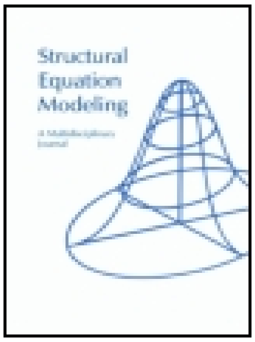

Routledge
Taylor & Francis Group

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: http://www.tandfonline.com/loi/hsem20

Using Bayesian Statistics to Model Uncertainty in Mixture Models: A Sensitivity Analysis of Priors

Sarah Depaoli, Yuzhu Yang & John Felt

Structural Equation Modeling: A Multidisciplinary Journal, 00: 1–18, 2016

ISSN: 1070-5511 print / 1532-8007 online
DOI: 10.1080/10705511.2016.1250640

Using Bayesian Statistics to Model Uncertainty in Mixture Models: A Sensitivity Analysis of Priors

Sarah Depaoli, Yuzhu Yang, and John Felt
University of California, Merced

The Bayesian estimation framework has specific benefits that can aid in the estimation of mixture models. Previous research has shown that the use of priors to capture (un)certainty in latent class sizes has the potential to greatly improve estimation accuracy of a mixture model. These priors can be beneficial in mixture modeling, but proper specification is key. A sensitivity analysis of priors is essential to understand the impact of the prior on the latent classes, whether diffuse or informed priors are implemented. We illustrate a full sensitivity analysis on Dirichlet priors for the class proportions of a latent growth mixture model. We show that substantive results can (drastically) shift as the prior setting is modified, even if only slightly. Math assessment data were used from the Early Childhood Longitudinal Study–Kindergarten class. We conclude with a discussion about final model interpretations when estimates are highly influenced by prior settings.

Keywords: Bayesian inference, mixture modeling, modeling uncertainty, priors, sensitivity analysis

Bayesian statistics has been making its way into the structural equation modeling (SEM) framework for several years now (see Kaplan & Depaoli, 2012; Lee, 2007; Lee & Song, 2012; Muthén & Asparouhov, 2012). Bayesian estimation carries many advantages. Perhaps the most flexible feature is the ability for researchers to capture degrees of (un) certainty about model parameters into the estimation process. Simulation work (e.g., Hox, van de Schoot, & Matthijsse, 2012; Kim, Suh, Kim, Albanese, & Langer, 2013) has shown this to be a great advantage for improving the estimation accuracy of structural equation models. Further, this feature of Bayesian statistics is particularly advantageous to latent variable mixture modeling (see Frühwirth-Schnatter, 2010; Lee, 2007).

Mixture models can be notoriously difficult to estimate because they carry complexities that can impede the ability to properly recover latent classes. The Bayesian estimation framework can mitigate some of these drawbacks to mixture modeling and provide a potentially more useful (and accurate) tool for applied researchers. Prior distributions (or *priors*) can be used to incorporate additional information about the mixture class solution. Overall, priors represent how (un)certain the researcher is about model parameters being estimated, and this level of certainty can be quite important to include in the model.[1] When accurate informed priors are specified, the amount of theory (via the prior) incorporated into the model estimation process increases. When all else is held equal (e.g., total sample size, amount of class separation, relative class sizes, etc.), then the use of informed (accurate) priors should be expected to outperform the frequentist or diffuse-prior Bayesian approach. The amount of theory incorporated into the model via priors

Correspondence should be addressed to Sarah Depaoli, Assistant Professor, Psychological Sciences, University of California, Merced, 5200 N. Lake Road, Merced, CA 95343. E-mail: sdepaoli@ucmerced.edu

[1] Uncertainty in Bayesian mixture modeling does not solely refer to the selection of priors. It can also refer to *model* uncertainty, which would fall under the issue of class enumeration with respect to mixture modeling. The issue of class enumeration is not directly addressed in this study, largely because the focus here is on the sensitivity analysis of prior settings and space does not allow for an assessment of both forms of uncertainty (i.e., model and prior uncertainty). However, the interested reader can see the following for more information on Bayesian class selection: Celeux, Forbes, Robert, and Titterington (2006) and Zhang, Lai, Lu, and Tong (2013). Bayesian model selection of mixture models is still a growing area that is in need of further research to develop tools that can properly select among competing Bayesian mixture models.

can also be tied to the sample size needed to properly estimate latent classes. Specifically, we can assume that with strong and accurate priors, less data would likely be needed to properly estimate parameter values. In contrast, more data would be needed in cases where no theory was incorporated (e.g., through frequentist estimation, or where diffuse priors are used).

Incorporating informed priors into the estimation process has the ability to improve convergence and improve the accuracy of latent class recovery. The additional information provided by priors can act to increase the power to detect small but substantively relevant latent classes. Priors can also aid in improving the estimation accuracy of latent growth trajectories.

Priors play an important role in mixture modeling, but their specification represents a main challenge in Bayesian modeling (MacCallum, Edwards, & Cai, 2012). In the case of mixture modeling, priors can incorporate additional information about the number of latent classes, the relative sizes of the latent classes, and class-specific parameter values. A researcher with no previous knowledge about these model features can display this uncertainty through noninformative (or diffuse) priors, those that represent a complete lack of knowledge about the model parameters and latent class structure.[2] In contrast, if a good deal of knowledge is present about the number of latent classes, their sizes, or other model parameter values, then the researcher can specify a prior with a lot of certainty (or information) about these elements in the model. This ability to capture degrees of certainty within the model is a major benefit to Bayesian mixture modeling.

Regardless of whether information is available, it is still quite common for researchers to rely on default prior settings in the software (van de Schoot, Winter, Zondervan-Zwijnenburg, Ryan, & Depaoli, in press). These priors are automatically determined (depending on software), and they typically represent complete uncertainty about the model parameters. Particularly within mixture modeling, this approach to specifying priors might not be the best solution. Implementation of informed or subjective priors has been shown to be ideal for mixture models. However, if inaccurate priors are specified, then results can display high degrees of bias and impaired efficiency (Baldwin & Fellingham, 2013; Depaoli, 2013). Thus, if subjective priors are implemented, then it is also recommended to conduct a thorough sensitivity analysis of those priors to understand the impact the additional information has on the final mixture model results (Kruschke, 2015; Muthén & Asparouhov, 2012).

There are currently very few examples (if any) for how to conduct a thorough sensitivity analysis on priors for a mixture model. We use a large-scale database to illustrate such an analysis for latent variable mixture models. We also provide specific guidelines for implementation of Bayesian mixture modeling and interpretation of disparate results across a variety of priors representing different degrees of (un)certainty about the mixture model.

Goals of This Article and Intended Audience

The main goals of this article include discussing how Bayesian inference can be applied to mixture modeling, illustrating the importance of a prior sensitivity analysis, and discussing how to interpret results that fluctuate with different prior settings. Ultimately, we aim to show how Bayesian estimation can be used to capture (un)certainty in latent classes. Our specific focus surrounds handling the (un)certainty in the size of the latent classes being estimated, and showing how sensitive results are to prior settings that incorporate this (un)certainty. The topics in this article are particularly relevant to researchers who are:

- Interested in Bayesian inference applied to mixture models.
- Working with mixture models where one or more latent classes are known to be relatively smaller than the others.
- Conducting a sensitivity analysis of priors for a mixture model.
- Wondering how to handle disparate results from a sensitivity analysis.

Structure of This Article

The remainder of this article is structured as follows. First, we include a brief discussion of the use of Bayesian statistics in SEM, and we expand this discussion to Bayesian inference in mixture modeling. Next, we highlight the main benefits of Bayesian inference applied to mixture models, which includes details of previous findings in the methodological (e.g., simulation) and applied literature. We then discuss the importance of prior sensitivity analysis in Bayesian mixture models. This discussion particularly focuses on the priors related to the mixture components of the model. A detailed empirical example using data from the Early Childhood Longitudinal Study–Kindergarten class (ECLS–K; National Center for Education Statistics [NCES], 2001) is included. We derive priors for the analysis of math assessment data and conduct a full sensitivity analysis on the prior for the latent class proportions. Prior settings are further examined as a function of sample size. We conclude with some general thoughts about Bayesian mixture modeling.

[2] The term *noninformative prior* refers to the case where researchers supply vague information about the population parameter value; the prior is typically defined with a very wide variance (Gill, 2008). Although noninformative is one term commonly used in the Bayesian literature to describe this type of prior (see, e.g., Gelman et al., 2013), other phrases such as *diffuse* (see, e.g., Gill, 2008), or *flat* (Jeffreys, 1961) are also used to describe this type of prior. We use noninformative and diffuse interchangeably in this article.

A BRIEF INTRODUCTION TO BAYESIAN STATISTICS IN SEM

Bayesian SEM has been increasing in popularity and use over the last decade. Van de Schoot et al. (in press) recently conducted a thorough systematic review examining the use of Bayesian statistics in psychology (and related fields) over the last 25 years. In this review, 1,579 Bayesian papers were reviewed, spanning theoretical (e.g., technical or tutorial), simulation, and empirical papers. The trends indicated that Bayesian SEM (including mixture modeling) was the second most common type of statistical model detailed in the technical papers (16.2%) and simulation papers (25.8%); the most common type of model in these categories was item response models. Within the empirical Bayesian papers, SEM was the most used model type (26.0%). There has been a steep increase in the number of SEM empirical papers, especially since 2012, whereas technical and simulation papers have been published at a relatively steady (but growing) rate since about 2006. The trends suggest that the use of Bayesian SEM in empirical work will continue to grow at a faster rate in the coming years.

Although there are many differences between the Bayesian and frequentist (e.g., maximum likelihood) estimation frameworks, the major distinction is in the use of background knowledge through priors. Priors represent an expectation or belief about the population parameters, and they are incorporated into the model using Bayes' theorem. Bayes' theorem states that the posterior distribution (i.e., the distribution representing the parameter estimate) is a product of the likelihood of the sample data and the prior divided by the marginal distribution; the marginal distribution is a constant in this equation because it does not contain model parameters. Equation 1 is written as:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{1}$$

where $p(\theta|y)$ represents the posterior of model parameter θ given data y, $p(y|\theta)$ represents the data likelihood, $p(\theta)$ is the prior distribution reflecting previous knowledge about model parameter θ, and $p(y)$ is the normalizing constant; Equation 1 reduces to $p(\theta|y) \propto p(y|\theta)p(\theta)$. For more information on Bayesian statistics in SEM, see Lee (2007), Lee and Song (2012), Kaplan and Depaoli (2012), MacCallum et al. (2012), Muthén and Asparouhov (2012), Rindskopf (2012), Yang and Dunson (2010), and Zhang, Hamagami, Wang, Grimm, and Nesselroade (2007).

