Examination of Mixed-Effects Models with Nonparametrically Generated Data

by

Kimberly L. Fine

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

Approved April 2019 to the
Graduate Supervisory Committee:

Kevin Grimm, Chair
Mike Edwards
Holly O'Rourke
Dan McNeish

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

Previous research has shown functional mixed-effects models and traditional mixed-effects models perform similarly when recovering mean and individual trajectories (Fine, Suk, & Grimm, 2019). However, Fine et al. (2019) showed traditional mixed-effects models were able to more accurately recover the underlying mean curves compared to functional mixed-effects models. That project generated data following a parametric structure. This paper extended previous work and aimed to compare nonlinear mixed-effects models and functional mixed-effects models on their ability to recover underlying trajectories which were generated from an inherently nonparametric process. This paper introduces readers to nonlinear mixed-effects models and functional mixed-effects models. A simulation study is then presented where the mean and random effects structure of the simulated data were generated using B-splines. The accuracy of recovered curves was examined under various conditions including sample size, number of time points per curve, and measurement design. Results showed the functional mixed-effects models recovered the underlying mean curve more accurately than the nonlinear mixed-effects models. In general, the functional mixed-effects models recovered the underlying individual curves more accurately than the nonlinear mixed-effects models. Progesterone cycle data from Brumback and Rice (1998) were then analyzed to demonstrate the utility of both models. Both models were shown to perform similarly when analyzing the progesterone data.

ACKNOWLEDGMENTS

Thank you to my committee members Drs. Kevin Grimm, Mike Edwards, Holly O'Rourke, and Dan McNeish for your help in developing this document. Your comments brought great insight into helping improve this project and future projects to come. Your sarcasm, humor, and sass will always be cherished.

Justin Fine, this document and culmination of work would not be possible without your endless love, support, and guidance. Thank you for believing in me and always pushing me. Thank you for keeping me rooted when I thought I was going to fall; I love you, my tree. Blair and Candace Bear, you two are the most inspirational women I know. Thank you for guiding, growing, falling, laughing, crying, and loving with me. Your love and support are a huge part of why I am who I am today. I never could have done this without my PolyFam.

Thank you to my family. Mom and Dad, thanks for always being there when I need you. Max, Maggie, and Kelly, thanks for being there to distract me with puppy pictures and make me laugh.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

A main component of research in the behavioral sciences involves studying variables collected across time. One of the goals of this research is to study individual and group change over time. The primary analytical approach to analyzing longitudinal data is growth curve modeling (McArdle, 1988; Meredith & Tisak, 1990; Raudenbush & Bryk, 2002; Singer & Willett, 2003). Growth curve modeling allows researchers to describe the underlying change process for each individual, as well as the overall change for the sample.

Traditionally, growth curve models have been estimated in two analytic frameworks: structural equation modeling and mixed-effects modeling. Structural equation modeling is a regression-based framework that encompasses path models, latent variable models, and latent variable path models. Researchers often use structural equation models to study outcomes that are not directly observed. Mixed-effects models, also referred to as random coefficient models, hierarchal linear models, and multilevel models, use fixed and random effects to describe aspects of change. Broadly, mixed-effects models for longitudinal data allow for the examination of individual change trajectories through estimation random effects parameters. The fixed effects parameters describe the average trajectory for the sample, and the random effects parameters capture how each person deviates from this average trajectory. While structural equation models are extremely useful, I focused on the mixed-effects framework for the estimation of growth curve models given the ability to directly estimate growth models that follow inherently nonlinear functions.

Historically, there are two types of mixed-effects models: linear mixed-effects models and nonlinear mixed-effects. Both types of models can be used to describe trajectories whose rate of change varies over time; however, the distinction between the two is how the fixed and random effects enter the model. Linear mixed-effects models only contain parameters that enter the model in a linear fashion. These types of models can be used to capture nonlinear trajectories via transformations of the timing variable (e.g., quadratic growth). Nonlinear mixed-effects models, on the other hand, have parameters that enter the model in a nonlinear fashion. Parameters that enter the model in a nonlinear fashion are contained in the mathematical function of time (e.g., contained within the exponent). These types of models are more flexible and can capture the nonlinearity in the data through inherently nonlinear functions, such as exponential growth or decline models.

Grimm, Ram, and Estabrook (2016) outlined two types of nonlinear mixed-effects models. First, there are nonlinear models that are nonlinear with respect to the *fixed effects* parameters. These statistical models have at least one fixed effect parameter that enters the model in a nonlinear fashion. However, the fixed effects that enter the model in a nonlinear fashion do not have an associated random effect since all random effects in these models enter the model in a linear fashion. These models have also been referred to as conditionally linear models (Blozis & Cudeck, 1999). An example of this kind of model would be an exponential growth model where the rate parameter is the same across participants. The second class of nonlinear mixed-effects models are nonlinear with respect to the *random effects*. These models are often referred to as *fully nonlinear*

models, and an example of such a model is the exponential growth model where the rate parameter varies across participants.

Traditionally, growth curve models, including both mixed-effects and structural equation models, are specified as parametric models. Parametric models assume the functional form of growth is known (e.g., linear, quadratic, exponential; for exceptions see Grimm, Steele, Ram, & Nesselroade, 2013) and the functional form is characterized by a relatively small number of parameters. The parameters in a parametric model have a meaningful interpretation in the context of study (Ram & Grimm, 2007). Parametric models work well when there is *a priori* knowledge about the functional form of growth or when there are few repeated assessments. However, they may be too restrictive if there is little or no information about the functional form of growth or when there are many repeated assessments with a complex change pattern. If this is the case, the researcher can specify a nonparametric model (Fitzmaurice, Davidian, Verbeke, & Molenberghs, 2009), where the specific functional form of growth is not assumed to be known. Nonparametric models allow researchers to describe temporal growth, without imposing any specific growth pattern on the data. The term *nonparametric* can be confusing as some models that fall into the nonparametric category have estimated parameters. However, the parameters in nonparametric models may not capture the same meaningful aspects of change as the parametric models provide (Eilers & Marx, 1996). Instead, the estimated parameters in nonparametric models are used to define a curve numerically in such a way that the curve can be evaluated at any specific point in time.

One such example of a nonparametric growth curve model is the functional mixed-effects model. A functional mixed-effects model is the functional extension of the

mixed-effects model. Within the functional framework, data are assumed to arise from smooth underlying curves that do not hold a specific functional form. Rather than constraining the data to follow a predetermined trajectory as in a linear or nonlinear mixed-effects model, the trajectory is estimated from the data through smoothing. There are several formulations of the functional mixed-effects model, this paper's focus was on the smoothing spline mixed-effects model formulation.

Functional mixed-effects models have seen scarce use in the behavioral sciences. Fine, Suk, and Grimm (2019) conducted a simulation study investigating how functional mixed-effects models compared to linear and nonlinear mixed-effects models when analyzing longitudinal data. The functional mixed-effects models performed very similarly to the linear and nonlinear mixed-effects models. However, the linear mixed-effects model estimated the mean and individual curves better than the functional mixed-effects model, and the nonlinear mixed-effects model estimated the mean curve better than the functional mixed-effects model. While the functional mixed-effects model estimated the individual curves better than the nonlinear mixed-effects model. The data generation process for the mean curves used in this simulation were parametric and the same parametric form, with strong starting values, were used in estimating the linear and nonlinear mixed-effects models. Therefore, it is not surprising the parametric models outperformed the nonparametric functional mixed-effects model.

The goal of this dissertation was to continue exposing the behavioral sciences to functional mixed-effects models and to extend the work of Fine et al. (2019). This project aimed to compare nonlinear and functional mixed-effects models when the data were generated from an inherently nonparametric process. More detail on nonlinear and

functional mixed-effects models are given in the next sections, followed by a comparison of both models. Finally, details and results of the proposed simulation and empirical example are presented

NONLINEAR MIXED-EFFECTS MODELS

Nonlinear mixed-effects models take on the general form

$$y_{ij} = f\left(t_{ij}, \boldsymbol{p}_i\right) + e_{ij}, \tag{1}$$

$$\boldsymbol{p}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i, \tag{2}$$

where y_{ij} is the outcome variable for person i ($i = 1, \dots, n$) at time j ($j = 1, \dots, n_i$). A total of $N = \sum_{i=1}^{n} n_i$ response are observed. The outcome is modeled by some function, $f()$, of the timing variable, t_{ij}, and individual specific parameters, $\boldsymbol{p}_i$. The individual specific parameters, $\boldsymbol{p}_i$, are composed of fixed effects, $\boldsymbol{\beta}$, and random effects, $\boldsymbol{b}_i$. The fixed effects, $\boldsymbol{\beta}$, model aspects of the trajectory that describe all participants. The vector $\boldsymbol{\beta}$ is length $(p \times 1)$, meaning there are p fixed effects in the model. The design matrix $\boldsymbol{X}_i$ is dimension $(n_i \times p)$. The random effects, $\boldsymbol{b}_i$, are individual specific deviations from the fixed effects. The vector $\boldsymbol{b}_i$ is length $(s \times 1)$, meaning there are s random effects in the model. The design matrix $\boldsymbol{Z}_i$ is dimension $(n_i \times s)$. The random effects are assumed to follow a normal distribution with mean 0 and covariance matrix $\boldsymbol{D}$, such that

$$\boldsymbol{b}_i \sim MVN(\mathbf{0}, \mathbf{D}) \tag{3}$$

where $\mathbf{D}$ has dimension $(s \times s)$. The individual specific residuals, e_{ij}, are also assumed to follow a normal distribution with mean 0 and covariance matrix $\mathbf{R}_i$

$$\boldsymbol{e}_i \sim N(0, \mathbf{R}_i). \tag{4}$$

The covariance matrix for the level-1 residuals, $\mathbf{R}_i$, is often assumed to be diagonal with constant variance; however, this assumption can be relaxed. The vector $\boldsymbol{e}_i$ has dimension $(n_i \times 1)$ and the covariance matrix $\mathbf{R}_i$ has dimension $(n_i \times n_i)$.

The parameters of a nonlinear mixed-effects model are often estimated using maximum likelihood. For ease of presentation, the above vectors and matrices are rearranged such that $\boldsymbol{y}$ is $(N \times 1)$, $\mathbf{X}_i$ is $(N \times p)$, $\mathbf{b}$ is $(sn \times 1)$, $\boldsymbol{e}$ is $(N \times 1)$, and $\boldsymbol{p}$ is $(N \times 1)$. Block diagonal matrices are formed with $\mathbf{D}$ $(sn \times sn)$, $\mathbf{Z}$ $(N \times sn)$, and $\mathbf{R}$ $(N \times N)$. Maximum likelihood attempts to solve for the parameters of the model by maximizing the likelihood function, which can be written as

$$L(\boldsymbol{\beta}, \boldsymbol{R}, \boldsymbol{D}) = \int p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{R}, \boldsymbol{D}) p(\boldsymbol{b}|\boldsymbol{D}, \boldsymbol{R}) d\boldsymbol{b}, \tag{5}$$

$$= \int \left[(2\pi)^{\frac{N}{2}} |\boldsymbol{R}|^{\frac{1}{2}}\right]^{-1} e^{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{p}))' \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{p}))} \left[(2\pi)^{\frac{N}{2}} |\boldsymbol{D}|^{\frac{1}{2}}\right]^{-1} e^{-\frac{1}{2}\boldsymbol{b}' \boldsymbol{D}^{-1} \boldsymbol{b}} d\boldsymbol{b} \tag{6}$$

$$= \int \left[(2\pi)^{\frac{N}{2}} |\boldsymbol{R}|^{\frac{1}{2}} (2\pi)^{\frac{N}{2}} |\boldsymbol{D}|^{\frac{1}{2}}\right]^{-1} e^{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{p}))' \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{p})) + \boldsymbol{b}' \boldsymbol{D}^{-1} \boldsymbol{b}} d\boldsymbol{b} \tag{7}$$

where $\boldsymbol{y}$ is the vector of responses, $\boldsymbol{\beta}$ is the vector of fixed effects, $\boldsymbol{b}$ is the vector of random effects, $\boldsymbol{D}$ is the covariance matrix of random effects, and $\boldsymbol{R}$ is the matrix of level-1 residuals. $L(\boldsymbol{\beta}, \boldsymbol{R}, \boldsymbol{D})$ is the likelihood, $p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{R}, \boldsymbol{D})$ is the conditional distribution of the data ($\boldsymbol{y}$) given the random effects ($\boldsymbol{b}$) and the parameters of the model ($\boldsymbol{\beta}, \boldsymbol{R}$, and $\boldsymbol{D}$), and $p(\boldsymbol{b}|\boldsymbol{D}, \boldsymbol{R})$ is the probability of the random effects ($\boldsymbol{b}$) given the random components ($\boldsymbol{D}$ and $\boldsymbol{R}$). In linear mixed-effects models, integrating out the random effects is relatively straightforward because the random effects enter the model in a linear fashion. In general for nonlinear mixed-effects models, it is impossible to find an

analytic solution to the likelihood function in Equation 7 due to the nonlinear fashion that the random effects enter the model (Wolfinger & Lin, 1997).

FUNCTIONAL MIXED-EFFECTS MODELS

Functional mixed-effects models are an extension of the standard mixed-effects model. These models contain both fixed and random effects (Searle, Casella, & McCulloch, 1992) and are used to analyze data with a nested structure. Their ability to model nested structures make them valuable tools for analyzing longitudinal data, where observations are nested within persons. Functional mixed-effects models extend the linear mixed-effects model by assuming the mean growth curve and person-specific deviation curves are represented by nonparametric smooth functions. These smooth functions can be constructed using kernel smoothing, regression splines, penalized splines, or smoothing splines (Wu & Zhang, 2006). The specific formulation of functional mixed-effects model I focused on is the smoothing spline mixed-effects models.

Smoothing spline mixed-effects models are functional mixed-effects models that use smoothing splines to construct the underlying curves. In order to define a smoothing spline, the definition of a *spline* is needed. A $(k+1)$-th order spline is a continuous piecewise polynomial function of degree k, that has continuous derivatives of orders up to $k-1$ at its knot points $\tau_1 < \cdots < \tau_{\mathrm{T}}$. There exists so-called *natural splines* that are defined just as splines, but with the additional constraints that outside the knots the polynomial is of degree $(k-1)/2$. The most common of the natural splines is the *cubic* natural spline. Cubic natural splines are constrained to be linear outside the knot points.

Smoothing splines are natural splines that place knot points at all unique time points and use a penalty parameter to control overfitting.

Splines are constructed using a set of basis functions. Basis functions can be thought of as an extension of basis vectors from the vector space to the function space. There are multiple sets of basis vectors that define each vector space. Every set of basis vectors can be used to express any vector within that vector space, with linear combinations of the basis vectors. Extending this concept to the function space, there are multiple sets of basis functions that define a function space. Functions can be expressed via linear combinations of any set of basis functions that define that function space. Fourier, wavelets, B-splines, polynomials, and truncated power are some examples of types of basis functions. While any set of basis functions can be used to construct smoothing splines, truncated power and B-splines are the most common. Technical details for B-splines and truncated power splines are provided in Appendix C and Appendix D, respectively.

As mentioned above, smoothing splines are natural splines that place knots at all unique time points and use a penalty parameter to control overfitting. Regardless of the choice of basis, knots are placed at each unique time point in the data. The penalty parameter controls the smoothness of the curve so that the smoothing spline is not over- or under-smoothed. A cubic smoothing spline can be obtained by minimizing the penalized sum of squares (PSS)

$$PSS = \sum_{i=1}^{n} \left[y_i - \hat{f}_i(t_i) \right]^2 + \lambda \int_{t_{min}}^{t_{max}} \hat{f}''(t)^2 dt. \tag{8}$$

Here, $\hat{f}(t)$ is the cubic smoothing spline fit to the underlying function $f(t)$. $\hat{f}''(t)$ is the second derivative of a smooth function $f(t)$, and $\int \left(\hat{f}''(t)\right)^2 dt$ is a roughness penalty. The penalized least squares approach aims to minimize the least squares criterion (i.e., sum of squared residuals) with the addition of a roughness penalty term. The first term in Equation 8 is the sum of squared residuals measuring the goodness-of-fit to the data. The second term is a roughness penalty, which is multiplied by a non-negative smoothing parameter λ. If $\hat{f}(t)$ is a straight line having no curvature, the second derivative ($\hat{f}''(t)$) will be zero over the entire range of t. If $\hat{f}(t)$ has a curvature at t, the second derivative of this function at t will deviate from 0. The squared second derivative of this function at t indicates the roughness of the function at t. Therefore, the integrated squared second derivative of this function indicates the overall roughness of $\hat{f}(t)$ across the entire range of t. The non-negative smoothing parameter λ controls the importance of the penalty term. When $\lambda = 0$, the penalty term vanishes and the function $\hat{f}(t)$ is estimated in such a way that it fits the data as closely as possible. When $\lambda = \infty$, the penalty term will dominate the criterion in Equation 8 and even a tiny amount of curvature in $\hat{f}(t)$ will yield a huge increase in the criterion. In other words, larger smoothing parameters will yield smoother curves.

Cubic smoothing splines are widely used because they are simple to construct (Green & Silverman, 1994; Wu & Zhang, 2006). By placing knots at each unique time point, smoothing splines hold the advantage of not needing to choose how many and where knots are placed. A disadvantage of smoothing splines is when the number of

unique time points in the data set is large because placing many knots can be computationally expensive.

The smoothing spline formulation of a functional mixed-effects models take the form

$$y_{ij} = \mu(t_{ij}) + \nu_i(t_{ij}) + e_{ij}, \tag{9}$$

where y_{ij} is the outcome variable for person *i* ($i = 1, \ldots, n$) at time t_{ij} ($j = 1, \ldots, n_i$). The outcome for person *i* at time t_{ij} is modeled as a combination of $\mu(t_{ij})$, a grand mean function evaluated at time t_{ij}, and $\nu_i(t_{ij})$, a person specific deviation function for person *i* evaluated at time t_{ij}, indicating how much person *i*'s growth curve deviates from the mean growth curve at time t_{ij}. Person-specific curve for person *i* can be obtained by $\mu(t_{ij}) + \nu_i(t_{ij})$. The residual, e_{ij}, shows how much person *i*'s observed response deviates from their specific trajectory at time t_{ij}.

Equation 9 can be rewritten in matrix notation as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\mu} + \boldsymbol{Z}_i\boldsymbol{\nu}_i + \boldsymbol{e}_i, \tag{10}$$

where $\boldsymbol{y}_i$ and $\boldsymbol{e}_i$ are both column vectors of length n_i containing the observed responses and residuals for person *i*, respectively. $\boldsymbol{\mu}$ is a column vector containing the values of the mean growth curve, $\mu(t_{ij})$, evaluated at all unique time points in the data. $\boldsymbol{\nu}_i$ is a column vector containing the values of the person-specific deviation curve, $\nu_i(t_{ij})$, evaluated at all unique time points in the data. $\mathbf{X}_i$ is an n_i by *U* incidence matrix, where *U* is equal to the number of unique time points. The j^{th} row of this matrix is a 1 by *U* indicator vector indicating which value of the distinct unique time points equals t_{ij}. That is, the u^{th} element of the indicator vector is 1 if the j^{th} time point, t_{ij}, equals the u^{th} unique time

point and all the other elements are 0. $\boldsymbol{\mu}$ contains fixed effects and $\mathbf{v}_i$ contains random effects. The random effects are assumed to follow multivariate normal distributions, $\mathbf{v}_i \sim MVN(\mathbf{0}, \mathbf{D})$ and $\mathbf{e}_i \sim N(0, \mathbf{R}_i)$. The covariance matrix for the level-1 residuals, $\mathbf{R}_i$, is often assumed to be diagonal with constant variance; however, this assumption can be relaxed.

