

International Journal of Behavioral Development

<http://jbd.sagepub.com>

Bayesian analysis of longitudinal data using growth curve models

Zhiyong Zhang, Fumiaki Hamagami, Lijuan Lijuan Wang, John R. Nesselroade and Kevin J. Grimm
International Journal of Behavioral Development 2007; 31; 374
DOI: 10.1177/0165025407077764

The online version of this article can be found at:
<http://jbd.sagepub.com/cgi/content/abstract/31/4/374>

Published by:

 SAGE Publications

<http://www.sagepublications.com>

On behalf of:



[International Society for the Study of Behavioral Development](http://www.sagepub.com/journalsPermissions.nav)

Additional services and information for *International Journal of Behavioral Development* can be found at:

Email Alerts: <http://jbd.sagepub.com/cgi/alerts>

Subscriptions: <http://jbd.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations (this article cites 25 articles hosted on the SAGE Journals Online and HighWire Press platforms):
<http://jbd.sagepub.com/cgi/content/abstract/31/4/374#BIBL>

Bayesian analysis of longitudinal data using growth curve models

Zhiyong Zhang, Fumiaki Hamagami, Lijuan Wang, and John R. Nesselroade
University of Virginia, USA

Kevin J. Grimm
University of California, USA

Bayesian methods for analyzing longitudinal data in social and behavioral research are recommended for their ability to incorporate prior information in estimating simple and complex models. We first summarize the basics of Bayesian methods before presenting an empirical example in which we fit a latent basis growth curve model to achievement data from the National Longitudinal Survey of Youth. This step-by-step example illustrates how to analyze data using both noninformative and informative priors. The results show that in addition to being an alternative to the maximum likelihood estimation (MLE) method, Bayesian methods also have unique strengths, such as the systematic incorporation of prior information from previous studies. These methods are more plausible ways to analyze small sample data compared with the MLE method.

Keywords: Bayesian analysis; growth curve models; informative priors; longitudinal data; pooling data; WinBUGS

Introduction

To study phenomena in their time-related patterns of constancy and change is a primary reason for collecting longitudinal data. All longitudinal data share at least three features: (1) the same entities are observed repeatedly over time; (2) the same measurements (including parallel tests) are used; and (3) the timing for each measurement is known (Baltes & Nesselroade, 1979; McArdle & Nesselroade, 2003). The need to analyze longitudinal data has stimulated the development of longitudinal data analytic techniques and models that, in turn, have advanced the collection of longitudinal data. Growth curve models (McArdle & Nesselroade, 2003; Meredith & Tisak, 1990) exemplify a widely used technique with a direct match to the objectives of longitudinal research described by Baltes and Nesselroade (1979) – to analyze explicitly intraindividual change and interindividual differences in change.

In the past decades, growth curve models have evolved from fitting a single curve for only one individual to fitting multi-level or mixed-effects models for multiple individuals and from linear to nonlinear models (Blozis, Conger, & Herring, 2007; McArdle, 2001; McArdle & Nesselroade, 2003; Meredith & Tisak, 1990; Tucker, 1958; Wishart, 1938). The use of these models in social and behavioral research has grown rapidly since Meredith and Tisak (1990) showed that growth curve models can be fitted as a restricted common factor model in the structural equation modeling framework (see also, McArdle, 1988; McArdle & Epstein, 1987). For a more comprehensive discussion of growth curve models, see McArdle and Nesselroade (2003).

For estimating growth curve models, the maximum likelihood estimation (MLE) method is commonly used (Demidenko,

2004; Laird & Ware, 1982). MLE for growth curve models is embedded in commercial statistical packages, such as SAS PROC MIXED and PROC NL MIXED and Splus LME and NLME. Recently, Bayesian methods have received more attention as useful tools for estimating a variety of models including growth curve models, in particular complex growth curve models which can be difficult or impossible to estimate in the current MLE-based software (Lee & Chang, 2000; Lee & Liu, 2000; McArdle & Wang, in press; Menzefricke, 1998; Pettitt, Tran, Haynes, & Hay, 2006; Seltzer, Wong, & Bryk, 1996).

Bayesian methods have become well established in the psychological literature since their introduction to psychology by Edwards, Lindman, and Savage (1963). Since then, Bayesian methods have been successfully applied to item-response models (Chang, 1996; Fox & Glas, 2001), factor analytic models (Bartholomew, 1981; S. Lee, 1981), structural equation models (Scheines, Hoijsink, & Boomsma, 1999), genetic models (Eaves & Erkanli, 2003), and multilevel models (Seltzer et al., 1996). Rupp, Dey, and Zumbo (2004) provided a valuable review of applications of Bayesian methods in social and behavioral research and proposed that neither the applied nor the theoretical measurement communities can afford to miss the opportunities opened up by Bayesian methods.

Even though Bayesian methods are powerful and can be used in a variety of analytic models, the strenuous programming and computational demands have discouraged many researchers from using them. Furthermore, the complexities of the models that usually need the application of Bayesian methods make the methods seem remote for empirical researchers. In this study, we aim to draw researchers' attention to Bayesian methods and provide an easy way to implement Bayesian analysis, especially for longitudinal data.

Correspondence should be sent to Dr Zhiyong Zhang, Department of Psychology, PO Box 400400, University of Virginia, Charlottesville, VA 22903-4400, USA; e-mail: zz5m@virginia.edu

First, we provide a brief review of basic Bayesian terms and methods, such as priors, posteriors, and the Markov chain Monte Carlo (MCMC) method. We then present the basic concepts of the latent basis growth curve model used in the empirical data analysis. Finally, we present an empirical example, step-by-step, to show how to implement Bayesian analysis practically using the software WinBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2003) and BAUW (Zhang & Wang, 2006). We show that Bayesian methods provide a direct alternative to MLE and also demonstrate their unique strengths for analyzing longitudinal data.

Basic ideas of Bayesian methods

This section aims to introduce the basic ideas of Bayesian methods (see also Walker, Gustafson, & Frimer, 2007) and is naturally redundant with extant publications. For more complete descriptions of Bayesian analysis, we recommend Box and Tiao (1992), Carlin and Louis (2000), and Lee (2004). Researchers primarily interested in the application of Bayesian methods may choose to skip the mathematical formulas and go directly to the verbal explanations.

Bayes' theorem

Bayesian analysis is based on the tenet that the concept of probability can be applied to the degree to which a person believes a hypothesis or a proposition. The degree of belief in proposition H can be represented as $\Pr(H)$, which is also called the prior degree of belief in H . A simple version of Bayes' theorem (also known as Bayes' rule) is,

$$\Pr(H|E) = \frac{\Pr(E \cap H)}{\Pr(E)} = \frac{\Pr(E|H)\Pr(H)}{\Pr(E)}, \quad (1)$$

which indicates that the degree of belief in proposition H given the observed evidence E is equal to the joint probability of H and E divided by the probability of E . $\Pr(H|E)$ is called the posterior degree of belief in H , in the sense of being the updated belief after observing the evidence.

