Assessing the Impact of Precision Parameter Prior in Bayesian Nonparametric Growth Curve Modeling

Xin Tong

University of Virginia

Zijun Ke

Sun Yat-sen University

Author Note

This paper is based upon work supported by the National Science Foundation under grant no. SES-1951038. Correspondence concerning this article should be addressed to Zijun Ke, Department of Psychology, Sun Yat-sen University, Higher Education Mega Center, Guangzhou, Guangdong 510006, China. Email: keziyun@mail.sysu.edu.cn.

Abstract

Bayesian nonparametric (BNP) modeling has been developed and proven to be a powerful tool to analyze messy data with complex structures. Despite the increasing popularity of BNP modeling, it also faces challenges. One challenge is the estimation of the precision parameter in the Dirichlet process mixtures. In this study, we focus on a BNP growth curve model and investigate how noninformative prior, weakly informative prior, accurate informative prior, and inaccurate informative prior affect the model convergence, parameter estimation, and computation time. A simulation study has been conducted. We conclude that the noninformative prior for the precision parameter is less preferred because it yields a much lower convergence rate, and growth curve parameter estimates are not sensitive to informative priors.

Keywords: Nonparametric Bayesian modeling, Growth curve modeling, Robust method, Dirichlet process mixture, Precision parameter prior.

Assessing the Impact of Precision Parameter Prior in Bayesian Nonparametric Growth Curve Modeling

Bayesian nonparametric (BNP) modeling, also called semiparametric Bayesian modeling in the literature, has been recognized as a valuable data analytical technique due to its great flexibility and adaptivity (e.g., Gershman & Blei, 2012; Müller & Mitra, 2004). It is rapidly gaining popularity among methodologists and practitioners and has been applied to a variety of models including regressions, latent variable models with complex structures, sequential models, etc. BNP models are on an infinite dimensional parameter space and the complexity of the models adapts to the data. One of the most popular BNP models is Dirichlet process (DP) mixtures. Being able to adapt the number of latent classes to the complexity of the data, DP mixtures are powerful in modeling empirical data. However, they also face technical challenges. One challenge is the estimation of the precision parameter in the DP mixture. In this study, we focus on the prior of precision parameter and investigate how it affects model convergence, parameter estimation, and computation time in BNP growth curve modeling.

Growth curve models are broadly used in longitudinal research (e.g., McArdle & Nesselroade, 2014; Meredith & Tisak, 1990). Many popular longitudinal models in social and behavioral sciences, such as multilevel models, some mixed-effects models, and linear hierarchical models, can be written as a form of growth curve models. In growth curve models, dependent variables are repeatedly measured and explained as a function of time and possible control variables. The mean function between the dependent variables and time is the mean growth. Random effects and measurement errors cause the individual growth trajectories to deviate from the mean growth curve. Traditional growth curve modeling is typically based on the normality assumption. That is, both the random effects and measurement errors are assumed to follow normal distributions. However, empirical data often violate the normality assumption (Cain et al., 2017; Micceri, 1989). Nonnormal population distributions and data contamination are two common causes of nonnormality. Although standard errors and test statistics have been corrected to reduce the adverse effect of distributional assumption violation (e.g., Chou et al.,

1991; Curran et al., 1996), normal-distribution-based maximum likelihood estimation may still yield inefficient or inaccurate parameter estimates, and thus misleading statistical inferences (e.g., Maronna et al., 2006; Yuan & Bentler, 2001). Therefore, researchers have developed robust methods to obtain accurate parameter estimation and statistical inference.

The ideas of robust methods can be divided into two types. For the first type, the key idea is to downweight extreme cases. To do so, this type of robust methods assign a weight to each subject in a dataset according to its distance from the center of the majority of the data (e.g., Pendergast & Broffitt, 1985; Silvapulle, 1992; Singer & Sen, 1986; Yuan & Bentler, 1998; Zhong & Yuan, 2010). For the second type, the key idea is to use nonnormal distributions that are mathematically tractable while building the statistical model. For example, latent variables and/or measurement errors are assumed to follow a t or skew-t distribution (Tong & Zhang, 2012; Zhang, 2016) or a mixture of certain distributions (Muthén & Shedden, 1999; Lu & Zhang, 2014). While being useful, these methods have limitations under certain conditions. For example, the downweighting method does not perform well when latent variables contain extreme scores (see Zhong & Yuan, 2011). Using a t distribution or a mixture of normal distributions still imposes restrictions on the shape of the data distribution.

The aforementioned issues are automatically resolved by BNP methods. BNP modeling relies on a building block, Dirichlet process (DP), to handle the nonnormality issue. DP is a distribution over probability measures that can be used to estimate unknown distributions. Consequently, the nonnormality issue can be addressed by directly estimating the unknown random distributions of latent variables or measurement errors (i.e., obtaining the posteriors of the distributions).

The advantages of using BNP methods with DP priors have been discussed in the literature (e.g., Fahrmeir & Raach, 2007; Ghosal et al., 1999; Hjort, 2003; Hjort et al., 2010; Müller & Mitra, 2004; MacEachern, 1999). They do not constrain models to a specific parametric form which may limit the scope and type of statistical inferences in many situations, especially when data are not normally distributed. Thus, a typical motivation of using BNP methods is that one is

unwilling to make somewhat arbitrary and unverified assumptions for latent variables or error distributions as in the parametric modeling. Meanwhile, BNP methods can provide full probability models for the data-generating process and lead to analytically tractable posterior distributions.

BNP methods have been applied to complex models. For example, Bush and MacEachern (1996), Kleinman and Ibrahim (1998), and Brown and Ibrahim (2003) used DP mixtures to handle nonnormal random effects. Burr and Doss (2005) used a conditional DP to handle heterogeneous effect sizes in the context of meta-analysis. Ansari and Iyengar (2006) included Dirichlet components to build a semiparametric recurrent choice model. Dunson (2006) used dynamic mixtures of DP to estimate the varied distributions of a latent variable which change nonparametrically across groups. Si and Reiter (2013) and Si, et al. (2015) used DP mixtures of multinomial distributions for categorical data with missing values. BNP approach has also been adapted to structural equation modeling to relax the normality assumption of the latent variables (e.g., Lee et al., 2008; Yang & Dunson, 2010). Tong and Zhang (2019) directly used a DP mixture to model nonnormal data in growth curve modeling.

Although the application of BNP modeling has increased dramatically since the theoretical properties of BNP methods were better understood and their computational hurdles were removed (e.g., Neal, 2000), BNP modeling is still unfamiliar to the majority of researchers in social and behavioral sciences. Additionally, there are technical issues that have not yet been fully addressed (Sharif-Razavian & Zollmann, 2009). The convergence issue is one of such unanswered questions. Nonconvergence can occur when BNP method is applied to complex models. Tong and Zhang (2019) found that nonconvergence was largely caused by the precision parameter of the mixing DP. The precision parameter is a critical hyperparameter that governs the expected number of mixture components. When a noninformative prior was used for the precision parameter, nonconvergence occurred or a longer computation time was observed (Tong & Zhang, 2019). Informative priors may help solve this issue. However, only a few studies have noticed and discussed the effect of the precision parameter in DP mixtures (e.g., Jara et al., 2011; Ohlssen

et al., 2007; West, 1992). Ishwaran (2000) was among the few that studied the informative prior for the precision parameter. Ishwaran (2000) suggested to use the $Gamma(2, 2)$ prior to encourage both small and large values of the precision parameter. In sum, despite its impact on the model convergence issue, no study has systematically investigated how the prior for the precision parameter should be specified.

Therefore, in this study, we evaluate and compare noninformative, weakly informative, accurate informative, and inaccurate informative priors for the precision parameter of DP mixtures. We study how these priors influence model convergence, model estimation, and computation time in BNP growth curve modeling. In the next section, we introduce BNP growth curve modeling. After providing the conditional posterior distribution of the precision parameter, we use a simulation study to assess the impact of four types of priors for the precision parameter. Recommendations are provided at the end of the article. We also provide a guideline about the implementation of BNP growth curve modeling using R (R Core Team, 2019) in the appendix.

Bayesian Nonparametric Growth Curve Modeling

We now introduce a typical growth curve model and a BNP method based on this model. Consider a longitudinal dataset with N subjects and T measurement occasions. Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ be a $T \times 1$ random vector with y_{ij} being a measurement from individual i at time j ($i = 1, \ldots, N; j = 1, \ldots, T$). A growth curve model without covariates can be written as

$$\begin{aligned} \mathbf{y}_i &= \mathbf{\Lambda}\mathbf{b}_i + \mathbf{e}_i, \\ \mathbf{b}_i &= \boldsymbol{\beta} + \mathbf{u}_i, \end{aligned}$$

where $\mathbf{\Lambda}$ is a $T \times q$ factor loading matrix that determines the growth curves, $\mathbf{b}_i$ is a $q \times 1$ vector of random effects, and $\mathbf{e}_i$ is a vector of measurement errors. The vector of random effects $\mathbf{b}_i$ varies around its mean $\boldsymbol{\beta}$. The residual vector $\mathbf{u}_i$ represents the deviation of $\mathbf{b}_i$ from $\boldsymbol{\beta}$. When

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & T-1 \end{pmatrix}, \mathbf{b}_i = \begin{pmatrix} L_i \\ S_i \end{pmatrix}, \text{ and } \boldsymbol{\beta} = \begin{pmatrix} \beta_L \\ \beta_S \end{pmatrix},$$

the model is reduced to a linear growth curve model with random intercept L_i and random slope S_i. The mean intercept and slope are denoted as β_L and β_S, respectively.

