

4-28-2020

## Logistic Growth Modeling with Markov Chain Monte Carlo Estimation

Jaehwa Choi

*The George Washington University, jaechoi@gwu.edu*

Jinsong Chen

*Sun Yen-Sat University, jinsong.chen@live.com*

Jeffrey R. Harring

*University of Maryland, harring@umd.edu*



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

---

### Recommended Citation

Choi, J., Chen, J., & Harring, J. R. (2019). Logistic growth modeling with Markov chain Monte Carlo estimation. *Journal of Modern Applied Statistical Methods*, 18(1), eP2997. doi: 10.22237/jmasm/1556669820

# Logistic Growth Modeling with Markov Chain Monte Carlo Estimation

**Jaehwa Choi**

The George Washington University  
Washington, DC

**Jinsong Chen**

Sun Yen-Sat University  
Guangzhou, China

**Jeffery R. Harring**

University of Maryland  
College Park, MD

---

A new growth modeling approach is proposed to can fit inherently nonlinear (i.e., logistic) function without constraint nor reparameterization. A simulation study is employed to investigate the feasibility and performance of a Markov chain Monte Carlo method within Bayesian estimation framework to estimate a fully random version of a logistic growth curve model under manipulated conditions such as the number and timing of measurement occasions and sample sizes.

*Keywords:* Growth modeling, latent growth modeling, nonlinear growth models, logistic functions, Markov chain Monte Carlo, Bayesian inference

---

## Introduction

Repeated measures data are common in social and behavioral sciences, especially when research questions revolve around developments or learnings over time. The linear latent growth model (LGM) or growth model (GM; Bollen & Curran, 2006; Preacher et al., 2008) has been an indispensable method for understanding individual differences in such longitudinal developmental processes. Nonlinear GM (Blozis & Harring, 2015; Browne, 1993; Grimm & Ram, 2009; Grimm et al., 2011) extend the GM framework for linear processes to nonlinear functions thought to more accurately represent complex response-time relations characterizing change in human behaviors, traits, and abilities. In contrast to their linear counterparts, nonlinear functions are flexible and can often be tailored so that the parameter of the function correspond to interesting and meaningful facets of the longitudinal process.

These facets might include asymptotic or limiting behavior (Browne, 1993; Browne & Du Toit, 1993), change points (knots) in processes that exhibit distinct,

---

or multiple phases (Cudeck & Klebe, 2002). An interesting nonlinear function that has garnered considerable attention in the applied and methodological literatures is the logistic (S-shape or sigmoidal) function. The logistic function is ideally suited for many change processes including skill acquisition and cognitive development (Grimm & Ram, 2009; Ram & Grimm, 2007) because it is characterized by natural lower and upper bounds, allows for gradual change near these bounds as well as a surge of more rapid change toward the center of the process (Choi et al., 2009). Consider the case of behavior cessation in which behavior prevalence decreases slowly at first followed by a period of steady decline, accelerates, gradually begins to slow and eventually plateaus as cessation of the behavior is approached. (Contrary to the backward S shape of such behavior cessation, skill acquisition would follow a forward S shape with the process progressing in reverse order).

Choi et al. (2009) proposed a reparameterization of a conventional logistic growth function that allowed the estimation of lower and upper asymptotes as well as a surge point (i.e., that time  $t$  where maximum change occurs) and the surge slope (i.e., the rate of change at that juncture). Choi et al. demonstrated how this new parameterized logistic function could be fit as an GM with structural equation modeling (SEM) software. While the nonlinear constraint feature in SEM software permits such a nonlinear growth model to be estimated (see, e.g., Preacher & Hancock, 2015), it has a number of limitations. Nonlinear GMs in general, and the logistic GM in particular, to be fitted as structural equation models must (1) be constrained so parameters that enter the function in a nonlinear manner are fixed across individuals (see, e.g., Blozis & Cudeck, 1999; Harring et al., 2006), or (2) the nonlinear function be linearized using analytical or numerical methods such as a first-order Taylor series (see, e.g., Browne, 1993) expansion of the expectation of the function.

The fully nonlinear form of the function is not permitted within the traditional estimation framework (see Blozis & Harring, 2015; Harring & Blozis, 2016; for a discussion of the computational and conceptual differences, respectively). Fixing intrinsically nonlinear parameters across individuals may seem too restrictive and/or theoretically implausible given the modeling situation. Preliminary fitting of individuals' curves might indicate that such variability in all function parameters needs to be accommodated. Furthermore, the linearized form of the original nonlinear function may not fit particularly well especially if there is substantial intra-individual variability (Davidian & Giltinan, 1995). Fortunately, other estimation approaches have emerged that are well-suited to handle the computational burden compelled by incorporating intrinsically nonlinear functions.

## LOGISTIC GROWTH MODELING WITH MCMC ESTIMATION

A Bayesian approach is outlined to estimating parameters of a fully nonlinear GM using a logistic function via Markov chain Monte Carlo (MCMC; Gelman et al., 2004) methods using OpenBUGS (Lunn et al., 2009; Spiegelhalter et al., 2002; Thomas, 2009). Although describing the repeated measures with a nonlinear function is an important analytic activity, it will be demonstrated how covariates can be introduced at a secondary stage to explain, in part, observed heterogeneity in growth characteristics. To investigate various practical issues of fitting a logistic GM, a Monte Carlo simulation will be carried out, in which the number and location of measurement occasions as well as sample size is manipulated. The required OpenBUGS code needed to fit the various models and source code for the final model can be found in the Appendix.

