

Nonlinear Mixed-Effects Growth Models: A Tutorial Using 'saemix' in R

Peter Boedeker¹

[1] *Department of Curriculum, Instruction, and Foundational Studies, Boise State University, Boise, ID, USA.*

Methodology, 2021, Vol. 17(4), 250–270, <https://doi.org/10.5964/meth.7061>

Received: 2021-07-01 • **Accepted:** 2021-10-05 • **Published (VoR):** 2021-12-17

Corresponding Author: Peter Boedeker, Education Building Office 406, 2133 Cesar Chavez Lane, Boise, ID 83725, USA. E-mail: peterboedeker@boisestate.edu

Supplementary Materials: Data, Materials [see [Index of Supplementary Materials](#)]



Abstract

Modeling growth across repeated measures of individuals and evaluating predictors of growth can reveal developmental patterns and factors that affect those patterns. When growth follows a sigmoidal shape, the Logistic, Gompertz, and Richards nonlinear growth curves are plausible. These functions have parameters that specifically control the starting point, total growth, overall rate of change, and point of greatest growth. Variability in growth parameters across individuals can be explained by covariates in a mixed model framework. The purpose of this tutorial is to provide analysts a brief introduction to these growth curves and demonstrate their application. The 'saemix' package in R is used to fit models to simulated data to answer specific research questions. Enough code is provided in-text to describe how to execute the analyses with the complete code and data provided in Supplementary Materials.

Keywords

nonlinear growth, Gompertz, tutorial, R

Modeling growth trajectories is often of interest in the behavioral and social sciences. Nonlinear mixed effects growth models (NLMEGMs) allow for variation in the rate of growth for an individual across time, such that an individual's growth may be characterized by periods of slower growth and periods of accelerated growth. While accurate modeling of the trajectory of within-individual growth is critical, answering questions about why individuals' growth patterns vary from one another can provide insight into the characteristics associated with tempered or delayed development. NLMEGMs in a mixed model (also, multilevel or hierarchical modeling, [Boedeker, 2017](#); [Raudenbush &](#)



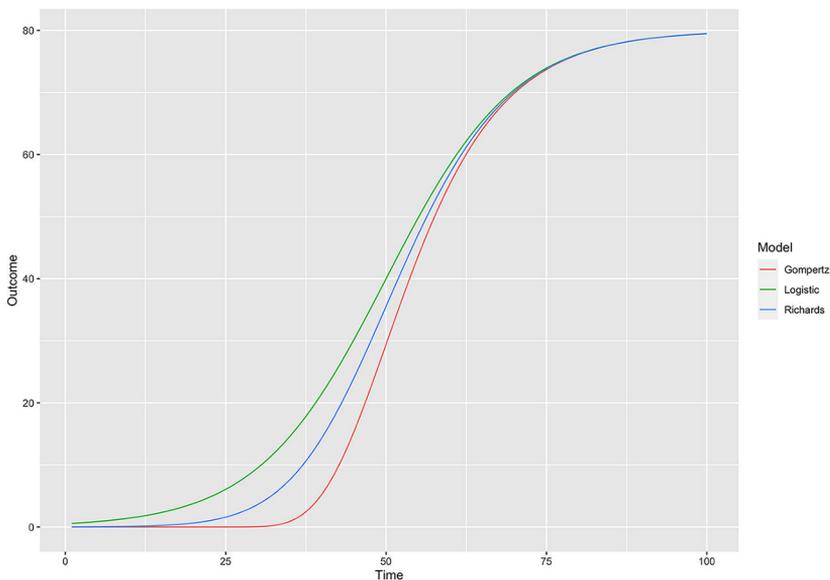
Bryk, 2002) framework allow for variability in the parameters of the nonlinear model to be explained by individual characteristics. Given the applicability of NLMEGMs to the study of development, we provide a tutorial on how to fit nonlinear growth models in R (R Core Team, 2021). We briefly describe three nonlinear growth curves (Logistic, Gompertz, Richards), provide an overview of the `saemix` (Comets et al., 2017; Version 2.4) package that can be used for fitting NLMEGMs, and demonstrate the estimation of NLMEGMs to answer specific research questions in the context of educational research.

Nonlinear Mixed Effects Growth Models

Nonlinear mixed effects growth models allow for the modeling of non-linear within-person change and between person differences in change (Grimm & Ram, 2009). A common nonlinear trajectory is S shaped (sigmoidal) and allows for a point of inflection at which rapid early growth reaches its highest rate of change and after which the rate of growth decreases, ultimately coming to a plateau. NLMEGMs allow the analyst to model growth, estimating parameters that have substantively meaningful interpretations, and provide opportunity for the evaluation of characteristics that may differentiate growth patterns across people. Figure 1 provides a graphical illustration of the Logistic, Gompertz, and Richards curves.

Figure 1

Gompertz, Logistic, and Richards Curves



Alternative methods of fitting curvilinear growth exist besides the Logistic, Gompertz, and Richards curves. Including polynomial terms (e.g., quadratic, cubic) for time can accomplish this. However, the parameters of the Logistic, Gompertz, and Richards curves allow for a more nuanced understanding of individual growth and the factors that may be associated with individual differences in growth. The researcher can ask substantively meaningful questions about specific aspects of growth that may not be possible with use of polynomial terms. Therefore, we focus on the Logistic, Gompertz, and Richards curves, describing each in terms of growth in student academic achievement.

Multilevel Framework

Within the multilevel framework, growth parameters can be estimated to have fixed and random effects. If a growth parameter is constrained to be equal across all individuals, then the fixed effect is the estimated value of this single parameter for all individuals. If a growth parameter is allowed to vary across individuals, then a fixed and random effect are estimated for the growth parameter. The fixed effect is essentially the weighted average of the individual parameter estimates and the random effect contains a variance that estimates individual variability about the fixed effect. Variability in a parameter can then be explained by adding predictors to the model for each parameter.

Latent growth modeling is an alternative framework for estimating non-linear growth. Each framework has its benefits and challenges (Ghisletta & Lindenberger, 2004; Grimm & Ram, 2009). For example, the latent growth modeling framework provides fit indices and the ability to model multiple dependent variables simultaneously; however, the statistical programs required for doing so are costly and certain constraints on the data, such as equally spaced time intervals of measurement, are required.