Traditionally, $\mathbf{e}_i$ and $\mathbf{u}_i$ are assumed to follow multivariate normal distributions with mean vectors of zero and covariance matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$, respsectively, so $\mathbf{e}_i \sim MN_T(\mathbf{0}, \mathbf{\Phi})$ and $\mathbf{u}_i \sim MN_q(\mathbf{0}, \mathbf{\Psi})$. Here MN denotes a multivariate normal distribution and its subscript indicates its dimension. Measurement errors are often assumed to be uncorrelated with each other and have equal variances across time. Statistically, this simplification means the covariance matrix of measurement error $\mathbf{\Phi}$ is reduced to $\mathbf{\Phi} = \sigma_e^2\mathbf{I}$ where σ_e^2 is a scale parameter. In linear growth curve models, $\mathbf{u}_i = (u_{Li}, u_{Si})'$. Its covariance matrix is then $\mathbf{\Psi} = cov(\mathbf{u}_i) = \begin{pmatrix} \sigma_L^2 & \sigma_{LS} \\ \sigma_{LS} & \sigma_S^2 \end{pmatrix}$. Here σ_L^2 and σ_S^2 represent the variances of the random intercept and slope across individuals, respectively, and σ_{LS} represents the covariance between the random intercept and slope.

BNP methods do not make arbitrary distributional assumptions as in the parametric modeling and thus are more flexible in handling nonnormal data (e.g., Lee et al., 2008; Tong & Zhang, 2019). Unlike conventional nonparametric methods such as permutation tests, BNP methods use full probability models to describe the data-generating process and thus can derive posterior distributions for model parameters.

Within the BNP modeling scope, the parametric distributions of latent variables and measurement errors in traditional methods are replaced by unknown random distributions. To estimate these unknown distributions, Dirichlet process is frequently used as the prior (Ferguson, 1973, 1974). Specifically, a random "sample" from a DP is a random distribution. Here we denote it as G. A DP has two hyperparameters, α and G_0. The base distribution, G_0, represents the central tendency or "mean" distribution in the distribution space. The precision parameter, α, quantifies how far away realizations of G deviate from G_0. According to Ferguson (1973), DP is a conjugate prior that has two desirable properties: (1) a sufficiently large support, and (2) analytically manageable posterior distributions. Ferguson further derived the posterior of G, $DP(\tilde{\alpha}, \tilde{G_0})$. Here $\tilde{\alpha} = \alpha + N$ and

$$\tilde{G_0} = \frac{\alpha}{\alpha + N} G_0 + \frac{N}{\alpha + N} G_N$$

with G_N being the empirical distribution of the data. Notably, the posterior point estimate of G, $E(G|data) = \tilde{G}_0$, is a weighted average of the base distribution or prior mean G_0 and the empirical distribution or data G_N. When $\alpha = 0$, the posterior point estimate is reduced to the empirical distribution G_N, which is pure nonparametric. When α approaches to infinity, the posterior point estimate gradually approximates G_0, which is parametric. A common practice is to specify a gamma prior for α, which would yield a posterior estimate that is neither 0 nor infinity.

In BNP growth curve modeling, latent variables and/or measurement errors can be modeled nonparametrically. In this article, we focus on the distributional assumption of measurement errors. When the normality of measurement errors is suspected, we assume that $\mathbf{e}_i \sim G_e$ where G_e is an unknown random distribution that is determined by the data. In the BNP framework, DP is typically adopted to specify G_e. Because the distribution of $\mathbf{e}_i$ is continuous but DP is essentially discrete, a DP mixture (DPM) can be used to model the the measurement errors such that

$$G_e = \begin{cases} D(\boldsymbol{\mu}_e^{(1)}, \boldsymbol{\Phi}^{(1)}), & with\ p = p_1 \\ D(\boldsymbol{\mu}_e^{(2)}, \boldsymbol{\Phi}^{(2)}), & with\ p = p_2 \\ \vdots & \vdots \\ D(\boldsymbol{\mu}_e^{(k)}, \boldsymbol{\Phi}^{(k)}), & with\ p = p_k \\ \vdots & \vdots \end{cases},$$

where D represents a predetermined multivariate distribution (e.g., multivariate normal, t, multinomial, etc.), and $\boldsymbol{\mu}_e^{(k)}$ and $\boldsymbol{\Phi}^{(k)}, k = 1, \ldots, \infty$ are means and covariances of the multivariate distribution in the kth component with probability p_k. Theoretically, given an arbitraty distriubtional shape, there could be infinite number of mixture components as k goes to infinity. In practice, a finite number of mixture components often can describe a distribution well and the number of mixture components is determined by the DP precision parameter α. Smaller α yields a smaller number of mixture components. If α approaches infinity, there would be N mixture components, one associated with each subject. Namely, the precision parameter α is an important parameter that can determine the complexity of the model and how well the model fits the data,

and thus may affect the convergence of the model. For the intraindividual measurement errors in the typical linear growth curve model, Tong and Zhang (2019) proposed that

$$\begin{aligned} \mathbf{e}_i|\mathbf{\Phi}_i &\sim MN_T(\mathbf{0}, \mathbf{\Phi}_i), \\ \mathbf{\Phi}_i|G &\sim G, \\ G &\sim DP(\alpha, G_0). \end{aligned}$$

That is, the unknown distribution G_e is approximated by a mixture of multivariate normal distributions where the mixing measure has a Dirichlet process prior, $G_e \sim DPM$. The DP prior $DP(\alpha, G_0)$ can be obtained using the truncated stick-breaking construction (e.g., Lunn et al., 2013; Sethuraman, 1994). Specifically, $DP(\cdot) = \sum_{j=1}^{C} p_j \delta_{z_j}(\cdot), 1 \leq C < \infty$, where C ($1 \leq C \leq N$,often set at a large number) is a possible maximum number of mixture components, $\delta_{z_j}(\cdot)$ denotes a point mass at z_j and $z_j \sim G_0$ independently. The random weights p_j can be generated through the following procedure. With $q_1, q_2, \ldots, q_C \sim Beta(1, \alpha)$, define

$$p_j^{'} = q_j \prod_{k=1}^{j-1}(1 - q_k), j = 1, \ldots, C.$$

Then, p_j is obtained by

$$p_j = \frac{p_j^{'}}{\sum_{k=1}^{C} p_k^{'}}, \tag{1}$$

to satisfy that $\sum_{j=1}^{C} p_j = 1$. In practice, the updating of $\mathbf{e}_i$ can proceed as in a typical DP mixture model and its distribution is a infinite mixture distribution [1].

In general, the distribution of $\mathbf{e}_i$ through the truncated stick-breaking construction is

$$G_e = \begin{cases} D(\boldsymbol{\mu}_e^{(1)}, \mathbf{\Phi}^{(1)}), & with\ p = p_1 \\ D(\boldsymbol{\mu}_e^{(2)}, \mathbf{\Phi}^{(2)}), & with\ p = p_2 \\ \vdots & \vdots \\ D(\boldsymbol{\mu}_e^{(C)}, \mathbf{\Phi}^{(C)}), & with\ p = p_C \end{cases},$$

[1] In practice, infinite-dimension means finite but unbounded dimension.

where D represents a predetermined multivariate distribution, $\boldsymbol{\mu}_e^{(j)}$ and $\boldsymbol{\Phi}^{(j)}, j = 1, \ldots, C$ are means and covariances of the multivariate distribution in the jth component, and p_j is obtained using Equation (1). Given that the mean of $\mathbf{e}_i$ is $\mathbf{0}$, we constrain $\sum_{j=1}^{C} p_j \boldsymbol{\mu}_e^{(j)} = \mathbf{0}$. For simplicity, in this study, we follow Tong and Zhang (2019) and use multivariate normal distributions for the mixing components and constrain $\boldsymbol{\mu}_e^{(j)}$ to be 0. We use inverse Wishart priors $p(\boldsymbol{\Phi}^{(j)}) = IW(n_0, W_0)$ for the covariance matrices of the mixture components, $\boldsymbol{\Phi}^{(j)}$, $j = 1, \ldots, C$. Following Lunn et al. (2013, page 294), we fix the shape parameter n_0 at a specific number and assign an inverse Wishart prior to the scale matrix W_0. With such a specification, the measurement error for individual i, $\mathbf{e}_i$, has a p_k probability of coming from the mixing component $MN(\mathbf{0}, \boldsymbol{\Phi}^{(k)})$. The measurement errors for other individuals may also come from the same mixing component. Let K denotes the number of mixing components or $MN(\mathbf{0}, \boldsymbol{\Phi}^{(k)})$ with $k = 1, \ldots, C$. In other words, K is the number of latent classes for $\mathbf{e}_i$ and K can be smaller than C, $K \leq C$. Within each class, $\mathbf{e}_i$s come from the same distribution.