Some benefits of Bayesian SEM include the use of approximate zeros to minimize specification errors involving fixed-zero loadings (Muthén & Asparouhov, 2012), improved estimation accuracy in cases of small sample sizes (see e.g., Muthén & Asparouhov, 2012; Zhang et al., 2007), and the ability to incorporate knowledge into the estimation process via priors (van de Schoot, Broere, Perryck, Zondervan-Zwijnenburg, & van Loey, 2015). Priors allow the user to specify degrees of (un)certainty about population parameters and then update this knowledge when new evidence is presented. The role of the prior is crucial in SEM, and especially within mixture models. As the amount of information (or precision) is increased, the prior plays a much stronger role in determining the final model estimate.[3] The prior can be used to increase statistical power under cases of small samples. Zhang et al. (2007) illustrated that reasonable estimates can be obtained in a latent growth model (i.e., a growth model that does not incorporate latent classes) with as few as 20 cases when informative priors are implemented. The selection of priors, and the subsequent investigation of their impact, is a key component to proper execution and understanding of the Bayesian analysis. The role of the prior distribution is a topic particularly relevant to mixture models, and we describe this issue in more detail in the next section.

BAYESIAN INFERENCE IN MIXTURE MODELING

The ability to estimate mixture models through the Bayesian framework has existed for some time now. Some initial developments of Bayesian mixture modeling are described in Diebolt and Robert (1994). This work illustrated the use of the Gibbs sampler (Geman & Geman, 1984) through the Markov chain Monte Carlo (MCMC) estimation algorithm, where the missing data structure of the mixture model is used during the estimation process. The Bayesian framework allows for the specification of priors with varying degrees of informativeness for the latent class proportions of a mixture model. By allowing for different levels of informativeness, the Bayesian framework offers a great deal of flexibility for estimating mixture models. This flexibility is particularly useful for modeling an unknown number of mixture components of unknown sizes (Frühwirth-Schnatter, 2001; Lee, 2007). In this section, we highlight some of the main areas of mixture modeling where Bayesian estimation has shown to be especially beneficial.

Model Complexity

Mixture models are plagued with issues related to assumption violation (e.g., properly specified covariance matrix; Bauer, 2007), convergence (Hipp & Bauer, 2006), and estimation accuracy (Depaoli, 2012, 2013; Tueller & Lubke, 2010). These issues are, at least in part, tied to the complexity of the model being estimated. Although the difficulties

[3] The term *precision* is a technical term in Bayesian statistics referring directly to the informativeness of the prior. In the case of a normal prior, for example, the precision represents the inverse of the variance hyperparameter.

surrounding mixture model estimation exist no matter the estimation framework being implemented, the Bayesian framework has been shown to have some distinct advantages over frequentist estimation. Some complex models cannot be estimated at all using conventional (frequentist) estimation methods (e.g., Kim et al., 2013). These models might be intractable due to high dimensional numerical integration needed to obtain the maximum likelihood estimates. The Bayesian framework can be used to circumvent the issues posed by numerical integration, and otherwise intractable models can be estimated.

Even when complex models can be estimated through conventional (frequentist) methods, the accuracy of the estimated model parameters might be quite poor. For example, Tueller and Lubke (2010) found that the recovery of latent classes in a mixture SEM was linked, in part, to the complexity of the model. In an empirical investigation of burn victims experiencing posttraumatic stress disorder, small but substantively important latent classes were recovered with the use of Bayesian methods (van de Schoot et al., 2015). In this case, maximum likelihood was underpowered and yielded poor coverage, whereas Bayesian methods with informed priors were able to properly extract and estimate latent classes.

As model complexity increases, additional difficulties can arise in latent class recovery. Incorporating prior knowledge into the estimation process through Bayesian statistics has been shown to be an important component to improving the estimation of mixture models. There are also added benefits beyond improved estimates. For example, Lu, Zhang, and Lubke (2011) illustrated the benefits of Bayesian estimation for handling nonignorable missing data in the context of latent growth mixture models (LGMMs). Bayesian methods were used to model a case where missing data were dependent on the latent class variable, as well as other explanatory variables in the model. Findings indicated that the Bayesian framework was able to perform well under conditions of nonignorable missing data for the LGMM. The ability for Bayesian methods to handle complex models, including mixture models, is a main benefit for using the estimation approach (Frühwirth-Schnatter, 2010; Lee, 2007).

Class Separation

Another topic closely tied to estimation issues within mixture modeling is how separate, or distinct, the latent classes are from one another. When classes are poorly separated, or more difficult to define from one another, then estimation of class-specific parameters (including the latent class proportions) can be quite poor. Much of the previous work examining class separation has been conducted in the frequentist framework. Findings generally point toward the need for larger sample sizes and greater class separation to obtain adequate parameter estimates and convergence (e.g., Depaoli, 2013; Tofighi & Enders, 2008; Tolvanen, 2007; Tueller & Lubke, 2010; van de Schoot et al., 2015). In other words, it might be difficult to identify a small or poorly separated latent class in the frequentist framework, despite substantive relevance.

It is typically not viable to control sample sizes or class separation in such a way that estimation accuracy is improved. Larger sample sizes are not always possible to obtain given restrictions to the study design, resources for data collection, or the use of small populations. Further, it is near impossible to control class separation. If populations are less separated from one another but still represent distinctly important groups, then the researcher must use statistical methodology that will help identify these groups and properly estimate model parameters. One method to circumvent these issues is to incorporate knowledge about the latent classes directly into the estimation process via priors.

Depaoli (2013) presented a simulation study looking at the impact of Bayesian estimation in the context of latent class separation for the LGMM. In this investigation, latent class separation was varied from extremely poor separation (i.e., classes with very similar growth trajectories, differing only slightly in the population) to high separation (i.e., completely distinct latent classes). Findings indicated that the inclusion of informed priors within the Bayesian framework allowed for proper latent class recovery and parameter estimation—especially when latent classes were poorly separated. Several different levels of prior informativeness were investigated, and many of the different Bayesian conditions outperformed the ability of frequentist estimation to properly recover the latent classes. The one exception, where recovery results were also poor, was with the Bayesian condition where noninformative priors were specified on all model parameters. We discuss this topic in more detail in the context of the example illustrated later. Sample sizes in the latent classes play an interesting role when crossed with class separation, and this topic is expanded on next.

Sample Sizes of Classes

Previous research on latent classes has indicated that estimation accuracy and proper class assignment is tied to sample size. Class assignment is negatively affected with smaller sample sizes when classes are poorly separated (e.g., Tueller & Lubke, 2010). This finding is not too surprising: Larger sample sizes, which is an issue tied to power, increase the ability to obtain proper estimates and class assignment. However, the literature also suggests that the relative size of the latent class can directly affect the accuracy of results. Several studies (e.g., Depaoli, 2013; Tofighi & Enders, 2008; Tolvanen, 2007; Tueller & Lubke, 2010) have found that latent classes with relatively fewer cases (e.g., 7% of the total sample size) are much more difficult to properly estimate. The estimation problems exist even if the absolute size of that class is considered large (e.g., 150 cases in a class might seem adequate, but if this class is much smaller than the others, there could be estimation accuracy problems). In other words, it is difficult to estimate a true minority class with relatively few cases, despite the total sample size. This difficulty increases further when classes are poorly separated.

When prior distributions are incorporated into the model estimation process, then the problematic issues resulting from small minority classes can be greatly diminished. This, in part, is due to manipulation of the prior distribution associated with the latent class proportions. It is well known that priors can have a large impact on final model estimates in cases of small sample sizes (see, e.g., Ghosh & Mukerjee, 1992; Zhang et al., 2007). However, it is also true that the prior can impact final estimates under moderate or large sample sizes (Depaoli & Clifton, 2015; Lambert, Sutton, Burton, Abrams, & Jones, 2005; Natarajan & McCulloch, 1998). Therefore, the impact of the prior might end up being very important in mixture modeling, regardless of how "small" or "large" the latent classes are.

THE IMPORTANCE OF PRIORS

Priors allow the researcher to incorporate knowledge and opinions directly into the model estimation process. This additional knowledge increases statistical power, making it an optimal choice when sample sizes are small. It can also be used as a method that amasses information across various studies, or collective knowledge in the field. In a sense, priors can act as the current state of knowledge regarding model parameters, and this information can be comprised of a collection of findings, expert knowledge, opinions, or even educated guesses.

Priors play a nontrivial role in Bayesian estimation. It follows that the specification or defining properties of the prior being implemented are key components in the final results obtained. Within the context of mixture models, the last statement cannot be underscored enough. Prior specifications have been shown to have a large and impactful role on the final model results for mixture models (see, e.g., Depaoli, 2012, 2013; van de Schoot et al., 2015). It could be that the opinion or knowledge expressed through the prior has a rather large impact on final model results. This artifact of Bayesian mixture modeling is quite concerning if left unaddressed in a research query. Without a full assessment of the impact and role of the prior, substantive results cannot be fully understood. If a researcher fails to report on the robustness of results to different forms of the prior (e.g., different opinions expressed through the prior), then it is possible that the mixture model results could be erroneously interpreted. As a result, a full assessment of the prior specified in a mixture model is a necessary step; this is also arguably true for any model (mixture or not) examined through the Bayesian framework (Depaoli & van de Schoot, 2015). If it is the case that the prior (e.g., opinion of the researcher) has a large impact on final model results, then the researcher must be clear about the instability of results and a new model might be warranted to explore.

In simulation work on latent mixture models, there are two main priors that have the largest impact on final model estimates: (a) the prior for the class proportions, and (b) the prior for the covariance matrix (if applicable to the model). We focus here on the former because it directly relates to all mixture models and is the component we feel has the most direct impact on model interpretation. For more information on the covariance structure prior, see Chung, Gelman, Rabe-Hesketh, Liu, and Dorie (2015) or Liu, Zhang, and Grimm (2016). We now turn our focus to the prior specified for the mixture class proportions.

Depaoli (2013) showed via simulation that when the mixture class proportions for latent variable models have diffuse priors, inaccurate estimates for the class proportions could be obtained. Depending on the setting, the prior can artificially pull class proportions to be equal in size, or they might reflect a single class for which almost all cases are assigned and the remaining classes are effectively empty—also a sign of label switching during estimation, which is discussed later. In contrast, when (accurate) informed priors are specified for the mixture class proportions, results obtained are typically reflective of true class proportions.

The focus of this article is to demonstrate how priors can be used to capture (un)certainty in the latent classes. We also show the impact of different prior settings on the mixture class proportions, and subsequently illustrate how final model results can substantively differ (sometimes substantially) due to the prior setting. Our focus is specifically on the need for a sensitivity analysis, as well as a discussion on how to interpret disparate findings that might result from the prior sensitivity analysis.