### Methodology

#### Logistic Growth Model

Choi et al. (2009) presented an extension of the logistic function presented by Verhulst (1845) who was interested in studying population growth to a longitudinal setting. For the  $i^{\text{th}}$  individual at time point  $t$ , the fitted logistic function can be specified as

$$\hat{y}_i = \gamma_{Li} + (\gamma_{Ui} - \gamma_{Li}) \frac{\exp(\beta_{0i} + \beta_{1i}t)}{1 + \exp(\beta_{0i} + \beta_{1i}t)}. \quad (1)$$

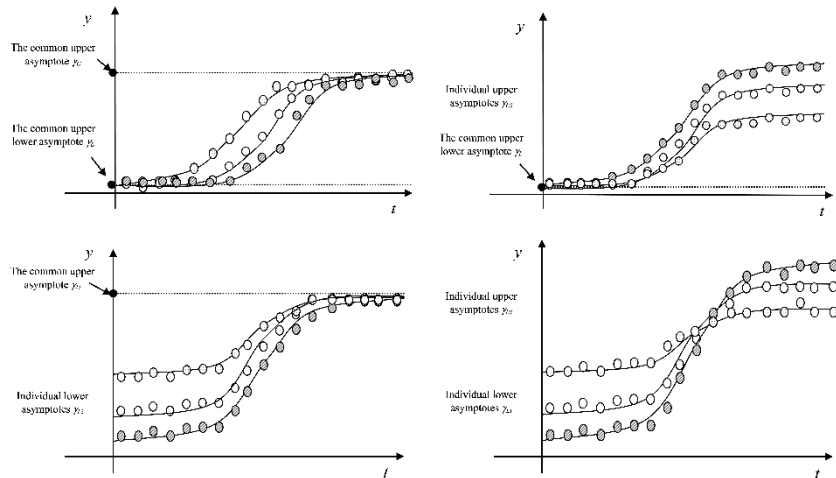
In this expression, the  $i^{\text{th}}$  subject has four individual-specific parameters: the lower asymptote  $\gamma_{Li}$ , the upper asymptote  $\gamma_{Ui}$ , the logistic intercept  $\beta_{0i}$  related to (but not equal to) the relative location of the individual's surge point, and the logistic slope  $\beta_{1i}$  related to (but not equal to) the slope of the individual's surge (further details regarding  $\beta_{0i}$  and  $\beta_{1i}$  are forthcoming). In Choi et al. (2009), the values of asymptotes  $\gamma_{Li}$  and  $\gamma_{Ui}$  were assumed to be known and common for all individuals (e.g., reflecting no knowledge and complete mastery, respectively). This modification facilitated a logit-type transformation of the individual-level data needed to convert Equation 1 to a model that is linear in its parameters, and one that could be estimated with SEM software by either using gradient-based methods such as Maximum Likelihood (ML) or limited information estimators such as Weighted Least Square (WLS). As will be illustrated in the later part of this paper, one of the advantages of an MCMC approach is one can model the logistic growth with equation (1) without the logit-type transformation.

**Logistic Model Reparameterization.** Although  $\beta_{0i}$  and  $\beta_{1i}$  have a sensible generalized linear model interpretation, other equivalent parameterizations may actually represent other key features of the sigmoidal change process that are more informative. Choi et al. (2009) proposed two factors, the surge point (*SP*) and surge slope (*SS*), as alternatives to the logistic intercept and slope growth factors which highlight important stages in the developmental process. The authors defined the *SP* factor as the location on the time axis where the maximum gain (maximum slope) for the response occurs. Mathematically, the surge point is the abscissa corresponding to the inflection point (change in concavity) on the logistic curve. The surge point for the  $i^{\text{th}}$  subject can be expressed as the quotient of an individual's logistic intercept and slope:

$$SP_i = -\beta_{0i} / \beta_{1i} . \tag{2}$$

An individual's slope at the surge point, the surge slope, can be expressed as a function of individual lower and upper asymptotes and their logistic slope (see Choi et al., 2009 for more details of the *SP* and *SS* factor derivation):

$$SS_i = 0.25(\gamma_{Ui} - \gamma_{Li}) \beta_{1i} . \tag{3}$$



**Figure 1.** Logistic growth curves of three subjects with various type of lower and upper asymptotes

## LOGISTIC GROWTH MODELING WITH MCMC ESTIMATION

Note, the  $SP$  and  $SS$  parameters are not directly expressed in the parameters in equation (1), yet because these coefficients are thought to represent change characteristics which are more fundamental to understanding the developmental S-shaped process, the function in equation (1) must be expressed in terms of these newly formed parameters as

$$\hat{y}_{ii} = \gamma_{Li} + (\gamma_{Ui} - \gamma_{Li}) \frac{1}{1 + \exp\left\{\frac{4SS_i(SP_i - t)}{\gamma_{Ui} - \gamma_{Li}}\right\}}. \quad (4)$$

Figure 1 depicts sample trajectories for various logistic growth scenarios based on equation (4).