We would like to note that a similar approach to BNP modeling is finite mixture modeling (FMM). FMM estimates or equivalently approximates an unknown distribution using a mixture of known distributions. A key difference between FMM and BNP modeling is that the number of mixture components is treated as known in FMM, whereas this number is treated as unknown and is freely estimated in BNP modeling. As a result, when FMM is used to handle nonnormality, additional analyses such as model comparison are needed to determine the unknown number of mixture components. BNP modeling therefore is believed to have the advantage of being more objective and data-driven, given that additional analyses such as model comparison that may be vulnerable to subjectivity are avoided.

Bayesian methods are applied to estimate BNP growth curve models. Bayesian methods are becoming increasingly popular in recent years because of their flexibility and powerfulness in estimating models with complex structures (e.g., Lee & Shi, 2000; Lee & Song, 2004; Lee & Xia, 2008; Serang et al., 2015; Zhang et al., 2007; Tong & Zhang, 2012). The key idea of Bayesian methods is to compute the posterior distributions for model parameters by combining the

likelihood function and the priors. As introduced previously, $\boldsymbol{\beta}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Psi}$ are the model parameters in traditional growth curve model. In a BNP growth curve model, $\boldsymbol{\beta}$ and $\boldsymbol{\Psi}$ remain model parameters. In contrast, the measurement error covariance matrix $\boldsymbol{\Psi}$ is not directly estimated. Instead, we obtain $\mathbf{e}_i$ based on which we can get $\boldsymbol{\Phi}$. Another important parameter in BNP growth curve modeling is the precision parameter α. Let $p(\boldsymbol{\beta}, \boldsymbol{\Psi}, \alpha)$ be the joint prior distribution of model parameters and let L be the likelihood function. The joint posterior distribution of model parameters is

$$p(\boldsymbol{\beta}, \boldsymbol{\Psi}, \alpha|\mathbf{y}_i) \propto \int p(\boldsymbol{\beta}, \boldsymbol{\Psi}, \alpha) \times L \, d\mathbf{b},$$

where $\mathbf{b} = (\mathbf{b}_1^{'}, \ldots, \mathbf{b}_N^{'})^{'}$. It is dfficult to solve for this integral in practice. Instead, Markov chain Monte Carlo (MCMC) methods (e.g., Gibbs sampling; Robert & Casella, 2004) are often used to obtain parameter estimates and statistical inferences. Specifically, we first derive the conditional posterior distribution for each of the parameters. We then iteratively draw samples from the derived conditional posteriors to obtain empirical marginal distributions of the model parameters. Finally, statistical inferences are made based on the empirical marginal distributions (Geman & Geman, 1984).

Precision Parameter in BNP Models

The convergence issue in BNP growth curve modeling is likely related to the precision parameter (Tong & Zhang, 2019). Here, we provide a theoretical discussion on how the prior of the precision parameter can influence the number of latent classes for $\mathbf{e}_i$.

The DP precision parameter α is the key to govern the expected number of latent classes. It directly determines the distribution of K, the number of latent classes of $\mathbf{e}_i$. With a larger K, measurement errors of different individuals are more likely to have different distributions. West (1992) found that K asymptotically follows a Poisson distribution

$$K = 1 + x, \; x \sim Poisson\left(\alpha\left(\gamma + logN\right)\right) \tag{2}$$

where γ is Euler's constant. Several percentiles of the distribution of K are given in Table 1. As

shown in the table, K increases as α and N increases.

As discussed previously, a gamm prior $Gamma(a_1, a_2)$ is often used for the hyperparameter α. Given such a prior, West (1992) derived the posterior of α as a mixture of two gamma densities

$$\alpha|\cdot \sim \pi_x Gamma(a_1 + K, a_2 - logx) + (1 - \pi_x)Gamma(a_1 + K - 1, a_2 - logx),$$

where x is an augmented variable $x|\cdot \sim Beta(\alpha + 1, N)$ and the weights π_x is defined by $\pi_x/(1 - \pi_x) = \dfrac{a_1 + K - 1}{N(a_2 - logx)}$. Although West (1992) also provided an approximation to the posterior of α, $p(\alpha|\cdot) \approx Gamma(a_1 + K - 1, a_2 + \gamma + logN)$, how good the approximation was has not been investigated.

A noninformative prior for α seems to be reasonable, especially when the information about number of latent classes are not available. However, a noninformative prior may cause nonconvergence of Markov chains. Therefore, it is worth evaluating different priors for the precision parameter.

A Simulation Study

We now present a simulation study to evaluate the influence of the prior for the precision parameter in BNP growth curve modeling when data are normally distributed and contain outliers[2]. The linear growth curve model in the previous section is used. Measurement errors are modeled nonparametrically to address the nonnormality. Based on the results of previous studies, the number of times points (T), the covariance between the random intercept and slope (σ_{LS}), and the measurement error variance (σ_e^2) have trivial effects on the performance of BNP growth curve modeling (e.g., Tong & Zhang, 2019). Therefore, we only consider a set of values for these parameters in this study. We follow the empirical data analysis results in Tong and Zhang (2019)

[2] Note that nonnormal data may be caused by nonnormal population distributions or data contaminations. We work with outliers in this simulation study because BNP methods are essentially infinite mixture modeling procedures. Generating and dealing with outliers from multiple different distributions are more manageable as we easily know the true number of underlying classes. It is worth verifying the conclusions of this paper for nonnormal population distributions in the future.

to select the population parameter values: the fixed effects are fixed at $\boldsymbol{\beta} = (\beta_L, \beta_S)' = (6.2, 0.3)'$; the number of measurement occation is $T = 4$; measurement error variance $\sigma_e^2 = 0.5$; variances of the random intercept and slope are 1 and 0.1, respectively; and the covariance between the random intercept and slope $\sigma_{LS} = 0$.

Three potentially influential factors are manipulated in the simulation study, including sample size, data distribution, and precision parameter prior. First, two sample sizes are considered, N =200 or 600, representing small and large sample sizes. Second, data are either normal or containing outliers. When generating outliers, three proportions of outliers are considered, $r\%$ =5%, 10%, or 20%. To generate outliers, we randomly select $r\%$ observations at each measurement occasion and replace them by extreme values. The extreme values are generated from 10 different distributions with a large mean of $L_i + S_i(j-1) + m\sigma_e$ where $m \geq 5$ is generated from a truncated Poisson distribution, and a variance of σ_e^2 which is the same as that of the normal data. As a result, the true distribution of the data is a mixture of 11 distributions. Outliers generated in this way conform to the definition of outliers (Tong & Zhang, 2017; Yuan & Zhong, 2008). See Figures 1 and 2 in the supplemental document to aid the understanding of the shape of generated normal data and data with outliers. Third, four priors for the precision parameter are investigated (see Figure 1): a diffuse prior $Gamma(.001, .001)$, a weakly informative prior $Gamma(2, 2)$ suggested by Ishwaran (2000), an accurate informative prior $Gamma(100, 100)$ and an inaccurate informative prior $Gamma(10, 100)$. Gamma(10,100) is an inaccurate informative prior because its mean is 0.1 and its variance is as small as 0.001. According to Table 1, the resulting number of latent classes ranges from 1 to 3 whereas the true number of mixed underlying distribution is 11. For all the other model parameters, conventional noninformative priors such as those in Zhang et al. (2013) are used. Specifically, fixed effects $\boldsymbol{\beta}$ have noninformative diffuse priors $N(0, 10^6)$. The covariance matrix of the random intercept and slope $\boldsymbol{\Psi}$ has an inverse-Wishart prior with an identity scale matrix and degrees of freedom being 2.

In each simulation condition, 500 datasets are generated. BNP growth curve modeling is

applied for each dataset using JAGS with the rjags package in software R (Plummer, 2017; R Core Team, 2019). The total length of Markov chains is set at 50,000 and the first half of iterations is the burn-in period. [3] We assess how different priors affect model convergence rate, parameter estimation, and computation time.

Geweke tests (Geweke, 1991) are used to perform the convergence diagnostics. After the burn-in period, if parameter values are sampled from the stationary distribution of the chain, the means of the first and last parts of the Markov chain (by default the first 10% and the last 50%) should be equal and Geweke's statistic asymptotically follows a standard normal distribution. A Markov chain converges when the Geweke's statistic is between -1.96 and 1.96. If none of the convergence diagnostics (i.e., Geweke tests) for all model parameters suggest non-convergence, the model is said to have converged. In each simulation condition, the convergence rate is defined as the proportion of converged models out of the total 500 generated replications.

For the assessment of model estimation, we obtain the parameter estimate bias, average standard error (ASE), empirical standard error (ESE), mean squared error (MSE), and coverage probability (CP) of the 95% highest posterior density (HPD) credible intervals for each parameter based on converged simulation replications[4].

In addition, the estimation time (in seconds) is recorded for each replication. The average estimation time (AET) is the average of the estimation time for all the converged replications.