THE MAIN PRIOR DISTRIBUTION(S) FOR MIXTURE MODELING

In this article, we use conjugate priors because they are commonly implemented in the context of Bayesian SEM and mixture modeling (e.g., Lee, 2007; Muthén & Asparouhov, 2012). However, this restriction can be relaxed given that recent software and computing advances have made it possible to easily specify nonconjugate priors in a variety of modeling contexts. The most common prior distribution used for the latent class proportions in mixture modeling is the Dirichlet distribution. Proportions ($p_1 \ldots p_c$) for C latent classes are modeled Dirichlet (D) such that:

$$(p_1 \ldots p_c) \sim \mathrm{D}(\delta_1, \ldots, \delta_c),$$

where all proportions sum to 1.0, and the δ elements represent the hyperparameters of the Dirichlet prior distribution. In this case, the hyperparameters reflect the amount of (un) certainty the researcher wishes to specify regarding the size of the C latent classes. When $C = 2$, the beta distribution can also be used, as the Dirichlet distribution is a generalization of the beta family of distributions when $C > 2$. The δ elements become very important in determining how much information will be included in the prior regarding the latent class proportions.

Specifically, the δ elements in the Dirichlet prior represent the number of participants (or cases) that will be added to each class according to the prior. When all δ elements are set to 1, then the prior signifies that a single participant must be assigned to the class. This setting indicates that the class must technically exist (i.e., with at least one case assigned) but no additional information about final class size is provided through the prior. In this case, the Dirichlet distribution reduces to a uniform distribution, where all proportions are by definition equally likely because there is no prior knowledge specified about how small or large the classes are with respect to one another. Treating the prior as a uniform distribution might seem initially appealing because no knowledge is imposed on the latent class proportions. However, this specification of the Dirichlet (e.g., D(1, …,1)) has some specific drawbacks in mixture modeling, which can make this setting potentially problematic because it might allow for very small classes to form that might not be substantively interpretable (Asparouhov & Muthén, 2010). The alternative approach is to incorporate some additional information into the Dirichlet prior and model the latent class proportions with at least a semi (or weakly) informative prior. An example of this approach is the setting of D(10, …,10), indicating that 10 participants (or cases) would be assigned to each class. In this prior setting, tiny classes are less likely to form because the prior specifies a larger number of cases for each class. However, if the researcher suspects a very small (and substantively real) minority class exists, then this might not be the best setting because the classes might be inappropriately forced to be equal in size (Depaoli, 2013).

THE IMPORTANCE OF A SENSITIVITY ANALYSIS

When priors are used to capture the amount of (un)certainty surrounding latent class proportions, or any other parameter for that matter, it is important to conduct a full sensitivity analysis of the prior setting to understand the role the prior is playing in the final model estimates (Depaoli & van de Schoot, 2015; Kruschke, 2015; Muthén & Asparouhov, 2012; Zhang et al., 2007). Priors can affect different parameters in different ways. It could be that a prior for a mean structure parameter will have little influence on the estimate, but a mixture component or variance component might be much more sensitive to the prior setting.

Whenever informative (subjective) or weakly informative priors are implemented, it is ideal to compare the impact of the priors to diffuse priors, as well as other forms of subjective priors (Depaoli & van de Schoot, 2015). A comparison to diffuse priors will help the researcher understand the subjectivity of the informed prior, and it can aid in understanding the full impact of the prior. A comparison to other variations of the subjective prior can help highlight discrepancies in final model estimates when the subjective prior is modified, even if only slightly so. In the sensitivity analysis, a completely different form of prior (i.e., a different probability distribution) can be implemented, or the hyperparameter settings of the current prior distribution can be altered systematically. The idea is to make systematic perturbations of the original comparison prior. Then one can identify the impact that small fluctuations in the prior might have on the final model results. If estimates are very sensitive to the prior settings, then this finding is imperative to report to highlight the potential instability of results due to particular prior settings.

There are different ways to conduct a sensitivity analysis of priors. Formal methods have been developed, many of which involve differential calculus for computing the rate of change in posterior estimates as priors are modified (e.g., Berger, 1990; Dey, Ghosh, Lou, & Verdinelli, 1996; Kass, Tierney, & Kadane, 1989; McCulloch, 1989). In addition, sequential Monte Carlo methods have been proposed, which reduce the computational expense of traditional sensitivity analyses (Bornn, Doucet, & Gottardo, 2010). However, these approaches have only recently been implemented with complex multivariate models (see, e.g., Müller, 2012; Roos, Martins, Held, & Rue, 2015; Zhu, Ibrahim, & Tang, 2011), and they have not yet been extended into the SEM framework. Informal approaches for conducting a sensitivity analysis are far more popular and feasible with SEMs and mixture models. These approaches involve systematically modifying hyperparameter values and then displaying summary statistics (e.g., posterior means, standard deviations, or credible intervals) in plots to visually assess the impact of the prior settings.

EXAMPLE: BAYESIAN LATENT GROWTH MIXTURE MODELING AND A SENSITIVITY ANALYSIS OF PRIORS

This section presents a thorough example of incorporating (un)certainty into the mixture model by deriving informed priors, estimating a Bayesian LGMM, conducting a sensitivity analysis on informed priors, and viewing the impact of the priors across sample sizes. Empirical data are used to illustrate the process of Bayesian implementation for a mixture model. This example is meant to show the impact of priors and is not meant to draw substantive conclusions because issues of substantive model building were not addressed when selecting the model to present here.

The Model and Associated Priors

In this section, we describe the LGMM, which is the model implemented here. We also list the associated prior distributions for the model parameters, where applicable; discussion of specific prior settings follows in subsequent sections. In the case of a mixture model, we can assume data are generated from a mixture distribution ($f(\mathbf{y}_i|\mathbf{\Psi})$). The following represents the mixture density function for class c:

$$f(\mathbf{y}_i|\mathbf{\Psi}) = \sum_{c=1}^{C} \boldsymbol{\pi}_c f_c(\mathbf{y}_i|\boldsymbol{\theta}_c), \quad (3)$$

where $\mathbf{y}_i$ represents the vector of repeated measure manifest variables for person i across T time points, $\boldsymbol{\pi}_c$ represents the unknown mixture class proportion for the cth latent class where $c = (1, 2, \ldots, C)$, and f_c are the densities across the C latent classes that are assumed multivariate normal such that $y|c \sim \text{MVN}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$, where $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ represent a mean vector and covariance matrix for the data structure, which can be allowed to vary across classes. $\mathbf{\Psi}$ is a vector of unknown parameters:

$$\mathbf{\Psi} = (\boldsymbol{\pi}, \mathbf{\Theta}')', \ \boldsymbol{\pi} \sim \text{D}(\delta_1, \ldots, \delta_c) \quad (4)$$

where $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_c)$ represents the latent class proportions. The conjugate prior for these class proportion parameters $\boldsymbol{\pi}$ is the Dirichlet distribution (D), with hyperparameters $\delta_1, \ldots, \delta_c$ controlling how uniform the distribution will be over class proportions. This prior is typically formulated to represent the proportion of cases (or absolute number of cases, in some software languages) that are in each of the latent classes, but the formulation of this prior can vary. The prior can be written in terms of the $C - 1$ elements of the Dirichlet (as the last class proportion is fixed to uphold the condition $\sum_{c=1}^{C} \pi_c = 1.0$), the prior can be specified on all elements (or all classes), or it can be specified in terms of the latent class proportion thresholds (e.g., as in M*plus*). In Equation 4, $\mathbf{\Theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2', \ldots, \boldsymbol{\theta}_c')$, and represents the model parameters with $\boldsymbol{\theta}_c$ denoting a vector of model parameters for latent class c. Any element with a c subscript is allowed to vary across the latent classes; in some cases we have fixed parameters to be homogeneous across classes, but these restrictions can be freed if desired.

The measurement model for the LGMM can be written as:

$$\mathbf{y}_i = \mathbf{\Lambda}\boldsymbol{\eta}_{ic} + \boldsymbol{\varepsilon}_i, \quad (5)$$

where $\mathbf{\Lambda}$ is the factor-loading matrix, with T rows (time points) and K columns (number of latent growth factors, in this case two). Within this growth model, the first column of $\mathbf{\Lambda}$ was fixed to all 1s and the remaining column represented constant time values. The $\boldsymbol{\eta}_{ic}$ term is a vector of the latent growth factors with K elements (i.e., intercept and slope), and $\boldsymbol{\varepsilon}_i$ is a vector of normally distributed residuals with diagonal covariance matrix $\mathbf{\Omega}_\varepsilon$. Residual variances can be allowed to vary across time or they can be fixed across time, and independence is typically assumed between the elements in $\mathbf{\Omega}_\varepsilon$. The prior for the residual variances found in $\mathbf{\Omega}_\varepsilon$ can be specified for individual elements in the matrix if residual variances are assumed independent (i.e., if there is a zero covariance among residual variances). To specify a prior for this matrix, the notation will be expanded to represent individual elements in the $J \times J$ matrix. Let $\omega_{\varepsilon jj}$ denote a single element in the covariance matrix $\mathbf{\Omega}_{\varepsilon jj}$. The conjugate prior for the individual residual variances is the inverse gamma (IG) distribution written as:

$$\omega_{\varepsilon_{jj}} \sim \text{IG}\left[a_{\omega_{\varepsilon_{jj}}}, b_{\omega_{\varepsilon_{jj}}}\right],$$

where a and b are the shape and scale hyperparameters for the IG distribution, respectively.

The structural model can be written as:

$$\boldsymbol{\eta}_{ic} = \boldsymbol{\alpha}_c + \boldsymbol{\varsigma}_i, \boldsymbol{\alpha}_c \sim \text{N}\left(\boldsymbol{\mu}_{\alpha_c}, \boldsymbol{\sigma}^2\right) \quad (6)$$

where $\boldsymbol{\alpha}_c$ is a vector of factor means, and deviations of parameters from their population means are in vector $\boldsymbol{\varsigma}_i$, which is assumed to be normally distributed: $\text{N}(0, \mathbf{\Omega}_\varsigma)$, where $\mathbf{\Omega}_\varsigma$ is the latent growth factor covariance matrix. The conjugate prior for $\boldsymbol{\alpha}_c$ (the growth factor means for the intercept and slope terms) is a normal distribution (N), where $\boldsymbol{\mu}_{\alpha_c}$ represents the expectation for the growth factor means, and $\boldsymbol{\sigma}^2$ represents the variance of the prior. The last prior to define is for the latent growth parameter covariance matrix $\mathbf{\Omega}_\varsigma$. The conjugate prior for the covariance matrix is the inverse Wishart (IW) distribution and is written as:

$$\mathbf{\Omega}_\varsigma \sim \text{IW}[\mathbf{\Omega}, d],$$

where $\mathbf{\Omega}$ is a positive definite matrix of size p, and d is an integer that can vary depending on the informativeness of the prior. For alternative forms of this prior, see Liu et al. (2016).