The parameters of the model in Equation 10 can be estimated by minimizing the following penalized generalized log likelihood criterion (PGLL; Wu & Zhang, 2006)

$$\begin{aligned} \text{PGLL} = \sum_{i=1}^{n} \{ & [\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\mu} - \mathbf{Z}_i\mathbf{v}_i]^T \mathbf{R}_i^{-1} [\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\mu} - \mathbf{Z}_i\mathbf{v}_i] + \mathbf{v}_i^T \mathbf{D}^{-1} \mathbf{v}_i \\ & + \log|\mathbf{D}| + \log|\mathbf{R}_i| \} + \lambda_\mu \boldsymbol{\mu}^T \mathbf{G} \boldsymbol{\mu} + \lambda_v \sum_{i=1}^{n} \{ \mathbf{v}_i^T \mathbf{G} \mathbf{v}_i \}, \end{aligned} \tag{11}$$

where the matrix **G** is the so-called roughness matrix as defined in Green and Silverman (1994, p.25, ***K*** matrix) and Wu and Zhang (2006, p.55, **G** matrix). The fixed effects ($\boldsymbol{\mu}$), random effects ($\mathbf{v}_i$), and variance components ($\mathbf{D}$ and $\mathbf{R}_i$) can be obtained by minimizing the PGLL criterion using an EM algorithm based on REML (Wu & Zhang, 2006, *p*.170). After obtaining $\hat{\boldsymbol{\mu}}$, $\hat{\mathbf{v}}_i$, $\widehat{\mathbf{D}}$, and $\widehat{\mathbf{R}}_i$ the estimated fixed effect curve, $\hat{\mu}(t)$, and the predicted random effect curves, $\hat{v}_i(t)$, can be obtained at any value of t using a formula given in Green and Silverman (1994, Chapter 2.4) or by simple interpolation.

Equation 8 for the single smoothing spline contains several similar components to Equation 11 for the mixed-effects smoothing spline. The first part of Equation 8, $\sum_{i=1}^{n} [y_i - \hat{f}_i(t_i)]^2$, is a similar structure to $[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\mu} - \mathbf{Z}_i\mathbf{v}_i]$ from Equation 11. Both are general "observed minus predicted" structures. The last part in Equation 8, $\lambda \int_{t_{min}}^{t_{max}} \hat{f}''(t)^2 dt$, is similar to the last two components in Equation 11, $\lambda_\mu \boldsymbol{\mu}^T \mathbf{G} \boldsymbol{\mu}$ and

$\lambda_v \sum_{i=1}^{n}\{\mathbf{v}_i^T \mathbf{G}\mathbf{v}_i\}$. All of these terms are used to control the smoothness of the smoothing spline.

Wu and Zhang (2006) used a model selection approach to determine the smoothing parameters λ_μ and λ_v. To begin, a grid of λ_μ and λ_v values are prespecified. At each pair of λ_μ and λ_v values, AIC, BIC, or the generalized cross validation (GCV) value can be obtained. The pair of λ_μ and λ_v values associated with either the lowest AIC, BIC, or GCV are selected as optimal smoothing parameters.

COMPARISON OF NONLINEAR AND FUNCTIONAL MIXED-EFFECTS MODELS

Scarce literature is available explicitly comparing these two methods of mixed-effects models. This chapter aimed to help fill that gap. A comparison on the structure of the models and on the criterions for the estimation of parameters of the models is presented, and is concluded with a discussion of the strengths and weakness of each model.

The nonlinear mixed-effects model (Equations 1 and 2) and the functional mixed-effects model (Equations 9 and 10) are presented again here for clarity. The nonlinear mixed-effects model is

$$y_{ij} = f(t_{ij}, \boldsymbol{p}_i) + e_{ij}, \tag{12}$$

$$\boldsymbol{p}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i, \tag{13}$$

and the functional mixed-effects model is

$$y_{ij} = \mu(t_{ij}) + \nu_i(t_{ij}) + e_{ij}, \tag{14}$$

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\mu} + \boldsymbol{Z}_i\boldsymbol{\nu}_i + \boldsymbol{e}_i. \quad (15)$$

Both of these models handle nested data so each subscript *i* and *j* index people ($i = 1, \ldots, n$) and time ($j = 1, \ldots, n_i$). Both models assume their random effects and residuals arise from a normal distributions with covariance matrices **D** and $\mathbf{R}_i$. Both models have residual variance matrices, $\mathbf{R}_i$, of dimension ($n_i \times n_i$), where n_i is the number of observations for person *i*. The two models differ in terms of their random effects matrix, $\boldsymbol{D}$. In the nonlinear mixed-effects model, this matrix is of size ($s \times s$), where s is the number of random effects in the model. For the functional mixed-effects model, the random effects matrix is of size (T $\times$ T), where T is the number of knots in the model. The residual vectors from both models, of dimension ($n_i \times 1$), represent deviations from each person's predicted trajectory and their observed data.

Both models utilize fixed and random effects to model average and person specific change. The fixed effects in the nonlinear mixed-effects model are contained in the ($p \times 1$) vector $\boldsymbol{\beta}$; this vector contains the values of the fixed effect parameters (p values in total). The fixed effects in the functional mixed-effects model are contained in the dimension (U $\times$ 1) vector $\boldsymbol{\mu}$; this vector contains the values of the mean growth curve ($\mu(t_{ij})$) evaluated at all unique time points in the data. The random effects in the nonlinear mixed-effects model are contained in the ($s \times 1$) vector $\boldsymbol{b}_i$; this vector contains the values of the random effect parameters (s values in total) for each person. The random effects for the functional mixed-effects model are contained in the (U $\times$ 1) vector $\mathbf{v}_i$; this vector contains the values of the person-specific deviation curve ($v_i(t_{ij})$) evaluated at all unique time points in the data. In practice, the value of the random effects are not estimated in both models. Rather, the random effects are assumed to follow a

normal distribution and the variance/covariance matrix of the random effects is estimated instead.

The last components of both models are the matrices $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$. In the nonlinear mixed-effects model, these matrices are of dimension $(n_i \times p)$ and $(n_i \times s)$, respectively. The first column of these matrices contain a column of ones. For the matrix $\boldsymbol{X}_i$ this column is for the fixed intercept. The matrix $\boldsymbol{Z}_i$ would also have its first column be a column of ones if there is a random intercept to be specified. The rest of the columns in these matrices correspond to predictors that have either fixed or random effects. In nonlinear mixed-effects models, these matrices are allowed to differ in the number of columns. In functional mixed-effects models, the design matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ must be identical. These matrices are each an n_i by U incidence matrix, where U is equal to the number of unique time points. The j^{th} row of this matrix is a 1 by U indicator vector indicating which value of the distinct unique time points equals t_{ij}. That is, the u^{th} element of the indicator vector is 1 if the j^{th} time point, t_{ij}, equals the u^{th} unique time point and all the other elements are 0.

The criterion for the nonlinear mixed-effects model (Equation 7) and the functional mixed-effects model (Equation 11) are presented here for clarity. The criterion for the nonlinear mixed-effects model is

$$= \int \left[(2\pi)^{\frac{N}{2}} |\boldsymbol{R}|^{\frac{1}{2}} (2\pi)^{\frac{N}{2}} |\boldsymbol{D}|^{\frac{1}{2}} \right]^{-1} e^{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{p}))' \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{p})) + \boldsymbol{b}' \boldsymbol{D}^{-1} \boldsymbol{b}} d\boldsymbol{b}, \qquad (16)$$

and the criterion for the functional mixed-effects model is

$$\text{PGLL} = \sum_{i=1}^{n} \{[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\mu} - \mathbf{Z}_i\mathbf{v}_i]^T\mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\mu} - \mathbf{Z}_i\mathbf{v}_i] + \mathbf{v}_i^T\mathbf{D}^{-1}\mathbf{v}_i + \log|\mathbf{D}| + \log|\mathbf{R}_i|\} + \lambda_\mu\boldsymbol{\mu}^T\mathbf{G}\boldsymbol{\mu} + \lambda_v\sum_{i=1}^{n}\{\mathbf{v}_i^T\mathbf{G}\mathbf{v}_i\}. \quad (17)$$

It is impossible to find an analytic solution to the likelihood function in Equation 16 due to the nonlinear fashion that the random effects enter the model (Wolfinger & Lin, 1997). Therefore, approximation methods are needed. With these approximations, maximum likelihood estimates can then be obtained. However, the functional mixed-effects model does not require any approximations to solve its criterion (Equation 11). The functional mixed-effects model can be solved using an EM algorithm that assumes fixed smoothing parameters (Wu & Zhang, 2006; Berk, 2012). Smoothing parameters are determined by fitting various models with different sets of smoothing parameters and picking the model with the lowest AIC, BIC, or generalized cross validation (GCV) value (Wu & Zhang, 2006).

Nonlinear mixed-effects models have advantages over functional mixed-effects models. There is more readily available software that has been thoroughly examined to estimate nonlinear mixed-effects models (e.g., PROC NLMIXED in SAS, *nlme* and *lme4* in R). The readily available software and frequency of use make nonlinear mixed-effects models more approachable to many researchers. Additionally, in terms of notation, nonlinear mixed-effects models are relatively straightforward extensions of linear mixed-effects models which makes these models more approachable to researchers.

However, functional mixed-effects models hold several advantages over nonlinear mixed-effects models. First, estimation of functional mixed-effects models is more straightforward as it does not require approximating integrals that are required when

estimating nonlinear mixed-effects models. It is possible to estimate nonlinear mixed-effects models using alternative estimation routines (e.g., Bayesian), however, maximum likelihood is the most common approach. Functional mixed-effects models have their biggest strength in that they impose no parametric assumptions on the form of growth of the data. These models utilize smoothing to allow the data to show the underlying trajectories without forcing the data to follow a predetermined shape, like the nonlinear mixed-effects model. Using smoothing techniques is not without its drawbacks, though. The selection of number and location of knots and of smoothing parameters is not an exact science. The best set of knots and smoothing parameters is often left to the interpretation of the analyst.

There are some major challenges for each of these models. The biggest challenge is the estimation of the nonlinear mixed-effects models. These models have difficulty converging when the sample size is not adequate, when several random effects present, and when poor starting values are provided. There are two big challenges for the functional mixed-effects models. First, is picking smoothing parameters. This is a nontrivial problem that only has vague guidelines to help in the process of selection. There can be multiple sets of smoothing parameters that visually appear to smooth the data well. It is often left to the discretion of the analyst to pick what smoothing parameters work best. The second major challenge for functional mixed-effects models is software. Since this is a relatively novel analytical approach, there are not many programs available to use. The software packages that are available have not been as thoroughly tested and used like the software packages for nonlinear mixed-effects models.

Fine et al. (2019) compared linear and nonlinear mixed-effects models to a functional mixed-effects model in a simulation study. This simulation involved data which were generated from a parametric process. The linear and nonlinear mixed-effects models were better at recovering the underlying mean curve. As yet, traditional mixed-effects models and functional mixed-effects models have not been compared on curve recovery when the data generation process is inherently nonparametric.

PROPOSED SIMULATION STUDY

This project aimed to compare nonlinear and functional mixed-effects models when the data were generated from an inherently nonparametric process. The goal was to examine which approach could more accurately recover the true underlying curves from observed data.

Data Generation

Individual curves in each data set were generated following Yao, Müller, and Wang (2005), such that

$$y_{ij} = x_i(t_{ij}) + e_{ij} \tag{18}$$

where y_{ij} is the observed response for the *i*th individual ($i = 1, \ldots, n$) at the *j*th time point ($j = 1, \ldots, N_i$), $x_i(t_{ij})$ is the true underlying person-specific curve for individual *i* evaluated at the *j*th time point for this person, t_{ij}; e_{ij} is the measurement error. The person-specific curve, $x_i(t)$, was generated by

$$x_i(t) = \mu(t) + v_i(t), \tag{19}$$

where $\mu(t)$ is the grand mean function and $v_i(t)$ is the person-specific effect function for the *i*th individual. Inserting Equation 18 into Equation 19 yields the following model

$$y_{ij} = \mu(t_{ij}) + v_i(t_{ij}) + e_{ij}, \quad (20)$$

which is identical to the model given in Equation 9.

The grand mean function, $\mu(t)$, was defined using the cubic B-spline basis in Figure 1; each colored line represents a different B-spline basis function. The B-splines in Figure 1 were weighted with coefficients to form a mean curve similar in structure to the data plotted in Figure 2. These data were (log) progesterone cycle data taken across 24 days from 51 women; each color represents a different woman's cycle. Figure 3 shows the grand mean function formed using the B-splines from Figure 1. This was the true mean curve for every simulated dataset in every condition.

The person-specific effect functions, $v_i(t)$, were generated following

$$v_i(t) = \sum_{m=1}^{3} \phi_m(t) c_{im}, \quad (21)$$

where $\phi_1(t)$, $\phi_2(t)$, and $\phi_3(t)$ were defined as three b-spline basis functions of degree three with equally spaced knots over [0, 20]. Figure 4 shows a graphical representation of these basis functions. The three basis functions have different shapes, which were used to create variability in the person-specific deviation curves at different locations of time. The first basis function $\phi_1(t)$ (solid black in line Figure 4) has higher values in early time points. Therefore, when different individual weights are given to this basis function, the individual curves have more variability in early time points and less variability toward the later time points. The second basis function $\phi_2(t)$ (dotted black in line Figure 4) has higher values in middle time points. Therefore, giving different weights to this basis function yields more variability in middle time points than earlier or later time points. Similarly, the third basis function (dashed black in line Figure 4) has higher values in later time points, and we can create more variability in later time points by

giving different weights to this third basis function. The three person-specific basis function coefficients, c_{i1}, c_{i2}, and c_{i3} were independently and identically generated from the multivariate normal distribution

$$\begin{bmatrix} c_{i1} \\ c_{i2} \\ c_{i3} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} .1 & 0 & 0 \\ 0 & .6 & 0 \\ 0 & 0 & 1.1 \end{bmatrix} \right). \quad (22)$$

The covariance matrix used to generate the coefficients of these basis functions had successively increasing variances from the first to the third (.1, .6, and 1.1). This indicates that the person-specific deviation curves were generated to have greater variability as time increases, which is common in many longitudinal studies. The measurement errors, e_{ij}, were randomly generated from a normal distribution of a mean of zero and a variance of $\sigma_e^2 = .1$.

Simulation Conditions

This project aimed to compare nonlinear and functional mixed-effects models when the data were generated from an inherently nonparametric process. Four factors were manipulated: sample size, number of time points per curve, measurement design, and model fit. Sample size was varied at three levels: 200, 500, and 1000. Number of time points per curve was varied at three levels: 5, 11, and 21. Both of the sets of values for the sample size and number of time points per curve were chosen to mimic values commonly seen in longitudinal research. There were three levels of measurement design: fixed, semi-random, and attrition. In the fixed measurement design condition, each curve was observed at the exact same time points. This mimics longitudinal studies where observations are collected at predetermined intervals and makes time points common to all subjects. An example of this would be every curve is observed at exact times [0, 5, 10,

15, 20]. In the semi-random measurement design condition, each curve was observed at time zero and time 20, and then could be observed at any increment of .5 from times .5 to 19.5. For example, in the five time points per curve condition, person A is observed at [0, 9.5, 11, 17.5, 20] while person B is observed at [0, .5, 2.5, 15, 20]. Every curve was observed at time zero and time 20, and then could be observed at any .5 increment between .5 and 19.5. Fine et al. (2019) highlighted the importance of collecting data at the boundaries of interest. The semi-random measurement design condition allows for measurement of end points while also allowing for semi-random assessments where each person is observed for the same amount of time. In the attrition measurement design condition, each curve was observed at time zero. Half of the curves could be observed at any increment of .5 from times .5 to 20. The other half of the curves could be observed at any increment of .5 from times .5 to 16. For example, in the five time points per curve condition, person A is observed at [0, 9.5, 11, 17.5, 20] while person B is observed at [0, .5, 2.5, 10.5, 13]. This mimics the attrition that is commonly seen in longitudinal research where some participant are not observed through the end of the study.

The structure of the semi-random and attrition time points were used for several reasons. First, when the number of unique time points was large, estimation of smoothing splines becomes computationally demanding when a knot is placed at each unique time point. To speed up the computation time, users can specify knots; however, the user is now using *penalized* splines as opposed to *smoothing* splines (refer to Fine et al. (2019) for more details). Semi-random and attrition time points, as defined here, keep the maximum number of unique knots to 41. Knots were not pre-specified which allowed for the use of smoothing splines and computation was not overly demanding. The reason for

this structure for semi-random time points comes from a result found in Fine et al. (2019) where curves which were not observed at the boundaries of interest resulted in poorer curve recovery. The lack of information at the boundaries resulted in estimated curves that deviated from the true underlying curve. Thus, in the semi-random condition, boundary information was collected for each curve (i.e., each curve observed at time zero and time 20). The same reasoning for the semi-random time point applies to the attrition time points. However, this condition each curve was only necessarily observed at one boundary point (time zero) to more closely mimic data commonly seen in longitudinal research.

Figure 5 displays the true underlying mean and individual curves, with the observed data, from a dataset with 200 people observed at 11 time points with each measurement design. Each panels of Figure 5 displays the true underlying individual curves in dashed gray, the true grand mean function in black, and the observed data as black circles. Each plot highlights the increasing variance in random effects across time; the spread of the curves at time zero is much smaller than the spread of the curves at time 20. Figure 5A shows the true underlying curves for the fixed measurement design; this panel highlights how each person was observed at the same points in time. Figure 5B shows the true underlying curves for the semi-random measurement design. This panel highlights the semi-random measurement design; each curve was observed at time zero and time 20, and then randomly in-between at any .5 increment between .5 and 19.5. Figure 5C shows the true underlying curves for the attrition measurement design. This panel highlights the attrition measurement design condition; curves were more densely observed at earlier time points, and become more sparsely observed at later time points.

The last factor manipulated was the statistical model used to analyze the data: the functional mixed-effects model and the cosine nonlinear mixed-effects model. The functional mixed-effects model is defined in Equation 9. The cosine nonlinear mixed-effects model is defined as

$$y_{ij} = \beta_{0i} + \beta_{1i} \cos\left(\beta_{2i}\left(t_{ij} + \lambda\right)\right) \quad (23)$$

where β_{0i} is a random intercept, β_{1i} is a random amplitude parameter, β_{2i} is a random frequency parameter, and λ is a fixed phase shift that shifts the curve horizontally. The amplitude parameter controls how vertically high the curves go. The frequency parameter controls how many cycles the curves complete in a given interval. Ideally, all three random effects are estimated and the model converges; however, if the models did not converge, random effects would be dropped one at a time until the model converged beginning with the random frequency parameter (β_{2i}), followed by the random amplitude parameter (β_{1i}).

The generated data do not have a known functional form. The cosine nonlinear mixed-effects model in Equation 23 is a functional form that most appropriately mimics the grand mean function (Figure 3). The cosine nonlinear mixed-effects model was estimated using the *nlme* package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018) in R. This package requires the user to specify starting values for each fixed effect in the model to be estimated. Potential fixed effects starting values were sampled to attempt at capturing the function in Figure 3. The set of fixed effect values that appeared to capture the mean trajectory (Figure 3) were

$$y_{ij} = 1.5 + 1.5 \cdot \cos\left(.25 \cdot \left(t_{ij} + 10\right)\right) \quad (24)$$

The values in Equation 24 were the starting values used in the simulation study. Figure 6 shows the function from the set of fixed effects in Equation 24. This curve approximated the grand mean function used in the simulation; however, the model did not capture a prominent feature of the curve. The grand mean function was relatively flat from time zero to roughly time five, while the cosine nonlinear mixed-effects model was curved (refer to Figures 3 and 6 for a comparison). This process of needing to select a functional form of change for data, even though it does not perfectly map onto the researcher's data, mimics what researchers do in practice. If the researcher wants to impose a functional form of growth onto the data, the researcher may need to sacrifice accurately modeling some aspects of the data.

The functional mixed-effects model was estimated using the *sme* package (Berk, 2013) in R. As noted in Equation 11, two smoothing parameters were involved in estimation - one for the fixed (λ_μ) and one for random effects (λ_ν). The choice of a smoothing parameter can be obtained in several ways, however, as Ramsay and Silverman (2005) noted, the selection of smoothing parameters can be subjective and is often best decided by examining various options. Berk (2012) introduced an efficient way of selecting the smoothing parameters by using Nelder-Mead simplex search algorithm to automatize the model comparison process for selecting the smoothing parameters. Berk's method to select smoothing parameters is implemented in the *sme* package (Berk, 2013). The *sme* package provides various measures that can be used for smoothing parameter selection, including AIC, AICc (corrected AIC), BIC, and BICn (BIC for longitudinal models with a sample size correction; Berk, 2013). For the simulation, smoothing

parameters for each dataset in each condition were chosen using the AIC, which aided automation and ensured consistent decisions regarding smoothing were made.

