, M. (2013). *JAGS Version 3.4.0 user manual*. http://www.stats.ox.ac.uk/~nicholls/MScMCMC14/jags_user_manual.pdf

Preacher, K.J., Wichman, A.L., MacCallum, R.C., Briggs, N.E. (2008). *Latent growth curve modeling*. Thousand Oaks, CA: Sage.

R Development Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-9000051-07-0). Retrieved from http://www.R-project.org.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2^{nd} ed.). Thousand Oaks, CA: Sage.

Rovine, M.J., & Molenaar, P. C. M. (2000). A structural equation modeling approach to a multilevel random coefficient model. *Multivariate Behavioral Research*, *35*, 51-88.

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.

Rupp, A. A., Levy, R., DiCerbo, K. E., Sweet, S., Crawford, A. V., Calico, T., Benson, M., Fay, D., Kunze, K. L., Mislevy, R. J., & Behrens, J. T. (2012). Putting ECD into practice: The interplay of theory and data in evidence models within a digital learning environment. *Journal of Educational Data Mining, 4(1), 49-110*.

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology, 22,* 249 –278.

Schaeffer, C.M., Petras, H., Ialongo, N., Poduska, J., & Kellam, S. (2003). Modeling growth in boys' aggressive behavior across elementary school: Links to later criminal involvement, conduct disorder, and antisocial personality disorder. *Developmental Psychology*, *39*(6), 1020-1035.

Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, *64*, 37–52.

Shevlin, M., & Millar, R. (2006). Career education: An application of latent curve modeling to career information-seeking behavior of school pupils. *British Journal of Educational Psychology*, *76*, 141-153.

Singer, J.D., & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Sinharay, S. (2005). Assessing fit of unidimensional item response theory models using a Bayesian approach. *Journal of Educational Measurement*, *42*, 375-394.

Sinharay, S., Johnson, M., & Stern, H.S. (2006). Posterior predictive model checking in item response theory models. *Applied Psychological Measurement, 30,* 298-321.

Spiegelhalter, D.J., Thomas, A., Best, N.G., & Lunn, D. (2007). *WinBUGS user manual: Version 1.4.3.* Cambridge, UK: MRC Biostatics Unit.

Stan Development Team. 2015. *Stan Modeling Language Users Guide and Reference Manual, Version 2.8.0.*

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Stoel, R. D., Galindo-Garre, F., Dolan, C., & Van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods*, *4*, 439–455.

Stout, W. F. (1987). A nonparametric approach for assessing latent trait dimensionality. *Psychometrika*, *52*, 589-617.

Titterington, D. M., Smith, A.F., and Makov, U.E. (1985). *Statistical analysis of finite mixture distributions*. New York: John Wiley & Sons.

Tucker, L.R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1-10.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distribu- tions. *Psychometrika, 48,* 465–471.

Verbeke, G., & Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, *59*, 254-262.

Vonesh, E. F. (1992). Nonlinear models for the analysis of longitudinal data. *Statistics in Medicine*, *11*, 1929-1954.

Vonesh, E.F., Chinchilli, V.M., & Pu, K. (1996). Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics*, *52*, 572-587.

Willett, J.B., & Sayer, A.G. (1994). Using covariance structure analysis to detect correlates of predictors of individual change over time. *Psychological Bulletin*, *116*, 363-381.

Wirth, R.J., & Edwards, M.C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58-79.

Wu, W., & West, S. G. (2010). Sensitivity of fit indices to misspecification in growth curve models. Multivariate Behavioral Research, 45, 420–452.

Wu, W., & West, S.G. (2013). Detecting misspecifications in mean structures for growth curve models: Performance of Pseudo R^2s and concordance correlation coefficients. *Structural Equation Modeling*, *20*(3), 4455-478.

Wu, W., West, S.G., & Taylor, A.B. (2009). Evaluating model fit for growth curve models: Integration of fit indices from SEM and MLM frameworks. *Psychological Methods, 14,* 183-201.

You, S., & Sharkey, J. (2009). Testing developmental-ecological model of student engagement: a multilevel latent growth curve analysis. *Educational Psychology*, *29*(6), 659-684.

Yuan, K.-H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research, 40,* 115–148.

Zhang, J., & Stout, W. F. (1999a). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika*, *64*¸ 129-152.

Zhang, J., & Stout, W. F. (1999b). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika*, *64*¸213-249.

APPENDIX A

THE MULTILEVEL MODELING AND STRUCTURAL EQUATION MODELING PARAMETERIZATIONS FOR GROWTH CURVE MODELS

The Multilevel Modeling Parameterization

A key assumption of ordinary least squares (OLS) regression modeling is that the units of analysis are *independent* observations from a homogeneous population. However, data in the social sciences often exhibit a nested (other common terms include clustered and hierarchical) structure among units of analysis; the implication of such structures is the potential for units at lower levels of analysis to be *dependent* on higher units of analysis. From the domain of education, students (level-1) are nested within schools (level-2) that are in turn nested within districts (level-3). In clinical settings, patients (level-1) are nested within clinicians (level-2). As described in greater detail below, repeated measures over time (level-1) can be viewed as nested within individuals (level-2). When the data exhibit a multilevel structure, the scores on an outcome measure at lower levels of analysis have the potential to be dependent on membership in higher units of analysis. For example, the scores among students within schools may be more similar to each other than they are to the scores of students in other schools, and therefore, the scores for students are dependent in part on their membership in a particular school. When the outcome of interest has variation at multiple levels of analysis (e.g., two levels), but a model with too few levels is employed (e.g., one level), the standard errors associated with estimates of model parameters will be too small (Raudenbush & Bryk, 2002; Singer & Willett, 2006). The ultimate impact is of employing a model with too few levels is overstatement of confidence in the inferences drawn on the basis of the model.

With standard OLS regression as the foundation, multilevel modeling (MLM; Raudenbush & Bryk, 2002) extends the model to account for clustering among units of

analysis. The foundational MLM for growth includes two levels of clustering. The level-1 units include the vector of repeated measures and level-2 units are individuals. That is, repeated measures are nested within individuals. The level-1 model is formally given by:

$$y_{it} = \pi_{0i} + \sum_{j=1}^{J} \pi_{ji} TIME_{it}^{j} + r_{it.} \quad \text{(A.1)}$$

Conceptually, the score for individual i at time t (y_{it}) is characterized by a linear combination of predictors and a residual term (r_{it}). The collection of predictors includes an intercept term (π_{0i}) and a series of terms relating the passage of time to scores on the outcome of interest (π_{1i}, π_{2i},…, π_{Ji}). Each additional predictor relating time to the outcome introduces complexity to the form of growth as represented by a polynomial of degree j (e.g., π_{1i} = linear & $TIME_{it}^{1}$, π_{2i} = quadratic & $TIME_{it}^{2}$, π_{3i} = cubic & $TIME_{it}^{3}$). As in single-level OLS regression models, residuals are typically assumed $r_{it} \sim N(0, \sigma^2)$ and uncorrelated for any two measurement occasions.

The level-2 model allows for the potential that the association between the passage of time and the outcome to (potentially) vary between individuals. This is captured in the level-2 model by a vector of means for each of the level-1 coefficients (γ_{00}, γ_{10},…, γ_{J0}) and a vector of deviation scores (u_{0i}, u_{1i},…, u_{Ji}) from these means for each individual i as follows:

$$\begin{bmatrix} \pi_{0i} \\ \pi_{1i} \\ \vdots \\ \pi_{Ji} \end{bmatrix} = \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \\ \vdots \\ \gamma_{J0} \end{bmatrix} + \begin{bmatrix} u_{0i} \\ u_{1i} \\ \vdots \\ u_{Ji} \end{bmatrix}, \quad \text{(A.2)}$$

such that $u_i = [u_{0i}, u_{1i},\ldots, u_{Ji}] \sim N(0, \sum_{JJ})$. The value of a level-1 coefficient can be forced to be the same for all individuals (i.e., $\pi_{ji} = \gamma_{j0}$) by fixing the variance of the coefficient to

zero $\left(\text{i.e.}, \sigma_{u_j} = 0\right)$. The level-2 model can be substituted into the level-1 model to yield a combined MLM model as follows:

$$y_{it} = (\gamma_{00} + u_{0i}) + \sum_{j=1}^{J} TIME_{it}^{j}(\gamma_{j0} + u_{ji}) + r_{it} \quad \text{(A.3)}$$

The level-2 model is the feature of MLM that generalizes the standard analysis of variance (ANOVA) approach for modeling longitudinal data. The ANOVA model allows individuals to vary at the intercept (i.e., $\pi_{0i} = \gamma_{00} + u_{0i}$) but restricts growth to be equivalent for individuals (i.e., $\pi_{1i} = \gamma_{10}$).

The Structural Equation Modeling Parameterization

Structural equation modeling (SEM; Kline, 2005) is a framework that provides tremendous flexibility for specifying relationships among observed and/or latent variables. In doing so, analysts can assess the fit of specific models that align with substantive theory against observed data. A model is deemed to fit the data well to the extent that features of the data (i.e., measures of central tendency, measures of dispersion, and measures of association) implied by the model reproduce the corresponding features of the observed data. Systematic discrepancies between the implied and observed features of data represent characteristics of the observed data that are not represented well by the model.

The SEM approach for modeling longitudinal data draws on confirmatory factor analysis (CFA), a class of measurement models that summarize the associations among dependent measures through one or more latent variables. Dependent measures usually include observed measures, but may also include other latent variables when higher-order latent variables are specified (Kline, 2005). With CFA as the basis, the parameters that

govern growth are viewed as independent latent variables that must be inferred from the set of dependent observed *T* repeated measures (Bollen & Curran, 2006; Meredith & Tisak, 1990; Preacher, Wichman, MacCallum & Briggs, 2008). Unlike typical applications of CFA models, which only include a covariance structure, the CFA model for growth also includes a mean structure among the latent variables that represent the growth parameters. The means of the repeated measures are fixed to zero to identify the model, and accordingly, the marginal mean structure is entirely determined by the growth factors. To differentiate this model from standard CFA models, and to be consistent with common terminology, the CFA model for growth will herein be referred as the latent curve model (LCM).

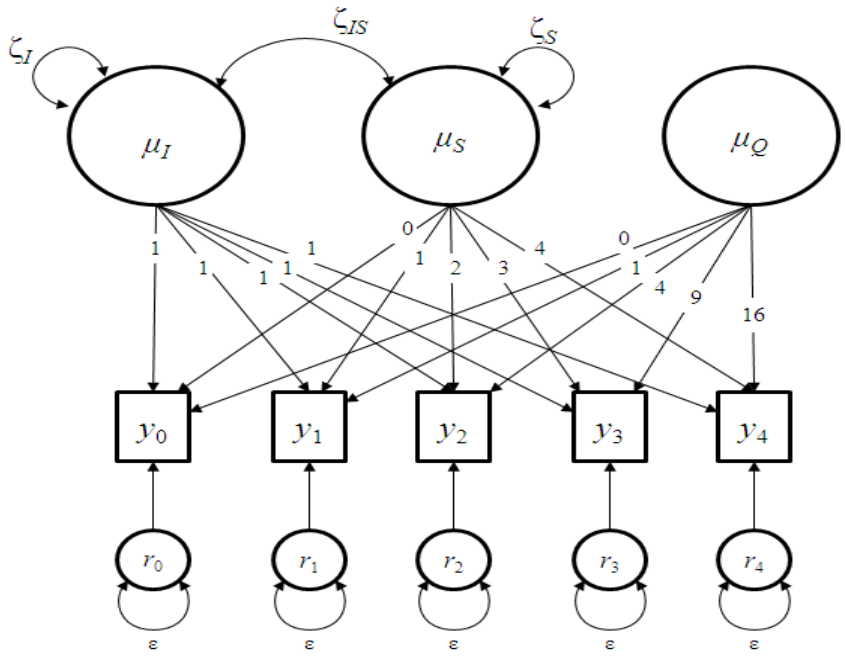

Figure A1. A path diagrammatic representing of a quadratic latent curve model.