The focus of the present paper is within the multilevel modeling framework given 1) modeling longitudinal growth and evaluating predictors of this growth can be done in a straightforward manner in the multilevel modeling framework and 2) open-source programs exist for doing so.

Logistic

The Logistic growth curve is defined as

$$Y_{it} = LwrAsy_i + \frac{TtlGrwth_i}{1 + e^{-Apprch_i(t - Timing_i)}} + u_{it}, \quad u_{it} \sim N(0, \sigma^2) \quad (1)$$

where $LwrAsy_i$ is the lower asymptote for student i , $TtlGrwth_i$ is the total change for student i over the period of time, $Apprch_i$ is the rate of approach to the upper asymptote at the point of inflection for student i , $Timing_i$ is the time that the rate of change is greatest (i.e., the point of inflection), and u_{it} is a residual for student i at time t . The Logistic curve is symmetric about the point of inflection. Each of these parameters is

substantively meaningful. The *LwrAsy* is the starting point of an individual, indicating where the student begins in achievement. Larger values of *TtlGrwth* indicate that a student has a greater total growth over time. The rate of approach parameter represents the rate of change across all time points, characterizing growth overall where larger values indicate quicker growth, once started, to the upper limit. Larger values of the timing parameter indicate that accelerated growth toward the upper limit occurs later in time.

Note that the parameters are allowed to vary across individuals (they each have an *i* subscript) and therefore have random effects with variances that are estimated. In models, this is specified as

$$\begin{aligned}
 TtlGrwth_i &= \Upsilon_{TtlGrwth} + r_{TtlGrwth,i} & r_{TtlGrwth,i} &\sim N(0, \tau_{TtlGrwth}^2) \\
 Apprch_i &= \Upsilon_{Apprch} + r_{Apprch,i} & r_{Apprch,i} &\sim N(0, \tau_{Apprch}^2) \\
 Timing_i &= \Upsilon_{Timing} + r_{Timing,i} & r_{Timing,i} &\sim N(0, \tau_{Timing}^2) \\
 LwrAsy_i &= \Upsilon_{LwrAsy} + r_{LwrAsy,i} & r_{LwrAsy,i} &\sim N(0, \tau_{LwrAsy}^2)
 \end{aligned} \tag{2}$$

Various modifications are possible. The researcher can evaluate predictors of each growth parameter. To do so, a predictor is added, for example, in the equation with *TtlGrwth* as outcome. A gamma coefficient accompanies the additional predictor and is evaluated. The interpretation of such a coefficient is demonstrated in a later example. The correlation between varying parameters can be constrained to zero or be estimated as an additional parameter. For example, constraining the relationship between *TtlGrwth* and *Timing* to zero implies that the total change of an individual (*TtlGrwth*) is unrelated to the point of inflection (*Timing*) of that individual's growth. If this relationship is instead freely estimated then the analyst may find, for example, that students who have greater overall growth reach a point of inflection earlier in time. Also, any of the parameters could be constrained to be equal across all cases. For example, if *LwrAsy_i* were assumed to be equal across all individuals then the *i* subscript would be dropped and no random effect (*r* term) would be estimated. The decision to constrain a parameter to equality across individuals would need to be defended by the researcher.

Gompertz

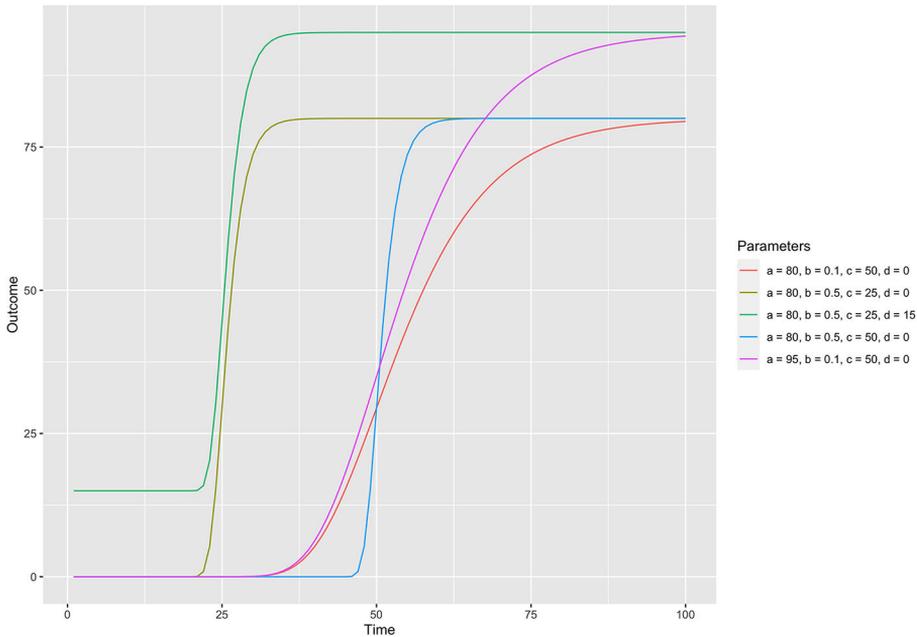
The Gompertz curve allows for asymmetric growth before and after the point of inflection. In an educational context, this would allow for a sharp increase as students enter school and then a more tempered approach to a maximum. The Gompertz function is

$$Y_{it} = LwrAsy_i + TtlGrwth_i e^{-Apprch_i(t - Timing_i)} + u_{it}, \quad u_{it} \sim N(0, \sigma^2) \tag{3}$$

where parameters are as defined above. The parameters, as described for the Logistic curve, can be specified with random effects. These equations are identical to the Logistic specification and therefore are not reiterated here. Figure 2 illustrates the Gompertz curve with different parameter specifications.