In sum, this simulation had a total of 54 conditions (3 number of time points x 3 sample sizes x 3 measurement designs x 2 statistical models); 100 data sets were generated per condition.

Evaluation

Analysis results were evaluated in the following way. First, to evaluate the accuracy of the estimated grand mean function, $\mu(t)$, a mean square error (MSE) was calculated under each conditions. The MSE was defined as

$$\mathrm{MSE}\big(\hat{\mu}(t)\big) = \frac{\sum_{s=1}^{100}(\hat{\mu}(t_s) - \mu(t_s))^2}{100}, \tag{25}$$

where $\hat{\mu}(t_s)$ is the estimated grand mean function for the evaluated at time t_s, t_s is the *s*th time point out of the 100 equally spaced time points over the range of [0, 20], and $\mu(t)$ is the true grand mean function in Figure 3. In the numerator, the term $\sum_{s=1}^{100}(\hat{\mu}(t_s) - \mu(t_s))^2$, measures how much the estimated grand mean function deviates from the true grand mean function across 100 equally spaced time points. Therefore the mean square error, $\mathrm{MSE}\big(\hat{\mu}(t)\big)$, measures how much the estimated grand mean function deviates from the true grand mean function across 100 equally spaced time points, on average across 100 data sets. Second, the accuracy of the predicted person-specific trajectories was assessed by calculating the average mean square error (AMSE) defined as

$$\mathrm{AMSE}\big(\hat{x}(t)\big) = \frac{\sum_{i=1}^{n}\sum_{s=1}^{100}\big(\hat{x}_i(t_s) - x_i(t_s)\big)^2}{(100)(n)} \tag{26}$$

where $\hat{x}_i(t_s)$ indicates the predicted curve for the *i*th individual evaluated at the *s*th time point, and $x_i(t_s)$ is the true curve for the *i*th individual evaluated at the *s*th time point. Similar to $\text{MSE}\big(\hat{\mu}(t)\big)$, a lower value of $\text{AMSE}\big(\hat{x}(t)\big)$ indicates that the estimated person-specific trajectories are closer to the true curves, indicating better recovery.

In addition to MSE and AMSE values, convergence percentage was reported for each dataset in every condition. The number of random effects used in each condition in each dataset were also reported. The number of random effects in each condition in each dataset could either be three (e.g., random amplitude, frequency, and intercept), two (e.g., random amplitude and intercept), or one (e.g., random intercept).

Hypotheses

The functional and the cosine nonlinear mixed-effects models are hypothesized to recover the mean curves better (i.e., lower MSE values) as the number of time points per curve and the sample size increase. Both models are hypothesized to recover the individual curves better (i.e., lower AMSE values) as the number of time points per curve increases, but not necessarily as sample size increases. Increasing the sample size does not inherently provide more information about a particular individual's curve. Based off results from Fine et al. (2019), the functional mixed-effects model is hypothesized to perform best, in terms of MSE and AMSE, with a fixed measurement design. The cosine nonlinear mixed-effects model is hypothesized to recover the mean curve better (i.e., lower MSE values) with a random measurement design, as more information will be available compared to fixed measurement design. Finally, it is hypothesized the functional mixed-effects model will recover the individual curves better (i.e., have lower

AMSE values) than the cosine nonlinear mixed-effect model. This is because the functional mixed-effects model is more flexible in modeling random effects.

SIMULATION RESULTS

This simulation aimed to compare nonlinear and functional mixed-effects models to determine how well each model recovered underlying curves. Sample size, number of time points per curve, and measurement design were manipulated and both a nonlinear and functional mixed-effects model were fit in each condition. The simulation achieved 100% convergence for every condition in both models. In the nonlinear mixed-effects modeling condition, all three random effects were estimable in every dataset in every condition.

Broadly, for both models, as the sample size and number of time points per curve increased, the mean curves were better recovered (i.e., lower MSE values). Increasing the number of time points per curve yielded better recovery for the individual curves (i.e., lower AMSE values). The mean curves were more accurately recovered than the individual curves, which was evident by lower MSE values compared to AMSE values. The MSE and AMSE results of the simulation are presented in Table 1 and 2.

In the following paragraphs, details of the fits from the functional and nonlinear mixed-effects model are presented. Fitted models from both the functional and nonlinear mixed-effects model are presented. Next, the effects of each factor manipulated in the simulation are detailed, starting with sample size, the number of time points per curve, measurement design. Lastly, a comparison of the estimation results for each model is presented.

Overview of Model Fit

Figures 7, 8, and 9 are model fits of the functional and nonlinear mixed-effects models to datasets from conditions with 11 time points per curve, sample size of 200 curves, and each measurement designs. Figure 7 displays plots for the fixed measurement design. Figure 8 displays plots for the semi-random measurement design. Figure 9 displays plots for the attrition measurement design. In all of these figures, panels A and B, are plots of the functional mixed-effects model fits, and panels C and D are plots of the cosine nonlinear mixed-effects model fits. Panels A and C are plots of the estimated mean (black line) and individual curves (dashed gray lines) fit by either model. Panels B and D are plots of a particular individual's observed data (black dots), that individual's estimated curve (dashed black), and the estimated mean curve (solid black).

Comparing panels A to panels C, across Figures 7, 8, and 9, highlights the difference in model fits. As expected, the cosine nonlinear mixed-effects model forced a curvature to the data at the beginning, while the functional mixed-effects model allowed the trajectory to be flatter initially. Figures 7B and 7D contain plots of a particular individual's observed data, individual estimated curve, and estimated mean curve for the fixed measurement design where each observed data point was at 11 equally spaced integers between [0, 20]. This person's observed data (black dots), and thus their estimated curve (dashed black line), were slightly lower than the estimated mean curve (solid black line). Figures 8B and 8D contain the same plots for the semi-random measurement design where this person was observed at time zero and time 20, and at a random set of time points in between. This person's observed data (black dots) were mostly above the estimated mean curve (solid black line). Thus, their estimated curve

(dashed black line) was mostly above the estimated mean curve. Figures 9B and 9D contain the same plots for the attrition measurement design where this person was observed at time zero but was only observed up to time 13. This person's observed data (black dots) was above the estimated mean curve (solid black line) at early time points and hovered close to the estimated mean curve at middle time points. In the functional mixed-effects model, Figure 9B, this person's estimated curve (dashed black line) was above the estimated mean curve at early time points then closely followed the estimated mean curve through the rest of time; however, for the nonlinear mixed-effects model, in Figure 9D, this person's estimated curve (dashed black line) deviated from the estimated mean curve at both early and later time points.

Figures 10, 11, and 12 contain the MSE and AMSE values from the simulation and can be used to compare these values from the nonlinear and functional mixed-effects models at each sample size level. Figure 10 displays results for the sample size of 200, Figure 11 displays results for the sample size of 500, and Figure 12 displays results for the sample size of 1,000. These figures are identical in structure. Panels A, B, and C are for MSE values, whereas panels D, E, and F are for AMSE values. The dashed line with squares correspond to the cosine nonlinear mixed-effects model and the solid line with circles correspond to the functional mixed-effects model. Panels A and D display results for the fixed measurement design, panels B and E display results for the semi-random measurement design, and panels C and F display results for the attrition measurement design.

The following paragraphs detail the effects of each factor manipulated in the simulation. The effect of sample size is discussed first, followed by the number of time points per curve, and measurement design.

Sample Size. For both models, as the sample size increased, the mean curves were better recovered (i.e., lower MSE values). This result was intuitive as increasing the sample size means more data are present to more accurately recover the mean. These results are most clearly seen in Table 1. For example, the last three rows of the first column of Table 1 contain MSE values from each sample size for five time points with the attrition measurement design fit with a functional mixed-effects model. For sample size $n = 200$, the MSE values is 0.0061, which decreases to a value of 0.0014 for a sample size of $n = 1{,}000$. These results can also be visualized by comparing corresponding lines within corresponding panels in Figure 10, 11, and 12. For example, comparing the solid line with the circles in Figure 10A to the same line in Figure 11A.

Across all conditions, as the sample size increased, the mean curves were better recovered. Given the results for the rest of the simulation are the similar across sample sizes, remaining results are focused on the sample size of $n = 200$.

Number of Time Points per Curve. For both models, as the number of time points per curve increased, the mean and individual curves were better recovered (i.e., lower MSE and AMSE values). These results are intuitive as increasing the number of time points means more information available to more accurately recover the mean and/or individual curves. These results are most clearly seen across all panels in Figure 10 (as well as in Figures 11 and 12). For example, in Figure 10D both the solid and dashed lines have decreasing slopes as the number of time points increase. In Figure 10D,

the nonlinear mixed-effects model (dashed line with squares) decreased from an AMSE value of 0.1048 to 0.0483 when the number of time points per curve and increased from five to 21. The decreasing MSE and AMSE values show the mean and individual curves were more accurately recovered as the number of time points per curve increased.

Measurement Design. Generally, the attrition measurement design had the highest MSE and AMSE values. This can be seen by comparing the dots/squares in Figure 10C to 10B and 10A, and by comparing the dots/squares in Figures 10F compared to 10E and 10D (the same comparisons can be made for Figures 11 and 12). The dots/squares were higher in the attrition measurement design indicating higher MSE and AMSE values than for both the fixed and semi-random measurement designs. For example, in Figure 10F, the AMSE value for the functional mixed-effects model (solid line with circles) with five time points per curve and attrition measurement design was 0.1801. The corresponding AMSE value for the semi-random measurement design was 0.1613 (Figure 10E) and 0.0741 (Figure 10D) for fixed measurement design. This result suggests the underlying curves were not recovered as well with the attrition measurement design.

The fixed measurement design had the lowest AMSE values for both the nonlinear and functional mixed-effects models, followed by semi-random, and attrition. This result can be seen by comparing Figures 10D through 10F. Here, the AMSE value for the nonlinear mixed-effects model with five time points per curve and the fixed measurement design was 0.1048. The corresponding AMSE value for the semi-random measurement design was 0.1184 and for the attrition measurement design was 0.1745. This result indicated that both models recovered the underlying mean and individual

curves best with the fixed measurement design. It was hypothesized that the functional mixed-effects model would recover the underlying individual curves best (i.e., lowest AMSE) with a fixed measurement design and this result supported that hypothesis.

The fixed measurement design had the lowest MSE values for the nonlinear mixed-effects models, followed by semi-random, and attrition. This result is found by examining Figures 10A through 10C. In these figures, the MSE value for the nonlinear mixed-effects model with 21 time points per curve and the fixed measurement design was 0.0288. The corresponding MSE value for the semi-random measurement design was 0.0306 and for the attrition measurement design was 0.0417. This results indicated the nonlinear mixed-effects model recovered the underlying mean curves better a with a fixed measurement design. The nonlinear mixed-effects model was hypothesized to recover the mean curve better (i.e., lower MSE values) with a random measurement design. This result, instead, showed the fixed measurement design was best.

The semi-random measurement design, generally, had the lowest MSE values for the functional mixed-effects models, followed by the fixed, and attrition designs. This result can be found by examining Figures 10A through 10C. That is, the MSE value for the functional mixed-effects model with 11 time points per curve and the semi-random measurement design is 0.0019, whereas, the corresponding MSE value for the fixed measurement design was 0.0020 and for the attrition measurement design was 0.0030. This results suggests that the functional mixed-effects model recovered the underlying mean curve better with a semi-random measurement design. It was hypothesized that the functional mixed-effects model would recover the underlying mean curve best (i.e.,

lowest MSE) with a fixed measurement design. This result did not support this hypothesis; the semi-random measurement design had the lowest MSE values.

Comparison of Models

Across all conditions, the functional mixed-effects models recovered the underlying mean curve better than the nonlinear mixed-effects models. This result can be seen by comparing the lines with the circles to the lines with the squares in Figures 10A through 10C. In these plots, the squares are above the circles, indicating the nonlinear mixed-effects models had higher MSE values compared to the functional mixed-effects model. For example, in Figure 10B, the nonlinear mixed-effects model had an MSE value of 0.0364 for 11 time points and a semi-random measurement design, whereas the functional mixed-effects model had an MSE value of 0.0019. This result highlights that the underlying mean curve was recovered more poorly using the cosine nonlinear mixed-effects model compared to a functional mixed-effects model.

It was hypothesized that the functional mixed-effects model would recover the individual curves better (i.e., have lower AMSE values) compared to the nonlinear mixed-effect model. This result was partially supported. There were 27 conditions (3 number of time points x 3 sample sizes x 3 measurement designs) where both models were fit. The functional mixed-effects model recovered the underlying individual curves better (i.e., had lower AMSE values) than the nonlinear mixed-effects model in 18 of 27 conditions. The functional mixed-effects model outperformed the nonlinear mixed-effects model with the fixed measurement design (9 conditions), when there was 11 and 21 time points per curve with the semi-random measurement design (6 conditions), and when there 21 time points per curve with the attrition measurement design (3 conditions). The

nonlinear mixed-effects model recovered the underlying individual curves better (i.e., had lower AMSE values) than the functional mixed-effects model in 9 of 27 conditions. The nonlinear mixed-effects model outperformed the functional mixed-effects model where there were five time points per curve with the semi-random measurement design (3 conditions), and when there were five and 11 time points per curve with the attrition measurement design (6 conditions).

These results can be seen comparing panels D through F in Figures 10, 11, and 12. Conditions where the nonlinear mixed-effects model had a higher AMSE value than the functional mixed-effects model are plotted with squares above circles. An example of this is in Figure 10E at 11 and 21 time points per curve. The squares are above the circles, meaning the functional mixed-effects model had lower AMSE values than the nonlinear mixed-effects model. Conversely, plots when circles are above the squares depict conditions where the functional mixed-effects model had a higher AMSE value than the nonlinear mixed-effects model. An example of this is in Figure 10E at five time points per curve.

EMPIRICAL EXAMPLE

The motivating empirical data for the simulation were progesterone cycle data taken from Brumbrack and Rice (1998). Their data were collected as part of continuing studies of early pregnancy loss conducted by the Institute for Toxicology and Environmental Health at the University of California, Davis in collaboration with the Reproductive Epidemiology Section of the California Department of Health Services, Berkeley. Their sample comes from patients with healthy reproductive function enrolled

in an artificial insemination clinic, where insemination attempts were well-timed for each menstrual cycle. As is standard practice in endocrinological research (Yen & Jaffe, 1991), progesterone profiles were pre-aligned by the day of ovulation, determined by serum luteinizing hormone (Brumback & Rice, 1998).

This project used a subset of the data from Brumback and Rice (1998). This project used a sample of 51 women's urinary metabolite progesterone measured across 24 days. Some missing data was present. The minimum number of days a woman's cycle was measured was 15 days; the average number of days measured was roughly 22. These data contain 29 nonconceptive and 22 conceptive menstrual cycles. The (log) progesterone cycles across both the conceptive and nonconceptive groups were presented in Figure 2. These data were analyzed using the functional mixed-effects model presented in Equation 9 and the cosine nonlinear mixed-effects model presented in Equation 23.

The cosine nonlinear mixed-effects model was fit using the R package *nlme*. Starting values for the fixed effects were specified to be

$$y_{ij} = -.5 + 2.5 \cdot \cos\left(.25 \cdot \left(t_{ij} + 15\right)\right) \quad (27)$$

Figure 13 shows the function from the set of fixed effects in Equation 27. This curve approximated the hypothesized mean curved for the progesterone cycle data. Note, the model in Equation 27 reflects more curvature in early and later time points than is seen in the data in Figure 2. Again, this process of needing to select a functional form of change for data, even though it does not perfectly map onto the researcher's data, mimics what researchers do in practice. If the researcher wants to impose a functional form of growth onto the data, the researcher may need to sacrifice accurately modeling some aspects of the data. Random effects were treated the same as in the simulation. If convergence

problems arose, random effects would be dropped one at a time, starting with the random frequency, until convergence was reached.

The functional mixed-effects model was fit using the R package *sme*. Cycles were observed from day -8 to day 15 (day zero is ovulation), and a knot was placed at each time point (i.e., each day). The smoothing parameter for the functional mixed-effects model was chosen using AIC, just as in the simulation.

The data were split into two separate datasets. The first dataset contained the observations from the odd numbered time points and the second dataset contained the observations from the even numbered time points. The functional and nonlinear mixed-effects model were estimated using the odd numbered time points. The predicted mean and individual trajectories from both models were used to examine misfit with the even time points. This separation of estimation versus evaluation was done to examine how well the model accounted for data that were not used to estimate the model. That is, examining how well a model can account for novel data is a better evaluation of the adequacy of the model than evaluating how well a model captures data that were used to estimate the model's parameters. Misfit of the mean curve was evaluated using the MSE value, calculated for each mixed-effects model as

$$\mathrm{MSE}\big(\hat{\mu}(t)\big) = \frac{\sum_{s=1}^{12}(\mu(\overline{t_s}) - \hat{\mu}(t_s))^2}{12}, \tag{28}$$

where $\mu(\overline{t_s})$ is the average of the observed data at a specific even numbered time point, $\hat{\mu}(t_s)$ is the estimated grand mean function evaluated at time t_s, t_s is the *s*th even numbered time point out of the 12 equally spaced even numbered time points over the range of [-8, 15]. In the numerator, the term $\sum_{s=1}^{12}(\mu(\overline{t_s}) - \hat{\mu}(t_s))^2$, measures how much

the estimated grand mean function deviated from the averaged observed data across 12 equally spaced even numbered time points. Therefore the mean square error, $\text{MSE}\big(\hat{\mu}(t)\big)$, measures how much the estimated grand mean function deviated from the mean of the observed data across even numbered time points.

Second, misfit of the individual curves was examined using the AMSE, which will be calculated for each mixed-effects model as

$$\text{AMSE}\big(\hat{x}(t)\big) = \frac{\sum_{i=1}^{n}\sum_{s=1}^{12}\big(x_i(t_s) - \hat{x}_i(t_s)\big)^2}{12n} \tag{29}$$

where $x_i(t_s)$ is the observed data for the *i*th individual evaluated at the *s*th even numbered time point, and $\hat{x}_i(t_s)$ is the value from the predicted curve for the *i*th individual evaluated at the *s*th even numbered time point. Similar to $\text{MSE}\big(\hat{\mu}(t)\big)$, a lower value of $\text{AMSE}\big(\hat{x}(t)\big)$ indicates that the estimated person-specific trajectories were closer to the observed data, indicating smaller misfit.

EMPIRICAL EXAMPLE RESULTS

The progesterone cycle data from Brumback and Rice (1998) contained 89 cycles from 51 women with 19 of these women being observed for more than one cycle. Analyses were conducted on one cycle per woman for a total of 51 cycles. If a woman was observed at more than one cycle, the cycle used in the analyses was randomly sampled from all of her cycles. These data were then split into data with odd and even numbered time points. The data with odd numbered time points were analyzed using a functional mixed-effects model and nonlinear mixed-effects model. Both models converged with no out of bounds parameter estimates.

Overview of Model Fit

Figures 14 contains plots of the predicted mean and individual trajectories of the functional and nonlinear mixed-effects models estimated using the odd numbered time points from the progesterone cycle data. Panels A and B contain plots from the functional mixed-effects model, and panels C and D contain plots of the nonlinear mixed-effects model fits. Panels A and C contain plots of the estimated mean (black line) and individual curves (dashed gray lines), and panels B and D contain the particular individual's observed data (black dots), this individual's predicted curve (dashed black), and the predicted mean curve for the sample (solid black).