**Model Specification.** Using equation (4) as the logistic function to move forward with individually-varying parameters, the response at the  $t^{\text{th}}$  measurement occasion ( $t = 1, \dots, m$ ) for individual  $i$  ( $i = 1, \dots, n$ ), can be written as

$$\begin{aligned} y_{ii} &= f_{\text{Logistic}}(t, \boldsymbol{\theta}_i) + \delta_{ii} \\ &= \gamma_{Li} + (\gamma_{Ui} - \gamma_{Li}) \left\{ 1 + \exp\left(\frac{4SS_i}{\gamma_{Ui} - \gamma_{Li}}(SP_i - t)\right) \right\}^{-1} + \delta_{ii}. \end{aligned} \quad (5)$$

The time-specific disturbances are assumed to be normally distributed,  $\delta_{ii} \sim N(0, \sigma_\delta)$ , where  $\sigma_\delta$  is the error standard deviation of the logistic growth model. Note that  $\sigma_\delta$  or variants of  $\sigma_\delta$  is the most fundamental measure for data-model fit in many types of modeling of continuous outcomes. Therefore, we will investigate  $\sigma_\delta$  as a measure of model estimation quality or appropriateness when we are evaluating different models and/or options in the later part of this paper. By imposing the Gaussian probability density functions for the outcome and other logistic growth parameters, the probability models for the logistic growth model can expressed as

$$\begin{aligned}
 y_{it} &\square \mathbf{N} \left[ \gamma_{Li} + (\gamma_{Ui} - \gamma_{Li}) \left\{ 1 + \exp \left( \frac{4SS_i}{\gamma_{Ui} - \gamma_{Li}} (SP_i - t_i) \right) \right\}^{-1}, \sigma_\delta \right], \\
 \gamma_{Li} &\sim \mathbf{N}(\kappa_{\gamma_L}, \sigma_{\gamma_L}), \\
 \gamma_{Ui} &\sim \mathbf{N}(\kappa_{\gamma_U}, \sigma_{\gamma_U}), \\
 SP_i &\sim \mathbf{N}(\kappa_{SP}, \sigma_{SP}), \text{ and} \\
 SS_i &\sim \mathbf{N}(\kappa_{SS}, \sigma_{SS})
 \end{aligned} \tag{6}$$

where  $\kappa_{\gamma_L}$ ,  $\kappa_{\gamma_U}$ ,  $\kappa_{SP}$ ,  $\kappa_{SS}$ ,  $\sigma_{\gamma_L}$ ,  $\sigma_{\gamma_U}$ ,  $\sigma_{SP}$ , and  $\sigma_{SS}$  are means and standard deviations (i.e., inverse of precisions in OpenBUGS setting) for the probability model of  $\gamma_L$ ,  $\gamma_U$ ,  $SP$ , and  $SS$ , respectively. Using information in (6), the likelihood function for this model can be expressed as

$$\begin{aligned}
 \prod_i \prod_t \mathbf{P}(y_{it} | \boldsymbol{\theta}_i, \sigma_\delta) &= \prod_i \prod_t \mathbf{P}(y_{it} | \gamma_{Li}, \gamma_{Ui}, SP_i, SS_i, \sigma_e) \\
 &= \prod_i \prod_t \mathbf{P}(y_{it}, \gamma_{Li}, \gamma_{Ui}, SP_i, SS_i | \sigma_\delta) \mathbf{P}(\gamma_{Li} | \kappa_{\gamma_L}, \sigma_{\gamma_L}) \\
 &\quad \times \mathbf{P}(\gamma_{Ui} | \kappa_{\gamma_U}, \sigma_{\gamma_U}) \mathbf{P}(SP_i | \kappa_{SP}, \sigma_{SP}) \mathbf{P}(SS_i | \kappa_{SS}, \sigma_{SS})
 \end{aligned} \tag{7}$$

It is worthwhile to note several characteristics of this likelihood function for the logistic growth model. First, this likelihood function approach can be categorized as an individual-level data approach (Choi & Levy, 2017; Levy & Choi, 2013), which requires individual-level data instead of summary-level data. Second, from the likelihood function in equation (7), full conditional distributions should be constructed that may require considerable programming. Third, in this model specification, there are  $4n + 9$  parameters ( $n$  individual parameters for  $\gamma_L$ ,  $n$  parameters for  $\gamma_U$ ,  $m$  individual parameters for the  $SP$  factor,  $n$  individual parameters for the  $SS$  factor, and  $\kappa_{\gamma_L}$ ,  $\sigma_{\gamma_L}$ ,  $\kappa_{\gamma_U}$ ,  $\sigma_{\gamma_U}$ ,  $\kappa_{SP}$ ,  $\sigma_{SP}$ ,  $\kappa_{SS}$ ,  $\sigma_{SS}$ , and  $\sigma_\delta$  to be estimated. For example, with  $n = 100$ , there are 409 parameters needed to be estimated! Estimating such large number of parameters with traditional estimation methods is evidently very challenging. However, by virtue of simulation-based estimation methods, such as MCMC within OpenBUGS, one can estimate such large amount of logistic growth model parameters without great difficulty merely by specifying the model in accordance with the OpenBUGS language.