Figure 2

The Gompertz Curve is Plotted With Different Values for $a = TtlGrwth$, $b = Apprch$, $c = Timing$, and $d = LwrAsy$



Richards

The Richards curve is more complex than either the Logistic or the Gompertz curve because the asymmetry is controlled by an additional parameter in the model. The functional form of the Richards curve is

$$Y_{it} = LwrAsy_i + \frac{TtlGrwth_i}{\left(1 + s_i \cdot e^{-Apprch_i(t - Timing_i)}\right)^{\frac{1}{s_i}}} + u_{it}, \quad u_{it} \sim N(0, \sigma^2) \tag{4}$$

where parameters are as previously defined and s is the symmetry parameter. The s parameter controls the amount of asymmetry in the slope of the curve before and after the point of inflection. A s equal to 1 simplifies to the Logistic, such that half of the change occurs prior to the point of inflection and half occurs after. A value of s less

than 1 indicates that less than half of the total change occurs prior to the point of inflection, approaching a Gompertz curve as s approaches 0. The s parameter greater than 1 indicates that more than half of the total change occurs prior to the point of inflection. As with the Logistic and Gompertz curves the growth model parameters can be allowed to vary across individuals with specification of random effects.

Applications in Education

[Cameron et al. \(2015\)](#) compared different growth trajectories of students in mathematics and reading achievement using large-scale educational datasets. Linear, quadratic, cubic, unstructured, Logistic, Gompertz, and Richards specifications were fit to the data and compared. Ultimately, the Gompertz model provided the best fit to the observed growth pattern. Using the Gompertz function to model growth, variability in three parameters of the model were then explained by sociodemographic characteristics of students. [Anthony and Ogg \(2020\)](#) extended this work by modeling science achievement growth with the Gompertz function and explaining variability in its parameters with components of executive functioning and learning related behaviors. These applications highlight the relevance of nonlinear growth modeling to estimating trajectories of growth in education.

Fitting NLMEGMs

Software

There are several programs for fitting NLMEGMs. [Stegman et al. \(2018\)](#) review four R packages including `nlme` ([Pinheiro et al., 2017](#)), `lme4` ([Bates et al., 2015](#)), `saemix`, and `brms` ([Buerkner, 2016](#)). [Comets et al. \(2017\)](#) made additional comparisons of `nlme` and `lme4` to `saemix`. Generally, `saemix` was found to converge to a solution more frequently than `nlme` and `lmer` ([Comets et al, 2017](#)) but may require longer processing time ([Stegman et al., 2018](#)). The `brms` package utilizes a Bayesian framework that allows incorporation of prior knowledge; however, such information may not be readily available for fitting NLMEGMs. NLMEGMs can also be fit using Mplus and SAS, as demonstrated by [Grimm and Ram \(2009\)](#). The `saemix` package was selected for demonstration because 1) R is free and open source, providing up-to-date advancements without financial burden and 2) `saemix` was found to perform better regarding convergence than rival R packages.

saemix

The stochastic approximation expectation maximization algorithm uses Markov Chain Monte Carlo processes to determine the likelihood and the model parameters that maximize it in an iterative fashion (see [Comets et al., 2017](#) for details). As a result,

model convergence can be evaluated using plots. The algorithm functions in such a way that the possible parameter space is explored to a point after which the iterations quickly converge to a final solution. Plots of these iterations can be used to evaluate convergence.

Comets et al. (2017) provide detailed documentation for conducting an analysis using `saemix`. We provide a brief overview here and, in our examples, explain details of how each specification changes to accommodate answering each research question. Prior to running an analysis, the package requires one function and two objects be specified: 1) the nonlinear growth function, 2) an `saemix` data object, and 3) an `saemix` model object. The nonlinear growth function is the mathematical formula of the nonlinear growth trajectory with parameters to be estimated. Once this object is specified, there is no need to modify it unless the nonlinear growth function is changed (e.g., from Gompertz to Logistic). The program requires a `saemix` data object rather than a standard data frame. A standard data frame contains variables of different scales as would be needed in modeling applications; the `saemix` data object in addition contains information specific for applications using `saemix`, such as designating which variable is the outcome, which are covariates, and what variable indicates the individual for whom repeated measures are available (an ID variable). The `saemix` data object should be oriented in long-format, with repeated measures for the same individual recorded in subsequent rows. After specifying the `saemix` data object, the `saemix` model object contains information concerning starting values, which of the growth parameters are to be explained by covariates, and whether variances and covariances of the growth parameters are to be estimated.

Starting Values

The starting values selected can have an impact on model estimates (Grimm & Ram, 2009). Good starting values provide the iterative estimation procedure with values that are close to the true value, enabling the algorithm to arrive at an accurate solution more quickly. One proposed method for determining starting values is to first fit a simplified version of the final model. The estimated values of growth parameters from the simpler model could then be used as starting values in the more complex model. Another possibility is to simply plot the data and, with knowledge of each parameter's meaning, make an educated guess of the value for each parameter.

Model Comparisons

Like other mixed models fit with maximum likelihood, model comparisons can be made using Akaike information criterion (AIC; Akaike, 1973), Bayesian information criterion (BIC; Schwarz, 1978), and a nested models likelihood ratio test (LRT). AIC and BIC are penalized likelihood values, such that the likelihood is increased for each additional

parameter estimated in the model. The model with the lower AIC and BIC is deemed the better fitting of two rival models. The LRT is used to evaluate the increase in fit when models are nested, producing a p-value, the null hypothesis of which is equal fit. Such nested models occur when covariates have been added to a model and the goal is to evaluate if the addition of covariates yields statistically significant improvement in model fit. Rejecting the null hypothesis indicates that the more complex model is preferred.

Diagnostics

The normalized prediction distribution error (NPDE) was developed for assessing non-linear mixed models. Comets et al. (2008) describe in detail the procedure for calculating NPDE but we provide here some guidance for evaluating the distribution of NPDE. If the model fits the data well, then the distribution of NPDE will be standard normal. Plots standard in evaluating ordinary least squares (OLS) residuals can be used, including Q-Q plots to evaluate normality of the NPDE and plots of predicted values versus NPDE values. Criteria are generally the same as for OLS regression; plots should indicate that the NPDE are normally distributed homogeneously across predicted values.

Example

We demonstrate the use of `saemix` for fitting NLMEGMs to a simulated dataset with 100 students, each with six measures of math achievement and two covariates. The two covariates are socioeconomic status (SES), a standardized measure with $M = 0$ and $SD = 1$, characterizing the economic resources of the student's family, and Sex, a dichotomous indicator where 0 = male and 1 = female, both of which are measured only at the first time point. For SES, higher scores indicate higher SES. Ultimately, the Gompertz model is retained for analyses; therefore, we provide in text the code for fitting the Gompertz model. Online [Supplementary Materials](#) contain full code and data. We constrain the `LwrAsy` parameter to be equivalent across individuals to demonstrate how the code differs for parameters that are allowed to vary and those that are not. Using the fictitious data, we answer four research questions:

RQ1: Which of the Logistic, Gompertz, or Richards curves models best the growth in achievement?