Usually, we have more than one hypothesis in our research. For example, if we have n different hypotheses, H_1, H_2, \dots, H_n , to explain a phenomenon, then Bayes' theorem is stated as,

$$\Pr(H_i|E) = \frac{\Pr(E|H_i)\Pr(H_i)}{\sum_{i=1}^n \Pr(E|H_i)\Pr(H_i)}, \quad (2)$$

This declares that our posterior belief on H_i not only depends on the observed evidence E but also depends on our prior beliefs regarding each hypothesis.

Bayes' theorem is useful because it provides a way to calculate the probability of a hypothesis based on the evidence or data. Given the evidence, the calculation of $\Pr(E|H_i)$ is straightforward. However, when we observe some evidence, we are interested in the probability of the hypotheses conditional on the evidence, $\Pr(H_i|E)$. Bayes' theorem provides a way to calculate this probability. But the calculation also depends on the prior probabilities $\Pr(H_i)$. Thus, Bayes' theorem provides a natural way to update our prior belief $\Pr(H_i)$ concerning the hypothesis to our posterior belief $\Pr(H_i|E)$ based on the evidence E .

So far, we have only discussed discrete probability applications. For continuous probability problems, the hypotheses are represented by one or more continuous parameters from a model denoted by θ . The evidence, represented by the data, is denoted by y . In this case, Bayes' theorem is written as,

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} = \frac{p(\theta)p(y|\theta)}{\int_0 p(\theta)p(y|\theta)d\theta} \quad (3)$$

in which $p(\theta)$ is the prior belief, prior probability distribution, or simply the prior of θ ; $p(\theta|y)$ is the posterior belief, posterior probability distribution, or the posterior of θ ; and $p(y|\theta)$ is the probability of the data which is also the likelihood $L(\theta;y)$ in MLE. Because $\int_0 p(\theta)p(y|\theta)d\theta$ is a normalizing constant, then

$$p(\theta|y) \propto p(\theta)p(y|\theta) = p(\theta)L(\theta;y), \quad (4)$$

which states that a posterior is proportional to the prior times the likelihood.

Choice of priors

Bayes' theorem shows how the prior belief is required for Bayesian analysis. A prior is the available information about the hypothesis and unknown parameters before the data are collected. The prior is classified as either an informative prior or a noninformative prior.

Noninformative priors. When no reliable prior information concerning the hypotheses or parameters exists, or an inference based only on the data at hand is desired, noninformative priors can be used. A noninformative prior does not favor any hypothesis or value of a parameter. For example, in the discrete case, the prior

$$\Pr(H_i) = 1/n, \quad i = 1, \dots, n, \quad (5)$$

is a noninformative prior because it assigns equal probability to each hypothesis H_i . For the continuous case, a similar prior is

$$\pi(\theta) = c, \quad \text{any } c > 0. \quad (6)$$

This prior is usually called an *improper prior* because its integration is infinity.

Priors with only a little information about the unknown parameters are also called noninformative priors. For example, to estimate tomorrow's temperature, we can specify a normal distribution prior with mean 0 and variance 10^6 . In this case, because of the large variance, the prior temperature information is very vague. In Bayesian analysis, the use of noninformative priors typically yields similar results to MLE.

Informative priors. Informative priors make Bayesian analysis more subjective because different priors can result in different conclusions, a situation that has been criticized by frequentists for a long time. An informative prior may be constructed from previous studies. For example, if we want to predict tomorrow's temperature, it is reasonable to use a normal distribution prior with the mean and variance equal to the mean and variance of the temperature on the same day over the past 20 years.

Clearly, the use of priors provides a way to pool and capitalize on extant scientific findings. For example, before any experiment is carried out, we may know nothing about a parameter and thus specify a noninformative prior $p(\theta)$. After

an experiment in which we obtain the data \mathbf{y}_1 , we update our knowledge about the parameter to $p(\boldsymbol{\theta}|\mathbf{y}_1)$. With an additional experiment, we obtain the data \mathbf{y}_2 , and can use the posterior $p(\boldsymbol{\theta}|\mathbf{y}_1)$ from the first experiment as the prior to update the knowledge about that parameter again.

Conjugate priors. Regardless of whether informative priors are used, one may try to use conjugate priors when they are appropriate to simplify computation. A conjugate prior is a prior from the family of probability density functions from which the derived posterior density functions have similar function forms to the priors. For example, if we use a prior with a normal distribution and derive a posterior also with a normal distribution based on the Bayes' theorem, then this prior is a conjugate prior (conjugate to the likelihood). The use of conjugate priors can usually reduce the computation complexity of the posterior distribution largely. The exponential family of distributions, which includes the normal distribution, gamma distribution, beta distribution, and so on, is the most often used family of distributions and has conjugate priors.

Statistical inference on posteriors¹

Once the posterior distribution of the parameters is obtained, statistical inference can be performed. Because we know the posterior distribution of the unknown parameters, we can plot their densities. However, such plots carry so much information that they become difficult to apprehend. Several statistics can be used to summarize the information of the posterior and are analogous to parameter estimates and standard errors from MLE. In particular, we consider point estimations and credible intervals.

Point estimation. Of the many point estimations, the mean is the most widely used. Given the posterior, the mean is calculated by

$$\bar{\boldsymbol{\theta}} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}, \tag{8}$$

which is the classical definition of the mean. Similarly, the associated variance can be obtained with

$$\text{Var}(\boldsymbol{\theta}) = \int (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) p(\boldsymbol{\theta}|\mathbf{y}) (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^t d\boldsymbol{\theta}. \tag{9}$$

These are also called the *posterior mean* and *posterior variance*, respectively.

Credible interval. In Bayesian statistics, credible intervals are used for purposes similar to those of confidence intervals in frequentist statistics. Many times, a credible interval is directly called a confidence interval. Formally, a $100 \times (1 - \alpha)\%$ credible interval (L, U) for $\boldsymbol{\theta}$ is obtained by

$$1 - \alpha = \int_L^U p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}, \tag{10}$$

with L and U are lower and upper bounds, respectively.

Because the parameter $\boldsymbol{\theta}$ is considered a random variable, we can interpret the credible interval as "The probability that $\boldsymbol{\theta}$ lies in the interval (L, U) given the observed data is at least $100 \times (1 - \alpha)\%$." In frequentist statistics, the confidence interval means that "If the experiment is repeated many times and the confidence interval is calculated each time, then overall $100 \times (1 - \alpha)\%$ of them contain the true parameter $\boldsymbol{\theta}$."

¹ Readers who are only interested in empirical studies may skip this section and proceed to the next section.

Thus, the credible interval has a more intuitively appealing interpretation.