Comparing panel 14A to 14C highlights the differences in the mean and individual predicted trajectories. The cosine nonlinear mixed-effects model forced the estimated curves to follow a smooth cosine curve, whereas, the functional mixed-effects model estimated curves that were mostly flat during the early and later time points, with a sharp increase at middle time points. Examining Figures 14B and 14D highlight the differences in a predicted trajectory for one particular person. In Figure 14D, the observed data at later time points were fairly stable (and potentially increasing); however, the predicted trajectory from the cosine nonlinear mixed-effects model began to decline. On the other hand, the predicted trajectory from Figure 14B, from the functional mixed-effects model, appeared to follow the trend seen in the observed data at the later time points where there is stability in the observed values.

Evaluation of MSE and AMSE

The data were split based on whether the observations were from the odd or even numbered time points. The functional mixed-effects model and the cosine nonlinear

mixed-effects model were fit to the odd numbered time points. Mean and individual predicted trajectories from both models were used to examine misfit with the even numbered time points. The MSE value for the functional mixed-effects model was 1.1573, and the MSE value for the cosine nonlinear mixed-effects model was 1.1655. Thus, both models performed similarly in terms of misfit of the mean curve to the even time points, with the functional mixed-effects model slightly outperforming the cosine nonlinear mixed-effects model. The AMSE value for the functional mixed-effects model was 2.2558, and the AMSE value for the cosine nonlinear mixed-effects model was 2.2768. These AMSE values indicate both models performed similarly in terms of misfit of the individual curves to the even time points, with the functional mixed-effects model slightly outperforming the cosine nonlinear mixed-effects model. The lower MSE values compared to the AMSE values indicate the individual predicted curves had more misfit to the even numbered time points than the mean predicted curves. MSE and AMSE values from both models were very similar, likely indicating no meaningful difference in the models with respect to these two outcomes.

DISCUSSION

Functional mixed-effects models are a useful tool for analyzing data with a nonlinear trajectory. They are particularly useful when the functional form of growth is unknown or is too complex for parametric models. This simulation investigated how functional and nonlinear mixed-effects models recovered mean and individual trajectories when the data were generated from an inherently nonparametric process. Sample size, number of time points per curve, and measurement design were manipulated. Results

showed as the sample size and number of time points per curve increased, the models recovered the underlying mean curve better. As the number of time points per curved increased, the models were able to more accurately recover the individual curves. These results were expected, as there was more information given to the models about the mean and each individual, leading to more accurate recovery of the mean and individual curves. Both models recovered the mean curves better than the individual curves. This result was expected as there is more information available to estimate a mean than an individual's curve. Generally, the models performed better with the fixed and semi-random measurement designs. Across all simulation conditions, the functional mixed-effects model recovered the mean curves more accurately than the nonlinear mixed-effects model, and the functional mixed-effects model recovered the individual curves more accurately than the nonlinear mixed-effects model in roughly two-thirds of the simulation conditions. Details of these results are discussed below.

Both models recovered the individual curves more accurately when data were observed at the fixed measurement design. This was expected for the functional mixed-effects model based off simulation work from Fine et al. (2019). The nonlinear mixed-effects model recovered the mean curves more accurately when the data were observed at the fixed measurement design. This result was unexpected; it was hypothesized the model would perform best with the semi-random measurement design as the model would be incorporating more information across time. This result likely occurred because the residual variances for the nonlinear mixed-effects model were fixed across time. The data were generated to have increasing variance across time. It is likely the variability in the semi-random time points was just added to the error term, adding to misfit of the model.

Generally, both models recovered the mean and individual curves the worst with the attrition measurement design. However, the differences between the fixed, semi-random, and attrition measurement designs were not so drastic that these models should not be used by applied researchers who have data that resemble the attrition measurement design. The attrition measurement design always results in the least accurately recovered individual curves. There were only four conditions where the attrition measurement design did not result in the least accurately recovered mean curves. All four of these conditions were for the functional mixed-effects model. One condition, with a sample size of 1,000 and 21 time points per curve, the mean curve was recovered least accurately with the semi-random measurement design; however, this difference was in the fifth decimal place and was likely not meaningful. The other three conditions, with five time points per curve at each sample size ($n = 200, 500, 1000$), the fixed measurement design resulted in the least accurately recovered mean curves. In these conditions, the curves were observed at five arbitrary points in time [0, 5, 10, 15, 20], and therefore had knots placed at those time points. It is likely the case the knots placed in the fixed measurement design did not capture meaningful areas of nonlinearity on the curve, while the spread of knots placed in both the attrition and semi-random measurement designs were able to.

The functional mixed-effects model recovered the mean curves more accurately with the semi-random measurement design. Previous simulation work suggested the fixed measurement deign would be better (Fine et al., 2019). There were three conditions where the functional mixed-effects model recovered the mean curve more accurately with the fixed measurement design. However, these difference were in the fifth decimal place and were likely not meaningful. The measurement design condition was used to mimic

different ways data can be collected in applied research (e.g., all participants collected at the same time, participants collected at various intervals, and participants collected at various intervals but not all the way to the end of the study). While it is useful to gain insight on a general 'best' ways to collected data, it might be more useful to focus on ways to collect data most accurately for different developmental processes. This result likely occurred because the semi-random measurement design had observations, and therefore knots placed, at important features of the curve. Placing knots on areas of the curve where there is nonlinear change seems to be more important than arbitrarily collecting data, and therefore placing knots, at uninformed points. Collecting data, and therefore placing knots, at meaningful time points needs to be balanced with not placing too many knots, especially where there is not data. A delicate balance needs to be struck between placing knots at important features of the curve and not placing so many knots that the number of knots placed becomes too close to the sample size.

The functional mixed-effects models recovered the individual curves more accurately than the nonlinear mixed-effects models in two-thirds of the simulation conditions. This result was expected as the functional mixed-effects models are more flexible in modeling the mean and random effects than the nonlinear mixed-effects models. However, in one-third of the conditions, the nonlinear mixed-effects models recovered the individual curves better than the functional mixed-effects models. The conditions in which this result occurred are conditions where the data did not contain rich information (e.g., five time points, attrition measurement design). It was likely the case that the flexibility of functional mixed-effects models struggled when the data did not

contain ample information, whereas the more rigid structure of the nonlinear mixed-effects models struggled less in these conditions.

The utility of the functional and nonlinear mixed-effects models were demonstrated on progesterone cycle data from Brumback and Rice (1998). The functional mixed-effects models captured the flatter beginning of the curve and the steep increase in the middle. It modeled the end of the curve to be fairly flat. The nonlinear mixed-effects models, as expected, imposed the cosine structure to the data and modeled a curvature at the beginning and end of the data and a gradual increase in the middle. The progesterone data contain two groups of women: conceptive and non-conceptive. The conceptive group show in increase in (log) progesterone at the end while the non-conceptive begin to decrease back down. Interestingly, it appeared that the functional mixed-effects models 'averaged' those effects and modeled the data as flattening out at the end. Additionally, the predicted mean and individual curves in Figure 14A were not as smooth as the curves in Figure 14C. The curves from the functional mixed-effects models were under-smoothed. This is related to the set of smoothing parameters selected by AIC.

In sum, functional mixed-effects models were found to be a useful tool for analyzing longitudinal data with an unknown nonlinear trajectory with complex interindividual differences in the patterns of change. These models are less established in the social sciences, but show great promise in their utility. These models are most useful when the data do not clearly map onto a parametric form and require flexibility to model them. These models are also useful when the random effects structure in a linear or nonlinear mixed-effects model is too restrictive. These models can be used in an exploratory manner to guide researchers to a proper parametric form to use. Additionally,

these models are useful when fitting a parametric model is not providing adequate fit to the data and requires a more flexible approach.

Challenges

While functional mixed-effects models show great promise for the social and behavioral sciences, there are three nontrivial challenges of the approach. The first challenge is deciding the number and location of knots. Ideally, smoothing splines would be used which removes the difficulty in deciding where and how many knots to place because a knot is automatically placed at each unique time point in the data. The smoothness of the curves is controlled by two smoothing parameters. However, the approach of using a knot at each unique time point suffers when the number of unique time points approaches or exceeds the sample size. The model quickly becomes too computationally intensive to be practical or the model may result in poor fit to the data. With functional mixed-effects models, the number and location of knots must be determined and there is no set guidelines on when there are too many or too few knots. However, it seems important for researchers to aim to collect data when individual trajectories are nonlinear (i.e., 2^{nd} derivative of the curves is non-zero). However, more research needs to be done in this area to provide guidelines to applied researchers.

The second area of challenge for functional mixed-effects models is in deciding the smoothing parameters. Just as in Fine et al. (2019), this simulation used the AIC to pick the set of smoothing parameters. Other approaches can be used to pick the set of smoothing parameters, such as BIC or generalized cross validation (GCV). Using this *hands off* approach of allowing AIC to pick the set of smoothing parameters does not always result in the set of smoothing parameters that best smooths the curves. This was

the case analyzing the progesterone data in the empirical example; the curves were under-smoothed with the set of smoothing parameters chosen by AIC. Functional mixed-effects models are built to be flexible. With this inherent flexibility, the researcher needs to be more *hands on* in deciding the smoothing parameters. The models can be fit with smoothing parameters selected using AIC (or BIC, GCV, etc); however, the resulting curves should be examined to ensure the curves are not over- or under-smoothed. If the curves appear to be over- or under-smoothed, the researcher needs to sample sets of smoothing parameter values until the curves no longer appear over- or under-smoothed. This is a subjective practice and there is no one 'right' way to ensure curves are not over- or under-smoothed.

The final challenge associated with functional mixed-effects model is software. Functional mixed-effects models were fit using the *sme* package in R (Berk, 2013). This package was developed to analyze gene expression data with few time points measured over huge numbers of variables. Fine et al. (2019) showed some of the default setting in the function might not be optimal for longitudinal data seen in the social sciences. In this dissertation, I followed the changes Fine et al. (2019) used to achieve model fits that converged to reasonable solutions. Other researchers may also need to adjust some of the default settings to achieve reasonable model fits.

Nonlinear mixed-effects models have two main challenges: choosing the correct functional form and choosing adequate starting values. Choosing the correct functional form requires researchers to not only have background knowledge of the developmental process under study, but also some knowledge of how to parameterize that change process. Choosing adequate starting values is a non-trivial problem. Nonlinear mixed-

effects models struggle to converge if good strong starting values are not provided. This may be a time intensive process of sampling sets of starting values until the model converges to an adequate fit.

Future Directions

This simulation used AIC to select the optimal set of smoothing parameters for the functional mixed-effects models. Future simulations should investigate other likelihood based approaches (e.g., BIC, GCV) to selecting the optimal set of smoothing parameters. Future research should also focus on the inclusion of covariates in the functional mixed-effects model to account for group differences.

REFERENCES

Berk, M (2013). sme: Smoothing-splines Mixed-effects Models. R package version 0.8. http://CRAN.R-project.org/package=sme

Berk, M. (2012). Statistical Methods for Replicated, High-Dimensional Biological Time Series (Doctoral dissertation). Retrieved from http://wwwf.imperial.ac.uk/~mab201/MauriceBerkPhDThesisFinal.pdf.

Blozis, S. A., & Cudeck, R. (1999). Conditionally linear mixed-effects models with latent variable covariates. *Journal of Educational and Behavioral Statistic*s, 24(3), 245. doi:10.2307/1165324

Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with B -splines and penalties. *Statistical Science,11*(2), 89-121. doi:10.1214/ss/1038425655

Fine, K. L., Suk, H. W., & Grimm, K. J. (2019). An examination of a functional mixed-effects modeling approach to the analysis of longitudinal data. *Multivariate Behavioral Research*. doi: 10.1080/00273171.2018.1520626

Fitzmaurice, G. M., Davidian, M., Verbeke, G., & Molenberghs, G. (Eds.). (2009). *Longitudinal data analysis*. Boca Raton, FL: CRC Press/Taylor & Francis.

Green, P., & Silverman, B. (1994). *Nonparametric regression and generalized linear models: A roughness penalty approach* (1st ed.). London: Chapman & Hall.

Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth modeling: structural equation and multilevel modeling approaches*. New York, NY: Guilford Press.

Grimm, K. J., Steele, J. S., Ram, N., & Nesselroade, J. R. (2013). Exploratory latent growth models in the structural equation modeling framework. *Structural Equation Modeling: A Multidisciplinary Journal*, 20, 568-591. doi:10.1080/10705511.2013.824775

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 561–614). Boston, MA: Springer. doi:10.1007/978-1-4613-0893-5_17

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107–122. doi:10.1007/bf02294746

Pinheiro J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team (2018). *nlme: Linear and nonlinear mixed effects models*. R package version 3.1-137, https://CRAN.R-project.org/package=nlme.

Ram, N., & Grimm, K. (2007). Using simple and complex growth models to articulate developmental change: Matching theory to method. *International Journal of Behavioral Development*, *31*, 303-316. doi:10.1177/0165025407077751

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*. New York: Springer.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis (2nd ed.)*. Thousand Oaks: Sage Publications.

Searle, S., Casella, G., & McCulloch, C. (1992). *Variance components*. New York: Wiley.

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

Wolfinger, R. D., & Lin, X. (1997). Two Taylor-series approximation methods for nonlinear mixed models. *Computational Statistics & Data Analysis*, 25(4), 465-490. doi:10.1016/s0167-9473(97)00012-1

Wu, H., & Zhang, J. (2006). *Nonparametric regression methods for longitudinal data analysis: Mixed-effects modeling approaches*. Hoboken, N.J.: Wiley-Interscience.

Yao, F., Müller, H., & Wang, J. (2005). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association,* 577-590.

APPENDIX A

FIGURES

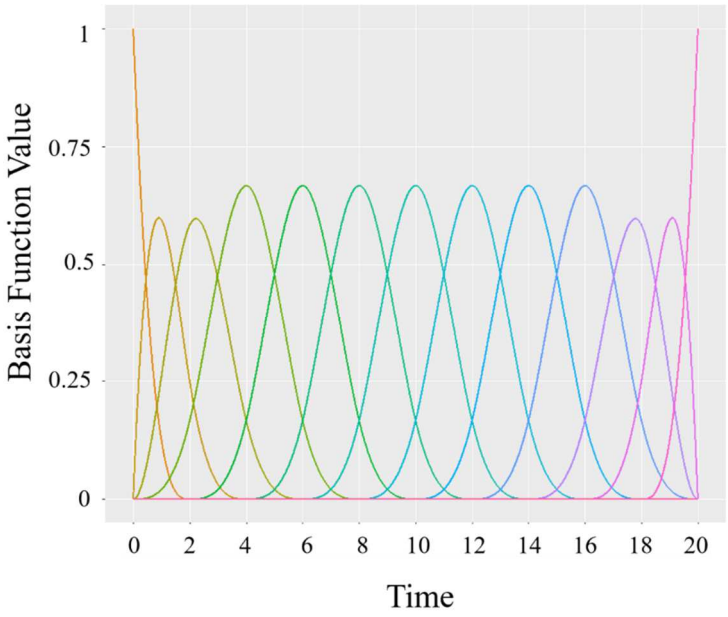


Figure 1. Cubic B-spline basis used to define the grand mean function in the simulation. Each color represents a different B-spline basis function

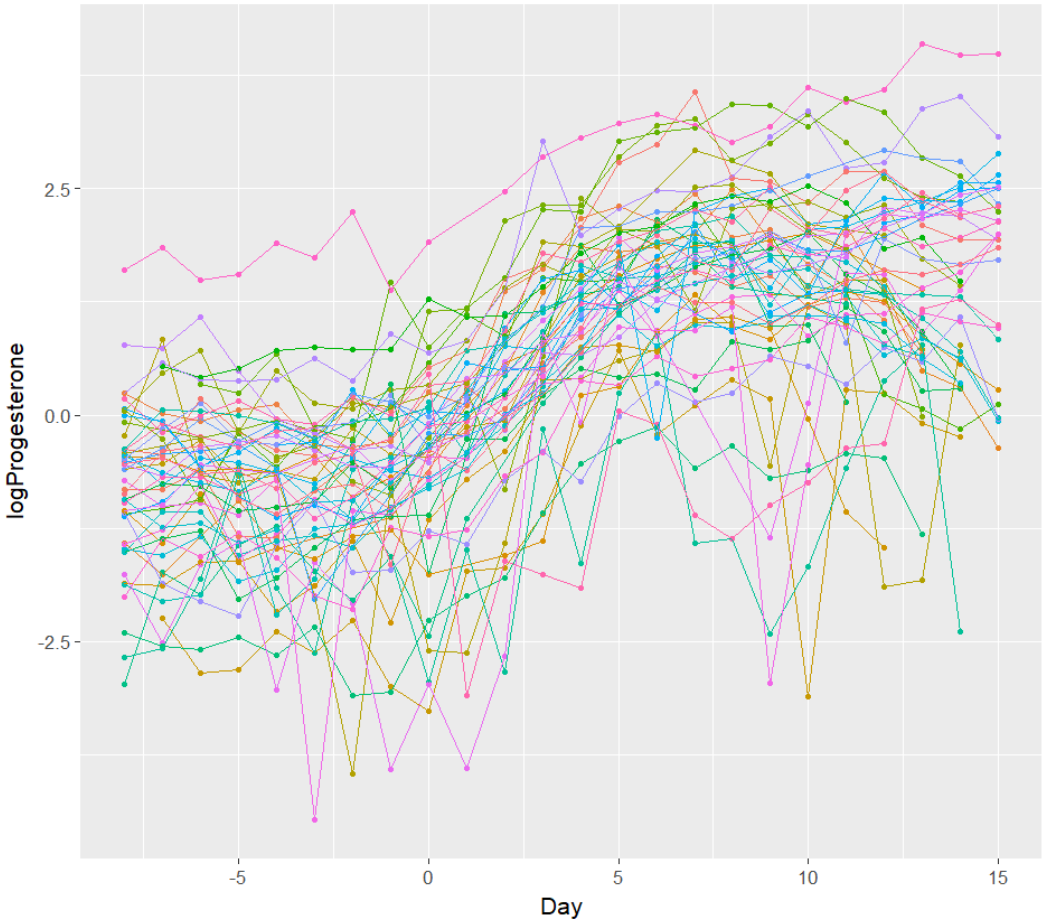


Figure 2. Progesterone cycle data (the log has been taken) taken across 24 days from 51 women.

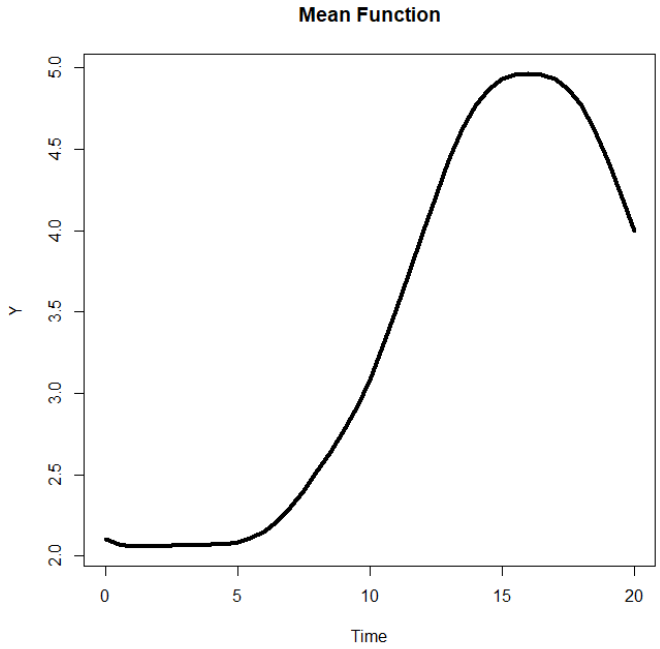


Figure 3. The grand mean function used for the simulation. This curve was formed using the B-spline from Figure 1.