RQ2: Does Sex have an association with the total growth, rate of approach to the upper asymptote, or point of inflection of the curve found optimal in answering the first research question?

RQ3: Does adding SES to the model in addition to Sex as a predictor of total growth, rate of approach to the upper asymptote, or point of inflection improve model fit? If so, how do SES and Sex relate to achievement growth?

RQ4: Does Sex moderate the association between SES and total growth, rate of approach to the upper asymptote, or point of inflection?

For RQ1, we compare the fit of the three unconditional NLMEGMS, use the NPDE to evaluate Gompertz model fit, interpret model parameters, and display relevant output and plots. For RQ2, a conditional model is fit wherein the growth parameters allowed to vary are predicted by the dichotomous Sex variable. To answer RQ3, SES is added as a predictor. RQ4 requires evaluation of an interaction between Sex and SES; this is done by multiplying the SES and Sex of each child and adding this product as a third variable in the model. Model comparisons are needed throughout to evaluate model improvement; therefore, likelihood ratio tests are conducted and comparisons using AIC and BIC are made. There are limitations to our demonstration. First, the data are simulated; therefore, any data cleaning that would need to occur with actual data is not conducted. Second, diagnostics of model fit should be conducted at each step though we only demonstrate this with the unconditional model.

Fitting a NLMEGM Using saemix

The Gompertz growth model is ultimately retained; therefore, the in-text example will demonstrate specification of the Gompertz growth model. Initial specifications of the Logistic and Richards functions needed for answering the first research question are available in [Supplementary Materials](#). The Gompertz function is specified with the lower asymptote constrained to be equal across cases and the remaining three parameters allowed to randomly vary. As mentioned previously, constraining the lower asymptote to be equal across cases is for demonstration purposes. We do not universally propose constraining the lower asymptote to be equal across cases; rather, such a decision should be made and defended by the researcher. Additionally, we do not specify any predictor of the lower asymptote throughout; this is similarly done for demonstration purposes. The following function sets up the Gompertz model.

```
gompertz.model <- function(psi,id,x) {
  t <- x[,1]
  TtlGrwth <- psi[id,1]
  Apprch <- psi[id,2]
  Timing <- psi[id,3]
  LwrAsy <- psi[id,4]
  ypred <- LwrAsy+TtlGrwth*exp(-exp(-Apprch*(t-Timing)))
  return(ypred)
}
```

The `psi`, `id`, and `x` components of the function are derived from the model and data objects that follow for each analysis. Note that the `ypred` line is the formula for the four

parameter Gompertz function, as previously defined. The following options are specified as a list:

```
NLMEGM.options <- list(seed=1234, displayProgress=FALSE)
```

Setting the seed allows for reproducibility. The estimation of model parameters is an iterative process; to see the iterations graphically and check convergence of the solution, `displayProgress = TRUE` is specified. We set it to false here to simplify the discussion of output but provide a plot of the result if set to TRUE for the unconditional Gompertz model.

RQ1: Comparing Gompertz, Logistic and Richards Curves

To answer RQ1 we must fit unconditional models using the data in the data file title `NLMEGMExData`. To fit the unconditional growth model, we specify the following data object:

```
GompertzData.RQ1 <- saemixData(name.data = NLMEGMExData, header = TRUE,
name.group = c("ID"), name.predictors = c("time"), name.response = c("Achievement"),
name.X = "time")
```

We specify the data file and that it contains variable names (`header = TRUE`). There are multiple observations of achievement for each student; therefore, we specify the variable used to group observations into students as `name.group = c("ID")`. The predictor (`time`) and outcome (`Achievement`) are provided next. The final entry `name.x = "time"` specifies the variable on the x-axis for diagnostic plots.

After creating the data object, the model object is specified:

```
GompertzModel.RQ1 <- saemixModel(model=gompertz.model,
description = 'Gompertz', psi0=c(TtlGrwth=0,Apprch=0,Timing=0,LwrAsy=0),
covariance.model = matrix(c(1,1,1,0,1,1,1,0,1,1,1,0,0,0,0,0), ncol = 4, byrow = TRUE),
transform.par=c(0,0,0,0))
```

The model uses the previously created `gompertz.model` function. The `psi0` vector provides starting values for each of the four growth parameters. Starting values of 0 were used and found to be adequate. As previously mentioned, other starting values could be used and were required for fitting the Richards curve. In the `covariance.model` matrix, a 1 on the main diagonal indicates that the variance for that parameter should be estimated and a 1 in the off diagonal entries indicates that the covariance between the two should be estimated. A 0 in either spot indicates the value is fixed to zero. Because there are four parameters, this is a 4X4 matrix. The `LwrAsy` parameter is constrained to equality across individuals; therefore, a variance and any covariance that would include `LwrAsy` is not estimated. The `transform.par` code allows the user to change the

distribution of each parameter. A 0 indicates a normal distribution but other options are log-normal, probit, and logit.

