Psychometrika (2026), **91**, 1599–1625
doi:10.1017/psy.2026.10119

PSYCHOMETRIC SOCIETY

LITERATURE REVIEW

Bayesian Estimation of Normal and Probit Psychometric Models

Edgar C. Merkle

University of Missouri, USA
Email: merklee@missouri.edu

(Received 12 June 2025; revised 30 January 2026; accepted 27 April 2026)

Abstract

This article synthesizes popular approaches to Bayesian estimation of psychometric models that rely on the multivariate normal distribution, including point estimation of posterior central tendency, Gibbs sampling, and Hamiltonian sampling. We place emphasis on the geometry of the posterior distribution, on building intuition by drawing connections to regression, and on providing background about how relevant matrix results are obtained. We also consider how the results can be extended to handle related psychometric models, including popular item response models. The goal is to provide researchers with ideas and tools for designing their own Bayesian estimation methods, which we hope will lead to continued developments and improvements.

Keywords: Bayesian estimation; Gibbs sampling; Hamiltonian Monte Carlo

1. Introduction

Bayesian methods have long received attention in psychometrics, and they have been increasingly used in recent times. Martin and McDonald (1975) were among early researchers who considered Bayesian methods, employing prior distributions to avoid Heywood cases in factor analysis. Other early approaches come from Bartholomew (1981), Lee (1981), and Press (1972). Lee, Song, and colleagues later made many additional developments involving Gibbs samplers (Lee, 2007; Lee & Song, 2002; Lee et al., 2007; Shi & Lee, 1997, 1998; Song & Lee, 2001, 2012) with other notable works, including Arminger and Muthén (1998), Scheines et al. (1999), Patz and Junker (1999a, 1999b), Edwards (2010), Fox (2010), and Fox and Glas (2001). Much of this research happened in the late 1990s and early 2000s, which was also the height of WinBUGS (Lunn et al., 2012). WinBUGS the first major piece of software to provide flexible and accessible MCMC methods to applied researchers, with specific psychometric applications coming from Curtis (2010), Lee et al. (2007), and Zhang et al. (2007), among others. WinBUGS also inspired psychometric model checking procedures from Glas and Meijer (2003) as well as Sinharay and colleagues (Sinharay, 2004, 2005; Sinharay et al., 2006).

Despite the above work, Bayesian psychometric modeling was a relatively niche research area through the early 2000s. The Bayesian skepticism held by many 20th century statisticians (e.g., Breiman, 1997) was also held by some psychometricians during that time. A notable change happened when Bayesian functionality was introduced to Mplus (Muthén & Muthén, 1998–2017) around 2010. The implementations and accompanying articles (Asparouhov & Muthén, 2010; Muthén & Asparouhov, 2012) helped researchers see the advantages that Bayesian methods hold for complex models, though

© The Author(s), 2026. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

some expressed appropriate levels of skepticism (e.g., MacCallum et al., 2012; Stromeyer et al., 2015). Today, Bayesian modeling appears to have a general level of acceptance among the psychometric community. Among other topics, researchers have been active in developing models for intensive time series (e.g., Asparouhov et al., 2018; Chen et al., 2023; Driver et al., 2017; Rast et al., 2022; Roman & Brandt, 2021; Woźniak, 2024), in studying prior distributions (e.g., Bainter, 2017; Kaplan et al., 2023; Liu et al., 2016; Lu et al., 2016; Padgett & Winter, 2025; Padgett et al., 2024; Pokropek et al., 2020; Van Erp & Browne, 2021; Van Erp et al., 2018; Van Zundert et al., 2022), and in developing model assessment tools (e.g., Cain & Zhang, 2018; Depaoli et al., 2021; Fox et al., 2017; Garnier-Villarreal & Jorgensen, 2020; Hoijtink & van de Schoot, 2018; Liu et al., 2022; Uanhoro, 2023; Verhagen & Fox, 2013; Winter & Depaoli, 2022, 2023; Zhang et al., 2019, 2022). Much of this work makes use of the fact that traditional Bayesian methods are often easily extended to handle incomplete datasets and to relax model assumptions (e.g., Cai et al., 2008; Jorgensen et al., 2018; Lee & Xia, 2008; Merkle, 2011; Tong & Zhang, 2020; Zhang et al., 2013, 2014).