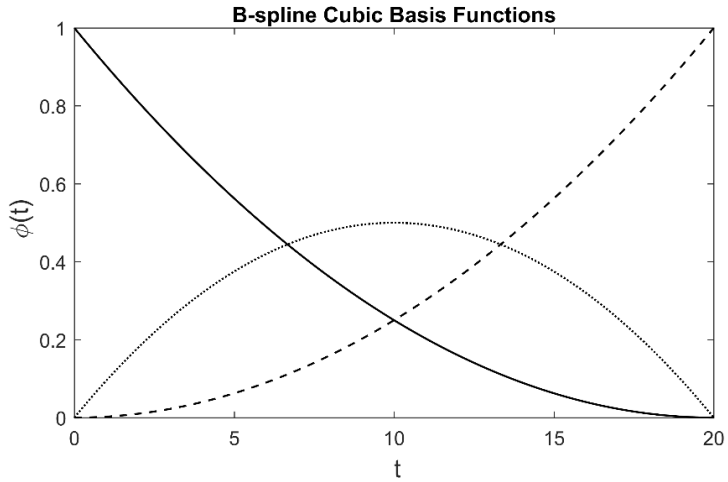


Figure 4. Three B-spline basis functions of degree 3 with equally spaced knots over [0, 20] used to generate the individual trajectories in the simulation.

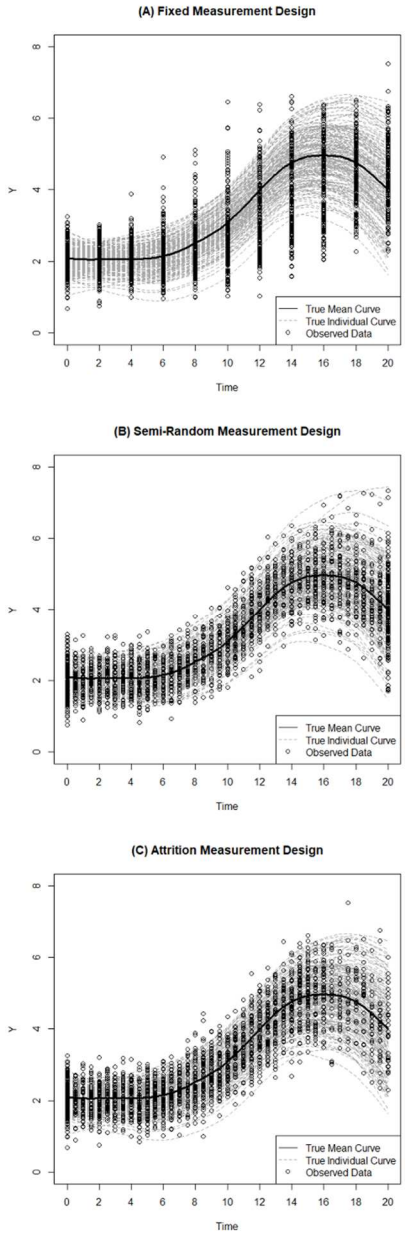


Figure 5. The true underlying mean and individual curves, with the observed data, from a dataset with 200 people observed at 11 time points with each measurement design. Each panels displays the true underlying individual curves in dashed gray, the true grand mean function in black, and the observed data as black circles. Figure 5A shows the true underlying curves for the fixed measurement design. Figure 5B shows the true underlying curves for the semi-random measurement design. Figure 5C shows the true underlying curves for the attrition measurement design.

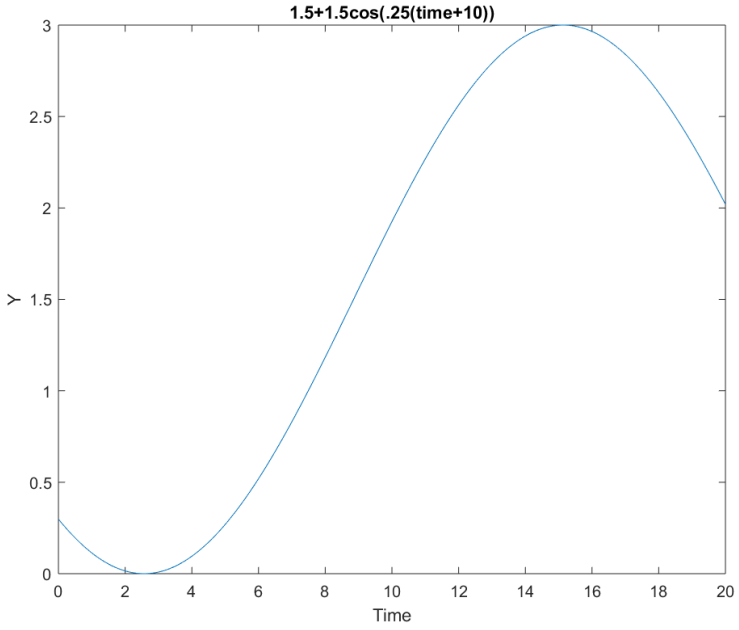


Figure 6. The function from the set of fixed effects in Equation 24 used for the simulation study. This is the closest functional form to the grand mean function used in the simulation.

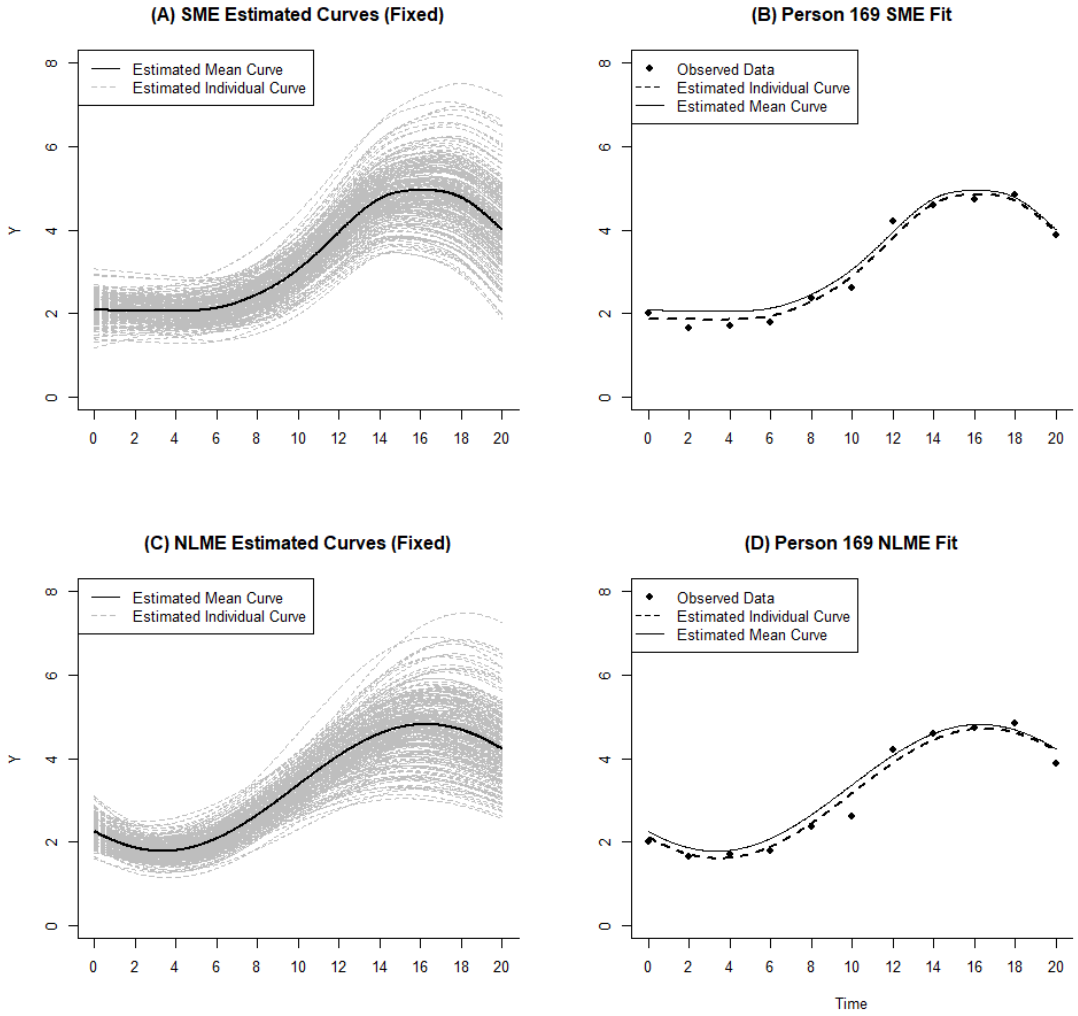


Figure 7. Model fits of the functional and nonlinear mixed-effects models to datasets from conditions with 11 time points per curve, sample size of 200 curves, for the fixed measurement designs. The first row (7A and 7B) is plots of the functional mixed-effects model fits and the second row (7C and 7D) is plots of the nonlinear mixed-effects model fits. The first column (7A and 7C) is plots of the estimated mean (black line) and individual curves (dashed gray lines) fit by either model. The second column (7B and 7D) is plots of a particular individual's observed data (black dots), that individual's estimated curve (dashed black), and the estimated mean curve (solid black).

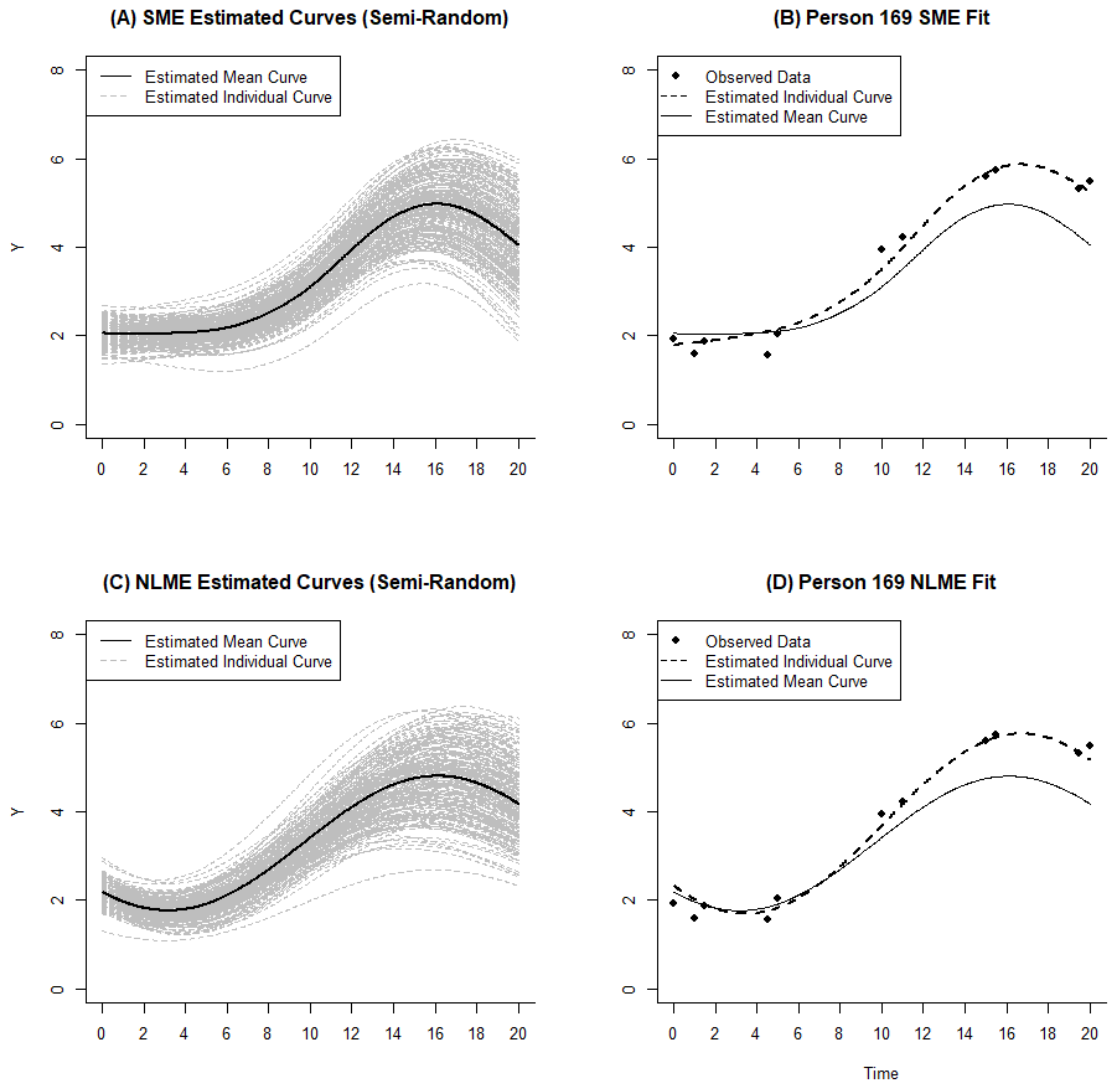


Figure 8. Model fits of the functional and nonlinear mixed-effects models to datasets from conditions with 11 time points per curve, sample size of 200 curves, for the semi-random measurement designs. The first row (8A and 8B) is plots of the functional mixed-effects model fits and the second row (8C and 8D) is plots of the nonlinear mixed-effects model fits. The first column (8A and 8C) is plots of the estimated mean (black line) and individual curves (dashed gray lines) fit by either model. The second column (8B and 8D) is plots of a particular individual's observed data (black dots), that individual's estimated curve (dashed black), and the estimated mean curve (solid black).

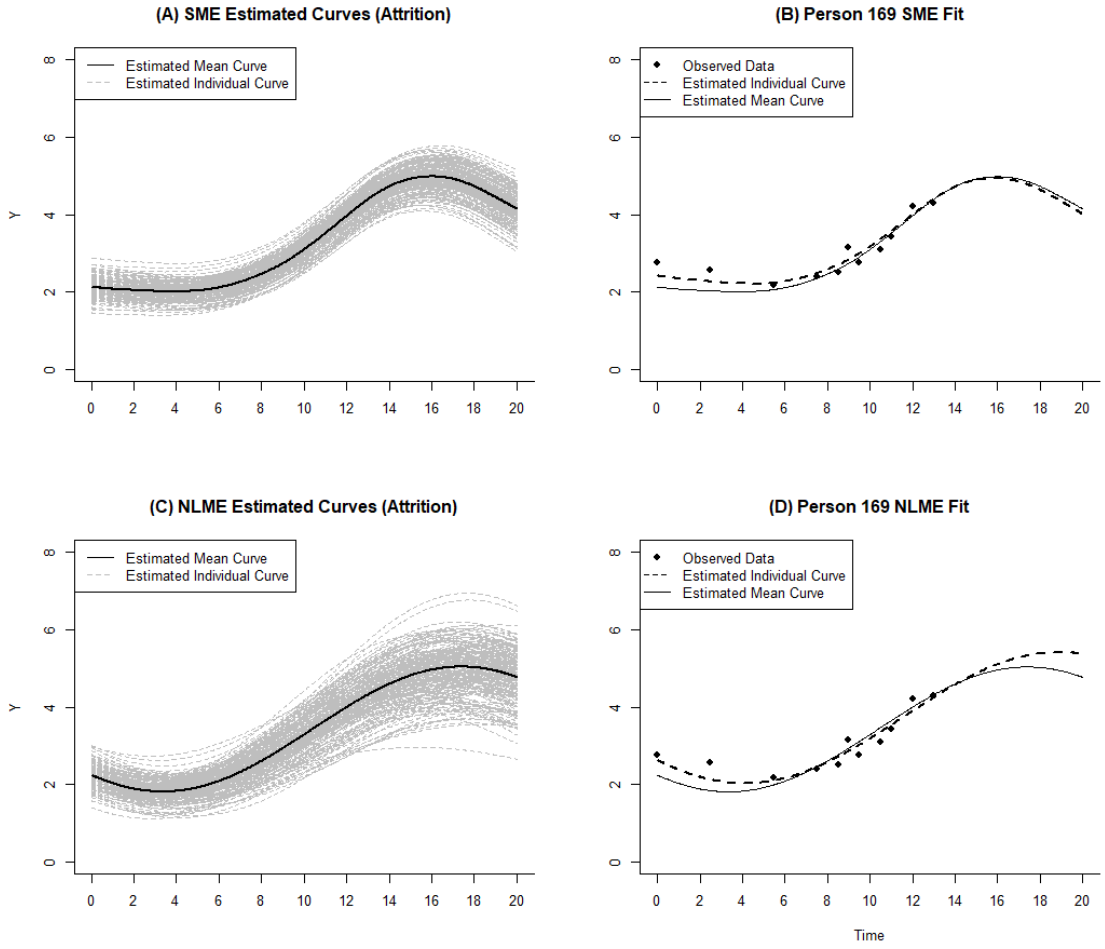


Figure 9. Model fits of the functional and nonlinear mixed-effects models to datasets from conditions with 11 time points per curve, sample size of 200 curves, for the attrition measurement designs. The first row (9A and 9B) is plots of the functional mixed-effects model fits and the second row (9C and 9D) is plots of the nonlinear mixed-effects model fits. The first column (9A and 9C) is plots of the estimated mean (black line) and individual curves (dashed gray lines) fit by either model. The second column (9B and 9D) is plots of a particular individual's observed data (black dots), that individual's estimated curve (dashed black), and the estimated mean curve (solid black).

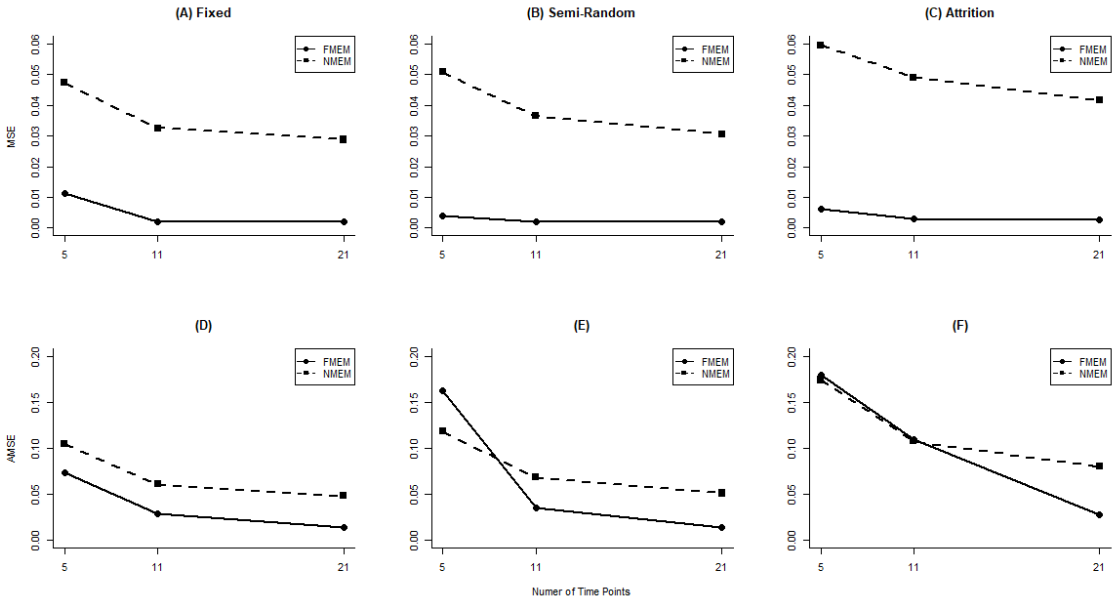


Figure 10. MSE and AMSE values from nonlinear and functional mixed-effects models at sample size 200. Panels A, B, and C correspond to MSE values. Panels D, E, and F correspond to AMSE values. The dashed line with squares correspond to the nonlinear mixed-effects model and the solid line with circles correspond to the functional mixed-effects model. Panels A and D display results for the fixed measurement design. Panels B and E display results for the semi-random measurement design. Panels C and F display results for the attrition measurement design.