To conduct the analysis:

```
GompertzFit.RQ1 <- saemix(GompertzModel.RQ1,GompertzData.RQ1, NLMEGM.options)
```

Abbreviated output is provided below. Note that the section `Fixed effects` provides the point estimates and standard errors for each of the four parameters as well as the within student residual standard deviation (`a.`). The `CV (%)` column is the coefficient of variation and is equal to $100 \times (\text{SE}/\text{Estimate})$. The section `Variance of random effects` provides estimates of variances and covariances. Recall, the `LwrAsy` was constrained such that its variance was zero and therefore is not reported. Correlations between parameters that were allowed to vary are reported in the section `Correlation matrix of random effects`. Finally, the estimated $-2 \times \log$ -likelihood, AIC, and BIC values are displayed.

```
-----
---- Results ----
Fixed effects
Parameter Estimate SE CV(%)
TtlGrwth 5.220 0.055 1.1
Apprch 1.760 0.072 4.1
Timing 1.987 0.020 1.0
LwrAsy -0.066 0.015 22.7
a. 0.192 0.008 4.0

Variance of random effects
Parameter Estimate SE CV(%)
omega2.TtlGrwth 0.236 0.039 16
omega2.Apprch 0.405 0.072 18
omega2.Timing 0.034 0.006 16
cov.TtlGrwth.Apprch 0.234 0.043 18
cov.TtlGrwth.Timing -0.012 0.010 -87
cov.Apprch.Timing 0.002 0.014 648

Statistical criteria

Correlation matrix of random effects
omega2.TtlGrwth omega2.Apprch omega2.Timing
omega2.TtlGrwth 1.0000 0.756 -0.133
omega2.Apprch 0.756 1.0000 0.019
omega2.Timing -0.133 0.019 1.0000

Likelihood computed by linearisation
-2LL= 359.088
AIC= 381.088
BIC= 409.745
```

Model Comparisons

Recall, to answer RQ1 the results for the Logistic, Gompertz, and Richards growth models are compared. Information criterion values are useful for this purpose, and in [Table 1](#) the $-2 \times \log$ -likelihood, AIC, and BIC values are shown for all three NLMEGM specifications. Looking within each column of [Table 1](#), the Gompertz specification is superior to the Logistic and Richards specifications. [Figure 3](#) shows the observed data overlaid by each of the fitted growth models. Notably, the Gompertz and Richards specifications have nearly identical curves; however, the Gompertz curve has one fewer parameter in its model. Given the visual fit of the Gompertz model and the values in [Table 1](#), the Gompertz model is retained. Further investigation of the diagnostic fit of the Gompertz model to the data is possible by evaluating model convergence and NPDE.

Table 1

Model Comparison Results

Model	-2LL	AIC	BIC
Logistic	417.525	439.525	468.182
Gompertz	359.088	381.088	409.745
Richards	359.141	383.141	414.403

Note. -2LL = $-2 \times \text{Log-Likelihood}$; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

Model Convergence

Each parameter estimate is derived iteratively; the plot of these iterations is used to evaluate convergence. See [Figure 4](#) for this application with the Gompertz growth model. There is a plot for each of the estimated growth parameters, the growth parameter variances, and the within student residual variance. A solution that converges will yield plots in which the parameter space is initially explored (the estimated values can bounce around substantially) but produces in the later portion of the plot a stable estimate. In this application, the solution converges well.

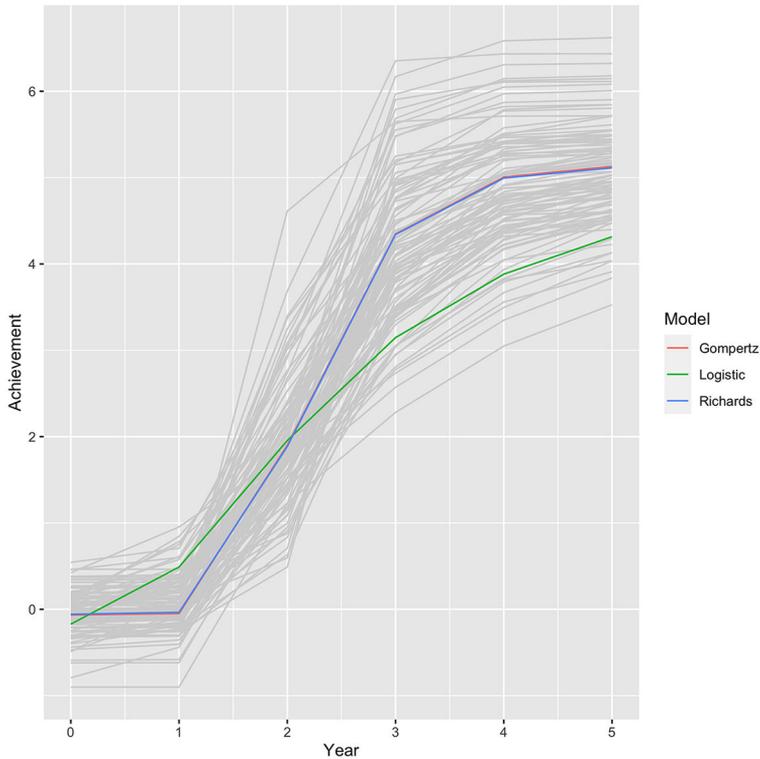
Diagnostics

Several diagnostic plots of model fit are available. First, simple plots of individual predictions overlaid on the observed values provides an indication of model fit. Such plots are presented in [Figure 5](#).

Four plots that utilize the NPDE are evaluated (see [Figure 6](#)). Recall, NPDE that is normally distributed with a mean of 0 and variance of 1 indicates good fit (see the Q-Q plot and histogram of NPDE values in the upper left and right of [Figure 6](#)). Additionally, the distribution of NPDE values across time (lower left plot of [Figure 6](#)) and across the