All program code and detailed results for the simulation study are available on our GitHub site: https://github.com/CynthiaXinTong/PrecisionParPrior_BNP_GCM.

---

[3] Multiple lengths of Markov chains were tested before the current setting was selected. The convergence results with 50,000 iterations were about the same as those for longer chains.

[4] ASE is the mean estimated standard error across replications. ESE is the standard deviation of the parameter estimates from all replications. MSE is computed as squared bias plus squared ESE. Posterior credible interval, also called credible interval, is the Bayesian counterpart of the frequentist confidence interval. A HPD interval is essentially the narrowest interval on a posterior that covers a given proportion of the probable posterior values.

Main results

Figure 2 shows the convergence rate for BNP growth curve modeling with different precision parameter priors when sample size is 200. This figure clearly shows that outliers harm model convergence. Note that the convergence rate for data with 5% outliers is the lowest. This may be because a small proportion of outliers (e.g., 5%) creates a steep and high-curvature region for the Markov chain to enter and thus more difficult to converge. As the outlier proportion increases, the curvature becomes smoother so the convergence rate is higher. Among the four studied priors, the noninformative prior for the precision parameter always leads to the lowest convergence rate, i.e., less than 30% across all the simulation conditions. Informative priors substantially increase the model convergence rate. Specifically, the convergence rate doubles when we switch from the noninformative prior to the the weakly informative prior suggested by Ishwaran (2000) in the condition with normal data. The incremental amount is about 30% of the original convergence rate in the conditions with outliers. Both accurate informative priors and inaccurate informative priors lead to higher convergence rates. The importance of using informative priors is more salient when data are not normal. Note that inaccurate informative priors yield slightly higher convergence rates than accurate informative priors because the variance of the inaccurate prior is lower and thus its precision is higher. When $N = 600$, model convergence results for BNP growth curve models follow the same pattern, and thus are not reported here.

For converged replications, we evaluate the impact of precision parameter priors on parameter estimation and computation time. Results for $N = 200$ are summarized in Tables 2-5. The relative performance of the four priors in conditions with a larger sample size ($N = 600$) has a similar pattern. Detailed results for $N = 600$ are available in the supplemental document.

From Tables 2-5, we obtain the following findings. First, the estimates of growth curve parameters ($\beta_L, \beta_S, \sigma_L^2, \sigma_S^2, \sigma_{LS}, \sigma_e^2$) are not affected by different priors. Estimation bias, standard errors, MSE, and coverage probability of the 95% HPD credible interval across different precision parameter prior conditions are very close to each other, respectively. Note that when outliers exist

(see Tables 3-5), the true population parameter value of the measurement error variance σ_e^2 is unknown. So, bias, MSE, and CP for this parameter cannot be calculated.

Second, the estimation of the hyperparameter α is greatly affected by different priors. When the noninformative prior is used, the estimated α can be very large (e.g., 28.284 in Table 3) or small (e.g., 0.019 in Table 5), associating with a large standard error. When $Gamma(2, 2)$ or $Gamma(100, 100)$ is used, estimated α is almost always close to 1. When $Gamma(10, 100)$ is used, estimated α is around 0.1. Different α values indicate a different total number of classes K. In general, a larger α value may yield a larger number of latent classes. Since the estimated α has a large standard error when the noninformative diffuse prior is used, the corresponding estimated K can be large or small, too. For the weakly informative and accurate informative priors, the estimated number of latent classes ranges from 4 to 6 for different data conditions, whereas for the inaccurate informative prior, the estimated number of latent classes is about 2 or 3. It is interesting to see that although distinctively different hyperparameter estimates are obtained leading to different number of latent classes, the estimated growth curve parameters are essentially similar. This is because although outliers are generated from 10 different distributions, the 10 different distributions are not separated far apart. With a low class separation, one distribution may be enough to describe several outliers generated from different distributions. Thus, even the inaccurate informative prior can yield a precision parameter that is adequate to model the measurement errors.

Third, BNP growth curve modeling with the inaccurate informative prior $Gamma(10, 100)$ requires the shortest computation time. This is because the inaccurate informative prior here has the smallest variance and thus is most "informative" among the four priors.

Fourth, outliers affect the performance of BNP growth curve modeling. When data contain a large proportion of outliers (e.g., 20%), estimation bias for the average of random intercepts β_L and variance of random intercepts σ_L^2 are much larger than those when outlier proportion is low. In addition, outliers influence computation time. It is worth mentioning that it is most time consuming when the outlier proportion is 5%. A possible reason is that a small proportion of

outliers creates a steep and high-curvature region for Markov chains to enter and thus takes longer time to converge. With more outliers, the curvature is smoother so the computation is faster.

Discussion

Restricting to a parametric probability family can delude investigators and falsely make an illusion of posterior certainty (Müller & Mitra, 2004). On the contrary, BNP methods are adaptive and powerful to discover complex patterns in real data. Although BNP growth curve modeling has been proposed, the effect of the precision parameter was not fully studied. In this article, we have conducted a simulation study to investigate how different types of precision parameter priors impact the convergence rate, model estimation, and computation time in BNP growth curve modeling. We found that the noninformative prior suffered from the lowest convergence rates while the inaccurate informative prior with the smallest prior variance yielded the highest convergence rates and the fastest computations. Furthermore, we found that the estimation of growth curve parameters was not affected by the prior of the precision parameter. Based on these results, we recommend to use informative priors with high precision in practice.

We would like to note that although it seems counterintuitive that the inaccurate informative prior for the precision parameter performed the best, such findings have been observed in the literature. For example, Finch and Miller (2019) found that slightly informative priors can be advantageous in small samples even when these priors are incorrect. Depaoli (2013) showed that growth mixture model estimations obtained with inaccurate priors were still more accurate than maximum likelihood or Bayesian estimation with diffuse priors. Zitzmann et al. (2020) explicitly discussed this issue for small samples. Our simulation results also supported the argument that the amount of information in the prior can be more important than the accuracy of the prior under certain circumstances.

We also want to point out that the estimation bias was relatively large in our simulation study, when compared to that in previous studies (Tong & Zhang, 2019). This is because we consider much higher outlier proportions. When the outlier proportion is low (i.e., 5%), parameter

estimates are very close to the true population values. As the outlier proportion increases, the bias increases. One possible way to improve the performance of BNP growth curve modeling when the outlier proportion is high is to use a nonnormal base distribution. In our simulation study, for simplicity, we used normal distributions with zero mean as the mixing components of BNP modeling. This cannot handle asymmetric nonnormal distributions, which may partly explain the less satisfactory performance of BNP modeling in the conditions with high outlier proportions. But BNP methods in general are very flexible. A nonnormal base distribution may overcome this limitation. While future studies may continue along this path, we want to emphasize that BNP modeling as in our study still outperforms traditional growth curve modeling and is recommended to use in general when data are suspected to be nonnormal (Tong & Zhang, 2019) no matter the nonnormality is caused by nonnormal population distribution or data contamination.

The convergence rate of BNP growth curve modeling was found to be higher in previous studies, i.e., close to one (Tong & Zhang, 2019). We would like to note that the difference is likely due to the list of parameters counted during convergence assessment. In Tong and Zhang (2019), the convergence rate was computed only for growth curve parameters. When only growth curve parameters are considered, nonconvergence rarely occurred in our study. The major problem is the precision parameter. As shown in the simulation study, nonconvergence frequently arose for this parameter (detailed Geweke tests results for each parameter are available on our GitHub site: https://github.com/CynthiaXinTong/PrecisionParPrior_BNP_GCM). Another possible reason why convergence rates were relatively low (below 70%) in our simulation is that Geweke tests often yield lower rates of convergence than other diagnostic methods (e.g., Jang & Cohen, 2020). However, as pointed out in Jang and Cohen, the pattern of convergence rates for model comparison was similar for different diagnostic tests. Namely, our conclusions about which precision parameter priors to use in BNP growth curve modeling will not be affected by the diagnostic tests. We further discuss the use of Geweke tests in the next paragraph. Notably, although the non-convergence for the precision parameter seemed not to impact parameter estimates for the growth curve parameters, such issue may mislead model fit assessment.

Although model assessment and model comparison methods have been proposed for various models, samples of different sizes, and data structures (e.g., Celeux et al., 2006), their performance in BNP analysis has not been studied. Therefore, future studies on how different precision parameter priors affect model fit assessment are encouraged.

In our study, model convergence diagnostics were conducted using Geweke tests. Although Geweke tests are commonly used in the Bayesian literature, it is impossible to say with certainty that a finite sample from an MCMC algorithm is representative of an underlying stationary distribution and a combination of strategies aiming at evaluating and accelerating MCMC sampler convergence is recommended (Cowles & Carlin, 1996). For our simulation study, Geweke tests were relatively easy to systematically implement. In empirical studies, we recommend using multiple strategies (e.g., trace plots, multiple chains) to check model convergence. In addition, since Zitzmann and Hecht (2019) pointed out that it is possible that the approximation of the Bayesian estimates is still not optimal even when a chain converges, we recommend substantive researchers conducting sensitivity analysis and evaluating how the length of the Markov chains affects the model estimation results.