Figure A1 shows a LCM with an intercept (μ_I), a linear slope (μ_S), and a quadratic effect (μ_Q); the intercept and slope parameters also include random effects (ζ_I, ζ_S,

respectively) and the potential for these effects to be correlated (ζ_{IS}). The single-headed arrows that emerge from the growth factors represent the unidirectional causal path to the observed repeated measures ($y_0, y_1,\ldots,y_{T-1}$). For the LCM shown, the intercept has an equal impact at each of the *T* occasions as indicated by a weight of 1 for each repeated measure. The integer weights for the slope factor reflect the assumption that the (a) interval between any two *adjacent* measurement occasions was the same and (b) all individuals had the same schedule of measurement occasions. The weights for the quadratic growth factor were obtained by simply squaring the weights for the linear slope growth factor. Finally, terms denoting residual scores ($r_0,\ldots, r_{T-1}$) are also shown as causal variables of observed measures. For the particular model shown, residuals are assumed to be homoskedasticly distributed (as shown by a common error variance, ε) and uncorrelated for all pairs of measurement occasions (as represented by the absence of curved arrows between error terms).

Defining the Data Model, Mean Structures, and Covariance Structures of LGM. The LCM can be viewed as consisting of three general structures that collectively give yield a model for describing longitudinal data. The three general structures include the data model, the mean structure, and the covariance structure; of these structure are described in turn in this section.

Data Model. Let y_i be the *T* (the total number of measurement occasions) × 1 vector of repeated measures, Λ_i the $T \times M$ (the number of growth factors) matrix of regression coefficients that define the metric for *TIME*, η_i the $M \times 1$ vector of scores on the growth factors, and ε_i is a $T \times 1$ vector of residuals. With these ingredients, the *data model* (Preacher et al., 2008) is given by:

$$\mathbf{y}_i = \boldsymbol{\Lambda}_i \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i. \quad \text{(A.4)}$$

Further, let μ_η be the $M \times 1$ vector of factor means (i.e., μ_I, μ_S, μ_Q) and ζ_i the $M \times 1$ vector of deviation scores for individual i (i.e., ζ_I, ζ_S, ζ_{IS}). With these elements, η_i is characterized by a mean and deviation term as follows:

$$\boldsymbol{\eta_i} = \boldsymbol{\mu_\eta} + \boldsymbol{\zeta}_i. \quad \text{(A.5)}$$

Mean Structures. There are two mean structures, both of which obtain by marginalizing over the data model. The first mean structure (herein called the *conditional mean structure*) can be thought of as a regression model between the vector of observed repeated measures and time for each individual. The vector of expected values on the outcome for a given individual depends on the metric used for clocking the passage of time (Λ_i) and scores on the growth factors (η_i) as follows:

$$E(\mathbf{y}_i|\boldsymbol{\Lambda}_i, \boldsymbol{\eta}_i) = \boldsymbol{\Lambda}_i(\boldsymbol{\eta}_i). \quad \text{(A.6)}$$

In essence, the model that defines the conditional mean structure allows for the possibility of unique growth trajectories (the deviation term for individual i is retained, ζ_i) but does not allow for measurement error at any measurement occasion (the vector of residuals is dropped, ε_i).

The second mean structure (herein called the *marginal mean structure*) can be thought of as the regression model between the vector of the repeated measures and time *averaging* over individuals. Mechanically, the marginal mean structure obtains by integrating over the latent variables that represent the growth factors and is fully defined by Λ_i and the means of the growth factors (μ_η) as follows:

$$E\left(\bar{\mathbf{y}}_i|\boldsymbol{\Lambda}_i, \boldsymbol{\mu_\eta}\right) = \boldsymbol{\Lambda}_i \boldsymbol{\mu_\eta}. \quad \text{(A.7)}$$

In the case of a fixed assessment schedule for all individuals (e.g., 0, 1,…, $T - 1$), the vector of expected values will be the same for all individuals. However, in the case of varying assessment schedules across individuals, the vector of expected values will only be the same for individuals with identical assessment schedules; it is possible that there is a unique assessment schedule for each individual. It important to note that even when assessment schedules do vary over all individuals, the marginal and conditional mean structures have the potential to differ if deviation terms on the growth factors are present. The former is dependent on the means of the latent growth factors but not a deviation term; the latter is dependent on the latent growth factors *and* the deviation term.

Defining the Covariance Structures. LGM also includes three covariance structures. The first covariance structure is a $T \times T$ matrix that includes the variances and covariances of residuals as follows:

$$\boldsymbol{\Theta}_{\varepsilon} = \begin{bmatrix} \sigma_{11}^2 & & & \\ \sigma_{21}^2 & \sigma_{22}^2 & & \\ \vdots & \vdots & \ddots & \\ \sigma_{T1}^2 & \sigma_{T2}^2 & \dots & \sigma_{TT}^2 \end{bmatrix}. \tag{A.8}$$

The second covariance structure is an $M \times M$ matrix that includes the variances and covariances among the growth factors as follows.

$$\boldsymbol{\Psi} = \begin{bmatrix} \sigma_{II}^2 & & & \\ \sigma_{SI}^2 & \sigma_{SS}^2 & & \\ \vdots & \vdots & \ddots & \\ \sigma_{MI}^2 & \sigma_{MS}^2 & \dots & \sigma_{MM}^2 \end{bmatrix}. \tag{A.9}$$

The third matrix is simply the overall sample covariance matrix of the observed repeated measures (Σ_i), which is obtained from Λ_i, Ψ, and Θ_{ε} as follows:

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_i \boldsymbol{\Psi} \boldsymbol{\Lambda}_i' + \boldsymbol{\Theta}_{\boldsymbol{\varepsilon}}. \tag{A.10}$$

Overlap Between the MLM and SEM Approaches for Modeling Growth. Although the surface features between the MLM and SEM approaches for modeling growth differ, there is a high degree of overlap between the two parameterizations (Bauer, 2003; Bauer & Curran, 2002; MacCallum & Kim, 2000; Preacher et al., 2008; Raudenbush, 2001; Rovine & Molenaar, 2000). In the description provided above for each parameterization, the equations and notation largely draw on the standard representations used for each in applied and methodological research. However, the two parameterizations have identical expressions despite the use of different notation. For instance, the level-1 regression coefficients in the MLM parameterization ($\pi_{0i} = \gamma_{00} + u_{0i}$; $\pi_{1i} = \gamma_{10} + u_{1i}$; $\pi_2 = \gamma_{20}$) are equivalent to the elements of η_i in the data model used for the SEM parameterization ($\eta_{Ii} = \mu_I + \zeta_{Ii}$; $\eta_{Si} = \mu_S + \zeta_{Si}$; $\eta_Q = \mu_Q$, respectively). Owing to these and other equivalent expressions, the two parameterizations often yield identical results for many types of models for describing growth.

Distinctions between the MLM and SEM Approaches for Modeling Growth. Any model for characterizing growth can be represented equally by the MLM and SEM parameterizations at the equation level. The key distinctions between the parameterizations largely come to differences in software. As a result, it is currently the case that the onus falls on the applied researcher to select the software package that is better suited to the research question at hand. The remainder of section identifies the unique advantageous of each approach to modeling growth.

Arguments for the Use of the SEM Approach. An attractive feature of growth modeling irrespective of the parameterization is the ability to incorporate predictors of the variation in growth parameters. However, when interest lies in using the growth

parameters (assuming that growth varies over individuals) as predictors of other variables, SEM is currently the preferred choice for addressing such research questions with respect to the selection of software. A hallmark of SEM is the ability to use latent variables as either (a) outcomes that are regressed on other variables (whether observed or latent) or (b) as predictors on to which other variables (whether observed or latent) are regressed on. This ability derives from the perspective that latent variables, such as growth factors, are variables in their own right (Preacher et al., 2008) and may be part of a larger system of observed and latent variables. As implemented in MLM-based software packages, growth parameters can either serve as (a) predictors of the repeated measures or (b) as outcomes of predictors. However, owing to the equivalence of mathematical expressions between the SEM and MLM-based parameterizations, MLM-based estimates of growth parameters that vary between individuals can *theoretically* be used as predictors of other variables within a larger variable system. However, it is currently that case that such models are either difficult or impossible to specify using dedicated MLM software packages.

Related to the perspective that latent variables may be part of a larger system of relationships, it is not necessary that the repeated measures be observed. Interest may lie in characterizing growth on a set of repeated *latent* variables that are in turn measured by a set of indicators that are repeatedly collected. Assuming the meaning of the latent factor remains unchanged over time, the capacity to specify such models makes it possible to describe growth on a set of variables that are *not* presumed to be free of measurement error (Preacher et al., 2008).

Arguments for the Use of MLM Approach. The primary advantage of the MLM approach to characterizing growth over the SEM approach is the ability to incorporate additional levels of analysis. From the MLM perspective, the repeated measures are viewed as level-1 units of analysis that are nested within individuals, the level-2 units of analysis. The MLM approach can readily accommodate, for instance, a dependence structure in which measurement occasions (level-1) are nested within individuals (level-2) who are in turn nested within schools (level-3). Generally speaking, constructing models with more than two levels using the SEM approach can be quite difficult and even impossible for some software packages dedicated to constructing structural equation models (Preacher et al., 2008).

Looking Beyond Software and Tradition. As mentioned above, the close interconnections between the MLM and SEM approaches have been widely acknowledged (Bauer, 2003; Bauer & Curran, 2002; MacCallum & Kim, 2000; Preacher et al., 2008; Raudenbush, 2001; Rovine & Molenaar, 2000). The distinctions can be regarded as *practical* differences, as opposed to *statistical differences*, that arise on account of (a) historical traditions for the use of the MLM and SEM approaches to modeling growth and (b) differences associated with the dedicated software packages that define the perceived boundaries for each approach. Historically, MLM has roots in the inherently nested structure associated with data in the tradition of education. The SEM approach to modeling growth is tied to the tradition of psychology, which largely draws on the use of latent constructs to represent unobserved processes that give rise to the relationships among a set of observed variables. Many of the perceived differences between the MLM and SEM approaches to modeling growth are tied to software.

However, these differences are beginning to blur as dedicated MLM software (such as MLM) incorporates features commonly associated with LGM and dedicated SEM software (such as M*plus*) becomes more capable of fitting MLM (Preacher et al., 2008).