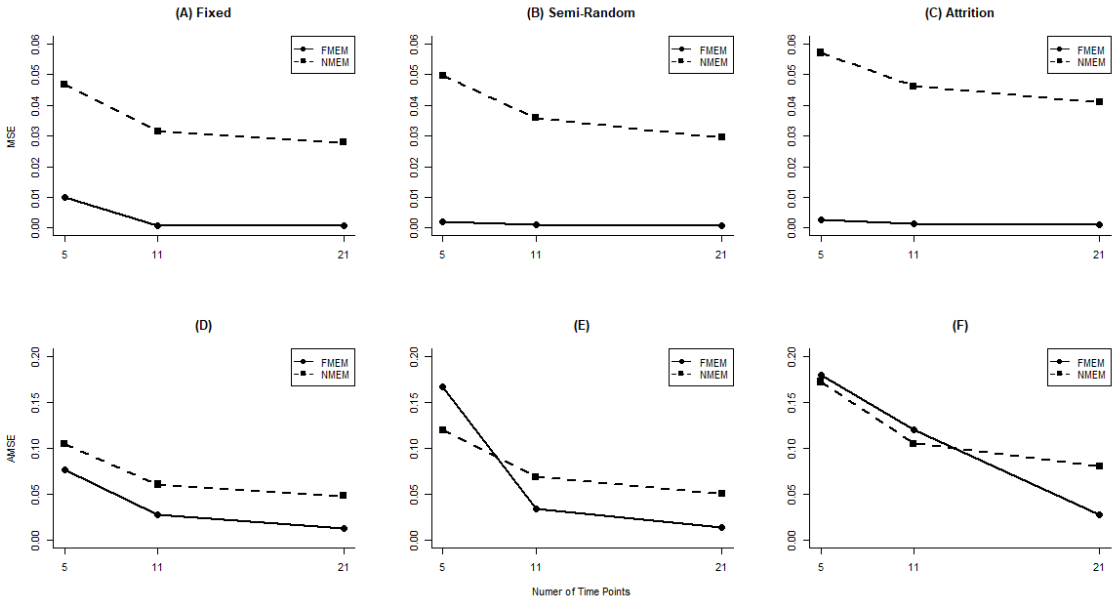


Figure 11. MSE and AMSE values from nonlinear and functional mixed-effects models at sample size 500. Panels A, B, and C correspond to MSE values. Panels D, E, and F correspond to AMSE values. The dashed line with squares correspond to the nonlinear mixed-effects model and the solid line with circles correspond to the functional mixed-effects model. Panels A and D display results for the fixed measurement design. Panels B and E display results for the semi-random measurement design. Panels C and F display results for the attrition measurement design.

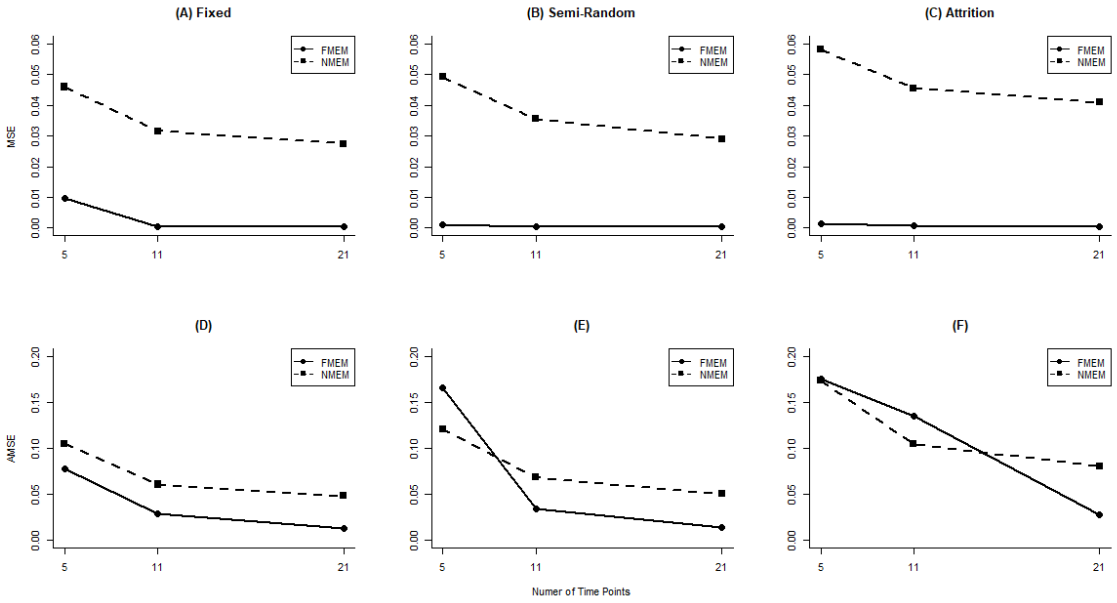


Figure 12. MSE and AMSE values from nonlinear and functional mixed-effects models at sample size 1,000. Panels A, B, and C correspond to MSE values. Panels D, E, and F correspond to AMSE values. The dashed line with squares correspond to the nonlinear mixed-effects model and the solid line with circles correspond to the functional mixed-effects model. Panels A and D display results for the fixed measurement design. Panels B and E display results for the semi-random measurement design. Panels C and F display results for the attrition measurement design.

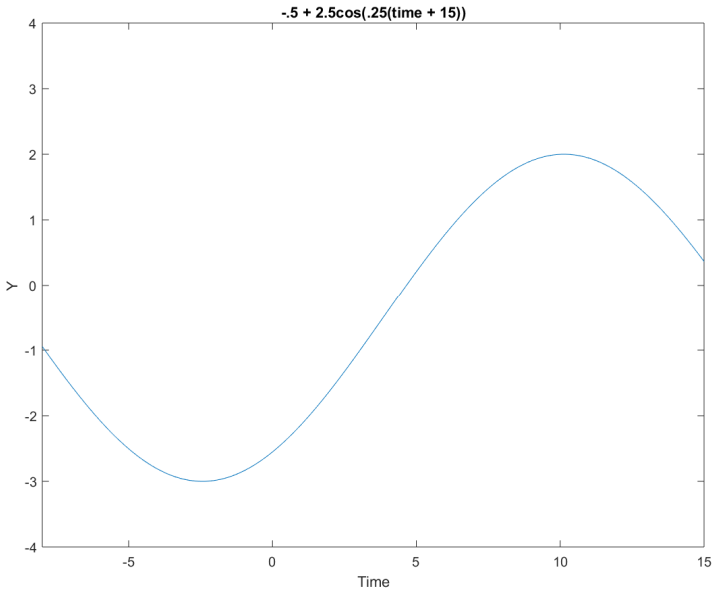
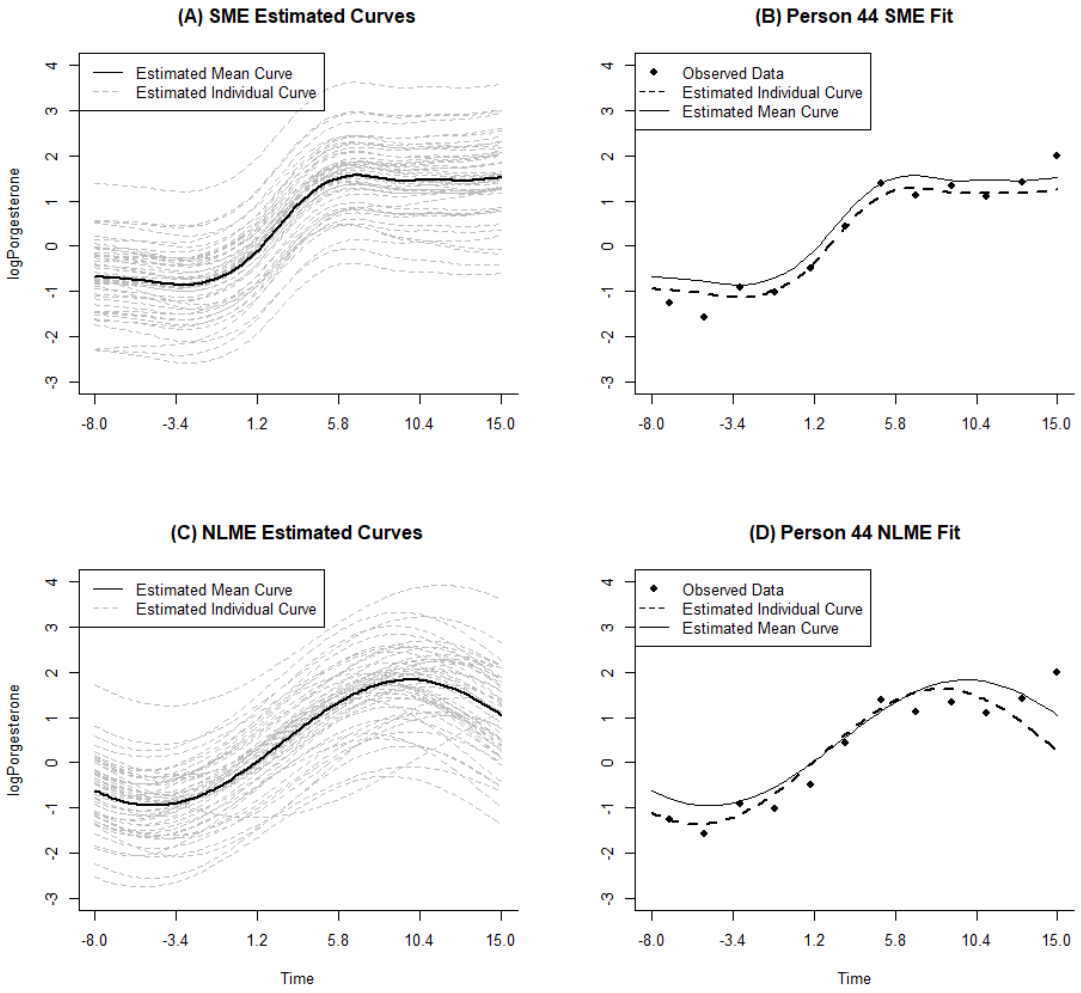


Figure 13. The function from the set of fixed effects in Equation 27 used for the progesterone cycle data.



Figures 14. Plot of model fits of the functional and nonlinear mixed-effects models to the odd time points from the progesterone cycle data. In this figures, panels A and B, are plots of the functional mixed-effects model fits, and panels C and D are plots of the nonlinear mixed-effects model fits. Panels A and C are plots of the estimated mean (black line) and individual curves (dashed gray lines) fit by either model. Panels B and D are plots of a particular individual's observed data (black dots), that individual's estimated curve (dashed black), and the estimated mean curve (solid black).

APPENDIX B

TABLES

Table 1

MSE values for the functional mixed-effects model (FMEM) and the cosine nonlinear mixed-effects model (NMEM) across the simulation conditions.

Measurement Design	Sample Size	Type of Model					
		FMEM			NMEM		
		$n_i = 5$	$n_i = 11$	$n_i = 21$	$n_i = 5$	$n_i = 11$	$n_i = 21$
Fixed	$n = 200$	0.0111	0.0020	0.0019	0.0473	0.0326	0.0288
	$n = 500$	0.0101	0.0009	0.0008	0.0469	0.0316	0.0279
	$n = 1000$	0.0098	0.0005	0.0003	0.0460	0.0316	0.0274
Semi-Random	$n = 200$	0.0038	0.0019	0.0019	0.0509	0.0364	0.0306
	$n = 500$	0.0019	0.0009	0.0008	0.0498	0.0359	0.0295
	$n = 1000$	0.0010	0.0005	0.0005	0.0493	0.0356	0.0290
Attrition	$n = 200$	0.0061	0.0030	0.0026	0.0596	0.0491	0.0417
	$n = 500$	0.0025	0.0013	0.0009	0.0571	0.0463	0.0410
	$n = 1000$	0.0014	0.0006	0.0004	0.0581	0.0455	0.0409

Table 2

AMSE values for the functional mixed-effects model (FMEM) and the cosine nonlinear mixed-effects model (NMEM) across the simulation conditions.

Measurement Design	Sample Size	Type of Model					
		FMEM			NMEM		
		$n_i = 5$	$n_i = 11$	$n_i = 21$	$n_i = 5$	$n_i = 11$	$n_i = 21$
Fixed	$n = 200$	0.0741	0.0286	0.0140	0.1048	0.0613	0.0483
	$n = 500$	0.0771	0.0284	0.0135	0.1047	0.0609	0.0484
	$n = 1000$	0.0780	0.0285	0.0135	0.1050	0.0609	0.0484
Semi-Random	$n = 200$	0.1631	0.0353	0.0144	0.1184	0.0685	0.0513
	$n = 500$	0.1671	0.0344	0.0139	0.1200	0.0689	0.0508
	$n = 1000$	0.1663	0.0344	0.0139	0.1213	0.0685	0.0507
Attrition	$n = 200$	0.1801	0.1096	0.0283	0.1745	0.1078	0.0801
	$n = 500$	0.1793	0.1205	0.0281	0.1724	0.1053	0.0809
	$n = 1000$	0.1753	0.1348	0.0280	0.1736	0.1048	0.0809

APPENDIX C

TECHNICAL DETAILS OF B-SPLINES

B-splines stands for "basis" splines. B-splines are best utilized for non-periodic data with continuous derivatives up to a certain order (Aguilera & Aguilera-Morillo, 2013). They are defined recursively. The i^{th} basis function for a B-spline of order M is defined as

$$B_{1,i}(t) = \begin{cases} 1, & if\ \tau_i \leq t < \tau_{i+1} \\ 0, & otherwise \end{cases}$$
$$B_{m,i}(t) = \frac{t - \tau_i}{\tau_{i+m-1} - \tau_i} B_{m-1,i}(t) + \frac{\tau_{i+m} - t}{\tau_{i+m} - \tau_{i+1}} B_{m-1,i+1}(t). \quad \text{(C.1)}$$

where τ_i is the i^{th} knot point and t is the time point the spline is being evaluated at. $B_{1,i}$ is the i^{th} B-spline of order one (e.g., degree zero). This spline takes a value of one if the time point (t) is between the knot points τ_i and τ_{i+1} and takes a value of zero everywhere else. Then, $B_{m,i}$, is the more general B-spline of order M $(m = 2, \dots, M)$ is defined using B-splines of order $M - 1$ $(B_{m-1,i})$. The terms $\frac{t-\tau_i}{\tau_{i+m-1}-\tau_i}$ and $\frac{\tau_{i+m}-t}{\tau_{i+m}-\tau_{i+1}}$ are constants multiplied by the B-splines of order $M - 1$. This paper focuses on cubic B-splines (i.e., B-splines of order four) since, in general, cubic splines are the lowest degree spline where the knot location cannot be visually detected. This makes cubic splines widely used in practice. Since B-splines are defined recursively, cubic B-splines are defined using B-splines of order one, two, three, and four (i.e., degrees zero, one, two, and three).

In addition to deciding on what order B-spline to use, one must determine the number and location of knots to use. Technically speaking, splines have *knots* and *breakpoints*. Breakpoints refer to the increasing sequence of time points where knots are located. It is possible to specify multiple knots at one time point. Multiple knots at a given time point may be needed when there are abrupt changes in the underlying curve or its derivatives (Ramsay & Silverman, 2005). Specifically for B-splines, one must place *M*

knots at each boundary breakpoint. In a cubic B-spline defined over the interval $[a, b]$, this means placing three knots at the first boundary breakpoint (e.g., three knots at time a) and placing three knots at the end boundary breakpoint (e.g., three knots at time b). Therefore, technically, knots refer to the sequence of time points at breakpoints. In practice, many use the terms interchangeably.

B-splines require specifying how many knots to use and where to place knots. For splines, in general, this is a nontrivial problem and does not have one absolute correct method. Specifying knots requires the user to sample various sets of knots and choosing which works best. Ramsay and Silverman (2005) suggest placing more knots where the curve shows complex variation and less knots where the curve is mildly nonlinear. Two commonly used knot selection methods are using equally spaced quantiles of time points as knots (Ruppert et al., 2003) or placing knots at equally spaced time points (Eilers & Marx, 1996). It is important to not under or over smooth the curves by placing too few or too many knots. However, whether the curve is under or over smoothed may be left to the interpretation of the user.

Figures C.1 through C.12 in Appendix D help demonstrate the recursive definition of the B-spline and to show placement of knots (Suk, 2015). As previously stated, this paper focuses on cubic B-spline. The following plots will show the definition of B-splines of order one, two, three, and four. Figure C.1 shows a time interval from [0,1] broken up into three subintervals (e.g., $\left[0, \frac{1}{3}\right], \left[\frac{1}{3}, \frac{2}{3}\right], [\frac{2}{3}, 1]$). Circled in light blue, there are two interior breakpoints at time $t = \frac{1}{3}$ (e.g., ξ_1) and time $t = \frac{2}{3}$ (e.g., ξ_2). Circled in dark blue, there are two boundary breakpoints at time $t = 0$ (e.g., ξ_0) and time $t = 1$

(e.g., ξ_3). The green dots are the knots placed at each interior breakpoint and the red dots are the knots placed at each boundary breakpoint. There is one knot at each interior breakpoint (e.g., τ_4 and τ_5), three knots at the first boundary breakpoint (e.g., τ_1, τ_2, τ_3), and three knots at the end boundary breakpoint (e.g., τ_6, τ_7, τ_8).

We can now start to define a B-spline of order 1 (i.e., degree 0) using the first equation from Equation C.1: $B_{1,i}(t) = \begin{cases} 1, & if\ \tau_i \leq t < \tau_{i+1} \\ 0, & otherwise \end{cases}$. This equation states a B-spline of order 1 (e.g., degree 0) will take on a value of one at time t if that value of t is greater than or equal to the value of the i^{th} knot (τ_i) and less than the value of the $i^{\text{th}} + 1$ knot (τ_{i+1}). We begin by defining $B_{1,1}(t)$, the order 1 (e.g., degree 0) B-spline at the $i = 1$ knot. To do so, we need τ_1 and τ_2; as Figure C.1 shows, both τ_1 and τ_2 equal zero and correspond with $t = 0$. Subbing $\tau_i = \tau_1 = 0$, $\tau_{i+1} = \tau_2 = 0$, and $t = 0$ into the first equation from Equation C.1 we get: $B_{1,1}(t) = \begin{cases} 1, & if\ 0 \leq 0 < 0 \\ 0, & otherwise \end{cases}$. Because $0 \not< 0$ the first B-spline function becomes $B_{1,1}(t) = 0$. The same logic applies to the second order 1 (e.g., degree 0) B-spline. By subbing $\tau_i = \tau_2 = 0$, $\tau_{i+1} = \tau_3 = 0$, and $t = 0$ into the first equation from Equation C.1, we see the second B-spline function becomes $B_{1,2}(t) = 0$ (again, because $t = 0 \not< \tau_{i+1} = \tau_3 = 0$). Making substitutions for $\tau_i = \tau_3 = 0$, $\tau_{i+1} = \tau_4 = \frac{1}{3}$, and $t = 0$ $B_{1,3}(t) = \begin{cases} 1, & if\ 0 \leq 0 < \frac{1}{3} \\ 0, & otherwise \end{cases}$. In this case, the statement $0 \leq 0 < \frac{1}{3}$ is true, and thus, the B-spline takes on a value of 1 between $\left[0, \frac{1}{3}\right]$ and takes on a value of 0 everywhere else. Figure C.2 displays the third B-spline of order 1 (e.g., degree 0).

Making substitutions for $\tau_i = \tau_4 = \frac{1}{3}$, $\tau_{i+1} = \tau_5 = \frac{2}{3}$, and $t = \frac{1}{3}$ $B_{1,4}(t) =$

$\begin{cases} 1, & if\ \frac{1}{3} \leq \frac{1}{3} < \frac{2}{3} \\ 0, & otherwise \end{cases}$. In this case, the statement $\frac{1}{3} \leq \frac{1}{3} < \frac{2}{3}$ is true, and thus, the B-spline takes on a value of 1 between $\left[\frac{1}{3}, \frac{2}{3}\right]$ and takes on a value of 0 everywhere else. Figure C.3 displays the fourth B-spline of order 1 (e.g., degree 0). Making substitutions for $\tau_i = \tau_5 = \frac{2}{3}$, $\tau_{i+1} = \tau_6 = 1$, and $t = \frac{2}{3}$ $B_{1,5}(t) = \begin{cases} 1, & if\ \frac{2}{3} \leq \frac{2}{3} < 1 \\ 0, & otherwise \end{cases}$. In this case, the statement $\frac{2}{3} \leq \frac{2}{3} < 1$ is true, and thus, the B-spline takes on a value of 1 between $\left[\frac{2}{3}, 1\right]$ and takes on a value of 0 everywhere else. Figure C.4 displays the fifth B-spline of order 1 (e.g., degree 0). The last two B-splines, $B_{1,6}$ and $B_{1,7}$, both take on values of zero. This occurs in the same way $B_{1,1}$ and $B_{1,2}$ take on the value of zero. When making the substitutions for $\tau_i = 1$, $\tau_{i+1} = 1$, and $t = 1$, $1 \not< 1$ and thus makes the B-splines take on a value of zero everywhere. Figure C.5 shows the order 1 (e.g., degree 0) B-spline basis.