Despite the popularity of Bayesian methods in psychometrics, the intuition underlying popular estimation methods is sometimes lacking, and psychometricians are sometimes reluctant to engage with the technical details underlying the estimation methods. Some of this reluctance is probably due to the relevant literature, which can present disparate results without providing context about where the results come from. This is understandable, as psychometrics researchers often apply established results from statistics to psychometric models. But it also leads to difficulties for researchers who are new to Bayesian psychometrics, because they must become familiar with a body of Bayesian statistics literature in addition to the psychometrics literature. The goal of this article is to synthesize popular procedures and results related to MCMC estimation of psychometric models, and to provide intuition about where the results come from and how they can be implemented. We discuss a unified regression-based framework that covers many popular psychometric models, including factor analysis, path models, growth models, structural equation models (SEMs), and some item response models. Near the end of the article, we also consider how this framework informs other psychometric models that are not covered.

In combining major results related to Gibbs samplers as well as Hamiltonian samplers, this article is intended to be a reference for psychometrics researchers on popular Bayesian estimation methods as well as an entrance to psychometric models for statistics researchers. We aim to provide the core results needed to implement popular MCMC methods while simultaneously providing enough intuition to understand popular MCMC methods. We assume some general background in statistical methods and linear algebra, with the article potentially being useful to different readers depending on their specific background. For example, readers who are familiar with SEM but not with Bayesian methods may elect to skip over some model definitions. Readers who are familiar with Bayesian methods but not SEM might spend more time on the model definitions, and they might find the comparisons between regression and SEM to be especially useful. Readers who are already familiar with Bayesian methods and SEM may skip directly to sections on Gibbs sampling or on Hamiltonian sampling, each of which provide a good deal of results for reference.

In the following pages, we first provide a conceptual overview of Bayesian estimation, emphasizing the geometry of the posterior distribution. This is helpful for building an intuition about how Bayesian estimation works. We then provide details about traditional Gibbs samplers for regression, which we build upon in describing Gibbs samplers for SEM. Next, we transition to a discussion of Hamiltonian Monte Carlo (HMC), focusing on key ideas and intuition that are important for implementing HMC. Next, we consider how our regression-based framework informs item response and other psychometric models.

2. Bayesian background

We begin with a brief overview of traditional Bayesian ideas. For a book-length, introductory treatment, see McElreath (2018) or Kaplan (2023). For book-length treatments focusing on Bayesian psychometric

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

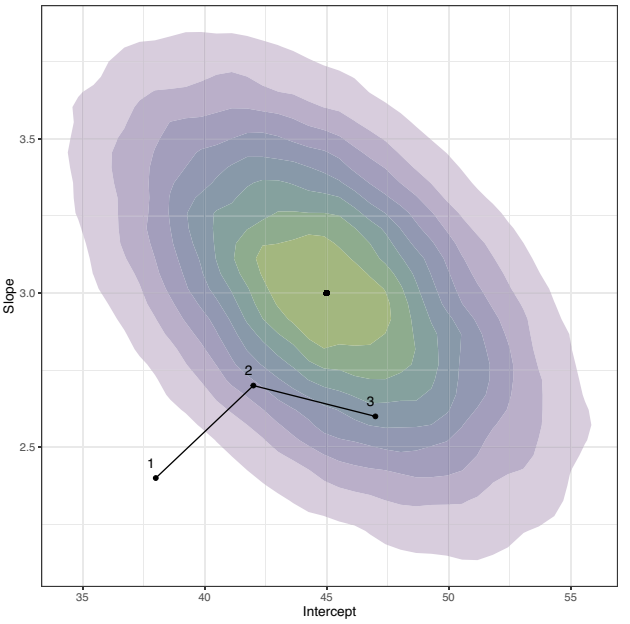


Figure 1. Plot of hypothetical posterior distribution. The black point in the center is the peak of the posterior distribution. The line is a sequence of three samples from the posterior distribution.