Markov chain Monte Carlo methods

Statistical inference discussed in the previous section can be done only when the integration in Equations 8–10 can be solved implicitly. However, this is usually impossible in practice especially when multiple unknown parameters are present. In practice, MCMC methods are generally used to circumvent the difficulty of multiple dimension integration. Many different MCMC methods have been proposed, such as Metropolis–Hastings sampling, Gibbs sampling, and slicing sampling. Because Gibbs sampling is widely used, we focus on this method.

Gibbs sampling is an algorithm to generate a data point from the conditional distribution of each parameter, conditional on the current values of the other parameters (Geman & Geman, 1984). Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_q)$ with q unknown parameters in the model of interest. The conditional distribution $\pi(\boldsymbol{\theta}_i | \boldsymbol{\theta}_{-i}, \dots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \dots, \boldsymbol{\theta}_q; \mathbf{y})$ for $\boldsymbol{\theta}_i$ can usually be obtained relatively easily using Bayes' theorem. Then we can use the following scheme to sample the data points from the conditional distributions.

At the $(i + 1)$ th iteration with current value $\boldsymbol{\theta}^{(i)} = (\boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, \dots, \boldsymbol{\theta}_q^{(i)})$, update $\boldsymbol{\theta}^{(i+1)} = (\boldsymbol{\theta}_1^{(i+1)}, \boldsymbol{\theta}_2^{(i+1)}, \dots, \boldsymbol{\theta}_q^{(i+1)})$ by means of sequentially generating

$$\begin{aligned} \boldsymbol{\theta}_1^{(i+1)} &\text{ from } \pi(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2^{(i)}, \boldsymbol{\theta}_3^{(i)}, \dots, \boldsymbol{\theta}_q^{(i)}; \mathbf{y}), \\ \boldsymbol{\theta}_2^{(i+1)} &\text{ from } \pi(\boldsymbol{\theta}_2 | \boldsymbol{\theta}_1^{(i+1)}, \boldsymbol{\theta}_3^{(i)}, \dots, \boldsymbol{\theta}_q^{(i)}; \mathbf{y}) \\ &\dots \\ \boldsymbol{\theta}_q^{(i+1)} &\text{ from } \pi(\boldsymbol{\theta}_q | \boldsymbol{\theta}_1^{(i+1)}, \boldsymbol{\theta}_2^{(i+1)}, \dots, \boldsymbol{\theta}_{q-1}^{(i+1)}; \mathbf{y}). \end{aligned}$$

Namely, the first parameter is updated based on values of parameters from the previous iteration. The second parameter is updated based on the just-updated first parameter estimate and the not-yet-updated third to q th parameters. The third parameter is then updated based on the just-updated first and second parameter estimates and the not-yet-updated fourth to q th parameters. This process of updating parameters is performed up to the q th parameter to finish one complete iteration. The iteration process above can be repeated I times. Geman and Geman (1984) showed that for sufficiently large I , $\boldsymbol{\theta}^{(I)}$ can be viewed as a simulated observation from the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y})$.

The simulated observations after I are then recorded for further analysis. For convenience, these observations are denoted by $\boldsymbol{\theta}^{(m)}$, $m = 1, \dots, M$. Sometimes, there are highly positive autocorrelations between the successive observations. To reduce autocorrelation and the computing storage space, one could pick observations with fixed interval (or thin) a indexed $1, 1 + a, 1 + 2a, 1 + 3a, \dots$ to perform further analysis (Zuur, Garthwaite, & Fryer, 2002). The point estimation is calculated by

$$\bar{\boldsymbol{\theta}} = \frac{1}{N} \sum_{m=0}^{N-1} \boldsymbol{\theta}^{1+ma}, \tag{11}$$

with variance

$$\text{Var}(\boldsymbol{\theta}) = \frac{1}{N-1} \sum_{m=0}^{N-1} (\boldsymbol{\theta}^{1+ma} - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta}^{1+ma} - \bar{\boldsymbol{\theta}})^t. \tag{12}$$

To construct the credible interval, we use the percentiles of the generated sequences. For example, the lower bound of the $100 \times (1 - \alpha)\%$ credible interval is equal to the $\alpha/2$ percentile of the sequence and the upper bound is equal to the $1 - \alpha/2$ percentile.

Two keys to using Gibbs sampling are to obtain the conditional posterior distributions and determine the convergence of the generated Markov chain (or determine I). If conjugate priors are used, conditional posteriors can usually be easily obtained. Furthermore, with the emergence of new software such as WinBUGS, one does not need to specify the conditional posterior distributions explicitly. However, the convergence diagnosis of the Markov chain is still under development although several ways have been suggested to determine I . In practice, the “eyeball” method, monitoring the convergence by visually inspecting the history plots of the generated sequences, is commonly used. Usually, if there is no change point or trend in the plot, the convergence of the generated sequence is accepted. To illustrate this, we plotted the history of generated sequences of three parameters in Figure 1. At the beginning, the sequences either increased quickly (Figure 1a), or declined quickly (Figure 1b), or fluctuated a lot (Figure 1c). However, after about 500 iterations, all sequences appear very flat and there is no noticeable trend or change. Thus, the sequences converged after 500 iterations, in other words, $I = 500$. Furthermore, the sequences for all parameters converged at approximately the same time and convergence should not be accepted until the sequences for all unknown parameters have converged. For example, although

the sequence in Figure 1b seemed to converge after 200 iterations, we cannot say this sequence was converged because the other two sequences had not yet converged.

Latent basis growth curve model

Growth curve models have been widely used in the analysis of growth processes in social and behavioral research (McArdle & Nesselroade, 2003; Meredith & Tisak, 1990). Figure 2 shows a path diagram for the growth curve model used in the present analyses. The observed variables are drawn as squares, unobserved or latent variables are drawn as circles, and the constant is represented by the triangle. The squares labeled y_1 through y_4 are the observed data on occasions 1 through 4, respectively. L in the circle is the individual latent initial level, μ_L is the mean of the initial level across all the participants, and σ_L^2 is the variability around the initial level representing interindividual differences in the latent initial level. S in the circle corresponds to the slope, μ_S is the mean of the slope across all the participants, and σ_S^2 represents its variability, or individual differences around the slope. The covariance, σ_{LS} , between level and slope is represented by the double-headed arrow between the latent variables. The circles labeled e_1