Now that the B-spline basis of order 1 (e.g., degree 0) has been defined, the B-spline basis of order 2 (i.e., degree 1) can be defined. The second equation in Equation C.1 is used to define B-spline bases of order 2 (i.e., degree 0) and above. Modifying the second equation in Equation C.1, making substitutions for m, gives: $B_{2,i}(t) =$ $\frac{t-\tau_i}{\tau_{i+2-1}-\tau_i} B_{2-1,i}(t) + \frac{\tau_{i+2}-t}{\tau_{i+2}-\tau_{i+1}} B_{2-1,i+1}(t) = \frac{t-\tau_i}{\tau_{i+1}-\tau_i} B_{1,i}(t) + \frac{\tau_{i+2}-t}{\tau_{i+2}-\tau_{i+1}} B_{1,i+1}(t)$. Again, finding the basis involves substitutions for τ_i and t; it now also involves substituting values of $B_{1,i}(t)$. Starting with $i = 1$, $B_{2,1}(t) = \frac{t-\tau_i}{\tau_{i+1}-\tau_i} B_{1,1}(t) + \frac{\tau_{i+2}-t}{\tau_{i+2}-\tau_{i+1}} B_{1,2}(t) =$ $\frac{0-0}{0-0} 0 + \frac{0-0}{0-0} 0 = 0$. So, the first B-spline of order 2 (i.e., degree 1) takes on a value of 0

everywhere. For $i = 2$, $B_{2,2}(t) = \frac{t-\tau_i}{\tau_{i+1}-\tau_i} B_{1,2}(t) + \frac{\tau_{i+2}-t}{\tau_{i+2}-\tau_{i+1}} B_{1,3}(t)$. The first part, $\frac{t-\tau_2}{\tau_3-\tau_2} B_{1,2}(t)$, equals zero because the only value $B_{1,2}(t)$ takes on is zero. The second part, $\frac{\tau_4-t}{\tau_4-\tau_3} B_{1,3}(t)$, is primarily used to define the second B-spline basis function of order 2 (i.e., degree 1). The value of $B_{1,3}(t)$ at τ_3 is 1, and maintains this value until just before τ_4. At τ_4, the value of $B_{1,3}(t)$ is 0. Plugging values of t between $\tau_3 = 0 < t < \tau_4 = \frac{1}{3}$, in $\frac{\tau_4-t}{\tau_4-\tau_3} B_{1,3}(t)$ creates the line seen in Figure C.6. For example, substituting $t = 0$ gives $\frac{\frac{1}{3}-0}{\frac{1}{3}-0}[1] = 1$. Substituting $t = \frac{1}{6}$ gives $\frac{\frac{1}{3}-\frac{1}{6}}{\frac{1}{3}-0}[1] = .5$. Substituting $t = \frac{1}{3}$ gives $\frac{\frac{1}{3}-\frac{1}{3}}{\frac{1}{3}-0}[0] = 0$.

For $i = 3$, $B_{2,3}(t) = \frac{t-\tau_i}{\tau_{i+1}-\tau_i} B_{1,3}(t) + \frac{\tau_{i+2}-t}{\tau_{i+2}-\tau_{i+1}} B_{1,4}(t) = \frac{t-0}{\frac{1}{3}-0} B_{1,3}(t) + \frac{\frac{2}{3}-t}{\frac{2}{3}-\frac{1}{3}} B_{1,4}(t)$.

Now values of t between $0 < t < 1$ will be substituted. Recall, the value of $B_{1,3}(t)$ is 1 between τ_3 and τ_4, and takes on a value of 0 everywhere else. Also recall, the value of $B_{1,4}(t)$ is 1 between τ_4 and τ_5, and takes on a value of 0 everywhere else. Substituting $t = 0$ into $\frac{t-0}{\frac{1}{3}-0} B_{1,3}(t) + \frac{\frac{2}{3}-t}{\frac{2}{3}-\frac{1}{3}} B_{1,4}(t) = \frac{0-0}{\frac{1}{3}-0}[1] + \frac{\frac{2}{3}-0}{\frac{2}{3}-\frac{1}{3}}[0] = 0$. Substituting $t = \frac{1}{6}$ into $\frac{t-0}{\frac{1}{3}-0} B_{1,3}(t) + \frac{\frac{2}{3}-t}{\frac{2}{3}-\frac{1}{3}} B_{1,4}(t) = \frac{\frac{1}{6}-0}{\frac{1}{3}-0}[1] + \frac{\frac{2}{3}-\frac{1}{6}}{\frac{2}{3}-\frac{1}{3}}[0] = .5 + 0 = .5$. Substituting $t = \frac{1}{3}$ into $\frac{t-0}{\frac{1}{3}-0} B_{1,3}(t) + \frac{\frac{2}{3}-t}{\frac{2}{3}-\frac{1}{3}} B_{1,4}(t) = \frac{\frac{1}{3}-0}{\frac{1}{3}-0}[0] + \frac{\frac{2}{3}-\frac{1}{3}}{\frac{2}{3}-\frac{1}{3}}[1] = 0 + 1 = 1$. Substituting $t = \frac{1}{2}$ into $\frac{t-0}{\frac{1}{3}-0} B_{1,3}(t) + \frac{\frac{2}{3}-t}{\frac{2}{3}-\frac{1}{3}} B_{1,4}(t) = \frac{\frac{1}{2}-0}{\frac{1}{3}-0}[0] + \frac{\frac{2}{3}-\frac{1}{2}}{\frac{2}{3}-\frac{1}{3}}[1] = 0 + .5 = .5$. Substituting $t = \frac{2}{3}$ into $\frac{t-0}{\frac{1}{3}-0} B_{1,3}(t) + \frac{\frac{2}{3}-t}{\frac{2}{3}-\frac{1}{3}} B_{1,4}(t) = \frac{\frac{2}{3}-0}{\frac{1}{3}-0}[0] + \frac{\frac{2}{3}-\frac{2}{3}}{\frac{2}{3}-\frac{1}{3}}[0] = 0$. Substituting any values over $t \geq \frac{2}{3}$ will

yield a value of zero because both $B_{1,3}(t)$ and $B_{1,4}(t)$ are zero at those values. The third B-spline of order 2 (i.e., degree 1) is shown in Figure C.7.

For $i = 4$, $B_{2,4}(t) = \frac{t-\tau_i}{\tau_{i+1}-\tau_i} B_{1,4}(t) + \frac{\tau_{i+2}-t}{\tau_{i+2}-\tau_{i+1}} B_{1,5}(t) = \frac{t-\frac{1}{3}}{\frac{2}{3}-\frac{1}{3}} B_{1,4}(t) + \frac{1-t}{1-\frac{2}{3}} B_{1,5}(t)$.

Now values of t between $0 < t < 1$ will be substituted. Recall, the value of $B_{1,4}(t)$ is 1 between τ_4 and τ_5, and takes on a value of 0 everywhere else. Also recall, the value of $B_{1,5}(t)$ is 1 between τ_5 and τ_6, and takes on a value of 0 everywhere else. Substituting any values over $t \leq \frac{1}{3}$ will yield a value of zero because both $B_{1,4}(t)$ and $B_{1,5}(t)$ are zero at those values. Substituting $t = \frac{1}{3}$ into $\frac{t-\frac{1}{3}}{\frac{2}{3}-\frac{1}{3}} B_{1,4}(t) + \frac{1-t}{1-\frac{2}{3}} B_{1,5}(t) = \frac{\frac{1}{3}-\frac{1}{3}}{\frac{2}{3}-\frac{1}{3}}[1] + \frac{1-\frac{1}{3}}{1-\frac{2}{3}}[0] = 0.$

Substituting $t = \frac{1}{2}$ into $\frac{t-\frac{1}{3}}{\frac{2}{3}-\frac{1}{3}} B_{1,4}(t) + \frac{1-t}{1-\frac{2}{3}} B_{1,5}(t) = \frac{\frac{1}{2}-\frac{1}{3}}{\frac{2}{3}-\frac{1}{3}}[1] + \frac{1-\frac{1}{2}}{1-\frac{2}{3}}[0] = .5 + 0 = .5.$

Substituting $t = \frac{2}{3}$ into $\frac{t-\frac{1}{3}}{\frac{2}{3}-\frac{1}{3}} B_{1,4}(t) + \frac{1-t}{1-\frac{2}{3}} B_{1,5}(t) = \frac{\frac{2}{3}-\frac{1}{3}}{\frac{2}{3}-\frac{1}{3}}[0] + \frac{1-\frac{2}{3}}{1-\frac{2}{3}}[1] = 0 + 1 = 1.$

Substituting $t = \frac{5}{6}$ into $\frac{t-\frac{1}{3}}{\frac{2}{3}-\frac{1}{3}} B_{1,4}(t) + \frac{1-t}{1-\frac{2}{3}} B_{1,5}(t) = \frac{\frac{5}{6}-\frac{1}{3}}{\frac{2}{3}-\frac{1}{3}}[0] + \frac{1-\frac{5}{6}}{1-\frac{2}{3}}[1] = 0 + .5 = .5$. Lastly, substituting $t = 1$ into $\frac{t-\frac{1}{3}}{\frac{2}{3}-\frac{1}{3}} B_{1,4}(t) + \frac{1-t}{1-\frac{2}{3}} B_{1,5}(t) = \frac{1-\frac{1}{3}}{\frac{2}{3}-\frac{1}{3}}[0] + \frac{1-1}{1-\frac{2}{3}}[1] = 0$. The fourth B-spline of order 2 (i.e., degree 1) is shown in Figure C.8.

For $i = 5$, $B_{2,5}(t) = \frac{t-\tau_i}{\tau_{i+1}-\tau_i} B_{1,5}(t) + \frac{\tau_{i+2}-t}{\tau_{i+2}-\tau_{i+1}} B_{1,6}(t)$. The first part, $\frac{\tau_{i+2}-t}{\tau_{i+2}-\tau_{i+1}} B_{1,6}(t)$, equals zero because the only value $B_{1,6}(t)$ takes on is zero. The first part, $\frac{t-\tau_i}{\tau_{i+1}-\tau_i} B_{1,5}(t)$, is primarily used to define the fifth B-spline basis function of order 2 (i.e., degree 1). Recalling $B_{1,5}(t)$ is zero everywhere except $\tau_5 = \frac{2}{3} < t < \tau_6 = 1$, only

values of t within this region are needed to define this basis function. Substituting $t = \frac{2}{3}$ into $\frac{t-\tau_i}{\tau_{i+1}-\tau_i}B_{1,5}(t) + \frac{\tau_{i+2}-t}{\tau_{i+2}-\tau_{i+1}}B_{1,6}(t) = \frac{\frac{2}{3}-\frac{2}{3}}{1-\frac{2}{3}}[1] + \frac{1-\frac{2}{3}}{1-1}[0] = 0$. Substituting $t = \frac{5}{6}$ into $\frac{t-\tau_i}{\tau_{i+1}-\tau_i}B_{1,5}(t) + \frac{\tau_{i+2}-t}{\tau_{i+2}-\tau_{i+1}}B_{1,6}(t) = \frac{\frac{5}{6}-\frac{2}{3}}{1-\frac{2}{3}}[1] + \frac{1-\frac{5}{6}}{1-1}[0] = .5 + 0 = .5$. Substituting $t = 1$ into $\frac{t-\tau_i}{\tau_{i+1}-\tau_i}B_{1,5}(t) + \frac{\tau_{i+2}-t}{\tau_{i+2}-\tau_{i+1}}B_{1,6}(t) = \frac{1-\frac{2}{3}}{1-\frac{2}{3}}[1] + \frac{1-1}{1-1}[0] = 1 + 0 = 1$. The fifth B-spline of order 2 (i.e., degree 1) is shown in Figure C.9.

The sixth order 2 (i.e., degree 1) B-spline basis function takes on values of zero across the range of t. This is because $B_{2,6}(t) = \frac{t-\tau_i}{\tau_{i+1}-\tau_i}B_{1,6}(t) + \frac{\tau_{i+2}-t}{\tau_{i+2}-\tau_{i+1}}B_{1,7}(t)$, both $B_{1,6}(t)$ and $B_{1,7}(t)$ only take on values of zero. There were seven order 1 (i.e., degree 0) B-spline basis functions defined. There are six order 2 (i.e., degree 1) B-spline basis functions. Figure C.10 displays the order 2 (i.e., degree 1) B-spline basis. In this plot, the second B-spline basis function of order 2 (i.e., degree 1) is in blue, the third is in green, the fourth is in red, and the fifth is in teal. The first and sixth all take on values of zero and are not plotted. The red dots are the knots.

The exact same set of steps can be taken to define the quadratic (e.g., order 3) and cubic (e.g., order 4) B-spline basis. To define the cubic B-spline, the definitions for the order 3 (i.e., degree 2) B-splines would be used. Again, knot values, values of the order 3 (i.e., degree 2) B-spline basis functions, and values of t would be plugged into the equations and solved. The calculations are not included here for either the order 3 or order 4 B-spline basis. The order 3 B-spline basis is plotted in Figure C.11. The first B-spline basis function of order 3 is in blue, the second is in green, the third is in red, the

fourth is in teal, and the fifth is in purple. The sixth B-spline basis function only takes on values of zero and is not plotted. The cubic (e.g., order 4) B-spline basis is plotted in Figure C.12. The first B-spline basis function of order 4 is in black, the second is in red, the third is in green, the fourth is in blue, the fifth is in teal, and the sixth is in purple.

FIGURES

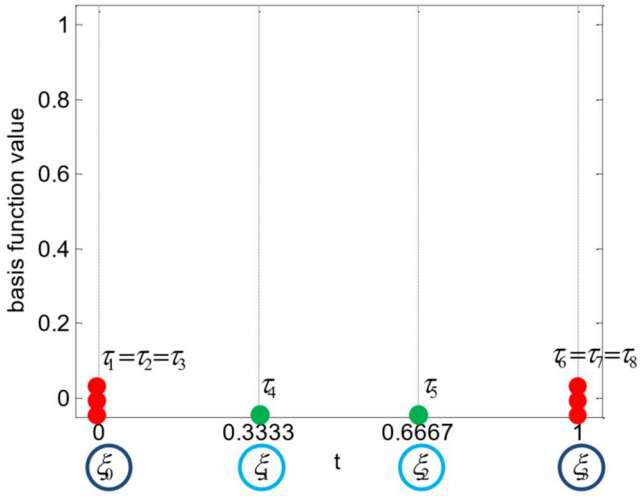


Figure C.1. This plot shows the subintervals (e.g., $\left[0,\ \frac{1}{3}\right]$, $\left[\frac{1}{3},\frac{2}{3}\right]$, $[\frac{2}{3},1]$), interior breakpoints (circled in light blue; e.g., ξ_1 and ξ_2), boundary breakpoints (circled in dark blue; e.g., ξ_0 and ξ_3), interior knots (green dots; e.g., τ_4 and τ_5), and boundary knots (red dots; e.g., τ_1, τ_2, τ_3 and τ_6, τ_7, τ_8) for defining a cubic B-spline.

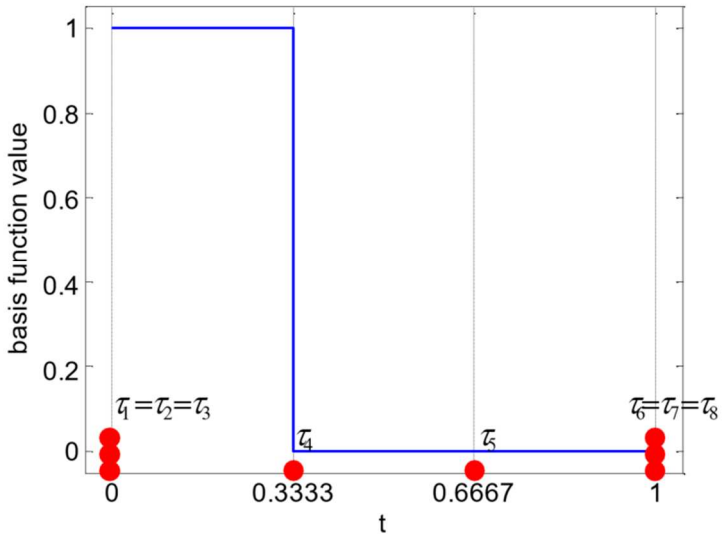


Figure C.2. This plot shows the third B-spline basis function of order 1 (i.e., degree 0) in blue. It takes on a value of 1 between $\left[0,\ \frac{1}{3}\right]$ and takes on a value of 0 everywhere else. The red dots are the knots.

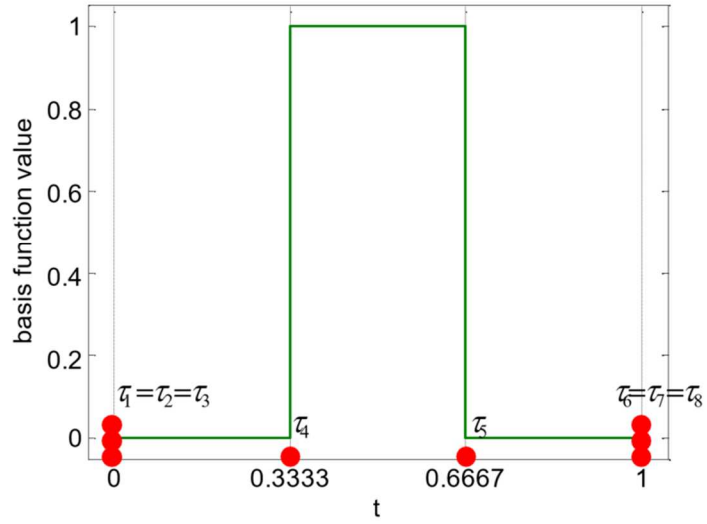


Figure C.3. This plot shows the fourth B-spline basis function of order 1 (i.e., degree 0) in green. It takes on a value of 1 between $\left[\frac{1}{3}, \frac{2}{3}\right]$ and takes on a value of 0 everywhere else. The red dots are the knots.

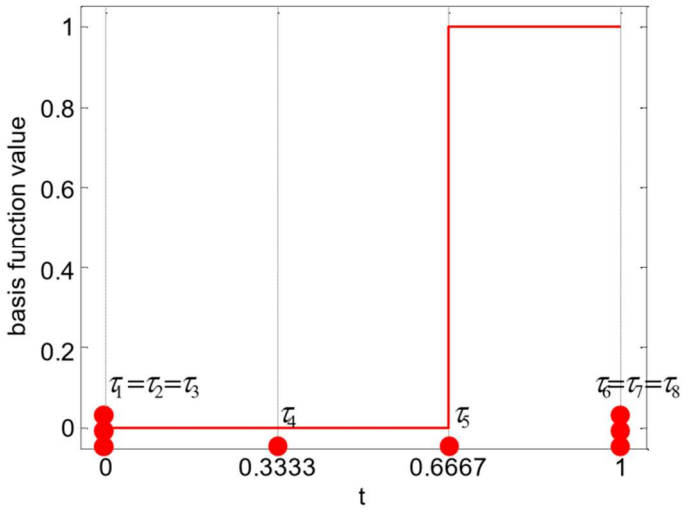


Figure C.4. This plot shows the fifth B-spline basis function of order 1 (i.e., degree 0) in red. It takes on a value of 1 between $\left[\frac{2}{3}, 1\right]$ and takes on a value of 0 everywhere else. The red dots are the knots.