models, see Depaoli (2021) or Levy and Mislevy (2017). Related computational approaches and resources include Bürkner (2021), Fazio and Bürkner (2025), and Padgett (2022), while Levy and McNeish (2023) provide an accessible overview of different Bayesian philosophies.

2.1. General ideas

In Bayesian modeling, we generally wish to learn about the posterior distribution of model parameters. The posterior distribution tells us about likely values of model parameters, and it can be summarized via point estimates, interval estimates, or other functions of parameters. Mathematically, the posterior distribution is proportional to the product of the model likelihood and the prior distribution. The model likelihood is the same as the likelihood used in frequentist estimation (i.e., it is the "likelihood" in "maximum likelihood estimation"). The prior distribution represents our uncertainty about model parameters before we observe data.

It is important to realize that the posterior distribution is just another probability distribution. We can visualize it for simple models, which can help us build an intuition for more complex models. Figure 1 shows a hypothetical posterior distribution for a model with two parameters, say, an intercept and a slope. Visualization of the distribution requires three dimensions: one for each parameter, and one for the value of the posterior distribution at each combination of parameter values. Figure 1 is arranged as if we are looking down on the posterior distribution; that is, we are flying over a mountain and looking down at it. The x-axis represents the intercept, the y-axis represents the slope, and the colors represent the height of the posterior distribution. Yellow-green is around the peak of the mountain, representing the most likely combinations of intercepts and slopes. Light purple is the bottom of the mountain, representing unlikely combinations of model parameters. Similar ideas apply to models with more than two parameters, but we cannot visualize the associated distributions because we live in a three-dimensional world.

To learn about the posterior distribution in Figure 1, we might want to know the value of the intercept and slope at the peak of the mountain. We could use popular optimization methods like Newton–

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

Raphson to find these values, and we might call the values "maximum a posteriori" (MAP) estimates or "posterior modes." In Figure 1, the maximum occurs at the black point, which is an intercept of 45 and a slope of 3.

2.2. Posterior sampling

While the MAP estimates are useful summaries of the posterior distribution, we often want to learn about the entire distribution, which requires multiple summaries. To accomplish this, we can employ Markov chain Monte Carlo (MCMC) methods. Instead of directly trying to find the intercept and slope at the peak, we draw thousands of random samples from the distribution in Figure 1, where the probability of drawing each sample is related to the height of the posterior distribution. The mean of those random samples (one mean for each parameter) provides a summary of the posterior distribution's center, and the standard deviation of the samples provides a summary of the posterior distribution's variability. The random samples can lead us to many other summaries. For example, to estimate the probability that the intercept is greater than 50, we would simply calculate the proportion of intercept samples that are greater than 50.

While it is easy to say that we will draw random samples from the posterior distribution, the exact procedures for doing this are complicated. For some simple models, we can derive analytic results telling us that we should draw samples from a particular normal (or other) distribution. For complicated models, there are no analytic results telling us how to do this. But we still want to use MCMC to learn about the posterior distributions of those models.

In the latter cases, MCMC sampling generally proceeds as follows. We start with some initial values of model parameters. Based on those initial values, we jump to a new point of the posterior distribution in a particular manner and record the parameter values at that point. Now we repeat the process: based on our current location in the posterior distribution, we jump to another point and record the parameter values at that point. Then we keep going. These ideas are shown by the line in Figure 1. We start at the point labeled 1, which is an intercept of 38 and a slope of 2.4. This is in a low-density region of the posterior distribution (i.e., "off the mountain") so represents unlikely parameter values. We then jump to the point labeled 2, which is an intercept of 42 and slope of 2.7 and is closer to the peak, as compared to point 1. We then jump to the point labeled 3, and we continue in this way until we visit thousands of points. Our posterior samples are the intercept and slope values that we visited along the way.