APPENDIX B

R CODE FOR DATA GENERATION

```
##############################################################################
###
## The following syntax generates growth curve data and estimates three models via
JAGS.
## The output from the three models are submitted to a PPMC analysis.
##############################################################################
###

################################################
## Load any necessary libraries.
################################################

library(MASS)

################################################
## Specify constants for the simulation.
################################################

# Number of measurement occasions.
J <- 5

# Number of trials.
trials <- 100

################################################
## Write all functions that are used.
################################################

# Simulating non-normal data.
fleishtarget <- function(x,a){
  b<-x[1]
  cc<-x[2]
  d<-x[3]
  g1<-a[1]
  g2<-a[2]
  (2 - ( 2*b^2 + 12*b*d + g1^2/(b^2+24*b*d+105*d^2+2)^2 + 30*d^2 ) )^2 +
    (g2 - ( 24*(b*d+cc^2*(1+b^2+28*b*d)+d^2*(12+48*b*d+141*cc^2+225*d^2)) )
)^2+
    (cc - (g1/(2*(b^2+24*b*d+105*d^2+2)) ) )^2
}

findbcd <- function(skew,kurtosis){
  optim(c(1,0,0),fleishtarget,a=c(skew,kurtosis),method="BFGS",
        control=list(ndeps=rep(1e-10,3),reltol=1e-10,maxit=1e8))
}
```

```
#############################################################################
#####
## The following specifies the conditions for generating data.
## Since they are not likely to be pursued in practice, conditions with a non-zero
quadratic ## variance will be removed if there is no quadratic mean.
#############################################################################
###

sample.size <- matrix(c("SN","MN","LN"))
quad.mean <- matrix(c("NQM", "SQM", "LQM"))
quad.sd <- matrix(c("NQV", "SQV", "LQV"))
res.sd <- matrix(c("SRV", "LRV"))
dist.shape <- matrix(c("N", "NN"))

conditions <- expand.grid(sample.size, quad.mean, quad.sd, res.sd, dist.shape)
colnames(conditions) <- c("samplesize", "quadmean", "quadsd", "ressd", "distshape")

attach(conditions)
conditions$remove <- ifelse(quadmean=="NQM" & (quadsd=="SQV"|quadsd=="LQV"),
1, 0)
detach(conditions)
conditions <- subset(conditions, remove==0, select=-remove)

# SPECIFY THE GENERAL FORMS OF DIRECTORIES TO CREATE AND WRITE
TO.
gendir <- "/Volumes/Seagate Backup Plus Drive/GCM Dissertation"
dir.create(gendir)

# Time the data generation and model estimation component.
system.time(

  # Open the loop over conditions.
  for(which.cond in 1:nrow(conditions)){

    # Specify directories that will be used for storing data.
    conddir <- paste(paste(gendir, "/", sep=""),
                     paste(conditions[which.cond,1], conditions[which.cond,2],
conditions[which.cond,3],
                           conditions[which.cond,4], conditions[which.cond,5], sep="_"), sep="")

    datadir <- paste(conddir, "/data", sep="")
    observeddatadir <- paste(datadir, "/", "ObservedData", sep="")

    dir.create(conddir)
```

```
   dir.create(datadir)
   dir.create(observeddatadir)

   # Specify directories that will be used for storing data.
   conddir <- paste(paste(gendir, "/", sep=""),
                paste(conditions[which.cond,1], conditions[which.cond,2],
conditions[which.cond,3],
                      conditions[which.cond,4], conditions[which.cond,5], sep="_"), sep="")

   datadir <- paste(conddir, "/data", sep="")
   observeddatadir <- paste(datadir, "/", "ObservedData", sep="")

   # Everything proceeds faster when the folders already exist.
   # The following folders are for later steps that are more time consuming.
   estimation.directory <- paste(conddir, "/", "Model Estimation Data", sep="")
   plot.directory <- paste(conddir, "/", "Diagnostic Plot", sep="")
   postpred.values.directory <- paste(datadir, "/", "Posterior Predictive Data", sep="")

   dir.create(estimation.directory)
   dir.create(plot.directory)
   dir.create(postpred.values.directory)

   # Open the loop over trials.
   for(which.trial in 1:trials){

     # Determine the number of records.
     if(conditions[which.cond,1]=="SN") N <- round(runif(1, min=225, max=275),
digits=0)
     if(conditions[which.cond,1]=="MN") N <- round(runif(1, min=475, max=525),
digits=0)
     if(conditions[which.cond,1]=="LN") N <- round(runif(1, min=975, max=1025),
digits=0)

     # Simulate a value for the mean intercept and slope parameters.
     imean <- runif(1, min=9, max=11)
     smean <- runif(1, min=.3*imean, max=.5*imean)

     # Simulate a value for the mean quadratic effect.
     if(conditions[which.cond,2]=="NQM") qmean <- 0
     if(conditions[which.cond,2]=="SQM") qmean <- runif(1, min=-.03*imean, max=-
.01*imean)
     if(conditions[which.cond,2]=="LQM") qmean <- runif(1, min=-.13*imean, max=-
.10*imean)

     # Simulate values of the quadratic variance parameter by specifying quadratic
```

```
standard deviations.
    if(conditions[which.cond,3]=="NQV") qsd <- 0
    if(conditions[which.cond,3]=="SQV") qsd <- runif(1, min=.1, max=.2)
    if(conditions[which.cond,3]=="LQV") qsd <- runif(1, min=.5, max=.6)
    qvar <- qsd^2


 # Simulate values of the residual variance parameter by specifying a residual standard
deviation and converting it to a variance.
    if(conditions[which.cond,4]=="SRV") rsd <- runif(1, min=.75, max=1.25)
    if(conditions[which.cond,4]=="LRV") rsd <- runif(1, min=2.25, max=2.75)
    rvar <- rsd^2

    # Simulate values of the skewness and kurtosis.
    if(conditions[which.cond,5]=="N") skew <- 0
    if(conditions[which.cond,5]=="N") kurtosis <- 0
    if(conditions[which.cond,5]=="NN") skew <- runif(1, min=.8, max=1.2)
    if(conditions[which.cond,5]=="NN") kurtosis <- runif(1, min=.8, max=1.2)

    # Create the matrix of means.
    means <- matrix(c(imean, smean, qmean))

    # Simulate a value for the intercept and slope variance and covariance parameters.
    isd <- runif(1, min=3, max=4)
    ssd <- runif(1, min=.3*isd, max=.5*isd)
    ivar <- isd^2
    svar <- ssd^2

    iscor <- runif(1, min=.25, max=.35)
    iscov <- iscor*(isd*ssd)

    # Create the covariance matrix among the growth parameters.
    psi <- matrix(c(ivar, iscov, 0,
                iscov, svar, 0,
                0, 0, qvar), ncol=3, nrow=3)

    # Create the diagonal matrix of residual variances.
    theta <- diag(rvar,J)

    # Using the growth parameter mean vector and covariance matrix, simulate level-1
growth parameters.
    level1 <- t(as.matrix(mvrnorm(N, means, psi)))

    # Using the residual covariance matrix, create a matrix of residual scores.
    resscores <- mvrnorm(N, matrix(rep(0,J),ncol=1,nrow=J), theta)
```

```
    # Specify the time matrix.
    time <- matrix(seq(from=0,to=4,by=1),ncol=1,nrow=J)

    # Create a null matrix to fill with simulated data.
    gcm <- matrix(NA, N, J)


# Now simulate data consistent with the particular combination of manipulations.
    gcm[ ,1] <- level1[1, ] + level1[2, ]*time[1,1] + level1[3, ]*(time[1,1]^2) + resscores[
,1]
    gcm[ ,2] <- level1[1, ] + level1[2, ]*time[2,1] + level1[3, ]*(time[2,1]^2) + resscores[
,2]
    gcm[ ,3] <- level1[1, ] + level1[2, ]*time[3,1] + level1[3, ]*(time[3,1]^2) + resscores[
,3]
    gcm[ ,4] <- level1[1, ] + level1[2, ]*time[4,1] + level1[3, ]*(time[4,1]^2) + resscores[
,4]
    gcm[ ,5] <- level1[1, ] + level1[2, ]*time[5,1] + level1[3, ]*(time[5,1]^2) + resscores[
,5]

    # If the data are to follow a non-normal distribution, alter the data accordingly.
    if(conditions[which.cond,5]=="NN") gcmcorr <- matrix(cor(gcm), ncol=ncol(gcm),
nrow=ncol(gcm))
    if(conditions[which.cond,5]=="NN") bcd <- as.matrix(findbcd(skew, kurtosis)$par)
    if(conditions[which.cond,5]=="NN") gcmcorrnew <- matrix(NA, ncol(gcm),
ncol(gcm))
    if(conditions[which.cond,5]=="NN") for(j in 1:nrow(gcmcorr)){
      for(jj in 1:ncol(gcmcorr)){
        gcmcorrnew[j,jj] <- gcmcorr[j,jj]*(bcd[1,1]^2 + (3*bcd[1,1]*bcd[3,1]) +
(3*bcd[3,1]*bcd[1,1]) + (9*bcd[3,1]*bcd[3,1])) +
          (gcmcorr[j,jj]^2*(2*bcd[2,1]*bcd[2,1])) + (gcmcorr[j,jj]*(6*bcd[3,1]*bcd[3,1]))
      }
    }
    if(conditions[which.cond,5]=="NN") gcmmeans <- matrix(rep(0,ncol(gcm)))
    if(conditions[which.cond,5]=="NN") gcmnew <- mvrnorm(nrow(gcm), gcmmeans,
gcmcorr)
    if(conditions[which.cond,5]=="NN") gcmfinal <- matrix(NA, nrow(gcm), ncol(gcm))
    if(conditions[which.cond,5]=="NN") for(wr in 1:nrow(gcmfinal)){
      for(wc in 1:ncol(gcmfinal)){
        a = -bcd[2,1]
        gcmfinal[wr,wc] <- a + bcd[1,1]*gcmnew[wr,wc] + bcd[2,1]*(gcmnew[wr,wc]^2)
+ bcd[3,1]*(gcmnew[wr,wc]^3)
      }
    }
    # The new non-normal data is not in the original scale; this following converts the
```

```
non-normal data to the scale of the original data.
     if(conditions[which.cond,5]=="NN")
     gcmfinal.new <- matrix(NA, N, J)
     for(wc in 1:ncol(gcmfinal)){
       gcmfinal.new[ ,wc] <- mean(gcm[,wc]) + ((gcmfinal[,wc]-
mean(gcmfinal[,wc]))*(sd(gcm[,wc])/sd(gcmfinal[,wc])))
     }

     if(conditions[which.cond,5]=="NN") gcmfinal.new <- as.data.frame(gcmfinal.new)
     if(conditions[which.cond,5]=="NN") colnames(gcmfinal.new) <- c("y1", "y2", "y3",
"y4", "y5")




# Provide column names for the GCM object.
     colnames(gcm) <- c("y1", "y2", "y3", "y4", "y5")

     # Write the data.
     setwd(observeddatadir)

     if(conditions[which.cond,5]=="N") write.table(round(gcm, digits=5),
                                       paste(observeddatadir, "/", "obs.trial", which.trial,
".dat", sep=""),
                                       row.names=F, col.names=F, quote=F, sep="\t")


     if(conditions[which.cond,5]=="NN") write.table(round(gcmfinal.new, digits=5),
                                       paste(observeddatadir, "/", "obs.trial", which.trial,
".dat", sep=""),
                                       row.names=F, col.names=F, quote=F, sep="\t")

     # Write out all simulated model parameters.
     setwd(datadir)

     if(which.trial==1) simulated.parameters <- matrix(NA, trials, 17)

     # Indicate which replication.
     simulated.parameters[which.trial,1] <- which.trial
     # Write out the number of records.
     simulated.parameters[which.trial,2] <- N
     # Write out the generating model parameters.
     simulated.parameters[which.trial,3] <- means[1,1]
     simulated.parameters[which.trial,4] <- means[2,1]
     simulated.parameters[which.trial,5] <- means[3,1]
```

```
    simulated.parameters[which.trial,6] <- psi[1,1]
    simulated.parameters[which.trial,7] <- psi[2,2]
    simulated.parameters[which.trial,8] <- psi[3,3]
    simulated.parameters[which.trial,9] <- psi[1,2]
    simulated.parameters[which.trial,10] <- isd
    simulated.parameters[which.trial,11] <- ssd
    simulated.parameters[which.trial,12] <- qsd
    simulated.parameters[which.trial,13] <- iscor
    simulated.parameters[which.trial,14] <- rvar
    simulated.parameters[which.trial,15] <- rsd
    # Skew and kurtosis information.
    simulated.parameters[which.trial,16] <- skew
    simulated.parameters[which.trial,17] <- kurtosis

    simulated.parameters <- round(as.data.frame(simulated.parameters), digits=3)
    colnames(simulated.parameters) <- c("rep", "N", "gamma0", "gamma1", "gamma2",
                                        "psi00", "psi11", "psi22", "psi01",
                                        "sdpsi00", "sdpsi11", "sdpsi22", "rho01",
                                        "resvar", "ressd",
                                        "skew", "kurtosis")

    write.table(simulated.parameters, paste(datadir,  "/SimulatedModelParameters.dat",
sep=""), sep="\t", col.names=T, row.names=F, quote=F)

   } # Closes the loop over trials.

  } # Closes the loop over conditions.

) # Closes the clocking of time.
```