We can combine the measurement and structural models into a reduced form:

$$\mathbf{y}_i = \mathbf{\Lambda}(\boldsymbol{\alpha}_c + \boldsymbol{\varsigma}_i) + \boldsymbol{\varepsilon}_i. \quad (7)$$

The model-implied mean and covariance of the reduced form is:

$$\boldsymbol{\mu}_c = \mathbf{\Lambda}\boldsymbol{\alpha}_c \quad (8)$$

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Omega}_\varsigma\mathbf{\Lambda}' + \mathbf{\Omega}_\varepsilon, \quad (9)$$

with $\boldsymbol{\mu}_c$ representing the mean vector of the observed repeated measures (ys) allowed to vary across latent classes, and $\boldsymbol{\Sigma}$ is the covariance matrix of the ys. In this example, the covariance matrix was treated as homogeneous across classes, which is evident because there is no longer a c subscript as presented earlier. However this restriction need not be present, and it is likely that empirical investigations would allow the matrix to vary across latent classes. We included the restriction on this matrix as an artifact of the sensitivity analysis presented.[4]

A picture of this model can be found in Figure 1. For more details surrounding LGMMs or advances and extensions of the model, see Bollen and Curran (2006), Dolan, Schmittmann, Lubke, and Neale (2005), Grimm and Ram (2009), Muthén (2004), Muthén and Shedden (1999), Palardy and Vermunt (2010), and Ram and Grimm (2009).

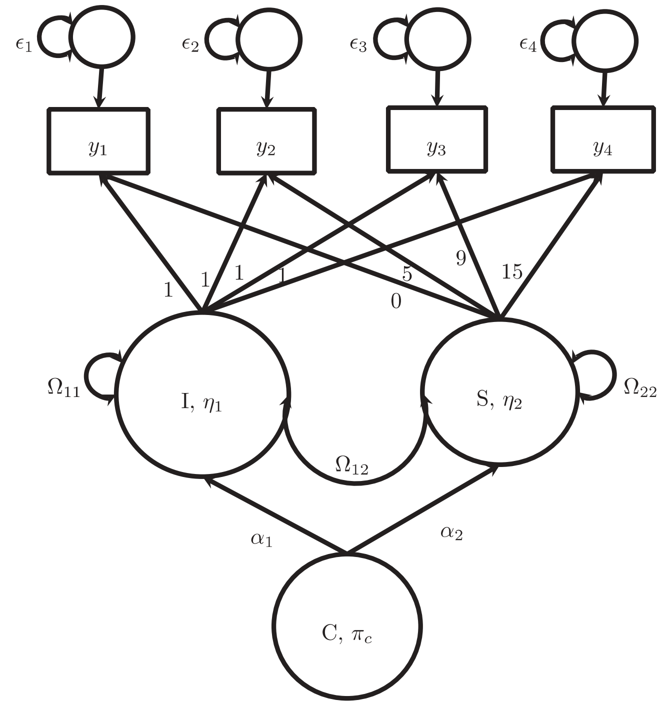

FIGURE 1 Linear latent growth mixture model (LGMM) used for ECLS–K math assessment data.

Data Source

In this example, we examined changes in math development throughout early childhood. Longitudinal data were based on children's item response theory scores on a math assessment, and data were obtained from the ECLS–K class of 1998–1999 (NCES, 2001). The math assessment was designed to measure conceptual, procedural, and problem-solving knowledge within specified content domains. Math development data were used here for children assessed in the fall kindergarten, spring kindergarten, fall first grade, and spring first grade sessions. The base year was fall 1998, and the unequal time spacing between the four waves was handled in the model coding by specifying the number of months between the waves in the model; see Kaplan (2002) for more detail on the spacing of data collection for the ECLS–K. The base sample size for this group of students is well over 3,000. However, the main goal of our example is to illustrate the impact of prior distributions. To accomplish this goal more effectively, we subsampled a much smaller group of students to better reflect sample sizes found in the behavioral and social sciences literature. The final model results are based on a sample consisting of 600 children, over four waves.

Main Analysis: Priors, Estimation Settings, and Results

Priors

There are many ways a researcher can derive informed priors for a Bayesian analysis. For ease of implementation, we opted to derive priors through a data-splitting technique.[5] In this case, one data set is randomly split. The first half is used to derive informed priors, and the second half is used for the full analysis with the informed priors.

We extracted a total of 1,200 random cases from the ECLS–K database. We randomly split this data set into two equally sized groups of children. The first group was used to derive subjective priors, and the second group was used to obtain final model estimates via the Bayesian

[4] Specifically, if the covariance matrix was allowed to vary across latent classes, then the IW priors placed on the class-specific matrices would need to be manipulated and studied through another sensitivity analysis. In fact, with a heterogeneous covariance matrix specified across classes, we found the IW prior too vague for the smaller class (C2). This resulted in the model not being properly estimated with the default IW prior for C2 using these data. In this case, the IW prior for C2 would require a sensitivity analysis because it would need to be altered for the model to properly estimate. Given that the focus in this article was on the prior for the latent class proportions (i.e., the Dirichlet prior), we opted to hold the priors for the covariance matrix constant and estimate the model with the covariance matrix restricted across classes. This restriction allowed us to fully investigate the impact of the Dirichlet prior with all other priors held constant. In an actual empirical investigation, one would have to conduct a full sensitivity analysis on all priors to fully understand how the combination of priors affects final model results. In this case, the covariance matrix restriction could be relaxed and the sensitivity analysis could also incorporate the IW prior (as well as all of the other priors specified).

[5] There are many additional methods that can be used for deriving priors. For example, one could conduct a meta-analysis (Ibrahim, Chen, & Sinha, 2001; Rietbergen, Klugkist, Janessen, Moons, & Hoijtink, 2011), consult with experts (Bijak & Wisniowski, 2010; Fransman et al., 2011; Howard, Maxwell, & Fleming, 2000; Martin et al., 2012; Morris, Oakley, & Crowe, 2014), use data-driven priors (Berger, 2006; Brown, 2008; Candel & Winkens, 2003; van der Linden, 2008), or use a data-splitting technique as we implemented here (Gelman, Bois, & Jiang, 1996; Moore, Reise, Depaoli, & Haviland, 2015).

estimation framework with these priors. All analyses were conducted using the M*plus* software program version 7.4 (Muthén & Muthén, 1998–2015); Bayesian mixture modeling can also be performed in a variety of other programs (e.g., via Jags [Plummer, 2015] in the R programming environment [R Core Team, 2016]).

A Bayesian linear two-class LGMM was estimated using the first half of the data with default diffuse priors (see Muthén & Muthén, 1998–2015, for more details about the default prior settings).[6] The sole purpose of the first half of the data was to derive subjective priors for future analysis. Posterior estimates and posterior standard deviations were used to form subjective priors for the second half of the data set. See Column 5 in Table 1 for the prior settings implemented; the note for Table 1 includes more detail about how these priors were derived from the first half of the data set. All subsequent analyses are based on the subjective priors implemented on the second half of the data set derived from the data-splitting technique.

Estimation and model settings

Once subjective priors were identified, the final analysis was estimated using the second data set derived from the data-splitting technique. The length of the Markov chain was set to have a range of iterations with a minimum and maximum number of iterations to reach convergence (as opposed to a single, fixed number of iterations). The minimum number of total iterations was set to 30,000 and the maximum number of iterations was set at 1 million; chains were allowed to run as long as necessary to reach convergence, as long as they did not exceed 1 million iterations. None of the cells in this article exceeded the minimum number of iterations. Thus, the burn-in phase was 15,000 iterations, with the remaining 15,000 iterations comprising the posterior distribution.[7]

[6] The purpose of this example is to illustrate how to incorporate (un) certainty into the mixture model and subsequently examine the impact of subjective priors. We decided to use a data-splitting technique, where Data Set 1 was used to estimate a LGMM with Bayesian diffuse (default) prior settings. Estimates from this first analysis were then converted into subjective priors implemented on select model parameters for Data Set 2. There are many different ways in which priors could have been obtained, and applied researchers should take great care when selecting the method(s) to implement when deriving priors. Even in the context of this data-splitting technique, we could have analyzed the first data set in a variety of ways (e.g., frequentist estimation or with subjective priors). The main point is to be purposeful when selecting priors, and then to always report the exact prior settings so that researchers can interpret results in the context of the prior.

[7] We reran the cells with increased chain lengths up to 300,000 iterations as the minimum number of iterations to check for stability when more iterations were requested. It is good practice to rerun analyses (especially mixture models, which can encounter increased problems with convergence) to ensure local convergence was not obtained (Depaoli & van de Schoot, 2015). Stable results across the runs were obtained so convergence was established.

One very important issue to address within Bayesian mixture modeling is label switching (see, e.g., Celeux, Hurn, & Robert, 2000; Frühwirth-Schnatter, 2001, 2010). Between-chain label switching occurs when there are multiple chains for each model parameter and one chain is estimating the parameter for Class 1 (C1) and the other chain is linked to Class 2 (C2). Averaging results across these chains gives nonsense information because the chains represent different latent classes. Using a single chain during MCMC monitoring removes this form of label switching altogether, which we did here. Within-chain label switching is when a single chain bounces between C1 and C2, but this can be identified if chain convergence is carefully monitored. To avoid within-chain label switching, we used an identifiability constraint on a model parameter believed to be disparate across the classes (i.e., C1 intercept mean < C2); this form of constraint will only work with a model parameter reflecting great separation between the classes. The chains were monitored for convergence using the Gelman and Rubin (1992) convergence diagnostic.

Results

The final model estimates for the Bayesian analysis using subjective priors are presented in Table 1. The posterior estimate, posterior standard deviation, and 95% credible interval are displayed for all model parameters. A picture of the final estimated growth trajectories for both classes is in Figure 2; shaded areas represent 95% credible regions for the growth parameters. The classes had relatively comparable growth rates, but different initial math assessment levels. Findings indicated there was a minority latent class (C2), with about 9% of cases, which represented a group of children with higher initial math assessment scores.

To capture the richness of Bayesian results, we have included highest density interval (HDI) plots in Figure 3 (C1) and Figure 4 (C2), which can also be used to depict final model estimates. These plots illustrate areas under the posterior distribution that are more likely for each of the model parameters. There is relatively small variability in each of the estimates in Figures 3 and 4. In addition, the final estimates for the class proportions came out to be very close to the subjective prior specified, with C1 and C2 proportions near 0.90 and 0.10, respectively.