This simple example of sampling helps provide intuition for a variety of issues surrounding MCMC. First, our initial value was not particularly good. In practice, we will run the sampler for an initial burn-in period so that any influence of the initial value wears off. We discard the locations that we visited during the burn-in period, so we can be sure that the remaining locations are good representations of the posterior distribution. Second, MCMC samplers differ in how quickly they can travel to different parts of the posterior distribution. For example, some samplers may be unable to jump directly from point 1 to point 2, requiring a large number of intermediate samples to make it that far. This gets at the idea of sampling efficiency, which includes metrics involving autocorrelation and effective sample size. Efficient samplers can quickly travel around the posterior distribution, whereas inefficient samplers may require many thousands of samples to fully explore the posterior distribution.

Finally, while we can occasionally update all parameters simultaneously, we more often sample individual parameters or subsets of parameters sequentially. Considering the points labeled 1 and 2 in Figure 1, a simultaneous update entails moving directly from (38, 2.4) to (42, 2.7) in one go. In a sequential update, we might first update the intercept, providing an intermediate move from (38, 2.4) to (42, 2.4). We would then update the slope value from 2.4 to 2.7, conditioned on the intercept equaling 42. This sequential updating is the main idea of Gibbs sampling, and it is especially helpful for models with many parameters. Gibbs sampling allows us to break up the full set of parameters into subsets that are easier to handle. For each parameter subset, we sample from a *conditional* posterior distribution that conditions on values of the remaining parameters (those not in the subset). This conditioning allows us to apply some key results to many types of models, an idea we further explore in the next section.

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

3. Gibbs sampling of regression models

For reasons that will become apparent later, Bayesian methods for SEM rely heavily on Bayesian methods for regression. This section's review of Bayesian regression results is intended to provide supports for the SEM results that come later. Our discussion generally follows that of Gelman et al. (2013).

3.1. Basics

A basic linear regression model of n responses can be written as

$$\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}), \tag{1}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of response variables, $\boldsymbol{X}$ is an $n \times m$ design matrix, and $\boldsymbol{\beta}$ is the $m \times 1$ vector of regression weights. The $n \times n$ model covariance matrix, $\sigma^2\boldsymbol{I}$, reflects the assumptions of independence and homogeneity: independence because all off-diagonal entries equal 0 (due to the identity matrix), and homogeneity because all diagonal entries equal σ^2.

We first consider a Bayesian version of the model that includes flat, noninformative priors on $\boldsymbol{\beta}$ and $\log(\sigma)$, that is,

$$p(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2}. \tag{2}$$

Under these priors, the conditional posterior distribution of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} \mid \sigma, \boldsymbol{y} \sim \mathrm{N}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, (\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2). \tag{3}$$

The mean of this distribution is especially intuitive, because it is the usual least squares estimate of $\boldsymbol{\beta}$. The posterior distribution of σ^2 can then be written in two ways, one that conditions on $\boldsymbol{\beta}$ and one that is marginal over $\boldsymbol{\beta}$:

$$\sigma^2 \mid \boldsymbol{y}, \boldsymbol{\beta} \sim \text{Inv-Gamma}\left(\frac{n}{2}, \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right) \tag{4}$$

$$\sigma^2 \mid \boldsymbol{y} \sim \text{Inv-Gamma}\left(\frac{n-k}{2}, \frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\right)'\left(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\right)\right). \tag{5}$$

The first line is most relevant to SEM, but the second line can lead to more efficient sampling. On the second line, the expression for the second parameter looks very involved, but it is just the usual sum of squared residuals in matrix form.