Our study echoed the previous literature in that using informative priors may help reduce computation time in Bayesian modeling. We would like to note that there are other approaches that can be used to further increase the computation efficiency. For example, Berger et al. (2020) and Daniels and Kass (1999) proposed shrinkage priors, and Hecht et al. (2020) proposed a model reformulation approach in which the sample covariance matrix was modeled instead of individual observations. This latter approach has been applied to the Bayesian continuous-time model (Hecht & Zitzmann, 2020) as well as the Bayesian STARTS model (Ludtke et al., 2018). Future research on BNP growth curve modeling could incorporate this approach and other potentially efficient approaches to reduce computation time.

The employment of BNP growth curve modeling is a field still in its early stage. New Dirichlet process variants and generalizations are being proposed every year to cater to specific applications. BNP modeling was only used to handle the nonnormality in intraindividual

measurement errors in our study. The similar strategy can be used for random effects, such as random intercepts and slopes. Also, although we worked with balanced data, BNP growth curve modeling should be able to handle unbalanced data (e.g., individually varying time points). However, as implied by previous studies (Tong, 2014), the convergence issue may be more challenging, thereby awaiting future studies.

Appendix. Implementation

To facilitate the application of BNP growth curve modeling, we illustrate how it can be implemented in free software R using the rjags package (Plummer, 2017; R Core Team, 2019). First, we specify the model using JAGS code and save it into a text file model.txt.

```
model{
  # Model specification for BNP linear growth curve model
        for (i in 1:N){
                LS[i,1:2]~dmnorm(muLS[i,1:2], Inv_cov[1:2,1:2])

                muLS[i,1]<-bL[1]
        muLS[i,2]<-bS[1]

                for (t in 1:T){
                        y[i, t] ~ dnorm(muY[i,t], taue[i])
                        muY[i,t] <- LS[i,1]+LS[i,2]*(t-1)
                }
                taue[i] <- taue.mix[groupe[i]]
                groupe[i] ~ dcat(pei[])

                for (j in 1:C){ ##C is the largest possible #
                   number of classes, can be set at a large
                   number, e.g., 20.
                        ginde[i,j] <- equals(j,groupe[i])
                }
        }

        #Priors for model parameter
        for (i in 1:1){
```

```
        bL[i] ~ dnorm(0, 1.0E-6)
                bS[i] ~ dnorm(0, 1.0E-6)
        }

        ## truncated stick breaking construction
        pe[1]<-qe[1]
        for (j in 2:C){
                pe[j] <- qe[j] * (1 - qe[j - 1]) * pe[j -1 ] / qe
                   [j - 1]
        }

        for (j in 1:C){
                qe[j] ~dbeta(1, alpha)T(0.0001,0.9999)
                pei[j] <- pe[j]/sum(pe[])
                taue.mix[j] ~ dgamma(aprece,bprece)
        }

        ##DP precision parameter,4 different priors were used in
           #our simulation study
        alpha~dgamma(100,100)

        aprece <- 2
        bprece ~dgamma(2,2)

        ##total clusters
        Ke <- sum(cle[])
        for (j in 1:20) {
                suminde[j] <- sum(ginde[,j])
                cle[j] <- step(suminde[j]-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

        para[1] <- bL[1]
        para[2] <- bS[1]
```

```
        Cov[1:2,1:2]<-inverse(Inv_cov[1:2,1:2])
        para[3] <- Cov[1,1]
        para[4] <- Cov[2,2]
        para[5] <- Cov[1,2]


        for (i in 1:N){
                for (t in 1:T){
                        par[i,t] <- y[i,t]-LS[i,1]-LS[i,2]*(t-1)
                }
        }

        for (t in 1:T){
                for(i in 1:N){
                err[(t-1)*N+i] <- par[i,t]
                }
        }

        para[6] <- sd(err[])*sd(err[])
        para[7] <- Ke
        para[8] <- alpha
        para[9] <- bprece
}
```

JAGS has been integrated with the R software environment. To run the above JAGS code in R, we first install and load the rjags package.

```
install.packages("rjags")
library(rjags)
```

Then, we prepare the data, the initial values, and run jags.

```
##prepare data
data <- read.table('data.txt')
N <- nrow(data)
jagsdata <- list(N=N, T=4, C=20, y=as.matrix(data))

##specify initial values
```

```
inits <- list(Inv_cov = structure(.Data = c(1.0,0.0,0.0,10.0), .
   Dim = c(2,2)),
  alpha = 1.0, bL = c(6.2),bS = c(0.3),bprece = 0.5,".RNG.name" =
      "base::Wichmann-Hill", ".RNG.seed" = 115)
#note that  we specified the random number generator and the seed
#so our study can be replicated.

##run jags
#save the start time
time0 <- proc.time()

#read the model, burn 25,000 iterations
model <- jags.model(file="model.txt", data=jagsdata, inits=inits,
    n.chains = 1, n.adapt=25000)

#run 25,000 iterations after the burn-in priord
model.samples <- coda.samples(model, c("para"), n.iter=25000)

#save results into model.res
model.res <- as.mcmc(do.call(rbind,model.samples))

#obtain the estimation time: end time - start time
time1 <- proc.time()-time0
```

Finally, we extract the model estimation results from model.res.

```
#parameter estimates
summary(model.res)

#HPD credible intervals
HPDinterval(model.res)

#geweke tests
geweke.diag(model.res)
```

References

Ansari, A. & Iyengar, R. (2006). Semiparametric Thurstonian models for recurrent choices: A Bayesian analysis. *Psychometrika*, 71, 631–657. DOI: 10.1007/s11336-006-1233-5.

Berger, J. O., Sun, D., & Song, C. (2020). Bayesian analysis of the covariance matrix of a multivariate normal distribution with a new class of priors. *Annals of Statistics*, 48, 2381–2403. DOI: 10.1214/19-AOS1891.

Brown, E. R. & Ibrahim, J. G. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59, 221–228. DOI: 10.1111/1541-0420.00028.

Burr, D. & Doss, H. (2005). A Bayesian semiparametric model for random-effects meta-analysis. *Journal of the American Statistical Association*, 100, 242–251. DOI: 10.1198/016214504000001024.

Bush, C. A. & MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika*, 83, 275–285. DOI: 10.1093/biomet/83.2.275.

Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, 49, 1716–1735. DOI: 10.3758/s13428-016-0814-1.

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, 1, 651–673. DOI: 10.1214/06-ba122.

Chou, C.-P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A monte carlo study. *British Journal of Mathematical and Statistical Psychology*, 44, 347–357. DOI: 10.1111/j.2044-8317.1991.tb00966.x.

Cowles, M. K. & Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91, 883–904.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16–29. DOI: 10.1037/1082-989x.1.1.16.

Daniels, M. J. & Kass, R. E. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, 94, 1254–1263. DOI: 10.2307/2669939.

Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods*, 18, 186–219. DOI: 10.1037/a0031609.

Dunson, D. B. (2006). Bayesian dynamic modeling of latent trait distributions. *Biostatistics*, 7, 551–568. DOI: 10.1093/biostatistics/kxj025.

Fahrmeir, L. & Raach, A. (2007). A Bayesian semiparametric latent variable model for mixed responses. *Psychometrika*, 72, 327–346. DOI: 10.1007/s11336-007-9010-7.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1, 209–230. DOI: 10.1214/aos/1176342360.

Ferguson, T. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics*, 2, 615–629. DOI: 10.1214/aos/1176342752.

Finch, W. H. & Miller, J. E. (2019). The use of incorrect informative priors in the estimation of MIMIC model parameters with small sample sizes. *Structural Equation Modeling: A Multidisciplinary Journal*, 26, 497–508. DOI: 10.1080/10705511.2018.1553111.

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. DOI: 10.1016/b978-0-08-051581-6.50057-x.

Gershman, S. J. & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56, 1–12. DOI: 10.1016/j.jmp.2011.08.004.

Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernado, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics 4* (pp. 169–193). Oxford, UK: Clarendon Press. DOI: 10.21034/sr.148.

Ghosal, S., Ghosh, J., & Ramamoorthi, R. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics*, 27, 143–158. DOI: 10.1214/aos/1018031105.

Hecht, M., Gische, C., Vogel, D., & Zitzmann, S. (2020). Integrating out nuisance parameters for computationally more efficient Bayesian estimation - An illustration and tutorial. *Structural Equation Modeling: A Multidisciplinary Journal*, 27, 483–493. DOI: 10.1080/10705511.2019.1647432.

Hecht, M. & Zitzmann, S. (2020). A computationally more efficient Bayesian approach for estimating continuous-time models. *Structural Equation Modeling: A Multidisciplinary Journal*, 27, 829–840. DOI: 10.1080/10705511.2020.1719107.

Hjort, N. L. (2003). Topics in nonparametric Bayesian statistics. In P. Green, N. L. Hjort, & S. Richardson (Eds.), *Highly Structured Stochastic Systems* (pp. 455–487). Oxford: Oxford University Press.

Hjort, N. L., Holmes, C., Müller, P., & Walker, S. G. (2010). *Bayesian nonparametrics*. Cambridge: Cambridge University Press. DOI: 10.1093/acprof:oso/9780199695607.003.0013.