To implement MCMC estimation within Bayesian framework in OpenBUGS, prior distributions should be specified. In this paper, uninformative conjugate priors

## LOGISTIC GROWTH MODELING WITH MCMC ESTIMATION

(distributions parameterized to have large variances) are specified for the means and standard deviations associated with the individually varying logistic growth parameters. Specifically,

$$\begin{aligned} \kappa_{\gamma_L}, \kappa_{\gamma_U}, \kappa_{SS}, \kappa_{SP} &\sim N(10, 10) \\ (\sigma_{\gamma_L}^2)^{-1}, (\sigma_{\gamma_U}^2)^{-1}, (\sigma_{SP}^2)^{-1}, (\sigma_{SS}^2)^{-1} &\sim \text{Gamma}(0.1, 0.1) \end{aligned} \quad (8)$$

These settings for the logistic growth parameters prior distributions will be used for all analyses throughout the remainder of the paper. It is also possible to include time-invariant covariate(s) as predictors of each logistic growth parameter. For example, if one is interested in including a covariate  $C$  for the  $SS$  factor, the probability model for  $SS$  factor and coefficients for the covariate can be specified as

$$\begin{aligned} SS_i &\sim N(\alpha_0 + \alpha_1 C_i, \sigma_{SS}) \\ \alpha_0 &\sim N(\kappa_{\alpha_0}, \sigma_{\alpha_0}) \\ \alpha_1 &\sim N(\kappa_{\alpha_1}, \sigma_{\alpha_1}) \end{aligned} \quad (9)$$

Although covariates play an important role in any latent growth model, for simplicity's sake, this component is not included in the upcoming analysis and simulation. The Appendix contains an OpenBUGS syntax for the logistic growth model.

### Simulation

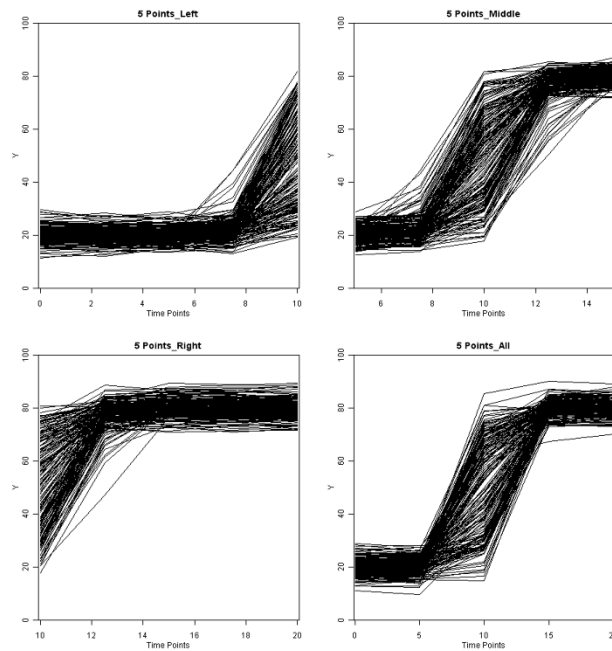
The logistic GM can be analyzed using MCMC methods. To empirically evaluate and compare the performance of these methods, a Monte Carlo simulation was employed considering various conditions thought to impact the accuracy and precision of the logistic growth parameters. All data were generated using language R (R Development Core Team, 2010). For MCMC, a general OpenBUGS model was formulated based on the model in equation (4) and prior distributions (see Appendix) using the OpenBUGS (Lunn et al., 2009) program. The first 5,000 MCMC iterations were discarded as burn-in iterations and estimates were based on the subsequent 10,000 iterations sampled from the marginal posterior distributions of all parameters of interest. Quantities characterizing these marginal posterior distributions, (e.g., median, weighted mean, 95% credibility intervals) can be

readily obtained. The R to OpenBUGS interface BRugs (Ligges et al., 2017) was used for communication between R and OpenBUGS.

**Table 1.** Simulation design conditions

Simulation factors	Levels
Growth type	Linear, Logistic
$n$	50, 100, 200
$p$	3, 5
$l$	Left (1), Middle (2), Right (3), All (4)
Replication #	200
$\mu(\gamma_L, \gamma_U, SS, SP)$	20, 80, 25, 10
$\text{var}(\gamma_L, \gamma_U, SS, SP)$	9, 9, 9, 1
Within subject error variance	1

Note:  $n$  = sample size,  $p$  = number of time points,  $l$  = locations of measurement points



**Figure 2.** Growth curves for 5 measurement points based on the simulation design ( $n = 200$ ) depicting trajectories of left measurement location ( $l = 1$ ), middle measurement location ( $l = 2$ ), right measurement location ( $l = 3$ ), all measurement location ( $l = 4$ )

## LOGISTIC GROWTH MODELING WITH MCMC ESTIMATION

**Table 2.** Measurement point locations ( $l$ ) over number of measurement points ( $p$ )

Location	$p = 3$	$p = 5$
Left (1)	0, 5, 10	0, 2.5, 5, 7.5, 10
Middle (2)	5, 10, 15	5, 7.5, 10, 12.5, 15
Right (3)	10, 15, 20	10, 12.5, 15, 17.5, 20
All (4)	0, 10, 20	0, 5, 10, 15, 20

Three sample sizes,  $n = 50, 100,$  and  $200$ ; two different number of measure occasions,  $p = 3$  and  $5$ ; four different types of measurement locations ( $l$ ), Left (1), Middle (2), Right (3), and All (4) were considered. All simulation conditions and population model parameters are summarized in Tables 1 and 2. For each condition, 200 datasets (i.e., replication number = 200) were generated, and each dataset was analyzed using MCMC with OpenBUGS. Figure 2 depicts sample trajectories from generated data from the given simulation conditions.