Now consider relaxing the independence and homogeneity assumptions of the covariance matrix. The model becomes

$$\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \tag{6}$$

where the structure of the $n \times n$ matrix $\boldsymbol{\Sigma}$ is unspecified. In this setting, the conditional posterior distribution of $\boldsymbol{\beta}$ now involves the full matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\beta} \mid \boldsymbol{\Sigma}, \boldsymbol{y} \sim \mathrm{N}\left((\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}, (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\right). \tag{7}$$

The conditional posterior distribution of $\boldsymbol{\Sigma}$, on the other hand, depends on its exact form (e.g., known up to a scalar, equal off-diagonal entries, and unrestricted). For some cases, this conditional posterior distribution will not have a form from which it is easy to sample.

3.2. Informative priors

The results of the previous section used noninformative priors, which are often used in psychometrics. This reflects researchers' reluctance to say that they know anything about model parameters before collecting data, perhaps because they view statistical models as yielding objective results that should not be unduly influenced. We think that researchers should use informative priors more often than they currently do. In considering reasonable prior distributions, it is often helpful to ask, "what are

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

ridiculous values of model parameters that would never occur in practice?" This question leads us to the idea of *weakly informative* prior distributions (e.g., Chung et al., 2015; Gelman et al., 2008), which place most density on plausible values of model parameters. There is not one "correct" weakly informative prior distribution, which may worry researchers who seek objective results. But based on past experience, so long as the prior approximates "weakly informative," its influence is less than other model choices, such as the number of latent variables, the nonzero entries of $\boldsymbol{\Lambda}$ and of $\boldsymbol{B}$, and assumptions about conditional independence between observed variables. We acknowledge that the definition of "weakly informative" can become tricky for some complex models where the data do not provide much information about certain parameters. See Levy and Mislevy (2017, Chapter 3) for further discussion of prior distributions, which emphasizes the relationship between exchangeability, conditional independence, and prior distributions.

The previous paragraph might lead us to ask "if the prior is not very influential, why not stick with some form of noninformative prior?" There are a few responses to this question. First, from a pragmatic point of view, mildly informative priors help prevent the sampler from straying into low-density regions of the posterior and getting stuck. This issue is especially relevant to HMC methods, which are typically better at exploring the full posterior distribution as compared to Gibbs samplers. Second, mildly informative priors can reduce noise in metrics that involve the model's log-likelihood, such as WAIC (Merkle et al., 2019). This can allow researchers to draw sharp inferences about their data using fewer posterior samples. Third, noninformative priors do not guarantee that the priors for all functions of parameters are noninformative, which can cause unexpected problems (Seaman et al., 2012). Finally, mildly informative priors help methodologists take responsibility for their analyses (e.g., Rouder et al., 2016). If we are to fit a model to data, we should know enough about the parameters beforehand to determine what values would be implausible.

Gelman et al. (2013) discuss how we can include informative priors (mildly informative or otherwise) for regression weights by augmenting the data with pseudo-observations representing the priors. Assume that the prior distribution is multivariate normal, that is,

$$\boldsymbol{\beta} \sim \mathrm{N}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_\beta),$$

which covers the situation where each individual regression weight is assigned a univariate normal prior distribution (in that case, $\boldsymbol{\Sigma}_\beta$ is diagonal). We can now define a new regression model

$$\boldsymbol{y}^* \sim \mathrm{N}(\boldsymbol{X}^*\boldsymbol{\beta}, \boldsymbol{\Sigma}^*),$$

where the vectors and matrices with $*$ augment the model with parts of the prior. Specifically, we have

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{\beta}_0 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{I} \end{pmatrix} \boldsymbol{\beta}, \begin{pmatrix} \boldsymbol{\Sigma} & 0 \\ 0 & \boldsymbol{\Sigma}_\beta \end{pmatrix} \right).$$

This formulation allows us to sample $\boldsymbol{\beta}$ using Equation (7), as if we had noninformative priors. We simply treat $\boldsymbol{y}^*$, $\boldsymbol{X}^*$, and $\boldsymbol{\Sigma}^*$ as $\boldsymbol{y}$, $\boldsymbol{X}$, and $\boldsymbol{\Sigma}$ in (7).