Ishwaran, H. (2000). Inference for the random effects in Bayesian generalized linear mixed models. In A. S. Association (Ed.), *ASA Proceedings of the Bayesian Statistical Science Section* (pp. 1–10).

Jang, Y. & Cohen, A. S. (2020). The impact of Markov chain covergence on estimation of

mixture IRT model parameters. *Educational and Psychological Measurement*, 80, 975–994. DOI: 10.1177/0013164419898228.

Jara, A., Hanson, T., Quintana, F., Müller, P., & Rosner, G. (2011). DP-package: Bayesian semi and nonparametric modeling in R. *Journal of Statistical Software*, 40, 1–30. DOI: 10.18637/jss.v040.i05.

Kleinman, K. P. & Ibrahim, J. G. (1998). A semiparametric Bayesian approach to the random effects model. *Biometrics*, 54, 921–938. DOI: 10.2307/2533846.

Lee, S. Y., Lu, B., & Song, X. Y. (2008). Semiparametric Bayesian analysis of structural equation models with fixed covariates. *Statistics in Medicine*, 27, 2341–2360. DOI: 10.1002/sim.3098.

Lee, S. Y. & Shi, J. Q. (2000). Joint Bayesian analysis of factor scores and structural parameters in the factor analysis model. *Annal of the Institute of Statistical Mathematics*, 52, 722–736. DOI: 10.1023/a:1017529427433.

Lee, S. Y. & Song, X. Y. (2004). Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology*, 57, 131–150. DOI: 10.1348/000711004849204.

Lee, S. Y. & Xia, Y. M. (2008). A robust Bayesian approach for structural equation models with missing data. *Psychometrika*, 73, 343–364. DOI: 10.1007/s11336-008-9060-5.

Lu, Z. & Zhang, Z. (2014). Robust growth mixture models with non-ignorable missingness: Models, estimation, selection, and application. *Computational Statistics and Data Analysis*, 71, 220–240. DOI: 10.1016/j.csda.2013.07.036.

Ludtke, O., Robitzsch, A., & Wagner, J. (2018). More stable estimation of the STARTS model: A Bayesian approach using Markov Chain Monte Carlo techniques. *Psychological Methods*, 23, 570–593. DOI: 10.1037/met0000155.

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2013). *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: CRC Press.

MacEachern, S. (1999). Dependent nonparametric processes. In A. S. Association (Ed.), *ASA Proceedings of the Section on Bayesian Statistical Science*.

Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). *Robust statistics: Theory and methods*. New York: John Wiley & Sons, Inc. DOI: 10.1002/0470010940.

McArdle, J. J. & Nesselroade, J. R. (2014). *Longitudinal data analysis using structural equation models*. American Psychological Association. DOI: 10.1037/14440-000.

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

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105, 156–166. DOI: 10.1037/0033-2909.105.1.156.

Müller, P. & Mitra, R. (2004). Bayesian nonparametric inference - why and how. *Bayesian Analysis*, 1, 1–33. DOI: 10.1214/13-ba811.

Muthén, B. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55, 463–469. DOI: 10.1111/j.0006-341X.1999.00463.x.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9, 249–265. DOI: 10.2307/1390653.

Ohlssen, D. I., Sharples, L. D., & Spiegelhalter, D. (2007). Flexible random-effects models using Bayesian semi-parametric models: applications to institutional comparisons. *Statistics in Medicine*, 26, 2088–112. DOI: 10.1002/sim.2666.

Pendergast, J. F. & Broffitt, J. D. (1985). Robust estimation in growth curve models. *Communications in Statistics: Theory and Methods*, 14, 1919–1939. DOI: 10.1080/03610928508829021.

Plummer, M. (2017). Jags version 4.3. 0 user manual.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robert, C. P. & Casella, G. (2004). *Monte Carlo Statistical Methods*. New York: Springer. DOI: 10.1007/978-1-4757-4145-2.

Serang, S., Zhang, Z., Helm, J., Steele, J. S., & Grimm, K. J. (2015). Evaluation of a Bayesian approach to estimating nonlinear mixed-effects mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 22, 202–215. DOI: 10.1080/10705511.2014.937322.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4, 639–650.

Sharif-Razavian, N. & Zollmann, A. (2009). An overview of nonparametric Bayesian models and applications to natural language processing. *[Online], available: http://www.cs.cmu.edu/?simzollmann/publications/nonparametric.pdf*.

Si, Y. & Reiter, J. P. (2013). Nonparametric Bayesian multiple imputation for incomplete categorical variables in large-scale assessment surveys. *Journal of Educational and Behavioral Statistics*, 38, 499–521. DOI: 10.3102/1076998613480394.

Si, Y., Reiter, J. P., & Hillygus, D. S. (2015). Semi-parametric selection models for potentially non-ignorable attrition in panel studies with refreshment samples. *Political Analysis*, 23, 92–112. DOI: 10.1093/pan/mpu009.

Silvapulle, M. J. (1992). On M-methods in growth curve analysis with asymmetric errors. *Journal of Statistical Planning and Inference*, 32, 303–309. DOI: 10.1016/0378-3758(92)90013-i.

Singer, J. M. & Sen, P. K. (1986). M-methods in growth curve analysis. *Journal of Statistical Planning and Inference*, 13, 251–261. DOI: 10.1016/0378-3758(86)90137-0.

Tong, X. (2014). Robust semiparametric bayesian methods in growth curve modeling. *(Unpublished doctoral dissertation). Notre Dame, IN: University of Notre Dame*.

Tong, X. & Zhang, Z. (2012). Diagnostics of robust growth curve modeling using Student's t distribution. *Multivariate Behavioral Research*, 47, 493–518. DOI: 10.1080/00273171.2012.692614.

Tong, X. & Zhang, Z. (2017). Outlying observation diagnostics in growth curve modeling. *Multivariate Behavioral Research*, 52, 768–788.

Tong, X. & Zhang, Z. (2019). Robust Bayesian approaches in growth curve modeling: Using Student's t distributions versus a semiparametric method. *Structural Equation Modeling: A Multidisciplinary Journal*, 27, 544–560. DOI: 10.1080/10705511.2019.1683014.

West, M. (1992). *Hyperparameter estimation in Dirichlet process mixture models*. Technical Report 92-A03, Duke University, ISDS.

Yang, M. & Dunson, D. B. (2010). Bayesian semiparametric structural equation models with latent variables. *Psychometrika*, 75, 675–693. DOI: 10.1007/s11336-010-9174-4.

Yuan, K.-H. & Bentler, P. M. (1998). Structural equation modeling with robust covariances. *Sociological Methodology*, 28, 363–396. DOI: 10.1111/0081-1750.00052.

Yuan, K.-H. & Bentler, P. M. (2001). Effect of outliers on estimators and tests in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, 54, 161–175. DOI: 10.1348/000711001159366.

Yuan, K.-H. & Zhong, X. (2008). Outliers, high-leverage observations and influential cases in factor analysis: Minimizing their effect using robust procedures. *Sociological Methodology*, 38, 329–368.

Zhang, Z. (2016). Modeling error distributions of growth curve models through Bayesian methods. *Behavior Research Methods*, 48, 427–444. DOI: 10.3758/s13428-015-0589-9.

Zhang, Z., Hamagami, F., Wang, L., Grimm, K. J., & Nesselroade, J. R. (2007). Bayesian

analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, 31, 374–383. DOI: 10.1177/0165025407077764.

Zhang, Z., Lai, K., Lu, Z., & Tong, X. (2013). Bayesian inference and application of robust growth curve models using Student's t distribution. *Structural Equation Modeling: A Multidisciplinary Journal*, 20, 47–78. DOI: 10.1080/10705511.2013.742382.

Zhong, X. & Yuan, K.-H. (2010). Weights. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 1617–1620). Thousand Oaks, CA: Sage. DOI: 10.4135/9781412961288.

Zhong, X. & Yuan, K.-H. (2011). Bias and efficiency in structural equation modeling: Maximum likelihood versus robust methods. *Multivariate Behavioral Research*, 46, 229–265. DOI: 10.1080/00273171.2011.558736.

Zitzmann, S. & Hecht, M. (2019). Going beyond convergence in Bayesian estimation: Why precision matters too and how to assess it. *Structural Equation Modeling: A Multidisciplinary Journal*, 26, 646–661. DOI: 10.1080/10705511.2018.1545232.

Zitzmann, S., Ludtke, O., Robitzsch, A., & Hecht, M. (2020). On the performance of Bayesian approaches in small samples: A comment on Smid, McNeish, Miocevic, and van de Schoot. *Structural Equation Modeling: A Multidisciplinary Journal*. Advance online publication. DOI: 10.1080/10705511.2020.1752216.