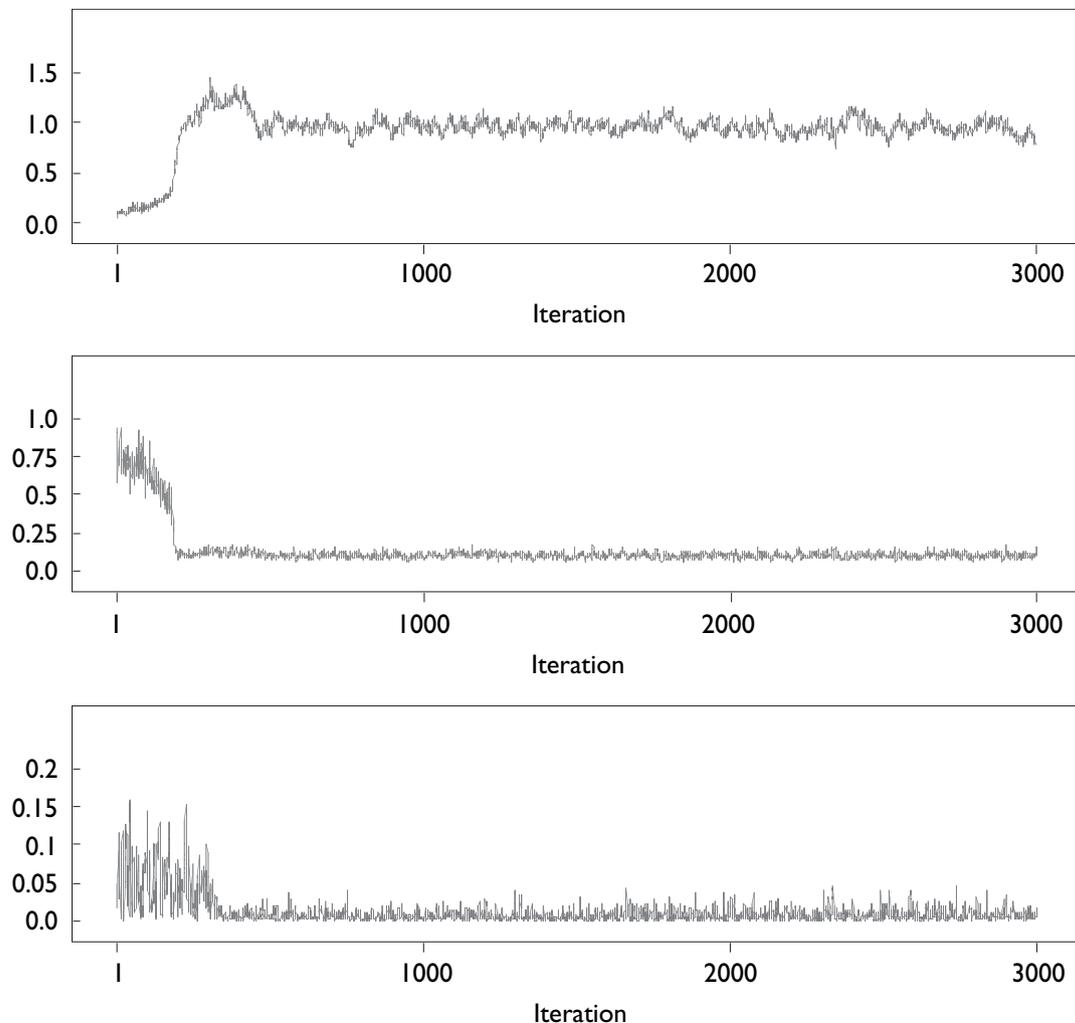


Figure 1. History plots of three parameters.

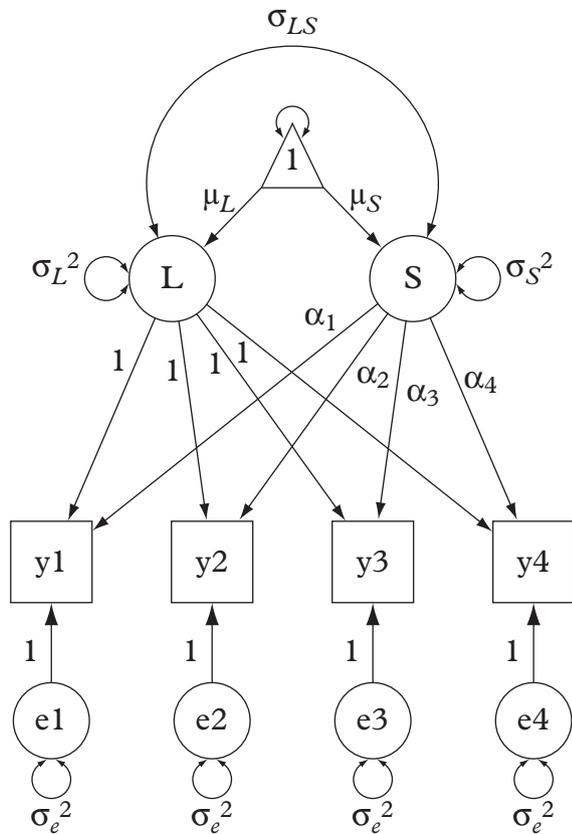


Figure 2. A growth curve model path diagram.

through e_4 are random errors, and their variance (σ_e^2) is assumed to be equal across time but it can vary across time. L and S are random-effects parameters that have different values for each individual, whereas μ_L and μ_S are fixed-effects parameters that are the same for all the participants. Mathematically, this model can be written as

$$\begin{aligned}
 y_{it} &= L_i + \alpha_i S_i + e_{it} \\
 L_i &= \mu_L + v_{Li} \quad i = 1, \dots, N; t = 1, \dots, T, \quad (13) \\
 S_i &= \mu_S + v_{Si}
 \end{aligned}$$

with N denoting the sample size and T denoting the number of occasions.

The model indicates that the observed variables y_1 – y_4 are determined by the initial level (L), the slope (S), and error (e). Different growth curve shapes can be produced by adjusting the weights of α_1 through α_4 . For example, assigning them the values 1 through 4 would lead to a linear growth curve. α_1 and α_4 can be fixed and the values of α_2 and α_3 can be estimated. This particular model, in which the weights or basis coefficients determine the shape of the growth curve, is known as a latent basis growth model, and has the advantage that the form of the function is determined by the data rather than specified a priori.

The latent basis growth model is very flexible because it can form many kinds of growth trajectories by assigning different values to α_i . The different growth curves for two individuals are plotted in Figure 3. In the first plot, $\alpha = (0, .33, .66, 1)$ indicates a linear growth trajectory. In the second plot, $\alpha = (0, .8, .95, 1)$ indicates an exponential growth trajectory. In the

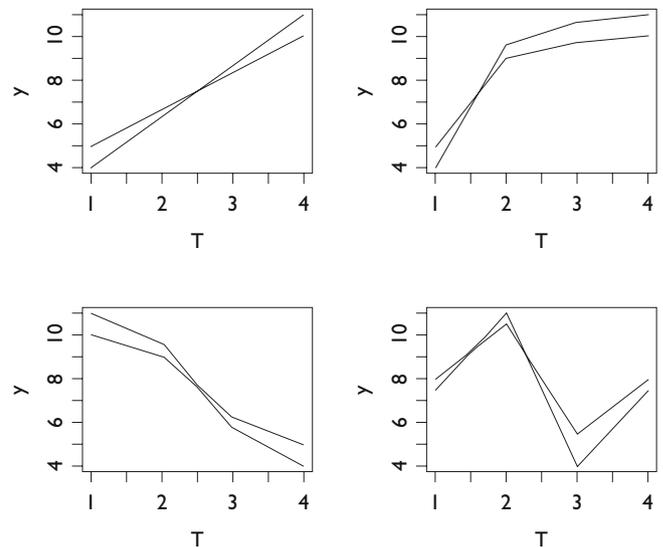


Figure 3. Manipulation of different growth curves with different basis coefficients.

third plot, $\alpha = (1, .8, .25, 0)$ indicates a nonlinear decline curve. Finally, $\alpha = (0, -1, 1, 0)$ indicates a fluctuating change trajectory. Furthermore, only two of the basis coefficients need to be fixed for the identification and scaling purposes. Illustrative is fixing the first and second coefficients to be 0 and 1 and estimating the other basis coefficients (McArdle & Nesselroade, 2003).