We now turn to the residual variance σ^2, when the model covariance matrix is specified as $\sigma^2\boldsymbol{I}$. The conjugate prior for σ^2 is the Inverse Gamma, where we denote the shape parameter as a_0 and the scale parameter as b_0. Under this prior, we have

$$\sigma^2 \mid \boldsymbol{y}, \boldsymbol{\beta} \sim \text{Inv-Gamma}\left(a_0 + \frac{n}{2}, b_0 + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right). \tag{8}$$

Now consider the case where the $n \times n$ residual covariance matrix $\boldsymbol{\Sigma}$ is unrestricted. The model now becomes overparameterized, but consideration of this situation becomes important for SEM. If $\boldsymbol{\Sigma}$ is assigned an inverse-Wishart prior with degrees of freedom ν_0 and scale matrix $\boldsymbol{S}_0$, the posterior distribution is

$$\boldsymbol{\Sigma} \mid \boldsymbol{y}, \boldsymbol{\beta} \sim \text{inv-Wishart}\left(\nu_0 + 1, \boldsymbol{S}_0 + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\right). \tag{9}$$

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

In this expression, we add 1 to the degrees of freedom because the $n \times 1$ response vector $\boldsymbol{y}$ is regarded as one observation that informs the $n \times n$ matrix $\boldsymbol{\Sigma}$.

4. Gibbs sampling of Bayesian SEM

Armed with our Gibbs sampler for heterogenous regression, we are now prepared to develop a Gibbs sampler for SEM. We first provide an explicit definition of the SEM, and we then see how the Gibbs sampler for SEM relies heavily on the Gibbs sampler for regression.

4.1. Model definition

We focus on the main *lavaan* (Rosseel, 2012) representation of an SEM, which is related to the LISREL "all-y" representation (e.g., Jöreskog & Sörbom, 1997). This representation differs from some others (notably, Asparouhov & Muthén, 2010) in how it handles exogenous observed variables. The other representations will separate exogenous observed variables from endogenous observed variables and assign them their own coefficient matrix (which is sometimes called $\boldsymbol{K}$ or $\boldsymbol{\Gamma}$). In contrast, *lavaan* typically includes exogenous observed variables alongside other observed variables. More detail about this distinction appears in the Appendix.

Let $\boldsymbol{y}_i$, $i = 1, \ldots, n$, be the p (continuous) observed variables associated with observation i. The traditional intent of the model is to describe the p observed variables by a smaller number of latent variables, similarly to how principal components compresses many observed variables into a single variable. But the SEM has a well-defined likelihood (as opposed to being an algorithm for data reduction), and it allows for regression relationships between latent variables. In traditional psychometric applications, the $\boldsymbol{y}_i$ are test scores and other observed variables about individuals, whereas the latent variables are unobserved traits such as "mathematics proficiency."

An SEM with $m < p$ latent variables may be represented by the equations

$$\boldsymbol{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \tag{10}$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i, \tag{11}$$

where $\boldsymbol{\eta}_i$ is an $m \times 1$ vector containing the latent variables; $\boldsymbol{\varepsilon}_i$ is a $p \times 1$ vector of measurement errors; and $\boldsymbol{\zeta}_i$ is an $m \times 1$ vector of structural errors. The vectors $\boldsymbol{\nu}$ and $\boldsymbol{\alpha}$ contain intercept parameters for the observed and latent variables, respectively; $\boldsymbol{\Lambda}$ is a matrix of factor loadings (regressions from latent to observed variables); and $\boldsymbol{B}$ encodes regressions between latent variables.