Table 1

Different percentiles (5%, 50%, 95%) of the distribution of the number of clusters K, given different values of precision parameter α and sample size N

	$\alpha = 0.1$			$\alpha = 1$			$\alpha = 2$		
	5%	50%	95%	5%	50%	95%	5%	50%	95%
$N = 200$	1	1	3	3	7	11	7	13	19
$N = 600$	1	2	3	4	8	13	9	15	21
$N = 1000$	1	2	3	4	8	13	10	16	23

Table 2

Model estimation for BNP growth curve modeling with different precision parameter priors when data are normal and $N = 200$

Prior		Est.	Bias	ASE	ESE	MSE	CP	AET
$Gamma(0.001, 0.001)$	β_L	6.204	0.004	0.082	0.084	0.007	0.957	539.332
	β_S	0.301	0.001	0.033	0.032	0.001	0.957	539.332
	σ_L^2	0.999	-0.001	0.138	0.142	0.020	0.936	539.332
	σ_S^2	0.118	0.018	0.021	0.018	0.001	0.922	539.332
	σ_{LS}	-0.010	-0.010	0.040	0.034	0.001	0.993	539.332
	σ_e^2	0.497	-0.003	0.024	0.036	0.001	0.816	539.332
	K	2.113	-	0.803	2.331	-	-	539.332
	α	11.134	-	18.109	104.805	-	-	539.332
$Gamma(2, 2)$	β_L	6.198	-0.002	0.082	0.080	0.006	0.958	740.331
	β_S	0.302	0.002	0.033	0.031	0.001	0.965	740.331
	σ_L^2	1.008	0.008	0.139	0.131	0.017	0.965	740.331
	σ_S^2	0.118	0.018	0.021	0.018	0.001	0.927	740.331
	σ_{LS}	-0.010	-0.010	0.040	0.034	0.001	0.983	740.331
	σ_e^2	0.499	-0.001	0.024	0.034	0.001	0.823	740.331
	K	4.106	-	2.415	0.776	-	-	740.331
	α	0.732	-	0.526	0.126	-	-	740.331
$Gamma(100, 100)$	β_L	6.200	0.000	0.083	0.083	0.007	0.948	1024.509
	β_S	0.299	-0.001	0.033	0.032	0.001	0.958	1024.509
	σ_L^2	1.014	0.014	0.139	0.133	0.018	0.967	1024.509
	σ_S^2	0.117	0.017	0.021	0.018	0.001	0.942	1024.509
	σ_{LS}	-0.010	-0.010	0.040	0.036	0.001	0.976	1024.509
	σ_e^2	0.499	-0.001	0.024	0.036	0.001	0.827	1024.509
	K	5.037	-	1.924	0.407	-	-	1024.509
	α	0.992	-	0.099	0.004	-	-	1024.509
$Gamma(10, 100)$	β_L	6.202	0.002	0.082	0.082	0.007	0.945	370.307
	β_S	0.301	0.001	0.033	0.031	0.001	0.971	370.307
	σ_L^2	1.001	0.001	0.138	0.129	0.017	0.971	370.307
	σ_S^2	0.117	0.017	0.021	0.018	0.001	0.942	370.307
	σ_{LS}	-0.012	-0.012	0.040	0.037	0.001	0.974	370.307
	σ_e^2	0.498	-0.002	0.024	0.035	0.001	0.835	370.307
	K	1.981	-	0.874	0.199	-	-	370.307
	α	0.099	-	0.031	0.001	-	-	370.307

Note. Est. = estimate; ASE = average standard error; ESE = empirical standard error; MSE = mean squared error; CP = coverage probability of the 95% HPD credible interval; AET = average estimation time.

Table 3

Model estimation for BNP growth curve modeling with different precision parameter priors when data contain 5% of outliers and $N = 200$

Prior		Est.	Bias	ASE	ESE	MSE	CP	AET
$Gamma(0.001, 0.001)$	β_L	6.300	0.100	0.092	0.083	0.017	0.793	841.706
	β_S	0.313	0.013	0.037	0.036	0.001	0.948	841.706
	σ_L^2	1.006	0.006	0.160	0.145	0.021	0.956	841.706
	σ_S^2	0.118	0.018	0.024	0.017	0.001	0.985	841.706
	σ_{LS}	-0.009	-0.009	0.046	0.044	0.002	0.985	841.706
	σ_e^2	3.133	-	0.125	0.138	-	-	841.706
	K	5.184	-	1.189	5.433	-	-	841.706
	α	28.284	-	51.922	75.124	-	-	841.706
$Gamma(2, 2)$	β_L	6.311	0.111	0.092	0.088	0.020	0.763	971.782
	β_S	0.311	0.011	0.037	0.034	0.001	0.957	971.782
	σ_L^2	1.007	0.007	0.161	0.146	0.021	0.967	971.782
	σ_S^2	0.117	0.017	0.024	0.018	0.001	0.976	971.782
	σ_{LS}	-0.008	-0.008	0.046	0.041	0.002	0.976	971.782
	σ_e^2	3.119	-	0.124	0.136	-	-	971.782
	K	6.515	-	2.905	0.981	-	-	971.782
	α	1.126	-	0.676	0.171	-	-	971.782
$Gamma(100, 100)$	β_L	6.298	0.098	0.091	0.090	0.018	0.794	1088.448
	β_S	0.314	0.014	0.037	0.034	0.001	0.944	1088.448
	σ_L^2	0.987	-0.013	0.158	0.134	0.018	0.964	1088.448
	σ_S^2	0.117	0.017	0.024	0.018	0.001	0.976	1088.448
	σ_{LS}	-0.004	-0.004	0.045	0.041	0.002	0.992	1088.448
	σ_e^2	3.133	-	0.124	0.130	-	-	1088.448
	K	6.161	-	1.930	0.467	-	-	1088.448
	α	1.003	-	0.100	0.005	-	-	1088.448
$Gamma(10, 100)$	β_L	6.311	0.111	0.091	0.090	0.020	0.767	561.074
	β_S	0.311	0.011	0.037	0.034	0.001	0.952	561.074
	σ_L^2	0.985	-0.015	0.158	0.144	0.021	0.960	561.074
	σ_S^2	0.119	0.019	0.024	0.018	0.001	0.964	561.074
	σ_{LS}	-0.009	-0.009	0.046	0.042	0.002	0.968	561.074
	σ_e^2	3.118	-	0.124	0.136	-	-	561.074
	K	2.903	-	0.872	0.240	-	-	561.074
	α	0.103	-	0.032	0.001	-	-	561.074

Note. Est. = estimate; ASE = average standard error; ESE = empirical standard error; MSE = mean squared error; CP = coverage probability of the 95% HPD credible interval; AET = average estimation time.

Table 4

Model estimation for BNP growth curve modeling with different precision parameter priors when data contain 10% of outliers and $N = 200$

Prior		Est.	Bias	ASE	ESE	MSE	CP	AET
$Gamma(0.001, 0.001)$	β_L	6.437	0.237	0.103	0.102	0.066	0.348	591.282
	β_S	0.335	0.035	0.043	0.039	0.003	0.917	591.282
	σ_L^2	1.018	0.018	0.187	0.173	0.030	0.977	591.282
	σ_S^2	0.126	0.026	0.028	0.022	0.001	0.917	591.282
	σ_{LS}	-0.007	-0.007	0.053	0.047	0.002	0.977	591.282
	σ_e^2	5.464	-	0.173	0.180	-	-	591.282
	K	3.112	-	1.106	1.728	-	-	591.282
	α	1.041	-	2.885	7.422	-	-	591.282
$Gamma(2, 2)$	β_L	6.424	0.224	0.103	0.105	0.061	0.426	938.496
	β_S	0.336	0.036	0.043	0.038	0.003	0.886	938.496
	σ_L^2	1.020	0.020	0.187	0.174	0.031	0.966	938.496
	σ_S^2	0.121	0.021	0.027	0.020	0.001	0.970	938.496
	σ_{LS}	-0.008	-0.008	0.053	0.044	0.002	0.979	938.496
	σ_e^2	5.448	-	0.172	0.171	-	-	938.496
	K	6.314	-	2.798	0.942	-	-	938.496
	α	1.090	-	0.652	0.163	-	-	938.496
$Gamma(100, 100)$	β_L	6.428	0.228	0.104	0.100	0.062	0.398	1045.439
	β_S	0.332	0.032	0.043	0.040	0.003	0.903	1045.439
	σ_L^2	1.020	0.020	0.188	0.172	0.030	0.964	1045.439
	σ_S^2	0.123	0.023	0.027	0.021	0.001	0.961	1045.439
	σ_{LS}	-0.009	-0.009	0.053	0.043	0.002	0.982	1045.439
	σ_e^2	5.459	-	0.174	0.175	-	-	1045.439
	K	6.091	-	1.911	0.409	-	-	1045.439
	α	1.002	-	0.100	0.004	-	-	1045.439
$Gamma(10, 100)$	β_L	6.426	0.226	0.103	0.102	0.062	0.395	389.282
	β_S	0.333	0.033	0.043	0.041	0.003	0.897	389.282
	σ_L^2	1.011	0.011	0.185	0.177	0.032	0.957	389.282
	σ_S^2	0.123	0.023	0.027	0.021	0.001	0.943	389.282
	σ_{LS}	-0.007	-0.007	0.052	0.045	0.002	0.975	389.282
	σ_e^2	5.457	-	0.172	0.169	-	-	389.282
	K	2.935	-	0.878	0.206	-	-	389.282
	α	0.103	-	0.032	0.001	-	-	389.282

Note. Est. = estimate; ASE = average standard error; ESE = empirical standard error; MSE = mean squared error; CP = coverage probability of the 95% HPD credible interval; AET = average estimation time.