### Outcome Statistics Examined

Several summary statistics were examined to evaluate four growth factors ( $\gamma_L, \gamma_U, SS,$  and  $SP$ ) over different simulation conditions. First, for evaluating the bias of point estimates, mean bias ( $MB$ ) and mean relative bias ( $MRB$ ) were employed (Bandalos & Leite, 2013):

$$\begin{aligned}
 MB &= \sum_{i=1}^{rep} (\hat{\theta}_i - \theta) / rep \\
 MRB &= \sum_{i=1}^{rep} [(\hat{\theta}_i - \theta) / \theta] / rep
 \end{aligned}
 \tag{10}$$

where  $\theta$  is the true value,  $\hat{\theta}_i$  is the  $i^{\text{th}}$  replication of the point estimate given the estimation method converged in that replication, and rep was the number of times the estimation method converged within the 200 replications. Second, for evaluating the variability and reliability of the point estimates, root mean squared error ( $RMSE$ ) was examined and is defined as:

$$RMSE = \left[ \sum_{i=1}^{rep} (\hat{\theta}_i - \theta)^2 / rep \right]^{1/2}
 \tag{11}$$

## Results

### Bias of Estimates

Statistics regarding the bias of estimators (*MB* and *MRB*) are presented in Table 3. Here are the common themes of the *MB* and *MRB* simulation results. First, the estimates of the upper asymptote,  $\gamma_U$ , showed poorer performance (e.g., large bias) compared to  $\gamma_L$ , *SS*, *SP*. The  $\gamma_U$  estimates were recovered with accuracy as the absolute value of *MRB* estimate was less than 0.05. This occurred, however, only when  $p = 5$ ;  $l =$  middle, right, or all; and  $n = 200$ . The estimates were particularly biased when the sample size was small (i.e.,  $n = 50$ ). Second, the biases were smaller as the number of time points increased. Specifically, the  $p = 5$  scenario clearly outperformed the  $p = 3$  scenario across the majority of simulation conditions except for the  $\gamma_U$  estimates. That is, increasing the number of measurement points

**Table 3.** *MB, MRB, RMSE* of estimates

<i>p</i>	<i>l</i>	<i>n</i>	<i>MB</i>				<i>MRB</i>				<i>RMSE</i>			
			$\gamma_L$	$\gamma_U$	<i>SS</i>	<i>SP</i>	$\gamma_L$	$\gamma_U$	<i>SS</i>	<i>SP</i>	$\gamma_L$	$\gamma_U$	<i>SS</i>	<i>SP</i>
3	1	50	18.23	24.13	17.64	7.97	-0.09	-0.70	-0.29	-0.20	2.57	56.02	7.59	2.15
		100	19.12	36.19	19.00	7.92	-0.04	-0.55	-0.24	-0.21	1.39	43.86	6.15	2.13
		200	19.29	44.53	19.45	7.83	-0.04	-0.44	-0.22	-0.22	1.33	35.50	5.71	2.21
	2	50	16.38	25.29	15.82	8.77	-0.18	-0.68	-0.37	-0.12	4.63	55.36	9.42	1.87
		100	17.79	72.79	19.49	9.05	-0.11	-0.09	-0.22	-0.09	3.22	10.70	6.02	1.54
		200	19.93	79.63	23.42	9.98	0.00	0.00	-0.06	0.00	0.26	0.45	1.82	0.08
	3	50	14.48	23.22	18.56	9.06	-0.28	-0.71	-0.26	-0.09	5.95	57.18	6.84	1.41
		100	18.21	79.31	23.35	9.93	-0.09	-0.01	-0.07	-0.01	2.13	0.81	1.98	0.13
		200	18.25	79.66	23.72	9.93	-0.09	0.00	-0.05	-0.01	2.10	0.39	1.56	0.11
4	50	17.05	25.33	14.60	8.62	-0.15	-0.68	-0.42	-0.14	3.98	55.26	10.72	2.01	
	100	19.26	77.19	20.41	9.78	-0.04	-0.04	-0.18	-0.02	1.12	4.19	5.07	0.42	
	200	19.94	79.63	22.94	9.99	0.00	0.00	-0.08	0.00	0.24	0.43	2.56	0.09	
5	1	50	18.96	20.88	22.08	8.48	-0.05	-0.74	-0.12	-0.15	3.04	59.17	5.23	2.36
		100	19.82	33.62	24.33	8.84	-0.01	-0.58	-0.03	-0.12	0.85	46.39	1.94	1.19
		200	20.07	44.91	25.86	8.79	0.00	-0.44	0.03	-0.12	0.24	35.10	1.04	1.21
	2	50	19.77	20.93	24.50	10.02	-0.01	-0.74	-0.02	0.00	0.47	60.32	0.76	0.13
		100	19.89	79.31	24.81	10.00	-0.01	-0.01	-0.01	0.00	0.36	0.76	0.45	0.09
		200	19.96	79.64	24.83	9.99	0.00	0.00	-0.01	0.00	0.24	0.42	0.36	0.07
	3	50	17.48	25.35	24.21	9.94	-0.13	-0.68	-0.03	-0.01	3.07	57.88	1.08	0.18
		100	18.68	79.36	24.48	9.96	-0.07	-0.01	-0.02	0.00	1.70	0.71	0.76	0.11
		200	19.03	79.68	24.75	9.97	-0.05	0.00	-0.01	0.00	1.31	0.38	0.48	0.09
	4	50	19.73	20.53	21.80	10.00	-0.01	-0.74	-0.13	0.00	0.47	60.51	3.40	0.15
		100	19.85	79.30	23.08	10.00	-0.01	-0.01	-0.08	0.00	0.34	0.77	2.14	0.10
		200	19.92	79.67	23.70	10.00	0.00	0.00	-0.05	0.00	0.24	0.39	1.56	0.08