Equation (11) is not always fully understood. It is peculiar because $\boldsymbol{\eta}_i$ appears on both the left and right side. To make sense of the equation, it is important to realize that the diagonal of $\boldsymbol{B}$ is fixed to 0, which means that no individual entry of $\boldsymbol{\eta}_i$ is predicting itself. And nonzero, off-diagonal entries of $\boldsymbol{B}$ represent one entry of $\boldsymbol{\eta}_i$ being predicted by another entry of $\boldsymbol{\eta}_i$. The researcher specifies which off-diagonal entries are free and which off-diagonal entries are fixed to 0; the free entries represent paths (directed arrows) between latent variables in a path diagram. Similarly, the free entries of $\boldsymbol{\Lambda}$ in Equation (10) represent paths (directed arrows) from latent variables to observed variables.

The residuals $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\zeta}_i$ are assumed to be multivariate normal:

$$\boldsymbol{\varepsilon}_i \sim N_p(0, \boldsymbol{\Theta}) \tag{12}$$

$$\boldsymbol{\zeta}_i \sim N_m(0, \boldsymbol{\Psi}), \tag{13}$$

where the associated covariance matrices are often diagonal. Taken together, these assumptions imply multiple distributions and likelihoods for the same model. Conditioned on the latent variables $\boldsymbol{\eta}_i$, we have

$$\boldsymbol{y}_i \mid \boldsymbol{\eta}_i \sim \mathrm{N}(\boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i, \boldsymbol{\Theta}). \tag{14}$$

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

If we instead integrate the $\boldsymbol{\eta}_i$ out of the model, we have

$$\boldsymbol{y}_i \sim \mathrm{N}\left(\boldsymbol{\nu} + \boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\alpha}, \boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{B}')^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}\right), \tag{15}$$

where we require that $(\boldsymbol{I}-\boldsymbol{B})$ is invertible.

There can be confusion about the conditional and marginal SEM distributions above. The confusion often stems from the idea that our one model has two different likelihoods associated with it. This means we could use either likelihood for Bayesian estimation, and we will obtain the same posterior distributions because both likelihoods come from the same model. For Bayesian estimation, we most commonly sample the latent variables $\boldsymbol{\eta}_i$ as model parameters, which relies on the conditional distribution in Equation (14). This is not absolutely required, though, and use of the conditional distribution can lead to inefficient HMC sampling (Merkle et al., 2021). For frequentist estimation, we usually rely on the marginal distribution from Equation (15) because the latent variables can lead to estimation difficulties, including the incidental parameter problem (e.g., Lancaster, 2000).

4.2. Gibbs sampling of SEM

We now discuss Gibbs samplers of the above model, which are highly related to Gibbs samplers of regression models. A key idea is this: if the latent variables $\boldsymbol{\eta}_i$ were known, our SEM could be written as a regression model. And if our model can be written as a regression model, then the regression Gibbs sampler can be used to estimate our model.

To make this work, we need to know how to sample the $\boldsymbol{\eta}_i$. This step is highly related to prediction of latent variables in frequentist models, and we review it in the next section. We then describe the full Gibbs sampler, which makes use of some results from Asparouhov and Muthén (2010).

4.2.1. Sampling latent variables

To sample latent variables $\boldsymbol{\eta}_i$, we use standard properties of the multivariate normal distribution. We have (e.g., Jöreskog & Sörbom, 1997, p. 34) that the joint distribution of $(\boldsymbol{y}_i' \; \boldsymbol{\eta}_i')'$ is multivariate normal with mean

$$\begin{pmatrix} \boldsymbol{\nu} + \boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\alpha} \\ (\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\alpha} \end{pmatrix} \tag{16}$$

and covariance matrix

$$\begin{pmatrix} \boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{B}')^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} & \boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{B}')^{-1} \\ (\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{B}')^{-1}\boldsymbol{\Lambda}' & (\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{B}')^{-1} \end{pmatrix}. \tag{17}$$