Empirical example

To illustrate how to practically implement Bayesian analysis, we present an empirical example.

Data

Data in this example are two subsets from the National Longitudinal Survey of Youth (NLSY).² The first subset includes repeated measurements of $N = 173$ children. At the first measurement in 1986, the children were about 6–7 years of age. The same children were then repeatedly measured at 2-year intervals for three additional measurement occasions (1988, 1990, and 1992). Missing data existed for some of the children. The second subset includes repeated measurements of $N = 34$ children. At their first measurement in 1992, the children were also about 6–7 years of age. The same children were also measured again at an approximate 2-year interval for another three times in years 1994, 1996, and 1998. Missing data also existed for several of the children. The children from both data sets were tested using the Peabody Individual Achievement Test (PIAT) Reading Recognition subtest that measured word recognition and pronunciation ability. The total score for this subtest ranged in value from 0 to 84. In the present study, this score was rescaled by 10.

² For a complete description of NLSY, visit <http://www.bls.gov/nls/>

Research questions

For the first set of data, the empirical question was “Is there systematic change in reading recognition and individual differences in this change over time (8 years)?”. The technical question to answer was “Is there difference between Bayesian estimation and MLE in parameter estimates?”. For the second set of data, the technical question to answer was “How does the use of prior information affect the parameter estimates and standard errors from a small sample?”.

Design of analyses

First, the first data set was analyzed using the Bayesian method with noninformative priors. Then the same data set was analyzed using MLE. The results from two methods were compared. Second, the second data set was first analyzed using the Bayesian methods with noninformative priors as the baseline. Then the informative priors were constructed from the results of the first data set and used in the analysis of the second data set. To further demonstrate the use of informative priors in the small sample study, we selected part of the data ($N = 20$) from the second data set. Both full informative priors and half informative priors as defined in the next section were used to analyze the selected data.

In all Bayesian analysis, conjugate priors were used. For the means of initial level and slope, the normal distribution prior was used. For the variance of measurement error, the inverse gamma distribution prior was used and for the covariance matrix of the random effects parameters, the inverse Wishart distribution prior was used. Finally, normal distribution priors were used for the basis coefficients.

Implementation of the analyses

All Bayesian analyses were implemented using WinBUGS and BAUW. WinBUGS is widely used free software for Bayesian analysis and is very flexible for both simple and complex models. Although WinBUGS can be used as menu-driven software, it is still not easy to use for at least two reasons. First, programming a model in WinBUGS requires understanding the details of the probability form for that model, which is usually missing in most textbooks especially for empirical researchers. Second, the data format of WinBUGS is similar to that of R or Splus, which makes the data conversion frustrating.

In order to advance the application of Bayesian analysis and make the implementation of the analysis easier, the software called BAUW (Zhang & Wang, 2006) was used. BAUW can be used to generate WinBUGS programs for many kinds of models, such as growth curve models, latent difference score models (McArdle & Hamagami, 2001), and IRT models (Embretson & Reise, 2000). BAUW is the menu-driven software and only needs a few inputs to generate full WinBUGS programs, including data conversion. For the analysis of the first data set by using the latent basis model, we need to input the sample size (173), the number of occasions (4), the missing data indicator (dot “.” in the data file for this example), and the data file. WinBUGS program generated from BAUW is given in Appendix A. The program can then be run in WinBUGS by clicking menus.³

³ For researchers who use WinBUGS for the first time, a video has been created to demonstrate the analytic process. To view it, go to <http://bauw.psychstat.org>.

As shown in Appendix A, a complete WinBUGS program has three parts, model specification, starting values, and data. In the model specification part, all codes for the latent basis model using WinBUGS syntax are put in `model{ }`. Then for each parameter in the model, a prior was specified. Notice that neither posterior nor conditional distributions were specified. WinBUGS calculated them based on the model and specified priors. Because the precision (reciprocity of the variance) was used in the normal distribution (*dnorm*), it was transformed back to variance in the transformation part. Finally, all the parameters were put into a parameter vector for convenience. All starting values were put together into `list()`. Usually, researchers need to change the starting values according to their data for fast convergence. In this example, we used the mean score at the first occasion as the starting value for the mean initial level parameter (μ_L) and used the difference between the mean scores from the first occasion and the fourth occasion as the starting value for the slope parameter (μ_S). Finally, the data were also put together into `list()`. In WinBUGS, the missing data are represented by “NA”. All missing data in the original data file indicated by dot “.” were replaced by “NA” in WinBUGS data. In the codes generated by BAUW, the residual variances are assumed to be equal at each time point and the first and last basis coefficients are fixed at 0 and 1. However, WinBUGS allows the residual variances to be estimated at each time point. Similarly, one can fix any two basis coefficients and estimate the others.⁴

BAUW always generates the WinBUGS scripts using noninformative priors. To use informative priors, the priors need to be modified. In the current example, the mean initial level parameter (μ_L) was initially given a noninformative prior `bL[1]-dnorm(0, 1.0E-6)`. To use the informative prior constructed from the first data set, this prior needs to be changed to `bL[1]-dnorm(2.047, 410)` with the precision 410 equal to 1 over the variance .0024. This prior can be called a full informative prior because it used all the parameter information from the first data set. A half informative prior was also used which was `bL[1]-dnorm(2.047, 200)`. In this case, the precision was only half of that from the first data set, which means only partial information from the first data set was incorporated in analyzing the second data set. Note that the half informative priors are not exactly same with those discussed by Ibrahim and Chen (2000). They generally used the power a_0 to specify informative priors.

Similarly, the priors for the other parameters need to be changed accordingly. The complete specifications of informative priors are given in Table 1. The complete WinBUGS codes for the use of the full informative prior are provided in Appendix B. For maximum likelihood estimation, SAS PROC NLMIXED was used to obtain the parameter estimates.

Results

The results from both Bayesian methods and MLE for the first data set are provided in Table 2. For Bayesian methods, the history plot showed that the generated sequences for all parameters converged after 1000 iterations. The iterations from 1000 to 5000 were used to calculate the parameter estimates and the credible intervals. The parameter estimates and standard errors using Bayesian estimation methods from

⁴ An example script file can be found at <http://bauw.psychstat.org>. These features will be incorporated into the next version of BAUW.