## LOGISTIC GROWTH MODELING WITH MCMC ESTIMATION

improved the accuracy in terms of bias of  $\gamma_L$ ,  $SS$ , and  $SP$ , but did not improve the performance of the estimation of the upper asymptote,  $\gamma_U$ . Third, the middle ( $l = 2$ ) and all ( $l = 4$ ) growth pattern scenarios (e.g., the absolute  $MRB$  values are all less than .1 when  $n = 200$  in Table 3) outperformed the left ( $l = 1$ ) and right ( $l = 3$ ) scenarios (e.g., the absolute  $MRB$  values of  $\gamma_U$ ,  $SS$ , and  $SP$  are all greater than .2 even when  $n = 200$  and  $p = 3$  in Table 3). That is, the estimates across these conditions were less biased. Furthermore, the  $\gamma_L$  estimates were more accurate for the  $l = 1$  scenario than the  $l = 3$  scenario, and the  $\gamma_U$  estimates were less biased for the  $l = 3$  scenario than the  $l = 1$  scenario. In other words, the estimation of the asymptote parameters,  $\gamma_L$  and  $\gamma_U$ , were sensitive to the growth pattern directionality (i.e., left or right side, respectively). Not surprisingly, the biases decreased as the sample size increased (e.g., the absolute  $MRB$  values are less than 0.02 when  $n = 200$ ,  $p = 5$ , and  $l = 2$ ). A closer examination of the sample size condition revealed that parameters were less biased for  $n = 100$  than for the scenario in which  $n = 50$ . For the  $n = 200$  case, the absolute value of  $MRB$  estimates were less than 0.1 for all simulation conditions except the  $l = 1$  case.

### Variability of Estimates

The statistics regarding the variability of estimates,  $RMSE$  are also presented in Table 3. The first trend is similar to the results reported on parameter bias—the  $\gamma_U$  estimates had larger values of  $RMSE$  compared to  $\gamma_L$ ,  $SS$  and  $SP$ . The  $RMSE$  for the estimates for  $\gamma_U$  were relatively small (i.e.,  $RMSE$  was less 5) when  $l =$  middle, right, or all, and  $n = 200$ . The  $SP$  estimates were more stable in terms of  $RMSE$  values compared to the  $RMSE$  values of the other three parameter estimates (e.g., the  $RMSE$  values of  $SP$  parameter is less than 2.37 for all simulation conditions). Second, the  $RMSE$  values decreased as the number of measurement occasions increased except when the sample size was small (i.e.,  $n = 50$ ). In other words, when the same size is small, the parameter estimates are not necessarily becoming more stable even when the number of measurement occasions increased. Also, similar to the  $MRB$  results, it seems that increasing the number of measurement points does not increase the performance of  $\gamma_U$  parameter estimation in terms of  $RMSE$ . For example, when  $p = 3$  and  $l = 4$ , the  $RMSE$  values of  $\gamma_U$  are 55.26 ( $n = 50$ ), 4.19 ( $n = 100$ ), and 0.43 ( $n = 200$ ). However, when  $p = 5$  and  $l = 4$ , the  $RMSE$  values are 60.51 ( $n = 50$ ), 0.77 ( $n = 100$ ), and 0.39 ( $n = 200$ ). Third, the middle ( $l = 2$ ) and all ( $l = 4$ ) growth pattern scenarios produced parameter estimates that had smaller  $RMSE$  values than those produced under the left ( $l = 1$ ) and right ( $l = 3$ ) scenarios and when  $p = 5$ . When  $p = 3$ , the  $RMSE$  values of all four parameters did

not differ substantially across the 4 growth patterns. Fourth, the RMSE values decreased as the sample size increased. Consistent with the parameter bias results in terms of parameter estimate stability, the estimates of the growth parameters were more stable when  $n = 100$  rather than when  $n = 50$ . The *RMSE* values of  $n = 50$  are all greater than those of  $n = 100$  and  $n = 200$  for all simulation conditions as seen in Table 3. For  $n = 200$ , the RMSE values were less than 3 for all simulation conditions except  $l = 1$  and for  $\gamma_U$ .