APPENDIX C

JAGS CODE FOR ESTIMATING GROWTH CURVE MODELS

Estimate a Linear Growth Curve Model

```
model{

  #/ Specify the likelihood of the data.
  #/ tau.res is the residual variance.
  for(i in 1:N){
    for(j in 1:J){
      y.prime[i,j] <- beta[i,1] + beta[i,2]*time[j,1];
      y[i,j] ~ dnorm(y.prime[i,j],tau.res);
    }
  }

  #/ Specify the priors for the average growth parameters.
  gamma[1] ~ dnorm(0,1);
  gamma[2] ~ dnorm(0,1);

  #/ Specify the priors for person-specific deviation terms.
  for(i in 1:N){
  beta[i,1:2] ~ dmnorm(gamma[ ], tau.psi[ , ])
  }

  #/ Specify the hyper priors for the person-specific deviation terms.
  tau.psi[1:2,1:2] ~ dwish(Omega[ , ],1);
  sigma.psi[1:2,1:2] <- inverse(tau.psi[ , ]);

  #/ Specify the priors for the residual variances and the person-specific growth
parameters.
  tau.res ~ dgamma(1,1);
  sigma.res <- 1/sqrt(tau.res);
  rho01 <- sigma.psi[1,2]/(sqrt(sigma.psi[1,1])*sqrt(sigma.psi[2,2]));

          }
```

Quadratic Growth Curve Model with Non-varying Functional Form Across Individuals

```
model{

  #/ Specify the likelihood of the data.
  #/ tau.res is the residual variance.
  for(i in 1:N){
    for(j in 1:J){
      y.prime[i,j] <- beta[i,1] + beta[i,2]*time[j,1] + gammaq*(time[j,1]*time[j,1]);
      y[i,j] ~ dnorm(y.prime[i,j],tau.res);
    }
  }

  #/ Specify the priors for the average growth parameters.
  gamma[1] ~ dnorm(0,1);
  gamma[2] ~ dnorm(0,1);
  gammaq ~ dnorm(0,1);

  #/ Specify the hyper priors for the person-specific deviation terms.
  tau.psi[1:2,1:2] ~ dwish(Omega[ , ],2);
  sigma.psi[1:2,1:2] <- inverse(tau.psi[ , ]);

  #/ Specify the person specific growth parameters.
  for(i in 1:N){
    beta[i,1:2] ~ dmnorm(gamma[ ], tau.psi[ , ])
  }


  #/ Specify the priors for the residual variances and the person-specific growth
parameters.
  tau.res ~ dgamma(1,1);
  sigma.res <- 1/sqrt(tau.res);
  rho01 <- sigma.psi[1,2]/(sqrt(sigma.psi[1,1])*sqrt(sigma.psi[2,2]));


}
```

Quadratic Growth Curve Model with Varying Functional Form Across Individuals

```
model{

  #/ Specify the likelihood of the data.
  #/ tau.res is the residual variance.
  for(i in 1:N){
    for(j in 1:J){
      y.prime[i,j] <- beta[i,1] + beta[i,2]*time[j,1] + betaq[i]*(time[j,1]*time[j,1]);
      y[i,j] ~ dnorm(y.prime[i,j],tau.res);
    }
  }

  #/ Specify the priors for the average growth parameters.
  gamma[1] ~ dnorm(0,1);
  gamma[2] ~ dnorm(0,1);
  gammaq ~ dnorm(0,1);

  #/ Specify the hyper priors for the person-specific deviation terms.
  tau.psi[1:2,1:2] ~ dwish(Omega[ , ],2);
  sigma.psi[1:2,1:2] <- inverse(tau.psi[ , ]);
  tau.psi.q ~ dgamma(1,1);
  sigma.psi.q <- inverse(tau.psi.q);

  #/ Specify the person specific growth parameters.
  for(i in 1:N){
    beta[i,1:2] ~ dmnorm(gamma[ ], tau.psi[ , ]);
    betaq[i] ~ dnorm(gammaq, tau.psi.q);
  }


  #/ Specify the priors for the residual variances and the person-specific growth
parameters.
  tau.res ~ dgamma(1,1);
  sigma.res <- 1/sqrt(tau.res);
  rho01 <- sigma.psi[2,1]/(sqrt(sigma.psi[1,1])*sqrt(sigma.psi[2,2]));

}
```

APPENDIX D

MODEL FITTING VIA JAGS AS INTERFACED THROUGH R

```
################################################################################
##############################################
## The purpose of this code is estimate all models for all all trials in all conditions in JAGS.
## Diagnostic plots will only be produced for the first two trials to conserve space and time.
## Significant preliminary investigations indicate that MCMC simulation parameters are
sufficient for all data generation by data analysis conditions.
################################################################################
################################################################################

##############################################
## Load any necessary libraries.
##############################################

library(mcmcplots)
library(rjags)

##############################################
## Specify constants for the simulation.
##############################################

# Number of measurement occasions.
J <- 5

# Number of trials.
trials <- 100

# Number of models. This is really only used to control looping over models.
nmodels <- 3

# SPECIFY THE PATHS FOR THE JAGS FILES THAT WILL BE USED FOR ESTIMATION.
model1.file <- "/Volumes/Derek Fay's Time
Caps/Dissertation/GrowthModel/code/model1_hierarchicalz.R"
model2.file <- "/Volumes/Derek Fay's Time
Caps/Dissertation/GrowthModel/code/model2_hierarchicalz.R"
model3.file <- "/Volumes/Derek Fay's Time
Caps/Dissertation/GrowthModel/code/model3_hierarchicalz.R"

# SPECIFY THE FILE THAT CONTAINS THE NUMBER OF BURN-IN ITERATIONS TO
USE FOR EACH DATA GENERATION/ANALYSIS COMBINATION.
nthin.file <- "/Volumes/Derek Fay's Time Caps/GCM Dissertation/Thin by Condition
NonNormal Fix.csv"
nthin.file <- read.csv(nthin.file, header=T)
#nthin.file <- subset(nthin.file, samplesize=='SN' & quadmean=='NQM' & quadsd=='NQV' &
ressd=='SRV' & distshape=='NN')
#nthin.file1 <- as.matrix(nthin.file)
#nthin.file <- as.data.frame(nthin.file1)

################################################################################
##############################################################################
## The following specifies the conditions for generating data.
```

```
## Since they are not likely to be pursued in practice, conditions with a non-zero quadratic
variance will be removed if there is no quadratic mean.
################################################################################
################################################################################

# Create the list of conditions.
sample.size <- matrix(c("SN","MN","LN"))
quad.mean <- matrix(c("NQM", "SQM", "LQM"))
quad.sd <- matrix(c("NQV", "SQV", "LQV"))
res.sd <- matrix(c("SRV", "LRV"))
dist.shape <- matrix(c("NN"))

conditions <- expand.grid(sample.size, quad.mean, quad.sd, res.sd, dist.shape)
colnames(conditions) <- c("samplesize", "quadmean", "quadsd", "ressd", "distshape")

attach(conditions)
conditions$remove <- ifelse(quadmean=="NQM" & (quadsd=="SQV"|quadsd=="LQV"), 1, 0)
detach(conditions)
conditions <- subset(conditions, remove==0, select=-remove)
#conditions <- subset(conditions, samplesize=='SN' & quadmean=='NQM' & quadsd=='NQV' &
ressd=='SRV' & distshape=='NN')
#conditions1 <- as.matrix(conditions)
#conditions <- as.data.frame(conditions1)

# This series of steps was necessary owing to an accidental stopping of the program.
# The program was accidently stopped at row 27 of the conditions matrix.
# The following keeps only those rows (there should be 16), converts to matrix, and then back to
data frame. This ensures that
# the levels in both files match.
#conditions <- conditions[27:nrow(conditions), ]
#conditions <- as.matrix(conditions)
#conditions <- as.data.frame(conditions)

# The following is the general form of a directory into which specific folders will be created and
into which data will be access and written to.
gendir <- "/Volumes/Seagate Backup Plus Drive/GCM Dissertation"

# Open the system time function.
system.time(

# Open the loop over conditions.
for(which.cond in 1:nrow(conditions)){

  condition.directory <- paste(gendir, "/", conditions[which.cond,1], "_",
conditions[which.cond,2], "_", conditions[which.cond,3], "_", conditions[which.cond,4], "_",
conditions[which.cond,5], sep="")
  estimation.directory <- paste(condition.directory, "/", "Model Estimation Data", sep="")
  plot.directory <- paste(condition.directory, "/", "Diagnostic Plot", sep="")

# Open the loop over trials within conditions.
```

```
for(which.trial in 1:trials){

  observeddatadir <- paste(condition.directory, "/data/ObservedData", sep="")
  setwd(observeddatadir)
  file <- paste("obs.trial",which.trial,".dat",sep="")
  y <- read.table(file, sep="\t", header=F)