Sensitivity Analysis: Method and Results

Method for creating prior sensitivity analysis conditions

As discussed previously, it is imperative to conduct a sensitivity analysis of priors when informed hyperparameter values have been implemented. The Dirichlet prior appears to have a particular impact on final model estimates for LGMMs, according to simulation findings (Depaoli, 2013). Additional simulation research found that informative Dirichlet prior settings are

TABLE 1
Results and Subjective Priors for a Linear Latent Growth Mixture Model Using Math Assessment Data

Parameter	*Posterior Estimate*	*Posterior SD*	*95% CI*	*Prior*
C1 proportion	0.909	0.011	[0.886, 0.928]	D(540, 60)
C2 proportion	0.091	0.011	[0.071, 0.114]	
Class 1 (C1)				
Intercept (I)	23.667	0.363	[22.936, 24.362]	N(23.471, 2.00704)
Slope (S)	4.145	0.061	[4.024, 4.267]	N(4.156, 0.04356)
Class 2 (C2)				
Intercept	41.590	1.471	[38.747, 44.598]	N(41.711, 20.896804)
Slope	4.629	0.254	[4.152, 5.153]	N(4.852, 0.45369)
Variances				
Intercept	26.843	3.638	[20.583, 34.955]	IW(0.000, −4)
Slope	0.983	0.106	[0.793, 1.209]	IW(0.000, −4)
Covariance(I,S)	4.794	0.449	[3.921, 5.688]	IW(0.000, −4)

Note. Posterior estimate is based on the median. *SD* = standard deviation; CI = credible interval; D = Dirichlet prior, placed on the class proportion threshold; N = normal prior; IW = inverse Wishart prior, placed on the covariance matrix, and represents default M*plus* settings. Findings represent the original model with subjective priors that were derived through a data-splitting technique using ECLS–K math achievement data. In the data-splitting technique, we randomly split 1,200 children into two groups. The first group was used to estimate an LGMM under Bayesian estimation with default priors from M*plus* (Analysis 1). We then pulled the estimates from Analysis 1 and created subjective priors for some model parameters, and we estimated the LGMM using these subjective priors (Column 5 of Table 1) and the second half of the data (Analysis 2, the final model results). Priors were constructed from Analysis 1 as follows. First, estimated latent class proportions from Analysis 1 were used directly in the informed Dirichlet prior for Analysis 2 to reflect Class 1 comprising 90% of cases and Class 2 comprising the remaining 10%. The growth parameter means for Analysis 2 received informed priors based on results from Analysis 1. The Analysis 1 final model estimates for the intercept and slope means were used as the mean hyperparameters for the normal priors for Analysis 2. To construct a relatively weakly informed prior, we then pulled the variance of the posterior from Analysis 1 for the intercept and slope means and multiplied that number by a constant of 10 to create the variance hyperparameter for Analysis 2. A multiple of 10 was selected after a small sensitivity analysis to ensure all priors would be considered weakly informed. For example, the final model estimate for the C1 intercept mean from Analysis 1 was 23.471 (posterior variance = 0.200704). We then used these results to create an informed prior for Analysis 2: N(23.471, 0.200704*10 = 2.00704) for the C1 intercept mean.

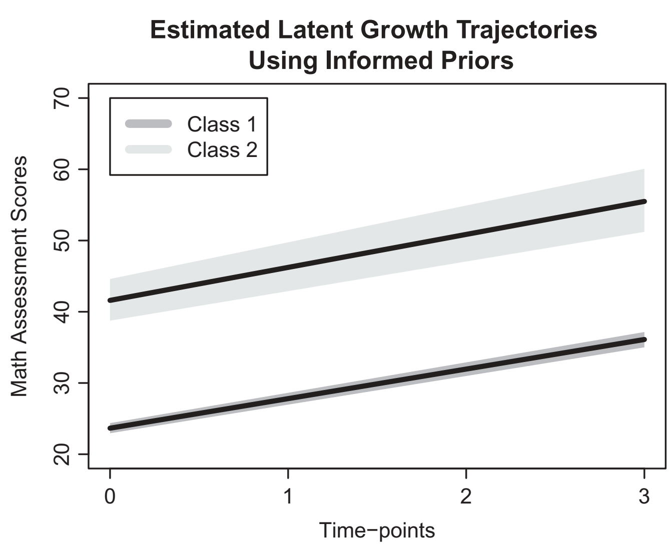

FIGURE 2 Estimated growth trajectories for math assessment data using informed (comparison) priors. Shaded regions represent 95% credible intervals for the latent growth parameters.

sometimes necessary to specify for model estimation to proceed for mixture models (see, e.g., Serang, Zhang, Helm, Steele, & Grimm, 2015). These findings indicate that it is essential to further assess the impact of this prior on final model results, especially if informed priors were needed. Even though this is an important component, it is rare to see a sensitivity analysis on the prior for the mixture class proportions, so there are very few (if any) informative guidelines for conducting this sort of assessment on the prior. Thus, we illustrate one way to conduct a sensitivity analysis on the class proportions. We identified several important categories of assessment regarding the Dirichlet prior. Specifically, it is important to assess (a) a variety of diffuse

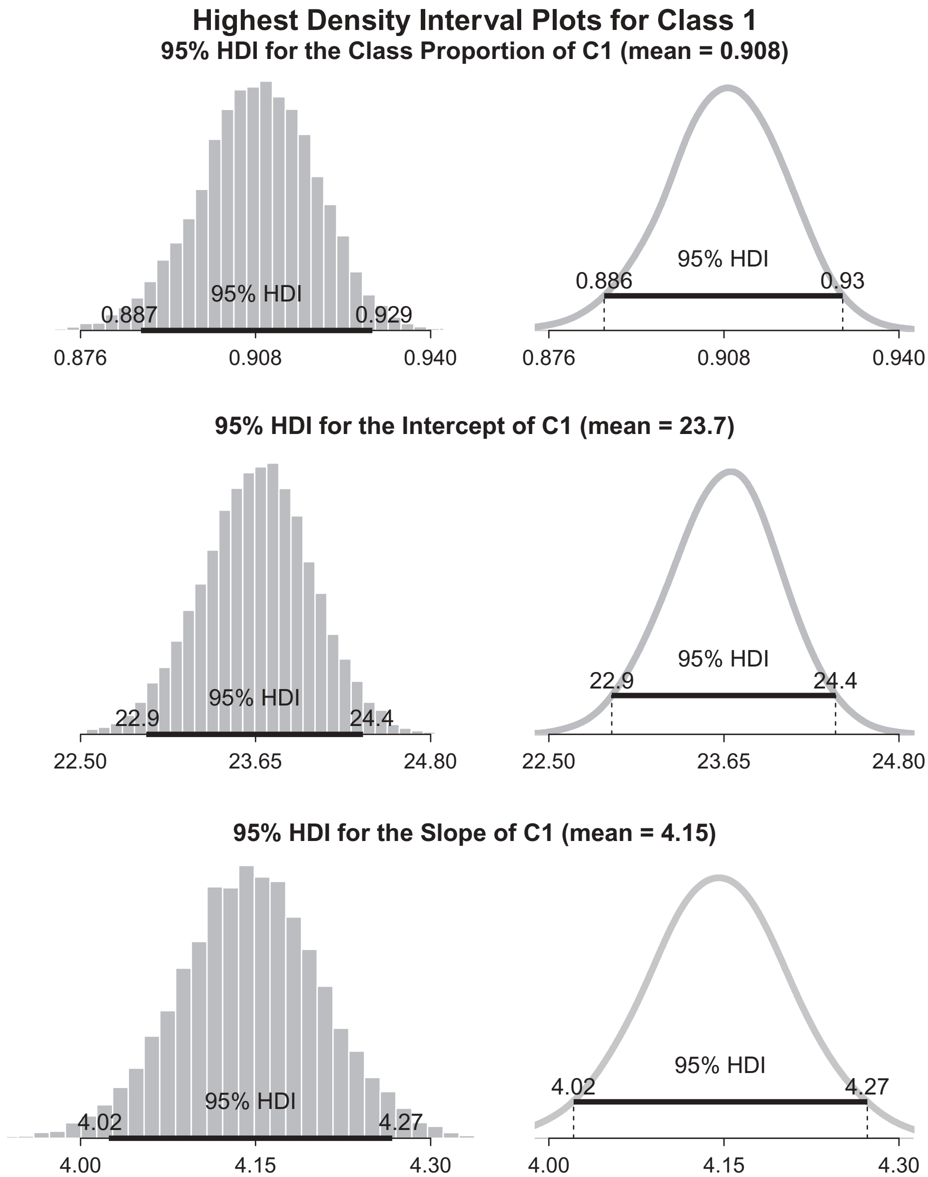

FIGURE 3 Highest density interval (HDI) plots for Class 1. Class proportion, intercept mean, and slope mean are included. These plots were created using the BEST package in R (Kruschke & Meredith, 2015).

or noninformative priors, (b) variations of the original subjective prior, and (c) variations of misinformed priors, which are specified as (potentially drastic) deviations from the original prior.

Column 1 of Table 2 provides labels for the different forms of the Dirichlet prior we implemented in this sensitivity analysis. The comparison, or original subjective model prior is listed first (Condition 0). Then there are three forms of diffuse priors (Conditions 1–3), two variations of the original prior (Conditions 4–5), and many misinformed priors (Conditions 6–14). The exact prior conditions for the sensitivity analysis are listed in the fourth column of Table 2 labeled (n = 600); we cover Columns 2 and 3 in the next main section. The Dirichlet prior was the only one modified in this sensitivity analysis illustration; all other prior settings mimicked those displayed in Table 1.

For the diffuse conditions, we looked at D(1,1), D(5,5), and D(10,10). The first condition of D(1,1) mimics the settings often used in the BUGS-like language (e.g., WinBUGS, OpenBUGS, and Jags). D(5,5) is slightly more informed in that 5 cases are specified for each class, and D (10,10) specifies 10 cases to each class. This last prior is the default setting in M*plus*, the program implemented here. All of these priors represent diffuse settings because very few cases are specified for each class, and it is important to examine a span of diffuse settings such as these to examine the full impact of noninformative variations.

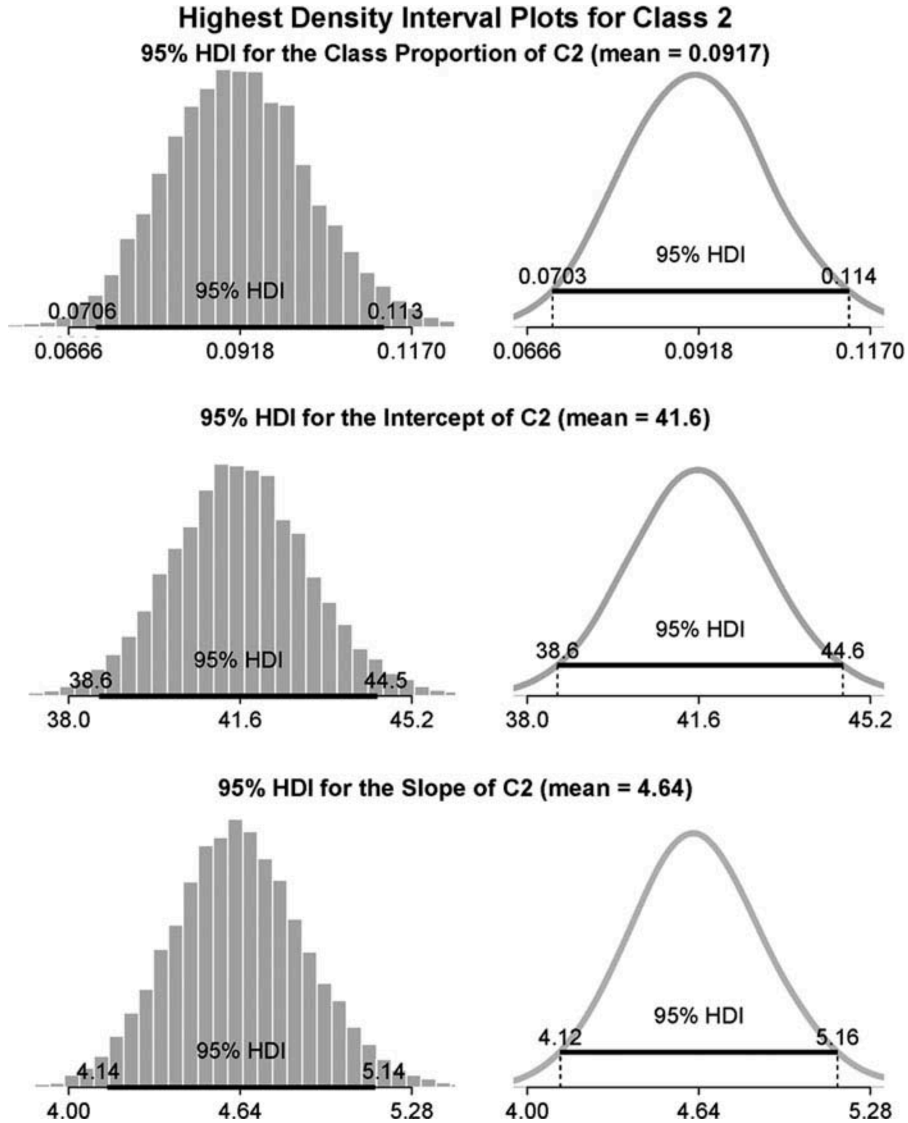

FIGURE 4 Highest density interval (HDI) plots for Class 2. Class proportion, intercept mean, and slope mean are included.