### **Recommendations for Practitioners**

Based on the simulation results, several recommendations can be made for the implementation of logistic growth models using an MCMC algorithm in a Bayesian estimation framework. First, one must secure at least a sample size of 100. In particular, if a researcher is not sure that he is modeling the middle or whole part of logistic growth trajectory, he should secure at least 200 samples. As can be seen from the results of this study, if the sample size is less than 100, the parameter bias and variability is uncomfortably larger than the case of  $n = 200$ . Second, to model the core characteristics of the logistic growth curve model, the overall growth trajectory must follow logistic growth and repeated measurements from at least five measurement occasions must be obtained. As can be seen from the results of this study, if the number of measurement occasions is only 3, parameter bias and variability is uncomfortably large. There is simply not enough information with which to estimate the logistic parameters with sufficient accuracy or precision. Specifically, allocating more measurement points on the asymptote sides is recommended. Without having enough measurement points at the beginning and/or end of the process, the lower and upper asymptotes will not be estimated with accuracy or stability. Third, researchers should be aware that the fact of what part of the logistic growth is modeled may have a substantial impact on the estimation of each parameter. Before modeling, one should first check the growth pattern of the raw repeated measures data. This can be accomplished through a series of spaghetti plots and by examining individuals' data. One should first check whether the growth pattern is an S-shape or sigmoidal pattern, especially if it is the whole growth trajectory or at least the middle part of the logistic growth pattern. Fourth, estimating the upper asymptote is relatively difficult. Especially when the sample size is small, the estimation accuracy of the parameters may show substantial bias, which could lead to erroneous inferences about the underlying process. Potential solutions for this might include (1) making the upper asymptote a known or fixed

point, or (2) applying an informative prior for the particularly problematic parameter.

### Discussion

In Choi et al. (2009), a treatment of growth following logistic (sigmoidal; S-shape) growth functions within a traditional structural equation model estimation framework was proposed by a reparameterization of the logistic function. Even though a nonlinear parameter is embedded in the overall structure, this latent growth model can be estimated as a kind of structural equation using conventional SEM software that has the facility of nonlinear constraints. However, the nonlinear growth function must be coerced into fitting into the software that allows only linear relations among latent variables. However, fixing intrinsically nonlinear parameters across subjects may seem too restrictive and/or theoretically implausible given the modeling situation. Preliminary fitting of individual curves might indicate that such variability in all parameters needs to be accommodated. Whether theoretically based or empirically driven, the GM can be extended to handle intrinsically nonlinear parameters. Unfortunately, the added complexity of the newer model precludes estimating parameters of the model within the conventional SEM estimation framework.

This research study introduced a specific, nonlinear GM—the logistic GM—and demonstrated how it could be estimated with MCMC estimation techniques which are well-suited to handle the computational burdens compelled by incorporating intrinsically nonlinear functions. The paper proposed new modeling approach that can fit inherently nonlinear (i.e., logistic) growth function without constraint nor reparameterization. And, this paper investigated various modeling issues for logistic GM using a Monte Carlo simulation study.

The current paper makes three contributions to the emerging nonlinear GM literature. First, this paper introduced intrinsically nonlinear GM without reparameterization and showed the estimation of the model within a Bayesian framework. Second, various practical issues associated with the logistic GM (e.g., number of measurement points, and locations of these points in the logistic trajectory, and sample sizes) were investigated using a Monte Carlo simulation study. Third, the study provides modeling and/or research design recommendations for successful implementation in practice. The recommendations can be summarized as follows. First, it is highly recommended to secure at least 100 sample size. Second, with sufficient number of measurement points (e.g., 5 measurement points), one can successfully fit a logistic GM. Third, one should set

the middle or full range measurement points as possible to fit a logistic GM with accuracy and stability.

Undertaking the study of the logistic GM, we acknowledge known limitations. First, the analyses in this paper are based on moderately diffuse prior distributions. As mentioned before, MCMC also allows us to explicitly investigate the sensitivity of prior distributions for parameters when used in the context of Bayesian inference. The capability of cooperating prior into the estimation procedure would be a distinctive advantage of Bayesian approach over the traditional approaches. Investigating the appropriate use and/or advantages of using prior in the context of growth modelling is unanswered and remains as future study. Second, our elaboration in this paper has only focused logistic growth on unbounded continuum from negative infinite to positive infinite. In the model proposed in this study, a mathematical concept of infinity is required to conceptualize/interpret the asymptote parameters. New type of logistic growth model has support on a bounded continuum would be practically and theoretically useful, but such development remains as future study. Third, this study did not compare the proposed model with other growth models (e.g., linear or quadratic growth) in terms of model fit. Model fit is an important factor in the evaluation of different growth models, and it is very needed to conduct a comparative analysis of the proposed model and other growth models from this point of view.

## Acknowledgements

An earlier version of this paper was presented at the 2018 meeting of the American Education Research Association, Denver, CO.