  N <- nrow(y)
  J <- ncol(y)

  for(which.model in 1:nmodels){

# Print out what is currently being estimated.
  print(paste(paste("Currently estimating model = ", which.model, " for ", sep=""),
            paste("Sample Size = ", conditions[which.cond,1], sep=""),
            paste("Quadratic Mean = ", conditions[which.cond,2], sep=""),
            paste("Quadratic Variance = ", conditions[which.cond,3], sep=""),
            paste("Residual Variance = ", conditions[which.cond,4], sep=""),
            paste("Distribution Shape = ", conditions[which.cond,5], " for trial = ", which.trial,
sep="")))

# Indicate which parameters will be monitored.
  if(which.model==1) params.to.monitor <- c("gamma", "beta", "sigma.psi", "rho01", "tau.res",
"sigma.res")
  if(which.model==2) params.to.monitor <- c("gamma", "gammaq", "beta", "sigma.psi", "rho01",
"tau.res", "sigma.res")
  if(which.model==3) params.to.monitor <- c("gamma", "gammaq", "beta", "betaq", "sigma.psi",
"sigma.psi.q", "rho01", "tau.res", "sigma.res")

  if(which.model==1) params.to.monitor.plot <- c("gamma", "sigma.psi", "rho01", "tau.res",
"sigma.res")
  if(which.model==2) params.to.monitor.plot <- c("gamma", "gammaq", "sigma.psi", "rho01",
"tau.res", "sigma.res")
  if(which.model==3) params.to.monitor.plot <- c("gamma", "gammaq", "sigma.psi",
"sigma.psi.q", "rho01", "tau.res", "sigma.res")

# Specify initial values.
  g00.inits1 <- runif(1, min=15, max=17)
  g00.inits2 <- runif(1, min=5, max=7.5)
  g10.inits1 <- runif(1, min=6, max=8)
  g10.inits2 <- runif(1, min=-1, max=1)
  if(which.model>1) g20.inits1 <- runif(1, min=0, max=2)
  if(which.model>1) g20.inits2 <- runif(1, min=-2, max=0)
  gamma.inits1 <- c(g00.inits1, g10.inits1)
  gamma.inits2 <- c(g10.inits2, g10.inits2)

  tau.res.inits1 <- runif(1, min=3, max=5)
  tau.res.inits2 <- runif(1, min=.03, max=.05)

# Combine the initial values for both chains into one object.
```

```
  if(which.model==1) inits1 <- list("gamma"=gamma.inits1, "tau.res"=tau.res.inits1)
  if(which.model==1) inits2 <- list("gamma"=gamma.inits2, "tau.res"=tau.res.inits2)
  if(which.model!=1) inits1 <- list("gamma"=gamma.inits1, "gammaq"=g20.inits1,
"tau.res"=tau.res.inits1)
  if(which.model!=1) inits1 <- list("gamma"=gamma.inits2, "gammaq"=g20.inits2,
"tau.res"=tau.res.inits2)
  inits <- list(inits1, inits2)

# Specify other key components that are needed for estimation.
  time <- matrix(c(0,1,2,3,4), nrow=J, ncol=1)

# The precision matrix for the person-specific growth parameters.
  Omega <- diag(1,2);

# Create list of the objects necessary for estimating in JAGS.
  jags.data <- list("y"=y, "N"=N, "J"=J, "time"=time, "Omega"=Omega)

# Specify MCMC simulation parameters.
  nchains <- 2
  nburnin <- 500
  if (which.model==1) nthin <- subset(nthin.file, samplesize==conditions[which.cond,1] &
                                quadmean==conditions[which.cond,2] &
                                quadsd==conditions[which.cond,3] &
                                ressd==conditions[which.cond,4] &
                                distshape==conditions[which.cond,5], select=model1)
  if (which.model==2) nthin <- subset(nthin.file, samplesize==conditions[which.cond,1] &
                                quadmean==conditions[which.cond,2] &
                                quadsd==conditions[which.cond,3] &
                                ressd==conditions[which.cond,4] &
                                distshape==conditions[which.cond,5], select=model2)
  if (which.model==3) nthin <- subset(nthin.file, samplesize==conditions[which.cond,1] &
                                quadmean==conditions[which.cond,2] &
                                quadsd==conditions[which.cond,3] &
                                ressd==conditions[which.cond,4] &
                                distshape==conditions[which.cond,5], select=model3)
  nthin <- as.numeric(nthin)

  nadapt <- 0
  iters.per.chain <- 150
  niters <- iters.per.chain*nthin

  # Determine which model is relevant.
  if(which.model==1) modelfile <- model1.file
  if(which.model==2) modelfile <- model2.file
  if(which.model==3) modelfile <- model3.file

  # Estimate the model.
  fit <- jags.model(file=modelfile, n.adapt=nadapt, n.chains=nchains, data=jags.data, inits=inits)
  update(fit, n.iter=nburnin, variable.names=params.to.monitor)
  fit.samples <- coda.samples(fit, variable.names=params.to.monitor, n.iter=niters, thin=nthin)
```

```
  # Save the posterior draws.
  posterior.draws <- as.matrix(fit.samples)
  beta0 <- subset(posterior.draws, select=noquote(paste("beta[", seq(from=1, to=N, by=1), ",",
"1]", sep="")))
    colnames(beta0) <- paste("beta0", seq(from=1, to=N, by=1), sep="_")
  beta1 <- subset(posterior.draws, select=noquote(paste("beta[", seq(from=1, to=N, by=1), ",",
"2]", sep="")))
    colnames(beta1) <- paste("beta1", seq(from=1, to=N, by=1), sep="_")
  if(which.model==3) beta2 <- subset(posterior.draws, select=noquote(paste("betaq[",
seq(from=1, to=N, by=1), "]", sep="")))
    if(which.model==3) colnames(beta2) <- paste("beta2", seq(from=1, to=N, by=1), sep="_")
  gammas <- subset(posterior.draws, select=noquote(paste("gamma[", seq(from=1, to=2, by=1),
"]", sep="")))
  if(which.model>1) gammaq <- subset(posterior.draws, select=noquote(gammaq))
    colnames(gammas) <- c("gamma00", "gamma10")
    if(which.model>1) colnames(gammaq) <- c("gamma20")
  sigma.res <- subset(posterior.draws, select=noquote("sigma.res"))
    colnames(sigma.res) <- c("theta")
  sigma.psi <- subset(posterior.draws, select=noquote(c("sigma.psi[1,1]", "sigma.psi[1,2]",
"sigma.psi[2,2]")))
  if(which.model==3) sigma.psi.q <- subset(posterior.draws, select=noquote("sigma.psi.q"))
    colnames(sigma.psi) <- c("sigma00", "sigma12", "sigma11")
    if(which.model==3) colnames(sigma.psi.q) <- "sigma22"

  # Write out the parameter files.
  setwd(estimation.directory)
  write.table(beta0, paste("beta0.model", which.model, ".trial", which.trial, ".dat", sep=""),
row.names=F, col.names=T, sep="\t", quote=F)
  write.table(beta1, paste("beta1.model", which.model, ".trial", which.trial, ".dat", sep=""),
row.names=F, col.names=T, sep="\t", quote=F)
  if(which.model==3) write.table(beta2, paste("beta2.model", which.model, ".trial", which.trial,
".dat", sep=""), row.names=F, col.names=T, sep="\t", quote=F)
  write.table(gammas, paste("gammas.model", which.model, ".trial", which.trial, ".dat", sep=""),
row.names=F, col.names=T, sep="\t", quote=F)
  if(which.model>1) write.table(cbind(gammas, gammaq), paste("gammas.model", which.model,
".trial", which.trial, ".dat", sep=""), row.names=F, col.names=T, sep="\t", quote=F)
  write.table(sigma.res, paste("theta.model", which.model, ".trial", which.trial, ".dat", sep=""),
row.names=F, col.names=T, sep="\t", quote=F)
  if(which.model<=2) write.table(sigma.psi, paste("psi.model", which.model, ".trial", which.trial,
".dat", sep=""), row.names=F, col.names=T, sep="\t", quote=F)
  if(which.model==3) write.table(cbind(sigma.psi, sigma.psi.q), paste("psi.model", which.model,
".trial", which.trial, ".dat", sep=""), row.names=F, col.names=T, sep="\t", quote=F)

  # Plot the diagnostics.
  trial.model.diagnostics <- paste(plot.directory, "/trial", which.trial, sep="")
  if(which.trial<2 & which.model==1) dir.create(trial.model.diagnostics)
  model.diagnostics <- paste(trial.model.diagnostics, "/model", which.model, sep="")
  if(which.trial<2) dir.create(model.diagnostics)
  if(which.trial<2) mcmcplot(fit.samples, parms=params.to.monitor.plot, dir=model.diagnostics,
```

```
style="plain",
                    filename=paste("diagnostics.model.", which.model, sep=""),
                    extension="html")

   } # Close the loop over models.

  } # Close the loop over trials.

} # Close the loop over conditions.

) # Closes the system time function.

#gc(rm(list=ls()))
```