The informed priors from the original analysis were varied to create two less informative versions. Specifically, we maintained the subjective prior class proportions (i.e., 0.90 and 0.10 for the two classes, respectively), but we created two versions of this prior with less information. The half-informed prior specifies only 300 of the 600 cases within the prior D(270,30). In this case, the hyperparameter values of the original prior (D(540,60)) were divided in half to create the half-informed prior. Likewise, the quarter-informed prior divided the original hyperparameter values into a quarter of the size (D(135,15)). The motivation behind this approach is to create deviating versions of the original prior that held the 0.90/0.10 proportions intact, allowing for the data to drive the class proportions to a greater degree compared to the original subjective prior setting.

The remaining misinformed priors implemented in the sensitivity analysis were categorized as full-, half-, and quarter-informed priors in the same manner. First, we have a misinformed prior that represents C1 and C2 class proportions of 0.80 and 0.20, respectively. Second, we examined prior settings representing 0.70 and 0.30, respectively. Finally, we specified priors representing a 0.60 and 0.40 split. Again, there were three versions of each of these settings. For example, the 0.70/0.30 full-informed prior was D(420,180), the half-informed prior decreased these hyperparameters by half to be D(210,90), and the quarter-informed prior had hyperparameters a quarter the size to be D(105,45). In total, these 14 additional prior conditions should provide insight into the impact of the prior setting specified on the class proportions.

TABLE 2
Dirichlet Hyperparameter Values Used in the Sensitivity Analysis

	Sample Size Conditions		
Condition Name	*n = 200*	*n = 400*	*n = 600*
0. Comparison/full informed	D(180, 20)	D(360, 40)	D(540, 60)
1. Diffuse 1	D(1, 1)	D(1, 1)	D(1, 1)
2. Diffuse 2	D(5, 5)	D(5, 5)	D(5, 5)
3. Diffuse 3	D(10, 10)	D(10, 10)	D(10, 10)
4. Half-informed	D(90, 10)	D(180, 20)	D(270, 30)
5. Quarter-informed	D(45, 5)	D(90, 10)	D(135, 15)
6. Full .80/.20	D(160, 40)	D(320, 80)	D(480, 120)
7. Full .70/.30	D(140, 60)	D(280, 120)	D(420, 180)
8. Full .60/.40	D(120, 80)	D(240, 160)	D(360, 240)
9. Half .80/.20	D(80, 20)	D(160, 40)	D(240, 60)
10. Half .70/.30	D(70, 30)	D(140, 60)	D(210, 90)
11. Half .60/.40	D(60, 40)	D(120, 80)	D(180, 120)
12. Quarter .80/.20	D(40, 10)	D(80, 20)	D(120, 30)
13. Quarter .70/.30	D(35, 15)	D(70, 30)	D(105, 45)
14. Quarter .60/.40	D(30, 20)	D(60, 40)	D(90, 60)

Note. These condition names can be interpreted as follows. The Half 0.80/0.20 condition would be interpreted as a Dirichlet prior representing 80% of cases in Class 1 and 20% of cases in Class 2 that is half-informed. The half-informed prior represents the instance where only half of the total sample size would be used to construct the prior. For example, if $n = 400$, then $(400 * 0.80)/2 = 160$ and $(400 * 0.20)/2 = 40$, so the prior for this condition would be D (160, 40). This prior holds the 0.80/0.20 class proportions, but only incorporates half of the sample size into the prior.

TABLE 3
Sensitivity Analysis Results by Sample Size for the Class 2 Proportion

	n = 200			*n = 400*			*n = 600*		
Prior Condition	*Posterior Estimate*	*SD*	*95% CI*	*Posterior Estimate*	*SD*	*95% CI*	*Posterior Estimate*	*SD*	*95% CI*
0. Comparison/full-informed	0.095	0.020	[.061, .138]	0.091	0.014	[.067, .120]	0.091	0.011	[.071, .114]
1. Diffuse 1 D(1,1)	0.069	0.051	[.012, .202]	0.047	0.025	[.017, .113]	0.047	0.017	[.022, .089]
2. Diffuse 2 D(5,5)	0.169	0.047	[.086, .271]	0.101	0.038	[.043, 190]	0.081	0.027	[.044, .150]
3. Diffuse 3 D(10,10)	0.214	0.047	[.129, .315]	0.160	0.038	[.088, .238]	0.122	0.029	[.075, .189]
4. Half-informative	0.090	0.026	[.046, .148]	0.084	0.018	[.053, .122]	0.085	0.014	[.062, .115]
5. Quarter-informative	0.082	0.032	[.034, .156]	0.071	0.021	[.039, .123]	0.076	0.016	[.049, .113]
6. Full .80/.20	0.181	0.025	[.135, .233]	0.179	0.017	[.147, .214]	0.180	0.015	[.153, .210]
7. Full .70/.30	0.259	0.028	[.207, .316]	0.262	0.020	[.224, .302]	0.269	0.017	[.237, .303]
8. Full .60/.40	0.341	0.033	[.282, .412]	0.353	0.025	[.307, .404]	0.362	0.019	[.324, .401]
9. Half .80/.20	0.168	0.031	[.112, .233]	0.164	0.022	[.123, .209]	0.165	0.019	[.130, .204]
10. Half .70/.30	0.233	0.034	[.169, .304]	0.235	0.025	[.189, .286]	0.245	0.022	[.204, .289]
11. Half .60/.40	0.299	0.039	[.229, .384]	0.309	0.030	[.252, .372]	0.327	0.025	[.279, .378]
12. Quarter .80/.20	0.150	0.036	[.087, .227]	0.143	0.027	[.093, .199]	0.140	0.023	[.100, .190]
13. Quarter .70/.30	0.201	0.039	[.129, .284]	0.199	0.029	[.146, .259]	0.209	0.026	[.161, .263]
14. Quarter .60/.40	0.250	0.043	[.172, .343]	0.253	0.032	[.194, .320]	0.275	0.030	[.220, .337]

Note. See Table 2 note for description of prior condition labels. Posterior estimate is based on the median. *SD* = posterior standard deviation; CI = credible interval.

Sensitivity analysis results

Modifying one prior distribution through a sensitivity analysis can have an impact on any of the model parameters being estimated. For illustration, we focus on the impact the prior settings had on a single model parameter —namely, the C2 proportion. Table 3 ($n = 600$) provides the estimated class proportion for C2, the posterior standard deviation, and the 95% credible interval of the posterior. Figure 5 presents the posterior densities for the C2 proportion across all sensitivity analysis cells with $n = 600$. For reference, the posterior resulting from the original, subjective prior (Condition 0) is embedded within all of the plots (called Comparison); note that the posterior estimate for this condition was 0.091 for the C2 class proportion. Each of the plots has the same x- and y-axis scaling for direct comparison. Posteriors provide information regarding the central tendency and variation of the estimate.

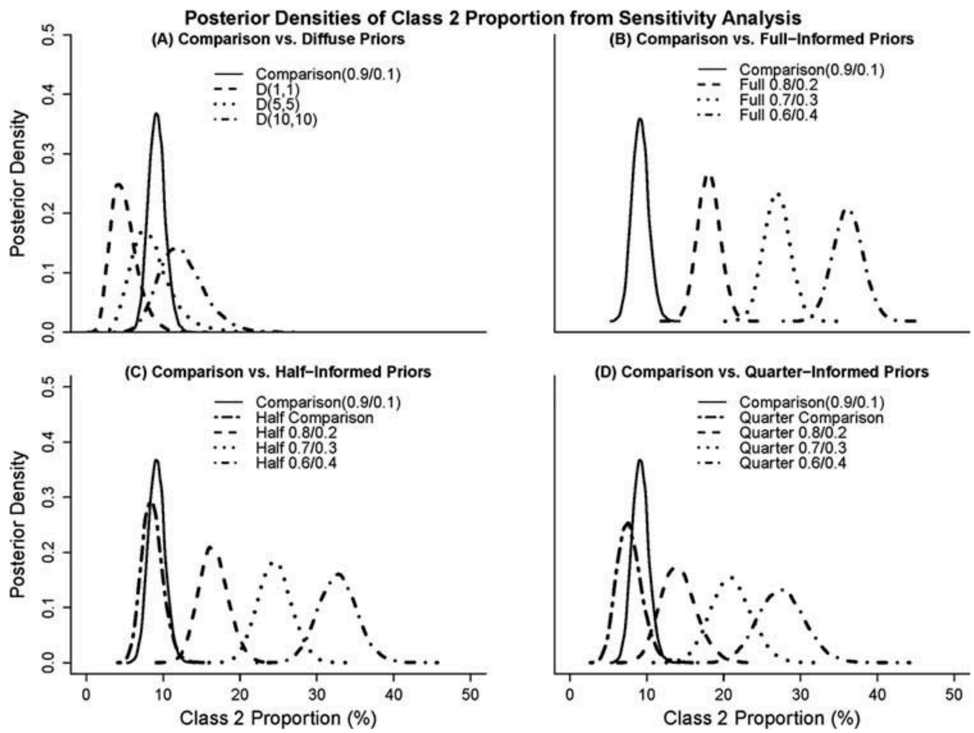

FIGURE 5 Posterior densities from sensitivity analysis of Class 2 proportion (n = 600).

The estimates resulting from the diffuse conditions (Plot A) ranged from 4.7% (D(1,1)) to 12.2% (D(10,10)), with D (5,5) producing an estimate of 8.1%, which was closest to the comparison condition. Notice that D(1,1) yielded an estimate pointing toward a more extreme minority class size with fewer than 5% of the cases assigned to C2. This posterior also had much less variability compared to those corresponding with the other diffuse priors. However, relative to the other plots in Figure 5, the diffuse prior settings all produced posteriors hovering relatively close to the informed comparison prior, with some substantial overlap with one another.