Table 1
(Non-)Informative priors used in WinBUGS for the analysis of the second data set

	<i>Noninformative priors</i>	<i>Full informative priors</i>	<i>Half informative priors</i>
α_2	$dnorm(0, 1.0E-6)$	$dnorm(.5, 6957)$	$dnorm(.5, 3500)$
α_3	$dnorm(0, 1.0E-6)$	$dnorm(.8, 5131)$	$dnorm(.8, 2500)$
μ_L	$dnorm(0, 1.0E-6)$	$dnorm(2.047, 410)$	$dnorm(2.047, 200)$
μ_S	$dnorm(0, 1.0E-6)$	$dnorm(3.542, 127)$	$dnorm(3.542, 60)$
$1/\sigma_e^2$	$dgamma(.001, .001)$	$dgamma(121.5, 20.4)$	$dgamma(60, 10)$
$\begin{pmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_S^2 \end{pmatrix}$	$dwish\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, 2\right]$	$dwish\left[\begin{pmatrix} 9.34 & 7.34 \\ 7.34 & 29.76 \end{pmatrix}, 40\right]$	$dwish\left[\begin{pmatrix} 4.67 & 3.67 \\ 3.67 & 14.88 \end{pmatrix}, 20\right]$

Note. All priors were constructed based on the results from the analysis of the first data set. Because of the multi-variate complexity of Wishart distribution, the Wishart priors are only an approximation.

WinBUGS were very similar to those using MLE from SAS PROC NLMIXED. There was slight difference in the estimates of covariance matrix of initial level and slope because the two programs deal with missing data in different ways.⁵ From our simulation study, we found that the two methods obtained exactly the same results when there was no missing data. When there were missing data, the two methods obtained similar results. However, Bayesian methods gave slightly smaller standard errors.⁶

In terms of the empirical question raised above, first, the average initial level of reading recognition was 2.05 (μ_L) and the variance for the initial level was .25 (σ_L^2). The variance was significantly larger than 0, so we can conclude there was individual variability in the initial level.⁷ Second, the overall change of reading recognition was about 3.54 (μ_S) across the eight years. There were also significant individual differences in this change because the variance of this change ($\sigma_S^2 = .80$) was significantly larger than 0. Furthermore, change in reading recognition was positively correlated with the initial level ($\sigma_{LS} = .18$). Finally, the change of reading recognition was nonlinear as approximately 50% of the overall change occurred between 6–7 years of age and 8–9 years of age.

The parameter estimates and associated standard errors from the second data set ($N = 34$) and its subset ($N = 20$) using noninformative prior are given in Table 3. The standard errors were very large and those for $N = 34$ were smaller than those for $N = 20$. Particularly when $N = 20$, the variance for the slope and the covariance between the initial level and slope were not significant any more. Based on the results of the first data set, the nonsignificant results were because of the small sample size.

When estimating the data again with the half and full informative priors, the standard errors were largely reduced. It was also clear that with more information in the priors the standard errors became smaller by comparing those obtained using full informative priors and half informative priors. This illustrates that using the informative prior can increase the

Table 2
Estimates of the parameters from the first data set for the latent basis growth model

	<i>Bayesian</i>			<i>MLE</i>	
	<i>Estimate</i>	<i>SE</i>	<i>CI</i>	<i>Estimate</i>	<i>SE</i>
α_1	0	= 0	—	= 0	—
α_2	0.50	0.012	0.48	0.52	0.012
α_3	0.80	0.014	0.77	0.83	0.014
α_4	= 1	—	—	= 1	—
μ_L	2.05	0.049	1.95	2.15	0.048
μ_S	3.54	0.089	3.37	3.72	0.088
σ_L^2	0.25	0.044	0.17	0.34	0.042
σ_S^2	0.80	0.139	0.55	1.10	0.138
σ_{LS}	0.18	0.054	0.07	0.28	0.056
σ_e^2	0.17	0.015	0.14	0.20	0.016

SE, standard error; CI, credible interval.

Table 3
Parameter estimates with different priors for the second data set

	<i>Noninformative</i>		<i>Half informative</i>		<i>Full informative</i>	
	<i>Estimate</i>	<i>SE</i>	<i>Estimate</i>	<i>SE</i>	<i>Estimate</i>	<i>SE</i>
<i>N = 34</i>						
α_2	0.48	0.035	0.49	0.014	0.50	0.011
α_3	0.80	0.037	0.80	0.017	0.80	0.012
μ_L	2.04	0.139	2.04	0.060	2.04	0.045
μ_S	3.24	0.225	3.46	0.109	3.49	0.081
σ_L^2	0.39	0.145	0.29	0.072	0.27	0.053
σ_S^2	1.06	0.379	0.98	0.233	0.92	0.176
σ_{LS}	0.27	0.155	0.26	0.093	0.24	0.070
σ_e^2	0.29	0.049	0.21	0.022	0.20	0.015
<i>N = 20</i>						
σ_2	0.50	0.054	0.50	0.015	0.50	0.011
σ_3	0.77	0.055	0.79	0.017	0.79	0.013
μ_L	2.18	0.228	2.07	0.065	2.06	0.047
μ_S	3.13	0.295	3.46	0.117	3.49	0.085
σ_L^2	0.66	0.305	0.36	0.104	0.31	0.067
σ_S^2	0.98	0.515	0.97	0.266	0.90	0.183
σ_{LS}	0.38	0.254	0.30	0.117	0.26	0.082
σ_e^2	0.36	0.077	0.22	0.024	0.20	0.016

SE, standard error.

⁵ SAS uses the full information maximum likelihood estimation method in which only the time point with data is used in estimating the parameters. WinBUGS takes every missing data as a parameter to sample from its posterior distribution.

⁶ An article on the simulation is still in development. Interested researchers may refer to the preliminary results at <http://bauw.psychstat.org>.

⁷ A strict way to test whether the variance is equal to 0 is to fit a reduced model and compare the difference in the change of fit statistics. Here, we simply used the t test that is usually employed in MLE.

statistical efficiency and power. The informative priors can be viewed as additional or extra data. Sometimes, researchers may “pool” data sets, which is analogous to the use of informative priors. Using informative priors is like to pooling one set of data with another.

After using the informative priors, the results from the first and second data sets were very consistent and the conclusions reached were the same for the two data sets. This is reasonable because the two data sets measured participants with the same age from two different cohorts (1979/1980 cohort and 1985/1986 cohort).

Discussion

The power of Bayesian methods in estimating complex models for complex data analysis is indisputable. Besides the capability for implementing estimation procedures, which generally cannot be done in MLE, the ease, flexibility, and computation time were also very acceptable (Arminger & Muthen, 1998; Dunson, 2003; McArdle & Wang, in press). Other merits of Bayesian methods have also been demonstrated throughout the present article. First, Bayesian methods interpret traditional statistics in a more intuitive way. For example, the meaning of the credible interval and p -value matches the common sense interpretation of these concepts. Second, Bayesian methods provide a clear way to incorporate prior information that both increases the statistical power of analysis and formulizes the accumulation of scientific findings. In the current study, even when the sample size was only 20, we still obtained reasonable results through using informative priors. Because Bayesian inference is not based on the asymptotic nature of the estimators as MLE is (Casella & Berger, 2001), it can be argued to be a more plausible way to analyze small sample data sets (Rindskopf, 2006).