APPENDIX E

R CODE FOR PERFORMING PPMC WITH THE ABSOLUTE FIT DISCREPANCY FUNCTIONS

```
################################################################################
################################################
## The purpose of the following code is to conduct PPMC.
## This writes out the set of model expectations and posterior predictive data files for each
## model for each trial corresponding to each data analysis model in each condition.
## Other dependencies include the covariance matrix among growth parameters, residual
variances, and the gammas.
################################################################################
################################################

# Specify any libraries that are needed.
library(TeachingDemos) # For setting the seed.
library(MASS)
library(psych)

# Specify the number of trials.
trials <- 100

# Specify the number draws that were used to construct the posterior distribution.
D <- 300

################################################################################
################################
## Generate and write out posterior predictive data.
## Conduct PPMC.
################################################################################
################################

# Create the list of conditions.
sample.size <- matrix(c("SN","MN","LN"))
quad.mean <- matrix(c("NQM", "SQM", "LQM"))
quad.sd <- matrix(c("NQV", "SQV", "LQV"))
res.sd <- matrix(c("SRV", "LRV"))
#dist.shape <- matrix(c("NN"))
dist.shape <- matrix(c("N", "NN"))

conditions <- expand.grid(sample.size, quad.mean, quad.sd, res.sd, dist.shape)
colnames(conditions) <- c("samplesize", "quadmean", "quadsd", "ressd", "distshape")

attach(conditions)
conditions$remove <- ifelse(quadmean=="NQM" & (quadsd=="SQV"|quadsd=="LQV"), 1, 0)
detach(conditions)
conditions <- subset(conditions, remove==0, select=-remove)

# Specify constants that will be used throughout for indexing through loops.
gendir <- "/Volumes/Seagate Backup Plus Drive/GCM Dissertation"

# Likelihood ratio - taken from Preacher, Wichman, MacCallum, & Briggs.
LR.f <- function(N, sigma, S, y, mu.hat){
  mean.diff <- matrix((colMeans(y) - mu.hat), ncol=1, nrow=5)
```

```
fml <- as.matrix((log(det(sigma))) - log(det(S)) + tr(S%*%solve(sigma)) - J +
((t(mean.diff)%*%solve(sigma)%*%mean.diff)))
  (N-1)*fml
  }

# Level-1 conditional concordance measure.
ccc1.f <- function(y, y.exp, mu.hat, N, J){
  diffs1 <- as.matrix(y - y.exp)
  diffs2 <- as.matrix(y - mean(y))
  diffs3 <- as.matrix(y.exp - mean(mu.hat))
  1 - ((sum(diag(t(diffs1)%*%diffs1)))/((sum(diag(t(diffs2)%*%diffs2))) +
(sum(diag(t(diffs3)%*%diffs3))) + (N*J)*((mean(y)-mean(mu.hat))^2)))
}

# Level-1 pseudo-R2.
r21.f <- function(y, y.exp, mu.hat){
  diffs1 <- as.matrix(y-mean(y))
  diffs2 <- as.matrix(y.exp - mean(mu.hat))

((sum(diag((t(diffs1)%*%diffs2))))^2)/((sum(diag(t(diffs1)%*%diffs1))*(sum(diag(t(diffs2)%*%
diffs2)))))
}

# Level-1 SGDDM.
sgddm1.f <- function(y, y.exp, J, N){
  diffs <- y-y.exp
  rescov <- as.matrix((cov(diffs)*(N-1))/N)
  sgddm.mat <- matrix(NA, J, J)
  for(j in 1:J){
    for(jj in 1:J){
      if(j!=jj) sgddm.mat[j,jj] <- rescov[j,jj]/(sqrt(rescov[j,j])*sqrt(rescov[jj,jj]))
    }
  }
  sum(abs(sgddm.mat[lower.tri(sgddm.mat, diag=F)]))/((J*(J-1))/2)
}

# Level-2 conditional concordance measure.
ccc2.f <- function(y, mu.hat, N, J){

  means.yexp <- matrix(NA, N, J)
  means.yexp[ ,1] <- mu.hat[1, ]
  means.yexp[ ,2] <- mu.hat[2, ]
  means.yexp[ ,3] <- mu.hat[3, ]
  means.yexp[ ,4] <- mu.hat[4, ]
  means.yexp[ ,5] <- mu.hat[5, ]

  diffs1 <- matrix(NA, N, J)
  diffs1[ ,1] <- as.matrix(y[ ,1]-mu.hat[1, ])
  diffs1[ ,2] <- as.matrix(y[ ,2]-mu.hat[2, ])
  diffs1[ ,3] <- as.matrix(y[ ,3]-mu.hat[3, ])
```

```
  diffs1[ ,4] <- as.matrix(y[ ,4]-mu.hat[4, ])
  diffs1[ ,5] <- as.matrix(y[ ,5]-mu.hat[5, ])
  diffs2 <- as.matrix(y - mean(y))
  diffs3 <- as.matrix(means.yexp - mean(mu.hat))
  1 - ((sum(diag(t(diffs1)%*%diffs1)))/((sum(diag(t(diffs2)%*%diffs2))) +
(sum(diag(t(diffs3)%*%diffs3))) + (N*J)*((mean(y)-mean(mu.hat))^2)))
}

# Level-1 pseudo-R2.
r22.f <- function(y, mu.hat, N, J){

  means.yexp <- matrix(NA, N, J)
  means.yexp[ ,1] <- mu.hat[1, ]
  means.yexp[ ,2] <- mu.hat[2, ]
  means.yexp[ ,3] <- mu.hat[3, ]
  means.yexp[ ,4] <- mu.hat[4, ]
  means.yexp[ ,5] <- mu.hat[5, ]

  diffs1 <- y - mean(y)
  diffs2 <- means.yexp - mean(mu.hat)


((sum(diag((t(diffs1)%*%diffs2))))^2)/((sum(diag(t(diffs1)%*%diffs1))*(sum(diag(t(diffs2)%*%
diffs2)))))
}

# Level-1 SGDDM.
sgddm2.f <- function(y, mu.hat, N, J){

  diffs <- matrix(NA, N, J)
  diffs[ ,1] <- as.matrix(y[ ,1]-mu.hat[1, ])
  diffs[ ,2] <- as.matrix(y[ ,2]-mu.hat[2, ])
  diffs[ ,3] <- as.matrix(y[ ,3]-mu.hat[3, ])
  diffs[ ,4] <- as.matrix(y[ ,4]-mu.hat[4, ])
  diffs[ ,5] <- as.matrix(y[ ,5]-mu.hat[5, ])
  rescov <- as.matrix((cov(diffs)*(N-1))/N)
  sgddm.mat <- matrix(NA, J, J)
  for(j in 1:J){
    for(jj in 1:J){
      if(j!=jj) sgddm.mat[j,jj] <- rescov[j,jj]/(sqrt(rescov[j,j])*sqrt(rescov[jj,jj]))
    }
  }
  sum(abs(sgddm.mat[lower.tri(sgddm.mat, diag=F)]))/((J*(J-1))/2)
}

#########################################################################
## For each model within each trial, open the parameter estimates.
#########################################################################

# use the system time function to time the computations.
```

```
system.time(

  # Open the loop over conditions.
  for(wc in 1:nrow(conditions)){

    # Specify paths to obtain requisite data.
    condition.directory <- paste(gendir, "/", conditions[wc,1], "_", conditions[wc,2], "_",
conditions[wc,3], "_", conditions[wc,4], "_", conditions[wc,5], sep="")
    observed.data.directory <- paste(condition.directory, "/data/", "ObservedData", sep="")
    estimation.directory <- paste(condition.directory, "/", "Model Estimation Data", sep="")
    postpred.directory <- paste(condition.directory, "/data/", "Alt Posterior Predictive Data",
sep="")
    expectation.directory <- paste(condition.directory, "/data/", "Model Expectation", sep="")
    results.directory <- paste(condition.directory, "/data/", "Results", sep="")
    dir.create(postpred.directory)
    dir.create(results.directory)

    print(conditions[wc, ])


# Set up empty matrices to compute posterior predictive p-values for each trial.
LRp <- matrix(NA, trials, 3)
CCC1p <- matrix(NA, trials, 3)
CCC2p <- matrix(NA, trials, 3)
R21p <- matrix(NA, trials, 3)
R22p <- matrix(NA, trials, 3)
SGDDM1p <- matrix(NA, trials, 3)
SGDDM2p <- matrix(NA, trials, 3)

# Set up the loop over trials.
for(wt in 1:trials){

    # Specify a null matrix for the draws for each fo the posterior predictive measures.
    LR <- matrix(NA, D, 6)
    CCC1 <- matrix(NA, D, 6)
    CCC2 <- matrix(NA, D, 6)
    R21 <- matrix(NA, D, 6)
    R22 <- matrix(NA, D, 6)
    SGDDM1 <- matrix(NA, D, 6)
    SGDDM2 <- matrix(NA, D, 6)

    # Open the observed data.
    setwd(observed.data.directory)
    y.obs <- as.matrix(read.table(paste("obs.trial", wt, ".dat", sep=""), header=F, sep="\t"))
    colnames(y.obs) <- NULL
    N <- nrow(y.obs)
    J <- ncol(y.obs)
    print(paste("Trial = ", wt, sep=""))

    # Model 1 files.
```

```
   setwd(estimation.directory)
   beta0.m1 <- as.matrix(read.table(file=paste("beta0.model1.trial", wt, ".dat", sep=""), header=T,
sep="\t"))
   beta1.m1 <- as.matrix(read.table(file=paste("beta1.model1.trial", wt, ".dat", sep=""), header=T,
sep="\t"))
   gammas.m1 <- as.matrix(read.table(file=paste("gammas.model1.trial", wt, ".dat", sep=""),
header=T, sep="\t"))
   psi.m1 <- as.matrix(read.table(file=paste("psi.model1.trial", wt, ".dat", sep=""), header=T,
sep="\t"))
   theta.m1 <- as.matrix(read.table(file=paste("theta.model1.trial", wt, ".dat", sep=""), header=T,
sep="\t"))


# Model 2 files.
   setwd(estimation.directory)
   beta0.m2 <- as.matrix(read.table(file=paste("beta0.model2.trial", wt, ".dat", sep=""), header=T,
sep="\t"))
   beta1.m2 <- as.matrix(read.table(file=paste("beta1.model2.trial", wt, ".dat", sep=""), header=T,
sep="\t"))
   gammas.m2 <- as.matrix(read.table(file=paste("gammas.model2.trial", wt, ".dat", sep=""),
header=T, sep="\t"))
   psi.m2 <- as.matrix(read.table(file=paste("psi.model2.trial", wt, ".dat", sep=""), header=T,
sep="\t"))
   theta.m2 <- as.matrix(read.table(file=paste("theta.model2.trial", wt, ".dat", sep=""), header=T,
sep="\t"))

   # Model 3 files.
   setwd(estimation.directory)
   beta0.m3 <- as.matrix(read.table(file=paste("beta0.model3.trial", wt, ".dat", sep=""), header=T,
sep="\t"))
   beta1.m3 <- as.matrix(read.table(file=paste("beta1.model3.trial", wt, ".dat", sep=""), header=T,
sep="\t"))
   beta2.m3 <- as.matrix(read.table(file=paste("beta2.model3.trial", wt, ".dat", sep=""), header=T,
sep="\t"))
   gammas.m3 <- as.matrix(read.table(file=paste("gammas.model3.trial", wt, ".dat", sep=""),
header=T, sep="\t"))
   psi.m3 <- as.matrix(read.table(file=paste("psi.model3.trial", wt, ".dat", sep=""), header=T,
sep="\t"))
   theta.m3 <- as.matrix(read.table(file=paste("theta.model3.trial", wt, ".dat", sep=""), header=T,
sep="\t"))

   # Create lambda matrices that will be used for computing the model implied mean vector and
covariance matrix.
   time <- as.matrix(seq(from=0, to=J-1, by=1))
   lambda1 <- as.matrix(cbind(rep(1,J), time))
   lambda2 <- as.matrix(cbind(rep(1,J), time, time^2))

   #for(wm in 1:models){

   for(d in 1:D){
```

```
    if(d==75|d==150|d==225) print(paste("Draw = ", d, sep=""))

    # Create null matrices to write posterior predictive and expected values.
    y.rep.m1 <- matrix(NA, N, J)
    y.rep.m2 <- matrix(NA, N, J)
    y.rep.m3 <- matrix(NA, N, J)