Plots B, C, and D represent the full-, half-, and quarter-informed priors for the remaining conditions, respectively. The posterior resulting from the original prior setting is still present in each of the plots for comparison. Note that across Plots C and D, the comparison model closely mimics the half- and quarter-informed versions of this prior. The same general pattern of results exists across Plots B, C, and D for the related conditions. Specifically, the 0.80/0.20 condition is pulled off to the right of the comparison prior condition in the full- (Plot B), half- (Plot C), and quarter-informed prior (Plot D) settings. The 0.60/0.40 condition is pulled drastically to the right of the comparison density. Even though the same general pattern of results is illustrated, there is more overlap in the respective posteriors as the prior becomes less informed. Specifically, increasing overlap in the posteriors can be seen moving from the full-informed priors in Plot B to the quarter-informed priors in Plot D. Notice also that the posterior for the 0.60/0.40 condition is shifted more toward the comparison (shifted left) in the quarter-informed setting (Plot D) compared to the full-informed setting of this prior (Plot B).

Sensitivity Analysis: How Results Vary by Sample Size

Next, we present the sensitivity analysis in terms of varying sample sizes. In an actual empirical investigation, the researcher would typically not conduct analyses akin to this section. The researcher would report an extensive sensitivity analysis on the full sample for each subjective prior. However, we wanted to illustrate how findings might remain constant or differ as sample sizes are altered. In some settings, the prior can have a large impact on final model results when sample sizes are decreased. This finding has not necessarily been the case with mixture models. In some forms of models, including mixture models, even relatively larger sample sizes can be affected by prior distribution settings in a similar way compared to lower sample sizes (see, e.g., Depaoli, 2013; Lambert et al., 2005; Natarajan & McCulloch, 1998).

In the following investigation, we apply the same sensitivity analysis conditions to lower sample sizes. We decreased the total sample size down to n = 400 and

n = 200 to show the impact of the sensitivity analysis prior settings when the number of cases (i.e., participants) is reduced. Along with the original number of cases n = 600, we might consider these conditions to represent relatively low, moderate, and high sample sizes. Notice that in the n = 200 condition, the C2 class proportion is presumably quite small ($n \approx 20$ or so); examples of class sizes this small can be found in van de Schoot et al. (2015) and Zhang et al. (2007), but informed priors are often necessary.

The same basic sensitivity analysis settings were used with these new sample sizes. We used the same three diffuse settings in their original form as before. Then we used the full-, half-, and quarter-informed priors for the following class proportion conditions: 0.90/0.10 (original, informed class proportions), 0.80/0.20, 0.70/0.30, and 0.60/0.40. We present solely on the C2 proportion to remain consistent with previous sections. However, any of the model priors could have been tracked in a similar manner, and the full body of sensitivity analysis results across sample sizes is available on request. Table 2 shows the exact hyperparameter values used for the latent class proportions across the different sample sizes examined; all other priors mimic those in Table 1, Column 5.

Sample size sensitivity analysis results

The results of the sensitivity analysis for the C2 proportion across different sample sizes are presented in Table 3 and Figure 6. Figure 6 has four main plots, separated by the type of prior. Plot A shows results for the three diffuse conditions. Plots B, C, and D illustrate results for the full-, half-, and quarter-informed prior settings, respectively. All of the plots contain results for the original, subjective (comparison) prior found in Table 2 (top row labeled Condition 0). Within each plot, the y axis represents the class proportion estimate for C2, the x axis represents the different sample size conditions, and the lines represent the different prior conditions. The data points for n = 600 conditions in these plots represent the final posterior estimates pulled from the posterior densities displayed in Figure 5.

It is striking to see that there are two main patterns of results across the plots. Notable, the sample size variations made virtually no impact on final model estimates in the prior conditions represented by Plots B, C, and D. These lines, although different from each other (implying a large impact of the prior), are basically flat. The lack of variation in estimates across sample sizes indicates that the results are essentially identical across the sample sizes included here. This is a rather interesting

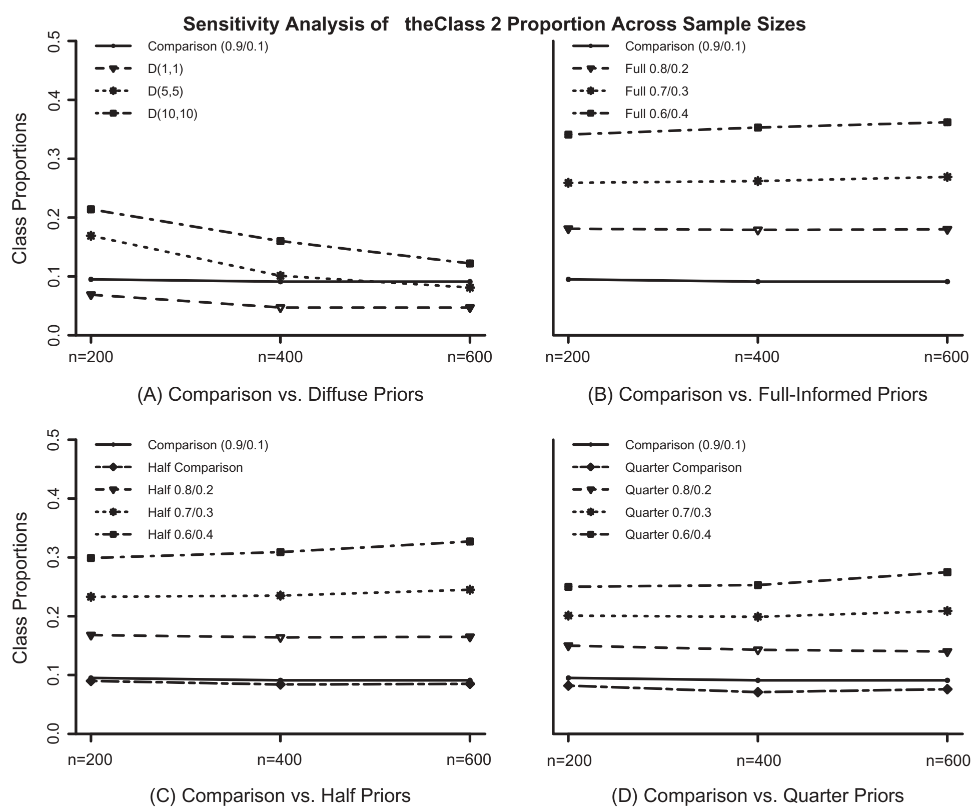

FIGURE 6 Sensitivity analysis results across sample sizes for Class 2 proportion (n = 600).

result because for the $n = 200$ condition, the minority class (C2) sample size is actually quite small and the impact was similar when this sample size was tripled to $n = 600$. The second notable pattern is in Plot A, where the diffuse prior conditions are presented. The D(5,5) and D(10,10) conditions varied more from the comparison condition and D(1,1) under the small sample sizes. Then as n increased, the estimates started to look more similar across the different conditions in this plot. This is the only plot exhibiting any impact of varying sample sizes, indicating that the prior settings in Plots B, C, and D might have swarmed the data information regardless of sample size.

How to Handle Disparate Results From a Sensitivity Analysis

This illustration was meant to act as a guide for capturing (un)certainty in a mixture model through the use of priors, and subsequently conducting a sensitivity analysis to better understand the impact of the prior settings. Specific to mixture modeling, even if a seemingly diffuse prior setting is used for the latent class proportions, results can be affected by the selection of that diffuse prior. Results showed that our different prior settings exhibited variability in how small the minority latent class really was. Depaoli (2013) and Zhang et al. (2007) similarly found that growth model results are highly variable and dependent on the prior setting. They subsequently discussed the importance of transparency in explicitly reporting the priors and drawing conclusions with the variability of results in mind.

Depaoli and van de Schoot (2015) expanded on this sentiment and recommended always conducting a thorough sensitivity analysis of priors, akin to what was presented here. Discrepancies in results across the sensitivity analysis cells are not necessarily troublesome. They simply indicate that the prior (or theory embedded within the prior) has a strong influence on final model estimates. This sort of finding can be quite interesting. If results from a subjective prior (e.g., the comparison model earlier) match those from other prior settings in the sensitivity analysis, then one can conclude that the prior (or theory) does not have much impact on final model results. However, if results are discrepant across prior settings, then this is a finding that will need more thought and attention. Further investigation regarding the impact of priors might even be warranted.

The researcher should note areas of instability in substantive conclusions, and discuss differences across results from different prior settings even if they are only minor. If results are highly variable across prior settings (as we saw here), then this could be an indication of model misspecification. However, if the original priors and model were specified correctly and results from the sensitivity analysis are still quite different, then this might just be the result of the study (Depaoli & van de Schoot, 2015). If this is the case, then the researcher must be completely transparent about the priors specified and the varying results obtained. It could be an indication of the instability of the theory-based model, the population, or other features of the study design. The key is to be clear about the findings and the discrepancies across different prior settings. These discrepancies might even lead toward a deeper understanding of the theory or population being studied.

CONCLUDING THOUGHTS

Bayesian methods are incredibly helpful for obtaining accurate and interpretable results in SEMs and mixture models. However, one of the main criticisms of Bayesian methods is the choice of priors. In this sensitivity analysis, we showed that the selection of priors could be highly influential in the case of mixture modeling. Our focus on the prior for the mixture class proportion was intentional given the impact that this prior can have on final latent class proportion estimates.

One important point to consider when estimating a Bayesian mixture model is whether previous knowledge can be incorporated into the model estimation process or not. There might be some substantive contexts where it is quite feasible to implement informed priors on class sizes. For example, in disease or clinical settings, there is often ample knowledge from previous research and experts to be able to construct informed priors regarding the number of latent classes and their respective sizes (see, e.g., Depaoli, van de Schoot, van Loey, & Sijbrandij, 2015; van de Schoot et al., 2015). However, in more exploratory inquiries, it might not be possible to construct informed priors about the class sizes because the number (and size) of the latent classes is unknown. In this case, diffuse or noninformed priors would likely be more appropriate to implement. In either situation, conducting a sensitivity analysis on the priors implemented is incredibly important for mixture models. We encourage researchers to experiment with the benefits of Bayesian estimation in mixture modeling, but we also caution them to be thorough and transparent in their investigation and reporting of the impact of the priors.

REFERENCES

Asparouhov, T., & Muthén, B. O. (2010). *Bayesian analysis using Mplus*. Los Angeles, CA: Muthén and Muthén.

Baldwin, S. A., & Fellingham, G. W. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods*, *18*, 151–164. doi:10.1037/a0030642

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, *42*, 757–786. doi:10.1080/00273170701710338

Berger, J. (1990). Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference*, *25*, 303–328. doi:10.1016/0378-3758(90)90079-A

Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, *3*, 385–402.