Figure 3

Overlay of Gompertz, Logistic, and Richards Curves on the Simulated Data



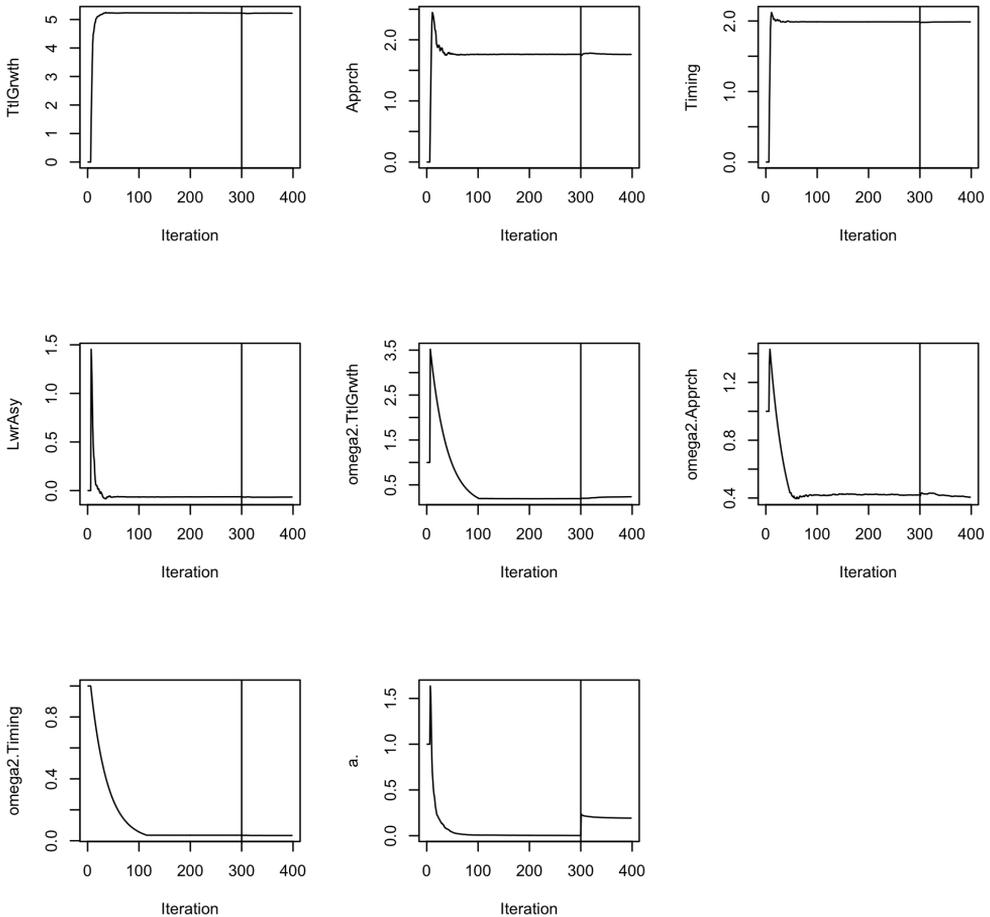
predicted values (lower right plot of Figure 6) should be normal with a mean of zero and variance of one to indicate good model fit to the data.

Interpretation of Parameter Estimates

Though not necessary for answering the first research question, we review the parameter estimates in the Fixed effects section. The Total Growth parameter is estimated to be 5.220, indicating approximately 5 units of average total growth in math achievement over the 6 years. The Rate of Approach parameter is estimated to be 1.760. This estimate is an aggregate across all time points and therefore provides a holistic description of the rate of growth. The Timing parameter is the time of most rapid change, on average, for all participants. Here, the estimate of Timing was 1.987. Year 0 was the first observed value of the outcome, therefore students on average reached their greatest acceleration between the second and third achievement measurements. The

Figure 4

Evaluation of Convergence



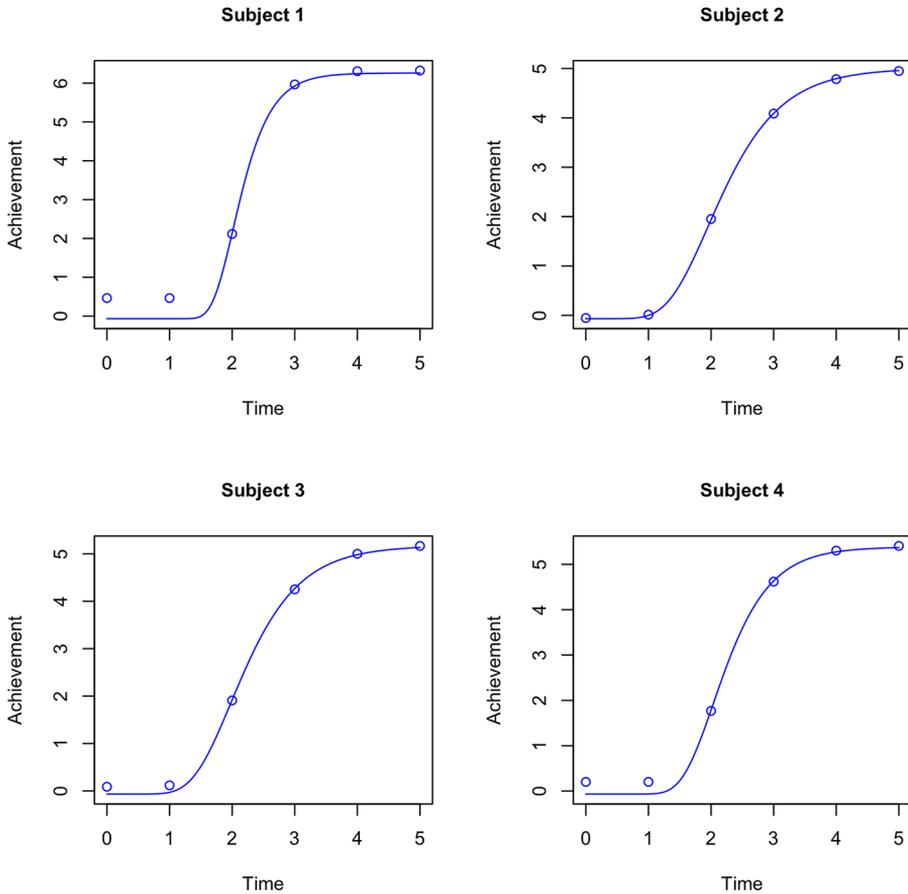
Lower Asymptote was constrained to be equal and estimated to be -0.066, indicating that the initial achievement measure was estimated to be -0.066 units for all students.

The Variance of random effects section provides the estimates of growth parameter variances across students. Larger variances indicate that differences in growth parameters exist across students that may be explained by student characteristics. Smaller variances indicate that students generally have similar estimated parameter values.