We believe that it is timely for empirical researchers to give serious consideration to trying out Bayesian methods for their initial data analysis. To facilitate such efforts, the primary goal of this article has been to provide the basics of Bayesian methods in the context of a popular modeling technique (latent growth curve models) and to render the application of Bayesian methods practical by using currently available software. To meet the special needs of longitudinal data analysts, we demonstrated the step-by-step application of Bayesian methods on a latent basis growth model using an empirical data set. The results showed that Bayesian methods can obtain similar parameter estimates to those from MLE when using noninformative priors. More importantly, they had unique strengths, such as the intuitive interpretation of the results and the efficient incorporation of prior information to empirical data analysis.

Although Bayesian methods can be used as the direct alternative to MLE for parameter estimation when using noninformative priors, we would like to emphasize the application of informative priors. Progress in scientific research rests on *accumulated* knowledge. The Bayesian methods and Bayes' theorem provide a natural “chain” for incorporating previous findings with current findings through the use of informative priors, which generally cannot be done formally in traditional statistics. Thus, informative priors should be used if reliable prior information is available. Furthermore, prior information is usually available for social and behavioral research. One type of prior information arises from theory. The

theory underlying an experiment can be used to build our model. This theory actually acts as prior information. A second type of prior has been widely used in quantitative models. For example, in factor analytic models and IRT models, factor scores are usually assumed to come from a normal distribution with means of 0 and 1. This kind of specification is equivalent to specifying the prior distribution in Bayesian framework. The third type of prior information arises from previous research outcomes. In the present study, we used the results from modeling growth of reading recognition of 6–7-year-old children from a 1979/1980 cohort to construct informative priors for the study of 6–7-year-old children from a 1985/1986 cohort.

Bayesian methods using informative priors can also be viewed as the alternative to meta-analysis (Wolf, 1986) and mega-analysis (McArdle & Horn, 2004). A meta-analysis combines the results of several closely related studies and a mega-analysis combines the raw data of several related studies directly. Because raw data from different studies are usually not available, the use of the research results from public sources is a more frequent occurrence than is the use of raw data. Bayesian methods using informative priors actually provide a feasible way to combine data of a current study with the results from related previous related studies, which is more practical than meta-analysis and mega-analysis. Our present analyses also demonstrate that we can use only partially available information in Bayesian analysis. If the prior information is not very reliable, a less informative prior can be constructed from all the available information.

Bayesian methods have been criticized for the choice of priors. For example, in the present analyses, when we analyzed the same subset of participants with noninformative priors, full informative priors, and half informative priors, the results were different. Discrepancies may become greater when different researchers use different prior information because of their accessibility to currently available information. Thus, when priors are used, these priors should be reported explicitly. Furthermore, when conclusions are drawn, the use of priors should be kept in mind.

To simplify the demonstration, we employed a univariate growth curve model. However, the Bayesian methods used here can be expanded to include multiple growth models to evaluate how intraindividual differences in intraindividual change covary across more than one set of variables (Hamagami, Zhang, & McArdle, submitted). Furthermore, Bayesian methods can be used to estimate more complex nonlinear models, such as univariate and bivariate nonlinear change points models (McArdle & Wang, in press). Besides DIC produced by WinBUGS (Spiegelhalter, Best, Carlin, & Linde, 2002), the other statistics can be constructed based on the log-likelihood statistics given by WinBUGS to compare models.

Programming and computational demands have limited the application of Bayesian methods in the past. However, the availability of WinBUGS and BAUW and other software now renders the programming process a series of mouse-clicking. Also, current computational power makes computing time no longer a significant concern. For example, in this study, it took a laptop with Celeron CPU = 1.7 GHZ and RAM = 512 less than 30 seconds to finish the computations. More importantly, complexities of models make the programming and computational demands for Bayesian methods even less than those for the traditional MLE methods (McArdle & Wang, in press).