    y.exp.m1 <- matrix(NA, N, J)
    y.exp.m2 <- matrix(NA, N, J)
    y.exp.m3 <- matrix(NA, N, J)

    # Generate expected values.
    y.exp.m1[ ,1] <- beta0.m1[d,] + beta1.m1[d,]*lambda1[1,2]
    y.exp.m1[ ,2] <- beta0.m1[d,] + beta1.m1[d,]*lambda1[2,2]
    y.exp.m1[ ,3] <- beta0.m1[d,] + beta1.m1[d,]*lambda1[3,2]
    y.exp.m1[ ,4] <- beta0.m1[d,] + beta1.m1[d,]*lambda1[4,2]
    y.exp.m1[ ,5] <- beta0.m1[d,] + beta1.m1[d,]*lambda1[5,2]

    y.exp.m2[ ,1] <- beta0.m2[d,] + beta1.m2[d,]*lambda2[1,2] + gammas.m2[d,3]*lambda2[1,3]
    y.exp.m2[ ,2] <- beta0.m2[d,] + beta1.m2[d,]*lambda2[2,2] + gammas.m2[d,3]*lambda2[2,3]
    y.exp.m2[ ,3] <- beta0.m2[d,] + beta1.m2[d,]*lambda2[3,2] + gammas.m2[d,3]*lambda2[3,3]
    y.exp.m2[ ,4] <- beta0.m2[d,] + beta1.m2[d,]*lambda2[4,2] + gammas.m2[d,3]*lambda2[4,3]
    y.exp.m2[ ,5] <- beta0.m2[d,] + beta1.m2[d,]*lambda2[5,2] + gammas.m2[d,3]*lambda2[5,3]

    y.exp.m3[ ,1] <- beta0.m3[d,] + beta1.m3[d,]*lambda2[1,2] + beta2.m3[d,]*lambda2[1,3]
    y.exp.m3[ ,2] <- beta0.m3[d,] + beta1.m3[d,]*lambda2[2,2] + beta2.m3[d,]*lambda2[2,3]
    y.exp.m3[ ,3] <- beta0.m3[d,] + beta1.m3[d,]*lambda2[3,2] + beta2.m3[d,]*lambda2[3,3]
    y.exp.m3[ ,4] <- beta0.m3[d,] + beta1.m3[d,]*lambda2[4,2] + beta2.m3[d,]*lambda2[4,3]
    y.exp.m3[ ,5] <- beta0.m3[d,] + beta1.m3[d,]*lambda2[5,2] + beta2.m3[d,]*lambda2[5,3]

    # Generate posterior predictive data.
    # Set the seed for generating posterior predictive data.
    set.seed(d+1000)
    y.rep.m1 <- y.exp.m1 + as.matrix(mvrnorm(N, rep(0,J), diag(theta.m1[d,1],J)))
    set.seed(d+2000)
    y.rep.m2 <- y.exp.m2 + as.matrix(mvrnorm(N, rep(0,J), diag(theta.m2[d,1],J)))
    set.seed(d+3000)
    y.rep.m3 <- y.exp.m3 + as.matrix(mvrnorm(N, rep(0,J), diag(theta.m3[d,1],J)))

    ## Write out the expected values and posterior predictive data.
    setwd(expectation.directory)
    write.table(y.exp.m1, paste("yexp_model1_trial", wt, "_draw", d, ".dat", sep=""),
col.names=F, row.names=F, quote=F, sep="\t")
    write.table(y.exp.m2, paste("yexp_model2_trial", wt, "_draw", d, ".dat", sep=""),
col.names=F, row.names=F, quote=F, sep="\t")
    write.table(y.exp.m3, paste("yexp_model3_trial", wt, "_draw", d, ".dat", sep=""),
col.names=F, row.names=F, quote=F, sep="\t")

    setwd(postpred.directory)
```

```
    write.table(y.rep.m1, paste("yrep_model1_trial", wt, "_draw", d, ".dat", sep=""),
col.names=F, row.names=F, quote=F, sep="\t")
    write.table(y.rep.m2, paste("yrep_model2_trial", wt, "_draw", d, ".dat", sep=""),
col.names=F, row.names=F, quote=F, sep="\t")
    write.table(y.rep.m3, paste("yrep_model3_trial", wt, "_draw", d, ".dat", sep=""),
col.names=F, row.names=F, quote=F, sep="\t")

    # Compute key functions to support the computations of discrepancy functions.

    # Model implied covariance matrix.
    psi.m1z <- matrix(c(psi.m1[d,1], psi.m1[d,2],
                psi.m1[d,2], psi.m1[d,3]), ncol=2, nrow=2)

    psi.m2z <- matrix(c(psi.m2[d,1], psi.m2[d,2], 0,
                psi.m2[d,2], psi.m2[d,3], 0,
                0, 0, 0), ncol=3, nrow=3)
    psi.m3z <- matrix(c(psi.m3[d,1], psi.m3[d,2], 0,
                psi.m3[d,2], psi.m3[d,3], 0,
                0, 0, psi.m3[d,4]), ncol=3, nrow=3)

    sigma.m1 <- lambda1%*%psi.m1z%*%t(lambda1) + diag(theta.m1[d,1], J)
    sigma.m2 <- lambda2%*%psi.m2z%*%t(lambda2) + diag(theta.m2[d,1], J)
    sigma.m3 <- lambda2%*%psi.m3z%*%t(lambda2) + diag(theta.m3[d,1], J)

    # Model implied mean vector.
    mu.hat.m1 <- t(as.matrix(gammas.m1[d, ]%*%t(lambda1)))
    mu.hat.m2 <- t(as.matrix(gammas.m2[d, ]%*%t(lambda2)))
    mu.hat.m3 <- t(as.matrix(gammas.m3[d, ]%*%t(lambda2)))


##############################################################################
###
    ## Compute the discrepancy functions.

##############################################################################
###

    LR[d,1] <- LR.f(N=N, sigma.m1, S=cov(y.obs), y=y.obs, mu.hat=mu.hat.m1)
    LR[d,2] <- LR.f(N=N, sigma.m1, S=cov(y.rep.m1), y=y.rep.m1, mu.hat=mu.hat.m1)
    LR[d,3] <- LR.f(N=N, sigma.m2, S=cov(y.obs), y=y.obs, mu.hat=mu.hat.m2)
    LR[d,4] <- LR.f(N=N, sigma.m2, S=cov(y.rep.m2), y=y.rep.m2, mu.hat=mu.hat.m2)
    LR[d,5] <- LR.f(N=N, sigma.m3, S=cov(y.obs), y=y.obs, mu.hat=mu.hat.m3)
    LR[d,6] <- LR.f(N=N, sigma.m3, S=cov(y.rep.m3), y=y.rep.m3, mu.hat=mu.hat.m3)

    CCC1[d,1] <- ccc1.f(y=y.obs, y.exp=y.exp.m1, mu.hat=mu.hat.m1, N=N, J=J)
    CCC1[d,2] <- ccc1.f(y=y.rep.m1, y.exp=y.exp.m1, mu.hat=mu.hat.m1, N=N, J=J)
    CCC1[d,3] <- ccc1.f(y=y.obs, y.exp=y.exp.m2, mu.hat=mu.hat.m2, N=N, J=J)
    CCC1[d,4] <- ccc1.f(y=y.rep.m2, y.exp=y.exp.m2, mu.hat=mu.hat.m2, N=N, J=J)
    CCC1[d,5] <- ccc1.f(y=y.obs, y.exp=y.exp.m3, mu.hat=mu.hat.m3, N=N, J=J)
    CCC1[d,6] <- ccc1.f(y=y.rep.m3, y.exp=y.exp.m3, mu.hat=mu.hat.m3, N=N, J=J)
```

```
  R21[d,1] <- r21.f(y=y.obs, y.exp=y.exp.m1, mu.hat=mu.hat.m1)
  R21[d,2] <- r21.f(y=y.rep.m1, y.exp=y.exp.m1, mu.hat=mu.hat.m1)
  R21[d,3] <- r21.f(y=y.obs, y.exp=y.exp.m2, mu.hat=mu.hat.m2)
  R21[d,4] <- r21.f(y=y.rep.m2, y.exp=y.exp.m2, mu.hat=mu.hat.m2)
  R21[d,5] <- r21.f(y=y.obs, y.exp=y.exp.m3, mu.hat=mu.hat.m3)
  R21[d,6] <- r21.f(y=y.rep.m3, y.exp=y.exp.m3, mu.hat=mu.hat.m3)

  SGDDM1[d,1] <- sgddm1.f(y=y.obs, y.exp=y.exp.m1, J=J, N=N)
  SGDDM1[d,2] <- sgddm1.f(y=y.rep.m1, y.exp=y.exp.m1, J=J, N=N)
  SGDDM1[d,3] <- sgddm1.f(y=y.obs, y.exp=y.exp.m2, J=J, N=N)
  SGDDM1[d,4] <- sgddm1.f(y=y.rep.m2, y.exp=y.exp.m2, J=J, N=N)
  SGDDM1[d,5] <- sgddm1.f(y=y.obs, y.exp=y.exp.m3, J=J, N=N)
  SGDDM1[d,6] <- sgddm1.f(y=y.rep.m3, y.exp=y.exp.m3, J=J, N=N)

  CCC2[d,1] <- ccc2.f(y=y.obs, mu.hat=mu.hat.m1, N=N, J=J)
  CCC2[d,2] <- ccc2.f(y=y.rep.m1, mu.hat=mu.hat.m1, N=N, J=J)
  CCC2[d,3] <- ccc2.f(y=y.obs, mu.hat=mu.hat.m2, N=N, J=J)
  CCC2[d,4] <- ccc2.f(y=y.rep.m2, mu.hat=mu.hat.m2, N=N, J=J)
  CCC2[d,5] <- ccc2.f(y=y.obs, mu.hat=mu.hat.m3, N=N, J=J)
  CCC2[d,6] <- ccc2.f(y=y.rep.m3, mu.hat=mu.hat.m3, N=N, J=J)

  R22[d,1] <- r22.f(y=y.obs, mu.hat=mu.hat.m1, N=N, J=J)
  R22[d,2] <- r22.f(y=y.rep.m1, mu.hat=mu.hat.m1, N=N, J=J)
  R22[d,3] <- r22.f(y=y.obs, mu.hat=mu.hat.m2, N=N, J=J)
  R22[d,4] <- r22.f(y=y.rep.m2, mu.hat=mu.hat.m2, N=N, J=J)
  R22[d,5] <- r22.f(y=y.obs, mu.hat=mu.hat.m3, N=N, J=J)
  R22[d,6] <- r22.f(y=y.rep.m3, mu.hat=mu.hat.m3, N=N, J=J)

  SGDDM2[d,1] <- sgddm2.f(y=y.obs, mu.hat=mu.hat.m1, J=J, N=N)
  SGDDM2[d,2] <- sgddm2.f(y=y.rep.m1, mu.hat=mu.hat.m1, J=J, N=N)
  SGDDM2[d,3] <- sgddm2.f(y=y.obs, mu.hat=mu.hat.m2, J=J, N=N)
  SGDDM2[d,4] <- sgddm2.f(y=y.rep.m2, mu.hat=mu.hat.m2, J=J, N=N)
  SGDDM2[d,5] <- sgddm2.f(y=y.obs, mu.hat=mu.hat.m3, J=J, N=N)
  SGDDM2[d,6] <- sgddm2.f(y=y.rep.m3, mu.hat=mu.hat.m3, J=J, N=N)

} # Close the loop over draws.

LRp[wt,1] <- mean(LR[,2]>=LR[,1])
LRp[wt,2] <- mean(LR[,4]>=LR[,3])
LRp[wt,3] <- mean(LR[,6]>=LR[,5])

CCC1p[wt,1] <- mean(CCC1[,2]>=CCC1[,1])
CCC1p[wt,2] <- mean(CCC1[,4]>=CCC1[,3])
CCC1p[wt,3] <- mean(CCC1[,6]>=CCC1[,5])

CCC2p[wt,1] <- mean(CCC2[,2]>=CCC2[,1])
CCC2p[wt,2] <- mean(CCC2[,4]>=CCC2[,3])
CCC2p[wt,3] <- mean(CCC2[,6]>=CCC2[,5])
```

```
   R21p[wt,1] <- mean(R21[,2]>=R21[,1])
   R21p[wt,2] <- mean(R21[,4]>=R21[,3])
   R21p[wt,3] <- mean(R21[,6]>=R21[,5])

   R22p[wt,1] <- mean(R22[,2]>=R22[,1])
   R22p[wt,2] <- mean(R22[,4]>=R22[,3])
   R22p[wt,3] <- mean(R22[,6]>=R22[,5])

   SGDDM1p[wt,1] <- mean(SGDDM1[,2]>=SGDDM1[,1])
   SGDDM1p[wt,2] <- mean(SGDDM1[,4]>=SGDDM1[,3])
   SGDDM1p[wt,3] <- mean(SGDDM1[,6]>=SGDDM1[,5])

   SGDDM2p[wt,1] <- mean(SGDDM2[,2]>=SGDDM2[,1])
   SGDDM2p[wt,2] <- mean(SGDDM2[,4]>=SGDDM2[,3])
   SGDDM2p[wt,3] <- mean(SGDDM2[,6]>=SGDDM2[,5])

   # Write out the values of the realized and posterior predictive values of the discrepancy
functions.
   setwd(results.directory)
   write.table(CCC1, paste("Level1_CCC_Draws_Trial", wt, ".dat", sep=""), row.names=F,
col.names=F, quote=F, sep="\t")
   write.table(CCC2, paste("Level2_CCC_Draws_Trial", wt, ".dat", sep=""), row.names=F,
col.names=F, quote=F, sep="\t")
   write.table(R21, paste("Level1_RSquared_Draws_Trial", wt, ".dat", sep=""), row.names=F,
col.names=F, quote=F, sep="\t")
   write.table(R22, paste("Level2_RSquared_Draws_Trial", wt, ".dat", sep=""), row.names=F,
col.names=F, quote=F, sep="\t")
   write.table(SGDDM1, paste("Level1_SGDDM_Draws_Trial", wt, ".dat", sep=""),
row.names=F, col.names=F, quote=F, sep="\t")
   write.table(SGDDM2, paste("Level2_SGDDM_Draws_Trial", wt, ".dat", sep=""),
row.names=F, col.names=F, quote=F, sep="\t")
   write.table(LR, paste("LR_Draws_Trial", wt, ".dat", sep=""), row.names=F, col.names=F,
quote=F, sep="\t")

   setwd(results.directory)
   if(wt==trials) CCC1p <- cbind(rep("CCC1", trials), seq(from=1, to=trials, by=1), CCC1p)
   if(wt==trials) CCC2p <- cbind(rep("CCC2", trials), seq(from=1, to=trials, by=1), CCC2p)
   if(wt==trials) R21p <- cbind(rep("RSquared1", trials), seq(from=1, to=trials, by=1), R21p)
   if(wt==trials) R22p <- cbind(rep("RSquared2", trials), seq(from=1, to=trials, by=1), R22p)
   if(wt==trials) SGDDM1p <- cbind(rep("SGDDM1", trials), seq(from=1, to=trials, by=1),
SGDDM1p)
   if(wt==trials) SGDDM2p <- cbind(rep("SGDDM2", trials), seq(from=1, to=trials, by=1),
SGDDM2p)
   if(wt==trials) LRp <- cbind(rep("LR", trials), seq(from=1, to=trials, by=1), LRp)

   if(wt==trials) results.to.write <- rbind(CCC1p, CCC2p, R21p, R22p, SGDDM1p, SGDDM2p,
LRp)
   if(wt==trials) colnames(results.to.write) <- c("Function", "Trial", "M1", "M2", "M3")
   if(wt==trials) write.csv(results.to.write, file="Posterior Predictive PValues.csv", na="",
row.names=F)
```

```
    } # Close the loop over trials.
} # Close the loop over conditions.
)
```