Three growth parameters were allowed to vary and correlate with one another. The Correlation matrix of random effects section can provide interesting

Figure 5

Gompertz Curve Trajectory Laid Over 4 Cases

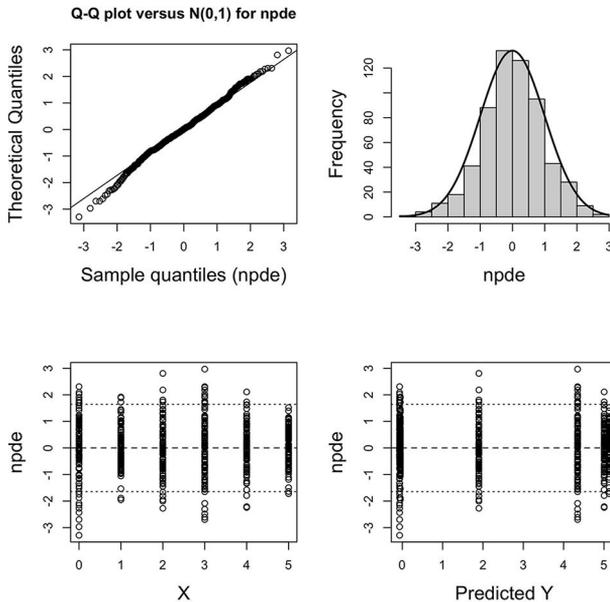


insights into the relationships between these parameters. For example, if the Total Growth parameter is negatively correlated with Timing and positively correlated with Rate of Approach, then students who grow the most tend to grow at a faster rate and change earlier than their peers. This was found in [Cameron et al. \(2015\)](#) when evaluating growth in reading achievement. In our fictitious example, such a relationship is not strongly supported because the correlation between Total Growth and Timing is only -0.133.

The remaining three research questions require variations of the code presented above. For each research question, the specific variation is provided in text with full code in [Supplementary Materials](#).

Figure 6

Plot for Evaluating the Fit of the Unconditional Gompertz Curve to the Simulated Data



RQ2: Conditional Growth Model (Sex Only)

Altering Code

The `saemixData` command now includes Sex as a covariate:

```
name.covariates = c("Sex").
```

To include Sex as a predictor of the Total Growth, Rate of Approach, and Timing parameters the following line to the `saemixModel` command is added:

```
covariate.model = matrix(c(1,1,1,0), ncol = 4, byrow = TRUE)
```

The above line of code indicates that the Sex covariate is a predictor of the `TtlGrwth`, `Apprch`, and `Timing` but not `LwrAsy`.

Results

Overall improvement in model fit after adding Sex was statistically significant ($\chi^2[3] = 32.416$, $p < .001$) and yielded improved AIC ($\Delta AIC = 25.058$) and BIC

($\Delta BIC = 17.242$). Sex was a statistically significant predictor of total growth (coef = 0.467, $SE = 0.095$, $p < .001$) and the rate of approach (coef = 0.697, $SE = 0.125$, $p < .001$); however, Sex was not a statistically significant predictor of the timing parameter (coef = 0.015, $SE = 0.041$, $p = .353$). These findings indicate that Females and Males do not differ statistically on their point of inflection but Females have greater total growth and are increasing at a faster rate at the point of inflection than Males.

RQ3: Conditional Growth Model (Sex and SES)

Altering Code

The `saemixData` command now includes Sex and SES as covariates:

```
name.covariates = c("Sex", "SES")
```

The command for creating the model object must be changed to include Sex as a predictor of the *TtlGrwth*, *Apprch*, and *Timing* parameters. This is done within the `saemixModel` command:

```
covariate.model = matrix(c(1,1,1,0,1,1,1,0), ncol = 4, byrow = TRUE)
```

where the first set of four values (1, 1, 1, 0) correspond to Sex as a predictor of the four parameters (*TtlGrwth*, *Apprch*, *Timing*, *LwrAsy*) and the second set of four values correspond to SES as a predictor of the four parameters.

Results

Overall improvement in model fit after adding SES to the model with Sex only as a predictor was statistically significant ($\chi^2[3] = 59.481$, $p < .001$) and yielded improved AIC ($\Delta AIC = 52.924$) and BIC ($\Delta BIC = 45.109$). The coefficient results for Sex are consistent with the previous model whereas SES is a statistically significant predictor of Total Growth and Rate of Approach but not Timing (see Table 2). Such a result indicates that students of higher SES tend to have greater total growth and faster growth, but do not differ in timing of the point of inflection when compared to lower SES peers.

Table 2

Predictors of Gompertz Model Parameters

Predictor	Total growth	Rate of approach	Timing
Sex	0.416 (0.079)***	0.676 (0.107)***	0.016 (0.040)
SES	0.253 (0.038)***	0.328 (0.053)***	-0.012 (0.020)

Note. SES = socioeconomic status.

*** $p < .001$.

RQ4: Conditional Growth Model (Moderation)

A test of moderation allows the investigator to determine if the relationship between a predictor and outcome is dependent on the value of a third variable. Moderation analyses can be used to determine for whom relationships hold or treatments are most effective. In growth modeling, moderation can be used to determine if, for example, the positive association between SES and the Total Growth parameter is equivalent for males and females or if for one sex the relationship is stronger. Next, the potential moderating role of Sex on the relationship between SES and Gompertz function parameters is investigated.

Altering Code

The `saemixData` command now includes Sex, SES, and the interaction of Sex and SES as covariates:

```
name.covariates = c("Sex", "SES", "SexSESmod")
```

The command for creating the model object must be changed to include the interaction of Sex and SES as a predictor of Total Growth, Rate of Approach, and Timing. This is done within the `saemixModel` command:

```
covariate.model = matrix(c(1,1,1,0,1,1,1,0,1,1,1,0), ncol = 4, byrow = TRUE)
```

Note that the first set of four values (1, 1, 1, 0) correspond to Sex as a predictor of the four parameters, the second set of four values correspond to SES as a predictor of the four parameters, and the third set of four values correspond to the interaction of Sex and SES as a predictor of the Gompertz model parameters. If the analyst only wanted to determine if Sex moderated the relationship between SES and Total Growth and not for the remaining two growth parameters, then the `covariate.model` command would be altered so that the third set of 1s and 0 is 1, 0, 0, 0, indicating that the moderator is evaluated only for the first growth parameter, Total Growth.

Results

Overall improvement in model fit after adding the interaction of Sex and SES to the previous model with Sex and SES as predictors was statistically significant ($\chi^2[3] = 22.232$, $p < .001$) and yielded improved AIC ($\Delta AIC = 15.77$ and BIC ($\Delta BIC = 7.955$). The interaction of Sex and SES is statistically significant for Total Growth and Rate of Approach but not for Timing (see Table 3). Such a result indicates that the relationship between SES and Total Growth is moderated by Sex, where the relationship between SES and Total Growth is stronger for females than for males. A similar result was found for the Rate of Approach. The relationship between SES and the Timing parameter does not differ statistically significantly for males and females.