References

- Arminger, G., & Muthen, B.O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika*, *63*, 271-300.
- Baltes, P.B., & Nesselroade, J.R. (1979). History and rationale of longitudinal research. In J.R. Nesselroade & P.B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1-39). New York: Academic Press.
- Bartholomew, D.J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology*, *34*, 93-99.
- Blozis, S.A., Conger, K.J., & Harring, J. (2007). Nonlinear latent curve models for multivariate longitudinal data. *International Journal of Behavioral Development*, *31*, 340-346.
- Box, G.E.P., & Tiao, G.C. (1992). *Bayesian inference in statistical analysis*. New York: Wiley.
- Carlin, B.P., & Louis, T.A. (2000). *Bayes and empirical Bayes methods for data analysis* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.
- Casella, G., & Berger, R.L. (2001). *Statistical inference* (2nd ed.). Pacific Grove, CA: Duxbury Press.
- Chang, H.-H. (1996). The asymptotic posterior normality of the latent trait for polytomous IRT models. *Psychometrika*, *61*, 445-463.
- Demidenko, E. (2004). *Mixed models: Theory and applications*. New York: Wiley.
- Dunson, D. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association*, *98*(463), 555-563.
- Eaves, L., & Erkanli, A. (2003). Markov Chain Monte Carlo approaches to analysis of genetic and environmental components of human developmental change and G X E interaction. *Behavior Genetics*, *33*, 279-299.
- Edwards, W., Lindman, H., & Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, *70*, 193-242.
- Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Fox, J.-P., & Glas, C.A.W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*, 271-288.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721-741.
- Hamagami, F., Zhang, Z., & McArdle, J.J. (submitted). Comparison of parameter estimations of the latent difference score model using structure equation model and Bayesian method.
- Ibrahim, J.G., & Chen, M.-H. (2000). Power prior distributions for regression models. *Statistical Science*, *15*, 46-60.
- Laird, N.M., & Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963-974.
- Lee, J.C., & Chang, C.H. (2000). Bayesian analysis of a growth curve model with a general autoregressive covariance structure. *Scandinavian Journal of Statistics*, *27*, 703-713.
- Lee, J.C., & Liu, K.-C. (2000). Bayesian analysis of a general growth curve model with predictions using power transformations and AR(1) autoregressive dependence. *Journal of Applied Statistics*, *27*, 321-336.
- Lee, P.M. (2004). *Bayesian statistics: An introduction* (3rd ed.). London: Arnold.
- Lee, S. (1981). A Bayesian approach to confirmatory factor analysis. *Psychometrika*, *46*, 153-160.
- McArdle, J.J. (1988). Dynamic but structural equation modeling of repeated measures data. In J.R. Nesselroade & R.B. Cattell (Eds.), *The handbook of multivariate experimental psychology*, *2* (pp. 561-614). New York: Plenum Press.
- McArdle, J.J. (2001). A latent difference score approach to longitudinal dynamic structural analyses. In R. Cudeck, S. du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future* (pp. 342-380). Lincolnwood, IL: Scientific Software International.
- McArdle, J.J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110-133.
- McArdle, J.J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 139-175). Washington, DC: American Psychological Association.
- McArdle, J.J., & Horn, J.L. (2004). *A mega analysis of the WAIS: Adult intelligence across the life-span*. Mahwah, NJ: Erlbaum.
- McArdle, J.J., & Nesselroade, J.R. (2003). Growth curve analysis in contemporary psychological research. In J. Schinka & W. Velicer (Eds.), *Comprehensive handbook of psychology: Research methods in psychology* (Vol. 2, p. 447-480). New York: Wiley.
- McArdle, J.J., & Wang, L. (in press). Modeling age-based turning points in longitudinal life-span growth curves of cognition. In P. Cohen (Ed.), *Turning points research*. Mahwah, NJ: Erlbaum.
- Menzeffricke, U. (1998). Bayesian prediction in growth-curve models with correlated errors. *Test*, *8*(1), 75-93.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107-122.
- Pettitt, A.N., Tran, T.T., Haynes, M.A., & Hay, J.L. (2006). A Bayesian hierarchical model for categorical longitudinal data from a social survey of immigrants. *Journal of Royal Statistical Society, A*, *169*, 97-114.
- Rindskopf, D. (2006, June). *Some neglected relationships between cognitive psychology and statistics: How people try to think using advanced statistical methods without realizing it*. Paper presented at the 71st Annual Meeting of the Psychometric Society, Montreal, Canada.
- Rupp, A.A., Dey, D.K., & Zumbo, B.D. (2004). To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*, 424-451.
- Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, *64*, 37-52.
- Seltzer, M.H., Wong, W.H.M., & Bryk, A.S. (1996). Bayesian analysis in applications of hierarchical models: Issues and methods. *Journal of Educational Statistics*, *21*, 131-167.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & Linde, A. v. d. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*, 583-639.
- Spiegelhalter, D.J., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS manual version 1.4. (MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, UK) <http://www.mrc-bsu.cam.ac.uk/bugs>
- Tucker, I.R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika*, *23*, 19-23.
- Walker, L.J., Gustafson, P., & Frimer, J.A. (2007). The application of Bayesian analysis to issues in developmental research. *International Journal of Behavioral Development*, *31*.
- Wishart, J. (1938). Growth rate determinations in nutrition studies with the bacon pig, and their analyses. *Biometrika*, *30*, 16-28.
- Wolf, F.M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: SAGE.
- Zhang, Z., & Wang, L. (2006). *BAUW: Bayesian analysis using WinBUGS, Version 1.0*. [Computer software and manual]. Available: <http://bauw.psychstat.org>
- Zuur, G., Garthwaite, P.H., & Fryer, R.J. (2002). Practical use of MCMC methods: Lessons from a case study. *Biometrical Journal*, *44*, 433-455.

Appendix A

WinBUGS program for the latent basis model generated from BAUW

```
#model specification
model {
  for (i in 1:N) {
    LS[i,1:2]~dmnorm(Mu[i,1:2], Inv_cov[1:2,1:2])
    Mu[i,1]<-bL[1]
    Mu[i,2]<-bS[1]
    for (t in 1:T) {
      y[i,t]~dnorm(MuY[i,t], Inv_sig_e)
      MuY[i,t]<-LS[i,1]+LS[i,2]*A[t]
    }
  }
}

#Prior distribution, can be changed to use informative prior
for (i in 1:1) {
  bL[i]~dnorm(0,1.0E-6)
  bS[i]~dnorm(0,1.0E-6)
}
A[1]<-0
for (t in 2:T-1) {
  A[t]~dnorm(0,1.0E-6)
}
A[T]<-1
Inv_cov[1:2,1:2]~dwish(R[1:2,1:2], 2)
R[1,1]<-1
R[2,2]<-1
R[2,1]<-R[1,2]
```

```

R[1,2]<-0
Inv_sig_e~dgamma(.001,.001)
#Transform the parameters
Cov[1:2,1:2]<-inverse(Inv_cov[1:2,1:2])
Sig_L<-Cov[1,1]
Sig_S<-Cov[2,2]
rho<-Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
Sig_e<-1/Inv_sig_e
#all parameter are put into Para
Para[1]<-Sig_L
Para[2]<-Sig_S
Para[3]<-Cov[1,2]
Para[4]<-rho
Para[5]<-Sig_e
Para[6]<-bL[1]
Para[7]<-bS[1]
Para[8]<-A[2]
Para[9]<-A[3]
} #end of model part

#Starting values
#You can change the starting values by yourself here.
list(Inv_cov= structure(.Data = c(1,0,0,1),.Dim=c(2,2)), Inv_sig_e=1,
     A=c(NA,0.333333,0.666667,NA),
     bL=c(2),bS=c(3))

#Data
list(N=173,T=4,y = structure(.Data = c(2.6,4.9,5.5,7.2,
.....
1.8,3.9,NA,NA), .Dim = c(173,4)))

```

Appendix B

WinBUGS program with informative priors

```

model{
  for (i in 1:N){
    LS[i,1:2]~dmnorm(Mu[i,1:2], Inv_cov[1:2,1:2])

```

```

    Mu[i,1]<-bL[1]
    Mu[i,2]<-bS[1]
    for (t in 1:T){
      y[i,t]~dnorm(MuY[i,t], Inv_sig_e)
      MuY[i,t]<-LS[i,1]+LS[i,2]*A[t]
    }
  }
#Prior distribution, can be changed to use informative prior
for (i in 1:1){
  bL[i]~dnorm(2.047, 410)
  bS[i]~dnorm(3.542, 127)
}
A[1]<-0
A[2]~dnorm(.5, 6957)
A[3]~dnorm(.8, 5131)
A[T]<-1
Inv_cov[1:2,1:2]~dwish(R[1:2,1:2], 40)
R[1,1]<-9.344603
R[2,2]<-29.765950
R[2,1]<-R[1,2]
R[1,2]<-7.343389
Inv_sig_e~dgamma(121.5, 20.4)
#Transform the parameters
Cov[1:2,1:2]<-inverse(Inv_cov[1:2,1:2])
Sig_L<-Cov[1,1]
Sig_S<-Cov[2,2]
rho<-Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
Sig_e<-1/Inv_sig_e
#all parameter are put into Para
Para[1]<-Sig_L
Para[2]<-Sig_S
Para[3]<-Cov[1,2]
Para[4]<-rho
Para[5]<-Sig_e
Para[6]<-bL[1]
Para[7]<-bS[1]
Para[8]<-A[2]
Para[9]<-A[3]
} #end of model part

```