APPENDIX F

R CODE FOR PERFORMING PPMC WITH THE LRT DISCPRENACY FUNCTION

```
#########################################################################
#########################################################
## The purpose of the following code is to conduct PPMC.
## This writes out the set of model expectations and posterior predictive data files for
each
## model for each trial corresponding to each data analysis model in each condition.
## Other dependencies include the covariance matrix among growth parameters, residual
## variances, and the gammas.
#########################################################################
#########################################################

# Specify any libraries that are needed.
library(TeachingDemos) # For setting the seed.
library(MASS)
library(psych)

# Specify the number of trials.
trials <- 100

# Specify the number draws that were used to construct the posterior distribution.
D <- 300

#########################################################################
#####################################
## Generate and write out posterior predictive data.
## Conduct PPMC.
#########################################################################
#####################################

# Create the list of conditions.
sample.size <- matrix(c("SN","MN","LN"))
quad.mean <- matrix(c("NQM", "SQM", "LQM"))
quad.sd <- matrix(c("NQV", "SQV", "LQV"))
res.sd <- matrix(c("SRV", "LRV"))
dist.shape <- matrix(c("NN"))
dist.shape <- matrix(c("N", "NN"))

conditions <- expand.grid(sample.size, quad.mean, quad.sd, res.sd, dist.shape)
colnames(conditions) <- c("samplesize", "quadmean", "quadsd", "ressd", "distshape")

attach(conditions)
conditions$remove <- ifelse(quadmean=="NQM" & (quadsd=="SQV"|quadsd=="LQV"),
1, 0)
detach(conditions)
conditions <- subset(conditions, remove==0, select=-remove)
```

```
#conditions <- droplevels(subset(conditions, samplesize=="SN" & quadmean=="NQM"
& quadsd=="NQV" & ressd=="SRV" & distshape=="N"))
#conditions <- droplevels(conditions[1, ])
# Specify constants that will be used throughout for indexing through loops.
gendir <- "/Volumes/Seagate Backup Plus Drive/GCM Dissertation"

# Negative log-likelihood for the model.
negtwoLL.f <- function(N, J, sigma, y, mu.hat){
  ind1 <- -1*(.5*(mahalanobis(y, mu.hat, sigma, inverted=FALSE)))
  constant <- -(((N*J)/2)*log(2*pi)) - ((N/2)*log(det(sigma)))
  ind2 <- constant + sum(ind1)
  -2*ind2
}

##########################################################################
##########################################################################
######
## For each model within each trial, open the parameter estimates.
##########################################################################
##########################################################################
######

# use the system time function to time the computations.
system.time(

  # Open the loop over conditions.
  for(wc in 1:nrow(conditions)){

    # Specify paths to obtain requisite data.
    condition.directory <- paste(gendir, "/", conditions[wc,1], "_", conditions[wc,2], "_",
conditions[wc,3], "_", conditions[wc,4], "_", conditions[wc,5], sep="")
    observed.data.directory <- paste(condition.directory, "/data/", "ObservedData", sep="")
    estimation.directory <- paste(condition.directory, "/", "Model Estimation Data",
sep="")
    postpred.directory <- paste(condition.directory, "/data/", "Posterior Predictive Data",
sep="")
    expectation.directory <- paste(condition.directory, "/data/", "Model Expectation",
sep="")
    results.directory <- paste(condition.directory, "/data/", "Results", sep="")
    #dir.create(results.directory)

    print(conditions[wc, ])

    # Set up empty matrices to compute posterior predictive p-values for each trial.
    NEGTWOLLp <- matrix(NA, trials, 6)
```

```
   # Set up the loop over trials.
   for(wt in 1:trials){

     # Specify a null matrix for the draws for each fo the posterior predictive measures.
     NEGTWOLL <- matrix(NA, D, 20)

     # Open the observed data.
     setwd(observed.data.directory)
     y.obs <- as.matrix(read.table(paste("obs.trial", wt, ".dat", sep=""), header=F,
sep="\t"))
     colnames(y.obs) <- NULL
     N <- nrow(y.obs)
     J <- ncol(y.obs)
     print(paste("Trial = ", wt, sep=""))

     # Model 1 files.
     setwd(estimation.directory)
     gammas.m1 <- as.matrix(read.table(file=paste("gammas.model1.trial", wt, ".dat",
sep=""), header=T, sep="\t"))
     psi.m1 <- as.matrix(read.table(file=paste("psi.model1.trial", wt, ".dat", sep=""),
header=T, sep="\t"))
     theta.m1 <- as.matrix(read.table(file=paste("theta.model1.trial", wt, ".dat", sep=""),
header=T, sep="\t"))

     # Model 2 files.
     setwd(estimation.directory)
     gammas.m2 <- as.matrix(read.table(file=paste("gammas.model2.trial", wt, ".dat",
sep=""), header=T, sep="\t"))
     psi.m2 <- as.matrix(read.table(file=paste("psi.model2.trial", wt, ".dat", sep=""),
header=T, sep="\t"))
     theta.m2 <- as.matrix(read.table(file=paste("theta.model2.trial", wt, ".dat", sep=""),
header=T, sep="\t"))

     # Model 3 files.
     setwd(estimation.directory)
     gammas.m3 <- as.matrix(read.table(file=paste("gammas.model3.trial", wt, ".dat",
sep=""), header=T, sep="\t"))
     psi.m3 <- as.matrix(read.table(file=paste("psi.model3.trial", wt, ".dat", sep=""),
header=T, sep="\t"))
     theta.m3 <- as.matrix(read.table(file=paste("theta.model3.trial", wt, ".dat", sep=""),
header=T, sep="\t"))

     # Create lambda matrices that will be used for computing the model implied mean
vector and covariance matrix.
```

```
    time <- as.matrix(seq(from=0, to=J-1, by=1))
    lambda1 <- as.matrix(cbind(rep(1,J), time))
    lambda2 <- as.matrix(cbind(rep(1,J), time, time^2))

    #for(wm in 1:models){

    for(d in 1:D){

      if(d==75|d==150|d==225) print(paste("Draw = ", d, sep=""))

      # Open the posterior predictive data.
      setwd(postpred.directory)
      y.rep.m1 <- read.table(paste("yrep_model1_trial", wt, "_draw", d, ".dat", sep=""),
header=F, sep="\t")
      y.rep.m2 <- read.table(paste("yrep_model2_trial", wt, "_draw", d, ".dat", sep=""),
header=F, sep="\t")
      y.rep.m3 <- read.table(paste("yrep_model3_trial", wt, "_draw", d, ".dat", sep=""),
header=F, sep="\t")

      # Compute key functions to support the computations of discrepancy functions.

      # Model implied covariance matrix.
      psi.m1z <- matrix(c(psi.m1[d,1], psi.m1[d,2],
                  psi.m1[d,2], psi.m1[d,3]), ncol=2, nrow=2)

      psi.m2z <- matrix(c(psi.m2[d,1], psi.m2[d,2], 0,
                  psi.m2[d,2], psi.m2[d,3], 0,
                  0, 0, 0), ncol=3, nrow=3)

      psi.m3z <- matrix(c(psi.m3[d,1], psi.m3[d,2], 0,
                  psi.m3[d,2], psi.m3[d,3], 0,
                  0, 0, psi.m3[d,4]), ncol=3, nrow=3)

      sigma.m1 <- lambda1%*%psi.m1z%*%t(lambda1) + diag(theta.m1[d,1], J)
      sigma.m2 <- lambda2%*%psi.m2z%*%t(lambda2) + diag(theta.m2[d,1], J)
      sigma.m3 <- lambda2%*%psi.m3z%*%t(lambda2) + diag(theta.m3[d,1], J)

      # Model implied mean vector.
      mu.hat.m1 <- t(as.matrix(gammas.m1[d, ]%*%t(lambda1)))
      mu.hat.m2 <- t(as.matrix(gammas.m2[d, ]%*%t(lambda2)))
      mu.hat.m3 <- t(as.matrix(gammas.m3[d, ]%*%t(lambda2)))


#########################################################################
###
```

```
      ## Compute the discrepancy functions.
#############################################################################
###

      NEGTWOLL[d,1] <- negtwoLL.f(N=N, J=J, sigma=sigma.m1, y=y.obs,
mu.hat=mu.hat.m1)
      NEGTWOLL[d,2] <- negtwoLL.f(N=N, J=J, sigma=sigma.m2, y=y.obs,
mu.hat=mu.hat.m2)
      NEGTWOLL[d,3] <- negtwoLL.f(N=N, J=J, sigma=sigma.m3, y=y.obs,
mu.hat=mu.hat.m3)

      NEGTWOLL[d,4] <- negtwoLL.f(N=N, J=J, sigma=sigma.m1, y=y.rep.m1,
mu.hat=mu.hat.m1)
      NEGTWOLL[d,5] <- negtwoLL.f(N=N, J=J, sigma=sigma.m2, y=y.rep.m1,
mu.hat=mu.hat.m2)
      NEGTWOLL[d,6] <- negtwoLL.f(N=N, J=J, sigma=sigma.m3, y=y.rep.m1,
mu.hat=mu.hat.m3)

      NEGTWOLL[d,7] <- negtwoLL.f(N=N, J=J, sigma=sigma.m1, y=y.rep.m2,
mu.hat=mu.hat.m1)
      NEGTWOLL[d,8] <- negtwoLL.f(N=N, J=J, sigma=sigma.m2, y=y.rep.m2,
mu.hat=mu.hat.m2)
      NEGTWOLL[d,9] <- negtwoLL.f(N=N, J=J, sigma=sigma.m3, y=y.rep.m2,
mu.hat=mu.hat.m3)

      NEGTWOLL[d,10] <- negtwoLL.f(N=N, J=J, sigma=sigma.m1, y=y.rep.m3,
mu.hat=mu.hat.m1)
      NEGTWOLL[d,11] <- negtwoLL.f(N=N, J=J, sigma=sigma.m2, y=y.rep.m3,
mu.hat=mu.hat.m2)
      NEGTWOLL[d,12] <- negtwoLL.f(N=N, J=J, sigma=sigma.m3, y=y.rep.m3,
mu.hat=mu.hat.m3)

    } # Close the loop over draws.

    # Realized values.
    NEGTWOLL[ ,13] <- NEGTWOLL[,1]-NEGTWOLL[,2]
    NEGTWOLL[ ,14] <- NEGTWOLL[,2]-NEGTWOLL[,3]

    # Posterior predictive values given model 1.
    NEGTWOLL[ ,15] <- NEGTWOLL[,4]-NEGTWOLL[,5]
    NEGTWOLL[ ,16] <- NEGTWOLL[,5]-NEGTWOLL[,6]

    # Posterior predictive values given model 2.
    NEGTWOLL[ ,17] <- NEGTWOLL[,7]-NEGTWOLL[,8]
    NEGTWOLL[ ,18] <- NEGTWOLL[,8]-NEGTWOLL[,9]
```

```
    # Posterior predictive values given model 3.
    NEGTWOLL[ ,19] <- NEGTWOLL[,10]-NEGTWOLL[,11]
    NEGTWOLL[ ,20] <- NEGTWOLL[,11]-NEGTWOLL[,12]

    # PPP-values given model 1.
    NEGTWOLLp[wt,1] <- mean(NEGTWOLL[,15]>=NEGTWOLL[,13])
    NEGTWOLLp[wt,2] <- mean(NEGTWOLL[,16]>=NEGTWOLL[,14])

    # PPP-values given model 2.
    NEGTWOLLp[wt,3] <- mean(NEGTWOLL[,17]>=NEGTWOLL[,13])
    NEGTWOLLp[wt,4] <- mean(NEGTWOLL[,18]>=NEGTWOLL[,14])

    # PPP-values given model 3.
    NEGTWOLLp[wt,5] <- mean(NEGTWOLL[,19]>=NEGTWOLL[,13])
    NEGTWOLLp[wt,6] <- mean(NEGTWOLL[,20]>=NEGTWOLL[,14])

    # Write out the values of the realized and posterior predictive values of the
discrepancy functions.

    setwd(results.directory)
    write.table(NEGTWOLL, paste("New_NEGTWOLL_Draws_Trial", wt, ".dat",
sep=""), row.names=F, col.names=F, quote=F, sep="\t")
    if(wt==trials) NEGTWOLLp <- cbind(rep("NEGTWOLL", trials), seq(from=1,
to=trials, by=1), NEGTWOLLp)
    if(wt==trials) colnames(NEGTWOLLp) <- c("Function", "Trial", "M1M2_M1",
"M2M3_M1", "M1M2_M2", "M2M3_M2", "M1M2_M3", "M2M3_M3")
    if(wt==trials) write.csv(NEGTWOLLp, file="New NEGTWOLL Posterior Predictive
PValues.csv", na=", row.names=F)

   } # Close the loop over trials.
  } # Close the loop over conditions.

)
```