Table 3*Moderation Results*

Predictor	Total growth	Rate of approach	Timing
Sex	0.414 (0.077)***	0.676 (0.100) ***	0.015 (0.041)
SES	0.114 (0.064)*	0.137 (0.065)*	-0.009 (0.033)
Sex x SES	0.216 (0.079)**	0.362 (0.097)***	-0.001 (0.041)

Note. SES = socioeconomic status.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Conclusions

NLMEGMS allow for the modeling of sigmoidal growth trajectories that resemble real-world phenomena. The parameters of NLMEGMS translate into easily understandable attributes of growth. Further extending the model to include predictors of these growth parameters can address specific hypotheses such as the existence of relationships between individual characteristics and growth or the effect of policy implementation on one or more attributes of growth. In this paper three non-linear growth trajectories were discussed and their accompanying parameters. This was followed by an example of how to use a specific package (`saemix`) to model non-linear growth trajectories and answer research questions. The goal was to provide a broad illustration of NLMEGMS and their application to an applied audience in a replicable manner; however, the presentation did not cover all aspects of non-linear growth nor all possible specifications within `saemix`.

An initial concern for the analyst when planning to model non-linear growth trajectories is the framework in which the modeling should be conducted. Non-linear growth models can be fit as multilevel models or as latent growth models. The multilevel modeling framework was demonstrated here because other demonstrations require use of costly programs (e.g., SAS in Grimm & Ram, 2009) and multilevel modeling is an accessible framework for testing multilevel hypotheses. Given limitations (Ghisletta & Lindenberger, 2004; Grimm & Ram, 2009) within each framework, the analyst must consider the hypotheses to be tested and the resources available to determine which framework to employ.

The demonstration utilized `saemix`, a freely available R package. Compared to other R packages (i.e., `nlmer`, `lme4`), `saemix` was found to have better convergence rates although may take longer to converge to a final solution (Comets et al., 2017; Stegman et al., 2018). Other for purchase software could be used to fit NLMEGMS, including SAS (see Grimm & Ram, 2009 for a demonstration using SAS). Though the example provided a demonstration of a broadly applicable set of analyses, the user is encouraged to understand well the program (`saemix` or otherwise) used in estimating NLMEGMS. Some specifications in `saemix` that were not discussed that could be altered were different

specifications of the error distribution and comparisons of alternative log-likelihoods. Some aspects of `saemix` are fixed and unalterable, such as the assumption that random effects are multivariate normally distributed.

The reader is encouraged to consider modeling non-linear growth trajectories to answer substantive questions about change over time. Extending beyond polynomial terms, sigmoidal trajectories may fit one's data more effectively while providing the opportunity to test interesting hypotheses about components of growth. This paper contributes to the NLMEGMs literature by providing an accessible introduction to the models with a reproducible example.

Funding: The authors have no funding to report.

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: The authors have declared that no competing interests exist.

Data Availability: The data for this article are freely available (see the [Supplementary Materials](#) section).

Supplementary Materials

For this article the following Supplementary Materials are available via PsychArchives repository (for access see [Index of Supplementary Materials](#) below):

- Dataset for replicating example analyses.
- Code for replicating all analyses.

Index of Supplementary Materials

Boedeker, P. (2021). *Supplementary materials to: Nonlinear mixed-effects growth models: A tutorial using 'saemix' in R* [Dataset and code]. PsychOpen GOLD. <https://doi.org/10.23668/psycharchives.5301>

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Proceedings of the second international symposium on information theory* (pp. 267-281). Akademiai Kiado.
- Anthony, C. J., & Ogg, J. (2020). Executive function, learning-related behaviors, and science growth from kindergarten to fourth grade. *Journal of Educational Psychology*, *112*(8), 1563-1581. <https://doi.org/10.1037/edu0000447>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. <https://doi.org/10.18637/jss.v067.i01>

- Boedeker, P. (2017). Hierarchical linear modeling with maximum likelihood, residual maximum likelihood, and fully Bayesian estimation. *Practical Assessment, Research, and Evaluation*, 22, 1-19. <https://doi.org/10.7275/5vvy-8613>
- Buerkner, P. C. (2016). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1-28. <https://doi.org/10.18637/jss.v080.i01>
- Cameron, C. E., Grimm, K. J., Steele, J. S., Castro-Shilo, L., & Grissmer, D. W. (2015). *Journal of Educational Psychology*, 107(3), 789-804. <https://doi.org/10.1037/edu0000009>
- Comets, E., Lavenu, A., & Lavielle, M. (2017). Parameter estimation in nonlinear mixed effects models using saemix, an R implementation of the SAEM algorithm. *Journal of Statistical Software*, 80(3), 1-41. <https://doi.org/10.18637/jss.v080.i03>
- Comets, E., Brendel, K., & Mentre, F. (2008). Computing normalized prediction distribution errors to evaluate mixed-effect models: The npde add-on package for R. *Computer Methods and Programs in Biomedicine*, 90(2), 154-166. <https://doi.org/10.1016/j.cmpb.2007.12.002>
- Ghisletta, P., & Lindenberger, U. (2004). Static and dynamic longitudinal structural analyses of cognitive changes in old age. *Gerontology*, 50(1), 12-16. <https://doi.org/10.1159/000074383>
- Grimm, K. J., & Ram, N. (2009). Non-linear growth models in Mplus and SAS. *Structural Equation Modeling A Multidisciplinary Journal*, 16(4), 676-701. <https://doi.org/10.1080/10705510903206055>
- Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D., & R Core Team (2017). nlme: Linear and nonlinear mixed effects models. Retrieved from <https://CRAN.R-project.org/package=nlme>
- R Core Team (2021). R: A language and environment for statistical computing. <https://www.R-project.org/>.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464. <https://doi.org/10.1214/aos/1176344136>
- Stegman, G., Jacobucci, R., Harring, J. R., & Grimm, K. J. (2018). Nonlinear mixed-effects modeling programs in R. *Structural Equation Modeling A Multidisciplinary Journal*, 25(1), 160-165. <https://doi.org/10.1080/10705511.2017.1396187>



Methodology is the official journal of the European Association of Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.