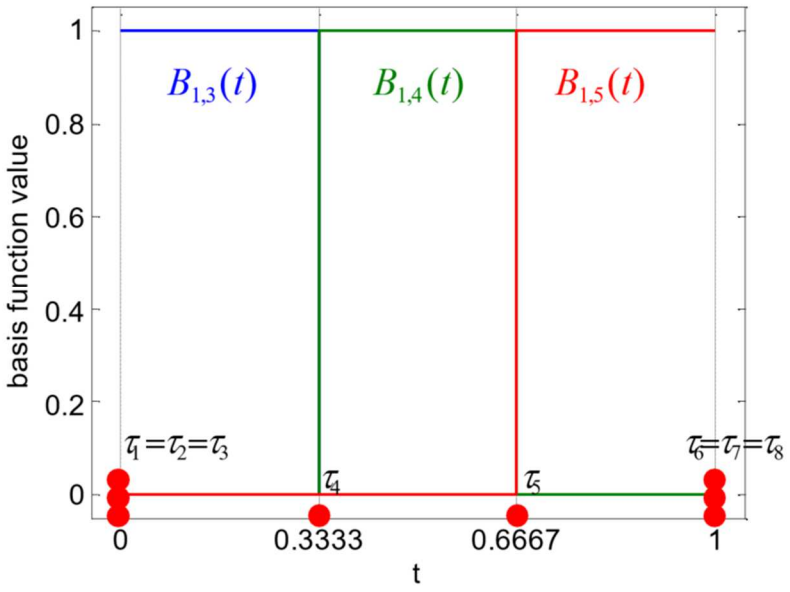


Figure C.5. This plot shows the order 1 (i.e., degree 0) B-spline basis. The third B-spline basis function of order 1 (i.e., degree 0) in blue, the fourth is in green, and the fifth is in red. The first, second, sixth, and seventh all take on values of zero and are not plotted. The red dots are the knots.

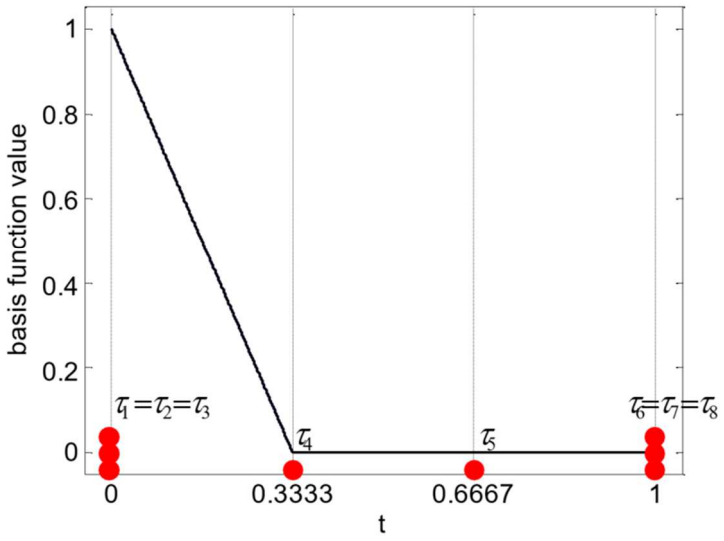


Figure C.6. This plot shows the second B-spline basis function of order 2 (i.e., degree 1) in black. The red dots are the knots.

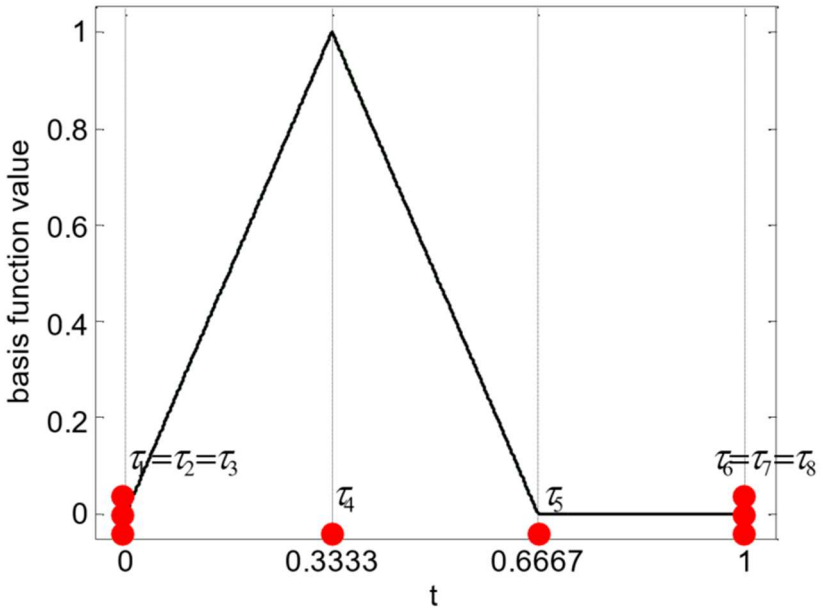


Figure C.7. This plot shows the third B-spline basis function of order 2 (i.e., degree 1) in black. The red dots are the knots.

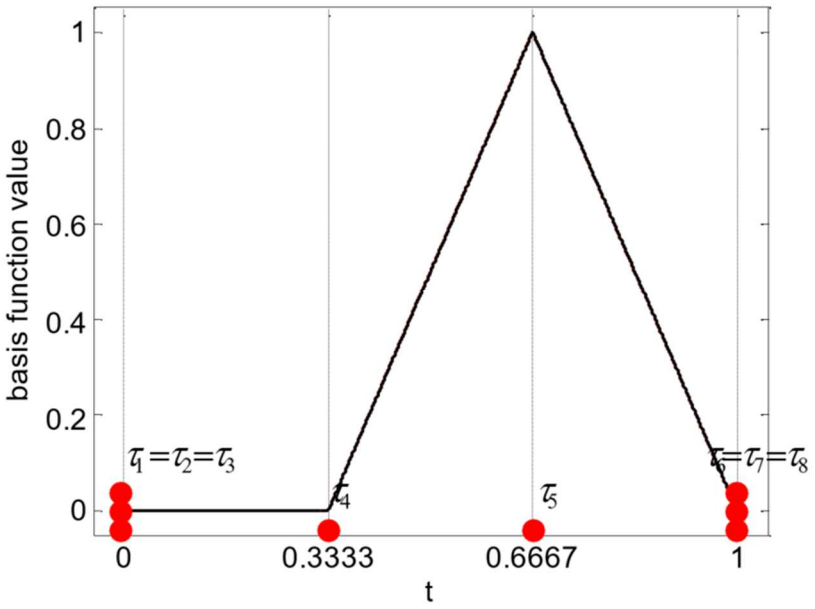


Figure C.8. This plot shows the fourth B-spline basis function of order 2 (i.e., degree 1) in black. The red dots are the knots.

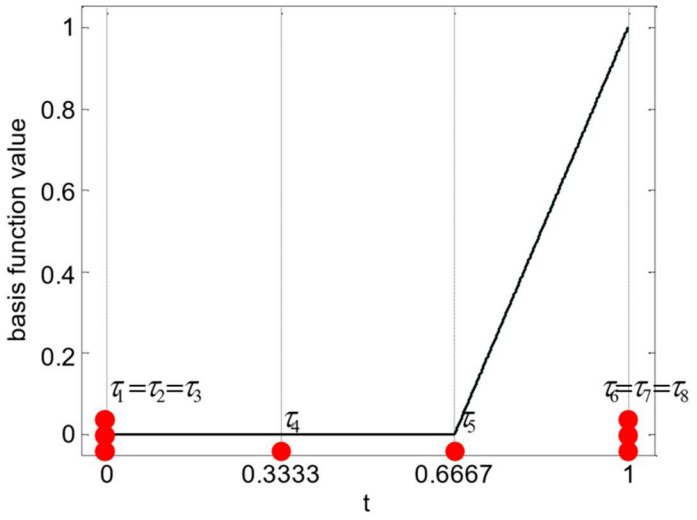


Figure C.9. This plot shows the fifth B-spline basis function of order 2 (i.e., degree 1) in black. The red dots are the knots.

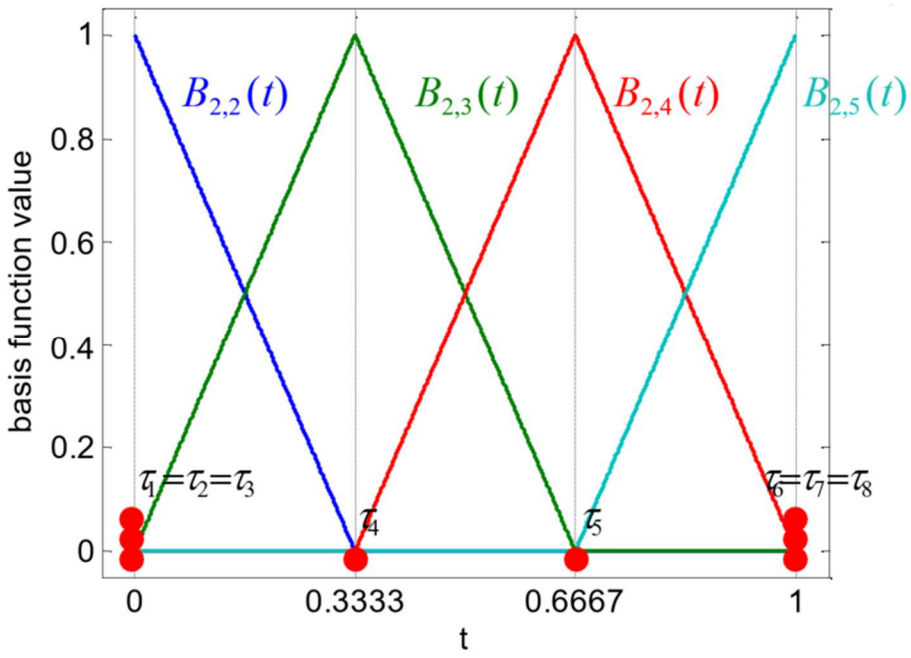


Figure C.10. This plot shows the order 2 (i.e., degree 1) B-spline basis. The second B-spline basis function of order 2 (i.e., degree 1) is in blue, the third is in green, the fourth is in red, and the fifth is in teal. The first, sixth, and seventh all take on values of zero and are not plotted. The red dots are the knots.

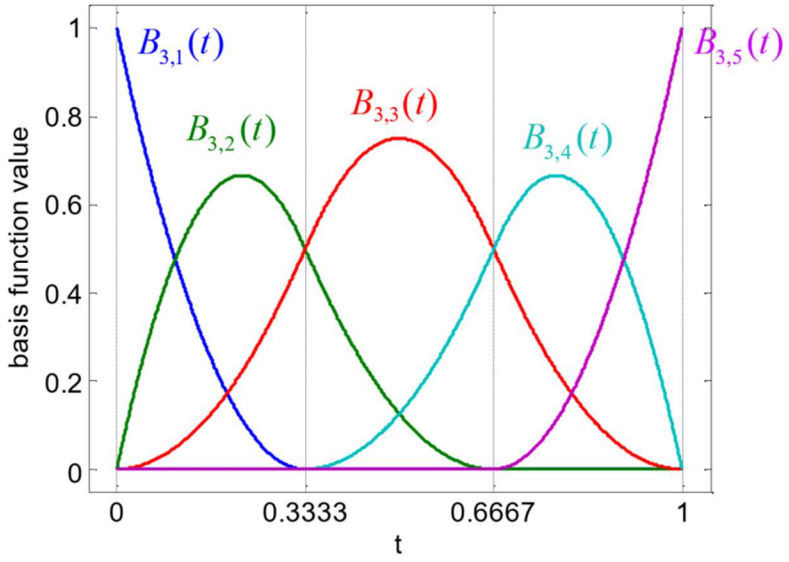


Figure C.11. This plot shows the order 3 (i.e., degree 2) B-spline basis. The first B-spline basis function of order 3 (i.e., degree 2) is in blue, the second is in green, the third is in red, the fourth is in teal, and the fifth is in purple. The sixth B-spline basis function only takes on values of zero and is not plotted.

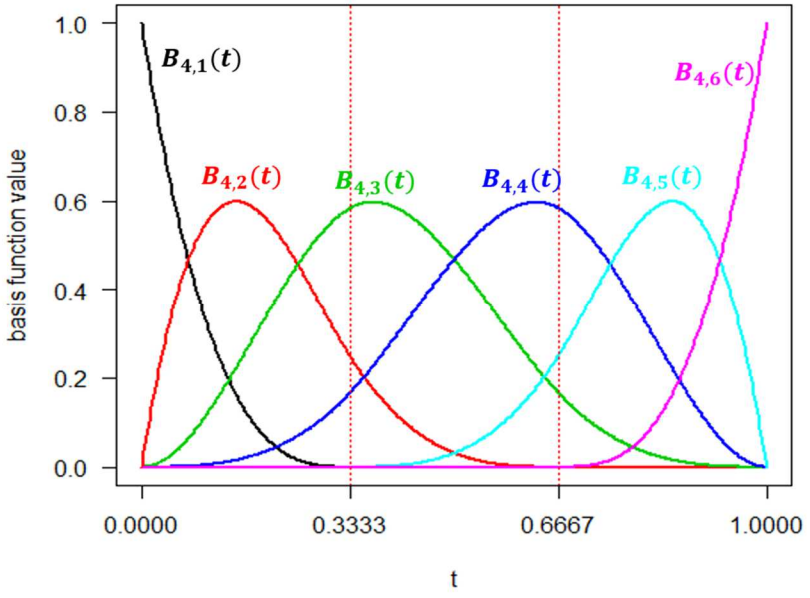


Figure C.12. This plot shows the order 4 (cubic; i.e., degree 3) B-spline basis. The first B-spline basis function of order 4 (cubic; i.e., degree 3) is in black, the second is in red, the third is in green, the fourth is in blue, the fifth is in teal, and the sixth is purple.

REFERENCES

Aguilera, A., & Aguilera-Morillo, M. (2013). Comparative study of different B-spline approaches for functional data. *Mathematical and Computer Modelling*, 58(7-8), 1568-1579. doi:10.1016/j.mcm.2013.04.007

Ruppert, D., Wand. M., Carroll, R. (2003). *Semiparametric Regression*. Cambridge, UK: Cambridge University Press.

Suk, H.W. (2015) Lecture 5- B-spline basis functions [PowerPoint slides]. Retrieved from Arizona State University Analysis of functional data Blackboard https://blackboard.asu.edu

APPENDIX D

TECHNICAL DETAILS FOR TRUNCATED POWER SPLINES

The truncated power basis is considered one of the simplest forms of splines. The truncated power basis with some set of knots $\tau_1 < \cdots < \tau_T$ is defined as

$$1, t, \ldots, t^k, (t-\tau_1)_+^k, \ldots, (t-\tau_T)_+^k, \quad \text{(D.1)}$$

where $[w]_+$ is the positive part of a function such that $[w]_+ = \max(0, w)$ (Wu & Zhang, 2006). In other words, in order for $(t-\tau_i)_+^k$ to be used to create the spline, the value of $t - \tau_i > 0$. Meaning the time point (t) must be larger than the knot point (τ_i). The phrase truncated power comes from these functions being truncated (i.e., forced to zero) to the left of the knot. k is the degree of the polynomial used within each subinterval. Typically, k is set to three for a cubic fit. The first $k+1$ basis functions of the truncated power basis (Equation D.1) are polynomials up to degree k and the rest are the truncated power functions of degree k (Wu & Zhang, 2006). The truncated power basis has continuous derivatives up to $k-1$; the cubic truncated power basis has continuous first and second derivatives. These basis functions are then multiplied by estimated basis coefficients to form the estimated curve.

The number of basis functions in the truncated power basis is determined by adding the number of knots (T) to the degree of the polynomial (k) plus one: $\mathrm{T} + (k+1)$. For example, let's say we want to construct a cubic truncated power basis with two knots $\tau_1 = 1$ and $\tau_2 = 2$. This means the basis would contain the following six (e,g., $\mathrm{T} + (\mathrm{k}+1) = 2 + (3+1) = 6$) basis functions: $1, t, t^2, t^3, (t-\tau_1)_+^3, (t-\tau_2)_+^3$. In order to construct some curve, x(t), the basis functions $(1, t, t^2, t^3)$ are used when $t < \tau_1$, when $\tau_1 \leq t < \tau_2$, $(1, t, t^2, t^3, (t-\tau_1)_+^3)$ are the basis functions used, and when $t \geq \tau_2$ $(1, t, t^2, t^3, (t-\tau_1)_+^3, (t-\tau_2)_+^3)$ are used. Figure D.1 shows the six basis functions used to define this truncated power basis with knots at $\tau_1 = 1$ and $\tau_2 = 2$. Time was arbitrarily

set to 100 equally spaced values between $[0, 3]$ in order to show smooth lines for the basis functions. The first truncated power basis function was left off as it is used for defining an intercept and only takes on a value of one. The second truncated power basis function (t) is plotted in red. The third truncated power basis function (t^2) is plotted in green. The fourth truncated power basis function (t^3) is plotted in blue. These three basis functions span the entire range of time. The fifth truncated power basis function (i.e., $(t - \tau_1)_+^3$) is plotted in purple. This basis function only takes on non-zero values after the first knot ($\tau_1 = 1$) represented by the vertical dashed black line at Time = 1. The sixth and final truncated power basis function (i.e., $(t - \tau_2)_+^3$) is plotted in yellow. This basis function only takes on non-zero values after the second knot ($\tau_2 = 2$) represented by the vertical dashed black line at Time = 2.

Just as with B-splines, the truncated power basis requires the user to specify the number and location of knots. Choice of knots can be specified using the same techniques discussed in the preceding section. Simplicity of construction and ease of parameter interpretation are two main advantages of using the truncated power basis. Two main disadvantages are associated with the truncated power basis. First, the basis matrix can be sparse since many (or even all) of the basis functions can be nonzero when evaluated at some time point. A sparse matrix can lead to slow computation times. Second, numerical precision problems can arise since the functions can grow without bound as time increases.

FIGURES

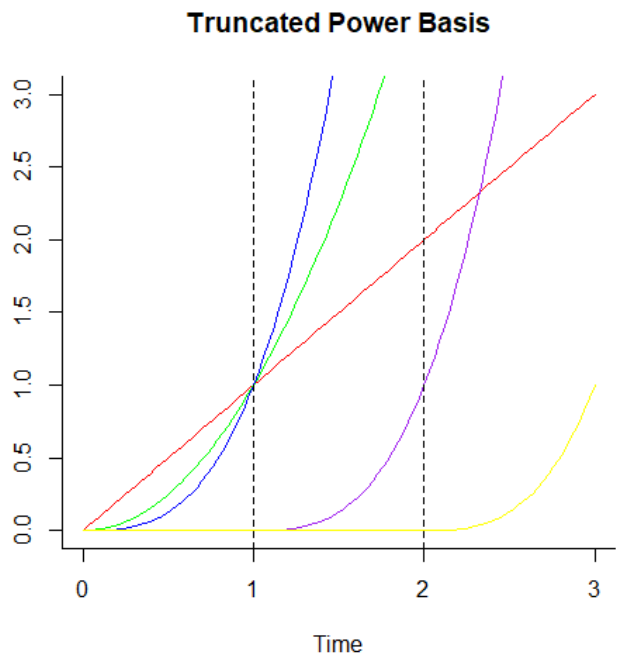


Figure D.1. This plot shows the six basis functions used to define this truncated power basis with knots at $\tau_1 = 1$ and $\tau_2 = 2$. The first truncated power basis function was left off as it is used for defining an intercept and only takes on a value of 1. The second truncated power basis function (t) is plotted in red. The third truncated power basis function (t^2) is plotted in green. The fourth truncated power basis function (t^3) is plotted in blue. These three basis functions span the entire range of time. The fifth truncated power basis function ($(t - \tau_1)_+^3$) is plotted in purple. This basis function only takes on non-zero values after the first knot ($\tau_1 = 1$) represented by the vertical dashed black line at Time = 1. The sixth and final truncated power basis function ($(t - \tau_2)_+^3$) is plotted in yellow. This basis function only takes on non-zero values after the second knot ($\tau_2 = 2$) represented by the vertical dashed black line at Time = 2.

REFERENCES

Wu, H., & Zhang, J. (2006). *Nonparametric regression methods for longitudinal data analysis: Mixed-effects modeling approaches*. Hoboken, N.J.: Wiley-Interscience.