Bijak, J., & Wisniowski, A. (2010). Bayesian forecasting of immigration to selected European countries by using expert knowledge. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*, 775–796. doi:10.1111/j.1467-985X.2009.00635.x

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach*. Hoboken, NJ: Wiley.

Bornn, L., Doucet, A., & Gottardo, R. (2010). An efficient computational approach for prior sensitivity analysis and cross-validation. *The Canadian Journal of Statistics*, *38*, 47–64.

Brown, L. D. (2008). In-season prediction of batting averages: A field test of empirical Bayes and Bayes methodologies. *The Annals of Applied Statistics*, *2*, 113–152. doi:10.1214/07-AOAS138

Candel, J. J. M., & Winkens, B. (2003). Performance of empirical Bayes estimators of level-2 random parameters in multilevel analysis: A Monte Carlo study for longitudinal designs. *Journal of Educational and Behavioral Statistics*, *28*, 169–194. doi:10.3102/10769986028002169

Celeux, G., Forbers, F., Robert, C., & Titterington, D. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, *1*, 651–674. doi:10.1214/06-BA122

Celeux, G., Hurn, M., & Robert, C. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, *95*, 957–970. doi:10.1080/01621459.2000.10474285

Chung, Y., Gelman, A., Rabe-Hesketh, S., Liu, J., & Dorie, V. (2015). Weakly informative prior for point estimation of covariance matrices in hierarchical models. *Journal of Educational and Behavioral Statistics*, *40*, 136–157. doi:10.3102/1076998615570945

Depaoli, S. (2012). Measurement and structural model class separation in mixture CFA: ML/EM versus MCMC. *Structural Equation Modeling*, *19*, 178–203. doi:10.1080/10705511.2012.659614

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

Depaoli, S., & Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling*, *22*, 327–351. doi:10.1080/10705511.2014.937849

Depaoli, S., & van de Schoot, R. (2015). Improving transparency and replication in Bayesian statistics: The WAMBS-checklist. *Psychological Methods*. Advance online publication. doi:10.1037/met0000065

Depaoli, S., van de Schoot, R., van Loey, N., & Sijbrandij, M. (2015). Using Bayesian statistics for modeling PTSD through latent growth mixture modeling: Implementation and discussion. *European Journal of Psychotraumatology*, *6*, 27516. doi:10.3402/ejpt.v6.27516

Dey, D. K., Ghosh, S. K., Lou, K.-R., & Verdinelli, I. (1996). On local sensitivity measures in Bayesian analysis [with discussion and rejoinder]. *Bayesian Robustness*, *29*, 21–39.

Diebolt, J., & Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*, 363–375.

Dolan, C. V., Schmittmann, V. D., Lubke, G. H., & Neale, M. C. (2005). Regime switching in the latent growth curve model. *Structural Equation Modeling*, *12*, 94–119. doi:10.1207/s15328007sem1201_5

Fransman, W., van Tongeren, M., Cherrie, J. W., Tischer, M., Schneider, T., Schinkel, J., … Tielemans, E. (2011). Advanced Reach Tool (ART): Development of the mechanistic model. *Annals of Occupational Hygiene*, *55*, 957–979. doi:10.1093/annhyg/mer083

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, *96*, 194–209. doi:10.1198/016214501750333063

Frühwirth-Schnatter, S. (2010). *Finite mixture and Markov switching models*. New York, NY: Springer.

Gelman, A., Bois, F., & Jiang, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*, *91*, 1400–1412. doi:10.1080/01621459.1996.10476708

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. New York, NY: Chapman & Hall/CRC.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472. doi:10.1214/ss/1177011136

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, *6*, 721–741. doi:10.1109/TPAMI.1984.4767596

Ghosh, J. K., & Mukerjee, R. (1992). Non-informative priors (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics* (Vol. 4, pp. 195–210). Oxford, UK: Oxford University Press.

Gill, J. (2008). *Bayesian methods: A social and behavioral sciences approach*. Boca Raton, FL: CRC.

Grimm, K. J., & Ram, N. (2009). Nonlinear growth models in Mplus and SAS. *Structural Equation Modeling*, *16*, 676–701. doi:10.1080/10705510903206055

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, *11*, 36–53. doi:10.1037/1082-989X.11.1.36

Howard, G. S., Maxwell, S. E., & Fleming, K. J. (2000). The proof of the pudding: An illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. *Psychological Methods*, *5*, 315–332. doi:10.1037/1082-989X.5.3.315

Hox, J. J., van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods*, *6*, 87–93.

Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). *Bayesian survival analysis*. New York, NY: Springer.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Kaplan, D. (2002). Methodological advances in the analysis of individual growth with relevance to education policy. *Peabody Journal of Education*, *77*, 189–215. doi:10.1207/S15327930PJE7704_9

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650–673). New York, NY: Guilford.

Kass, R. E., Tierney, L., & Kadane, J. B. (1989). Approximate methods for assessing influence and sensitivity in Bayesian analysis. *Biometrika*, *76*, 663–674. doi:10.1093/biomet/76.4.663

Kim, S.-Y., Suh, Y., Kim, J.-S., Albanese, M., & Langer, M. M. (2013). Single and multiple ability estimation in the SEM framework: A non-informative Bayesian estimation approach. *Multivariate and Behavioral Research*, *48*, 563–591. doi:10.1080/00273171.2013.802647

Kruschke, J. K. (2015). *Doing Bayesian analysis: A tutorial with R, Jags, and STAN*. San Diego, CA: Academic.

Kruschke, J. K., & Meredith, M. (2015). *Package 'BEST'*. Retrieved from https://cran.r-project.org/web/packages/BEST/BEST.pdf

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., & Jones, D. R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine*, *24*, 2401–2428. doi:10.1002/(ISSN)1097-0258

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Hoboken, NJ: Wiley.

Lee, S.-Y., & Song, X.-Y. (2012). *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences*. Hoboken, NJ: Wiley.

Liu, H., Zhang, Z., & Grimm, K. J. (2016). Comparison of inverse Wishart and separation-strategy priors for Bayesian estimation of covariance parameter matrix in growth curve analysis. *Structural Equation Modeling*, *23*, 354–367. doi:10.1080/10705511.2015.1057285

Lu, Z. L., Zhang, Z., & Lubke, G. (2011). Bayesian inference for growth mixture models with latent class dependent missing data. *Multivariate Behavioral Research*, *46*, 567–597. doi:10.1080/00273171.2011.589261

MacCallum, R. C., Edwards, M. C., & Cai, L. (2012). Hopes and cautions in implementing Bayesian structural equation modeling. *Psychological Methods*, *17*, 340–345. doi:10.1037/a0027131

Martin, T. G., Burgman, M. A., Fidler, F., Kuhnert, P. M., Low-Choy, S., Mcbride, M., & Mengersen, K. (2012). Eliciting expert knowledge in conservation Science. *Conservation Biology*, *26*, 29–38. doi:10.1111/j.1523-1739.2011.01806.x

McCulloch, R. E. (1989). Local model influence. *Journal of the American Statistical Association*, *84*, 473–478. doi:10.1080/01621459.1989.10478793

Moore, T. M., Reise, S. P., Depaoli, S., & Haviland, M. G. (2015). Iteration of partially specified target matrices: Applications in exploratory and Bayesian confirmatory factor analysis. *Multivariate Behavioral Research*, *50*, 149–161. doi:10.1080/00273171.2014.973990

Morris, D. E., Oakley, J. E., & Crowe, J. A. (2014). A web-based tool for eliciting probability distributions from experts. *Environmental Modelling & Software*, *52*, 1–4. doi:10.1016/j.envsoft.2013.10.010

Müller, U. K. (2012). Measuring prior sensitivity and prior informativeness in large Bayesian models. *Journal of Monetary Economics*, *59*, 581–597. doi:10.1016/j.jmoneco.2012.09.003

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335. doi:10.1037/a0026802

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

Muthén, L., & Muthén, B. (1998–2015). *M*plus *user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Natarajan, R., & McCulloch, C. E. (1998). Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference? *Journal of Computational and Graphical Statistics*, *7*, 267–277.

National Center for Education Statistics. (2001). *Early childhood longitudinal study: Kindergarten class of 1998–99: Base year public-use data files user's manual* (Tech. Rep. No. NCES 2001-029). Washington, DC: U.S. Government Printing Office.

Palardy, G., & Vermunt, J. K. (2010). Multilevel growth mixture models for classifying groups. *Journal of Educational and Behavioral Statistics*, *35*, 532–565. doi:10.3102/1076998610376895

Plummer, M. (2015). *JAGS Version 4.0 user manual*. Retrieved from http://www.uvm.edu/~bbeckage/Teaching/DataAnalysis/Manuals/manual.jags.pdf

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from https://www.r-project.org/

Ram, N., & Grimm, K. J. (2009). Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development*, *33*, 565–576. doi:10.1177/0165025409343765

Rietbergen, C., Klugkist, I., Janssen, K. J., Moons, K. G., & Hoijtink, H. J. (2011). Incorporation of historical data in the analysis of randomized therapeutic trials. *Contemporary Clinical Trials*, *32*, 848–855. doi:10.1016/j.cct.2011.06.002

Rindskopf, D. (2012). Next steps in Bayesian structural equation models: Comments on, variations of, and extensions to Muthén and Asparouhov (2012). *Psychological Methods*, *17*, 336–339. doi:10.1037/a0027130

Roos, M., Martins, T. G., Held, L., & Rue, H. (2015). Sensitivity analysis for Bayesian hierarchical models. *Bayesian Analysis*, *10*, 321–349. doi:10.1214/14-BA909

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*, *22*, 202–215. doi:10.1080/10705511.2014.937322

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelson (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Charlotte, NC: Information Age.

Tolvanen, A. (2007). *Latent growth mixture modeling: A simulation study* (Unpublished doctoral dissertation). University of Jyvaskyla, Jyvaskyla, Finland.

Tueller, S., & Lubke, G. (2010). Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling*, *17*, 165–192. doi:10.1080/10705511003659318

van de Schoot, R., Broere, J. J., Perryck, K., Zondervan-Zwijnenburg, M., & van Loey, N. (2015). Analyzing small data sets using Bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology*, *6*. doi:10.3402/ejpt.v6.25216

van de Schoot, R., Winter, S., Zondervan-Zwijnenburg, M., Ryan, O., & Depaoli, S. (in press). A systematic review of Bayesian papers in psychology: The last 25 years. *Psychological Methods*.

van der Linden, W. J. (2008). Using response times for item selection in adaptive testing. *Journal of Educational and Behavioral Statistics*, *33*, 5–20. doi:10.3102/1076998607302626

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

Zhang, Z., Hamagami, F., Wang, L., Grimm, K., & 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*, *20*, 47–78. doi:10.1080/10705511.2013.742382

Zhu, H., Ibrahim, J. G., & Tang, N. (2011). Bayesian influence analysis: A geometric approach. *Biometrika*, *98*, 307–323. doi:10.1093/biomet/asr009