Table 5

Model estimation for BNP growth curve modeling with different precision parameter priors when data contain 20% of outliers and $N = 200$

Prior		Est.	Bias	ASE	ESE	MSE	CP	AET
$Gamma(0.001, 0.001)$	β_L	6.890	0.690	0.149	0.120	0.490	0.000	460.170
	β_S	0.385	0.085	0.061	0.054	0.010	0.735	460.170
	σ_L^2	1.321	0.321	0.315	0.284	0.183	0.884	460.170
	σ_S^2	0.141	0.041	0.038	0.027	0.002	0.952	460.170
	σ_{LS}	0.019	0.019	0.080	0.062	0.004	0.980	460.170
	σ_e^2	9.238	-	0.242	0.258	-	-	460.170
	K	2.713	-	0.810	0.307	-	-	460.170
	α	0.019	-	0.047	0.089	-	-	460.170
$Gamma(2, 2)$	β_L	6.890	0.690	0.150	0.120	0.490	0.000	949.186
	β_S	0.381	0.081	0.061	0.052	0.009	0.787	949.186
	σ_L^2	1.358	0.358	0.321	0.279	0.206	0.879	949.186
	σ_S^2	0.143	0.043	0.038	0.024	0.002	0.962	949.186
	σ_{LS}	0.011	0.011	0.082	0.064	0.004	0.983	949.186
	σ_e^2	9.167	-	0.245	0.265	-	-	949.186
	K	5.458	-	2.392	0.564	-	-	949.186
	α	0.941	-	0.566	0.095	-	-	949.186
$Gamma(100, 100)$	β_L	6.882	0.682	0.149	0.121	0.480	0.000	1056.953
	β_S	0.381	0.081	0.061	0.054	0.010	0.774	1056.953
	σ_L^2	1.323	0.323	0.314	0.284	0.185	0.878	1056.953
	σ_S^2	0.143	0.043	0.038	0.026	0.003	0.944	1056.953
	σ_{LS}	0.010	0.010	0.081	0.062	0.004	0.981	1056.953
	σ_e^2	9.172	-	0.243	0.256	-	-	1056.953
	K	5.695	-	1.811	0.321	-	-	1056.953
	α	0.998	-	0.099	0.003	-	-	1056.953
$Gamma(10, 100)$	β_L	6.897	0.697	0.150	0.116	0.499	0.000	391.429
	β_S	0.379	0.079	0.061	0.052	0.009	0.803	391.429
	σ_L^2	1.354	0.354	0.319	0.280	0.204	0.861	391.429
	σ_S^2	0.141	0.041	0.038	0.026	0.002	0.956	391.429
	σ_{LS}	0.014	0.014	0.081	0.064	0.004	0.980	391.429
	σ_e^2	9.166	-	0.242	0.255	-	-	391.429
	K	2.880	-	0.855	0.151	-	-	391.429
	α	0.103	-	0.032	0.001	-	-	391.429

Note. Est. = estimate; ASE = average standard error; ESE = empirical standard error; MSE = mean squared error; CP = coverage probability of the 95% HPD credible interval; AET = average estimation time.

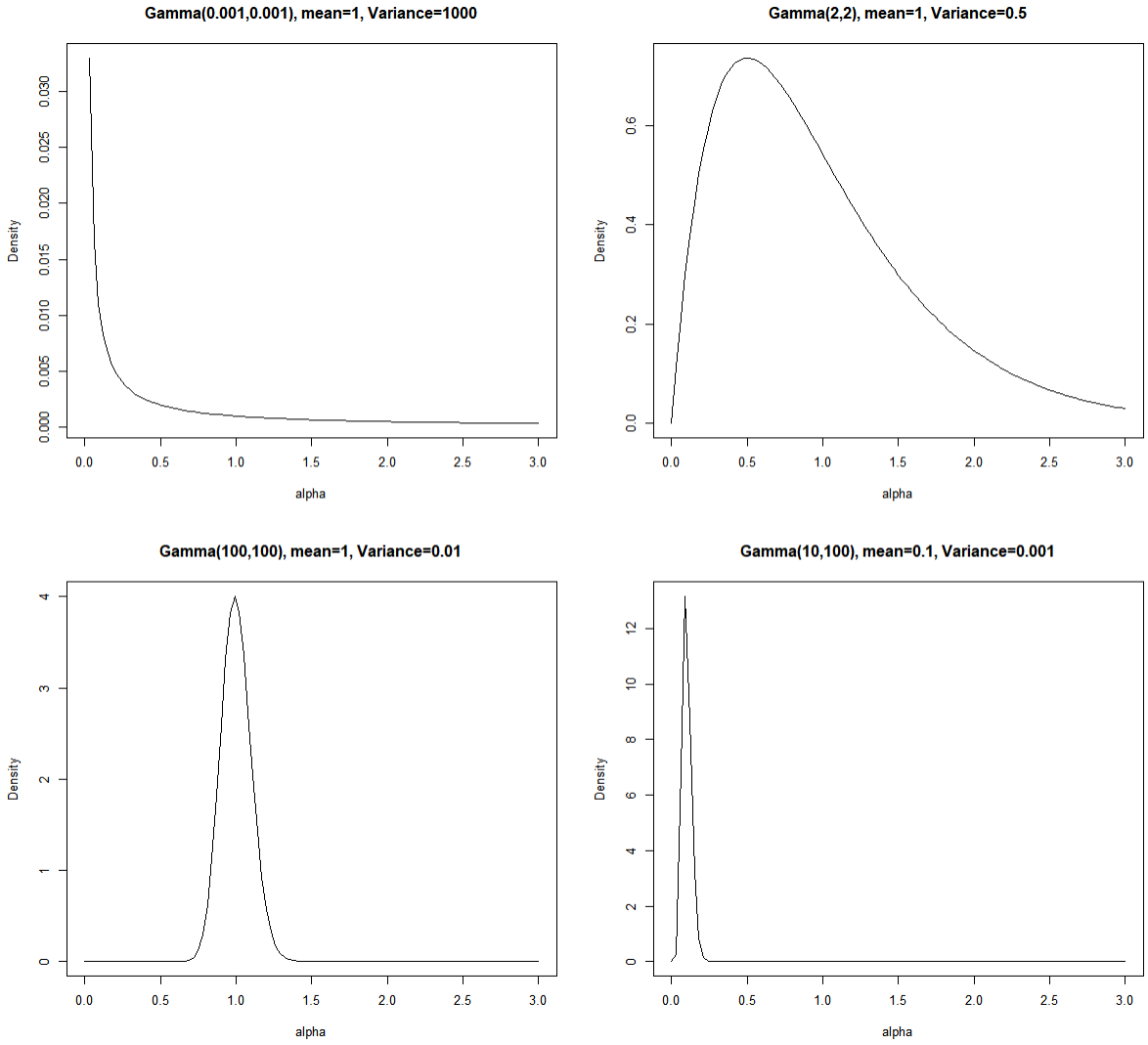

Figure 1. Density curves for the four precision parameter priors used in the simulation study

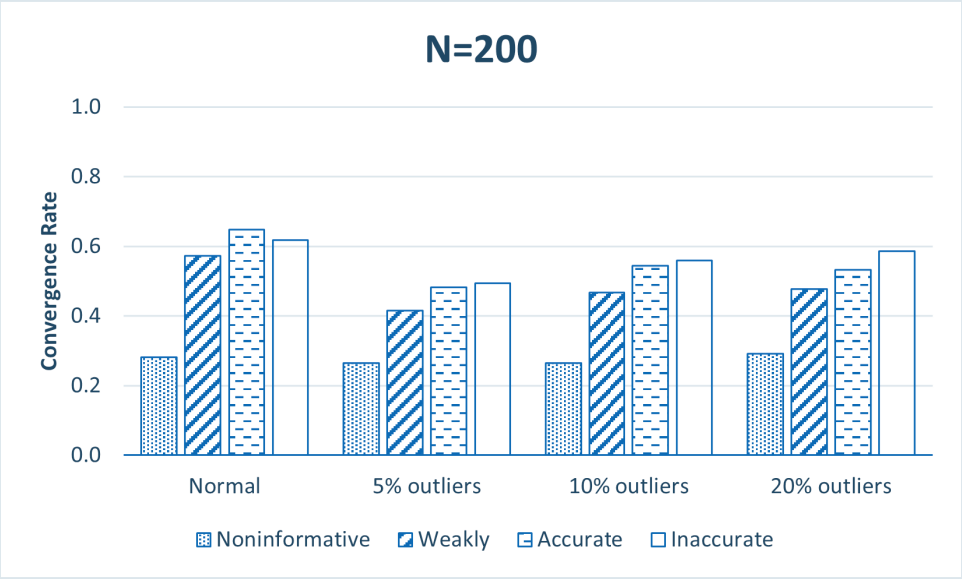

Figure 2. Convergence rate for different priors when $N = 200$