## References

- Bandalos, D. L., & Leite, W. (2013). The use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2<sup>nd</sup> edition) (pp. 385-427). Greenwood, CT: Information Age Publishing, Inc.
- Blozis, S. A., & Cudeck, R. (1999). Conditional linear mixed-effects models with latent covariates. *Journal of Educational and Behavioral Statistics*, 24(3), 245-270. doi: 10.3102/10769986024003245

## LOGISTIC GROWTH MODELING WITH MCMC ESTIMATION

- Blozis, S. A., & Harring, J. R. (2017). Understanding individual-level change through the basis functions of a latent curve model. *Sociological Methods & Research, 46*(4), 793-820. doi: 10.1177/0049124115605341
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley. doi: 10.1002/0471746096
- Browne, M. W. (1993). Structured latent curve analysis. In C. M. Cudras & C. R. Rao (Eds.), *Multivariate analysis: Future directions 2* (pp.171-197). Amsterdam, Netherlands: Elsevier Science. doi: 10.1016/B978-0-444-81531-6.50016-7
- Browne, M. W., & Du Toit, S. H. C. (1993). Models for learning data. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 47-68). Washington, DC: American Psychological Association. doi: 10.1037/10099-004
- Choi, J., Harring, J. R., & Hancock, G. R. (2009). Latent growth modeling for logistic response functions. *Multivariate Behavioral Research, 44*(5), 620-645. doi: 10.1080/00273170903187657
- Choi, J., & Levy, R. (2017). Markov chain Monte Carlo estimation methods for structural equation modeling: A comparison of subject-level data and moment-level data approaches. *Biometrics and Biostatistics International Journal, 6*(5), 463-474. doi: 10.15406/bbij.2017.06.00182
- Cudeck, R., & Klebe, K. J. (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods, 7*(1), 41-63. doi: 10.1037/1082-989X.7.1.41
- Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. New York: Chapman and Hall.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2<sup>nd</sup> edition). Boca Raton, FL: Chapman & Hall/CRC.
- Grimm, K. J., & Ram, N. (2009). Nonlinear growth models in Mplus and SAS. *Structural Equation Modeling, A Multidisciplinary Journal, 16*(4), 676-701. doi: 10.1080/10705510903206055
- Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development, 82*(5), 1357-1371. doi: 10.1111/j.1467-8624.2011.01630.x
- Harring, J. R., & Blozis, S. A. (2016). A note on recurring misconceptions when fitting nonlinear mixed models. *Multivariate Behavioral Research, 51*(6), 805-817. doi: 10.1080/00273171.2016.1239522

Harring, J. R., Cudeck, R., & du Toit, S. H. C. (2006). Fitting partially nonlinear random coefficient models as SEMs. *Multivariate Behavioral Research*, *41*(4), 579-596. doi: [10.1207/s15327906mbr4104\\_7](https://doi.org/10.1207/s15327906mbr4104_7)

Levy, R., & Choi, J. (2013). An introduction to Bayesian structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2<sup>nd</sup> edition). Greenwood, CT: Information Age Publishing, Inc.

Ligges, U., Sturtz, S., Gelman, A., Gorjanc, G., & Jackson, C. (2017, June 26). Package ‘BRugs’: Interface to the ‘OpenBUGS’ MCMC software [R software package]. Retrieved from <https://cran.r-project.org/package=BRugs>

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, *28*(25), 3049-3067. doi: [10.1002/sim.3680](https://doi.org/10.1002/sim.3680)

Preacher, K. J., & Hancock, G. R. (2015). Meaningful aspects of change as novel random coefficients: A general method for reparameterizing longitudinal models. *Psychological Methods*, *20*(1), 84-101. doi: [10.1037/met0000028](https://doi.org/10.1037/met0000028)

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

R Development Core Team. (2010). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Ram, N., & Grimm, K. J. (2007). Using simple and complex growth models to articulate developmental change: Matching theory to method. *International Journal of Behavioral Development*, *31*(4), 303-316. doi: [10.1177/0165025407077751](https://doi.org/10.1177/0165025407077751)

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2003, January). *WinBUGS user manual: Version 1.4*. Cambridge, UK: University of Cambridge Medical Research Council Biostatistics Unit. <https://www.mrc-bsu.cam.ac.uk/software/bugs/>

Thomas, N. (2009, December 14). *Overview*. OpenBUGS. <https://www.openbugs.net/w/FrontPage>

Verhulst, P. F. (1845). La loi d'accroissement de la population. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, *18*, 14-54.

## LOGISTIC GROWTH MODELING WITH MCMC ESTIMATION

### Appendix

BUGS code:

```
model
{
  for (i in 1:n) {
    for (j in 1:t) {
      y[i, j] ~ dnorm(my[i, j], ty)
      my[i,j] <- b[i,1] + (( b[i,2] - b[i,1] )
/ ( 1 + exp(4*b[i,3] *(b[i,4]-x[j]) / (b[i,2] - b[i,1]))))
    }

    for (k in 1:4) {
      b[i,k]~dnorm(mu[k],tau[k])
    }
  }
  ty ~ dgamma(1.0E-1, 1.0E-1)
  vy <- 1 / ty
  for (k in 1:4) {
    mu[k] ~ dnorm(10, 1.0E-1)
    tau[k] ~ dgamma(1.0E-1, 1.0E-1)
    var[k] <- 1 / tau[k]
  }
}
```