From this joint distribution, we can obtain the multivariate normal conditional distribution of $\boldsymbol{\eta}_i$ given $\boldsymbol{y}_i$. Using the Woodbury identity and others to simplify the expressions (e.g., Asparouhov & Muthén, 2010), the mean of the conditional distribution is

$$\left(\boldsymbol{\Lambda}'\boldsymbol{\Theta}^{-1}\boldsymbol{\Lambda} + \left((\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{B}')^{-1}\right)^{-1}\right)^{-1}\left(\boldsymbol{\Lambda}'\boldsymbol{\Theta}^{-1}\left(\boldsymbol{y}_i - \boldsymbol{\nu}\right) + (\boldsymbol{I}-\boldsymbol{B}')\boldsymbol{\Psi}^{-1}\boldsymbol{\alpha}\right) \tag{18}$$

and the covariance matrix is

$$\left(\boldsymbol{\Lambda}'\boldsymbol{\Theta}^{-1}\boldsymbol{\Lambda} + \left((\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{B}')^{-1}\right)^{-1}\right)^{-1}. \tag{19}$$

This distribution is one piece of the Gibbs sampler. Conditioned on all model parameters except the $\boldsymbol{\eta}_i$, we take draws from the above distribution. If we are using MCMC with the marginal likelihood (Equation (15); see Merkle et al., 2021), we could use the above distribution to sample the $\boldsymbol{\eta}_i$ as a post-estimation step, after sampling the other model parameters. As discussed later, this idea can be especially useful in HMC.

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

4.2.2. Sampling location parameters

Once the $\boldsymbol{\eta}_i$ have been sampled, we can use them to set up our regression model. The idea is to express Equations (10) and (11) in such a way that the free parameters in $\boldsymbol{\nu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\alpha}$, and $\boldsymbol{B}$ are the regression weights.

Let $\boldsymbol{\xi}$ be a vector containing these free parameters. Then we can write Equations (10) and (11) as

$$\begin{pmatrix} \boldsymbol{y}_i \\ \boldsymbol{\eta}_i \end{pmatrix} \sim \mathrm{N}(\boldsymbol{H}_i\boldsymbol{\xi}, \boldsymbol{V}),$$

where our design matrix $\boldsymbol{H}_i$ includes 1s for intercept parameters and entries of $\boldsymbol{\eta}_i$ that get multiplied by loadings and regression weights. So, for example, a CFA with three observed variables and one latent variable could be arranged as

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ \eta_{i1} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 1 & \eta_{i1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \eta_{i1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \eta_{i1} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \nu_1 \\ \lambda_1 \\ \nu_2 \\ \lambda_2 \\ \nu_3 \\ \lambda_3 \end{pmatrix}, \boldsymbol{V} \right),$$

where the final row consists of all zeros because the mean of the $\boldsymbol{\eta}_i$ is fixed to zero (which is typically the case, but not always). For structural models that include nonzero terms in $\boldsymbol{\alpha}$ and/or in $\boldsymbol{B}$, this final row may contain other latent variables that get paired with regression weights in $\boldsymbol{B}$. Our multivariate normal covariance matrix, $\boldsymbol{V}$, is then given as

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{\Theta} & 0 \\ 0 & \boldsymbol{\Psi} \end{pmatrix}. \tag{20}$$

With our model now set up as a regression, we can begin to see how to sample the model intercepts, loadings, and regression weights. But the model above is only for a single case i, and we want to use all cases in the data. To handle all cases, we keep adding to our regression model. We obtain

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{\eta}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{y}_n \\ \boldsymbol{\eta}_n \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{H}_1 \\ \boldsymbol{H}_2 \\ \vdots \\ \boldsymbol{H}_n \end{pmatrix} \begin{pmatrix} \nu_1 \\ \lambda_1 \\ \nu_2 \\ \lambda_2 \\ \nu_3 \\ \lambda_3 \end{pmatrix}, \boldsymbol{I} \otimes \boldsymbol{V} \right),$$

or more concisely,

$$\boldsymbol{z} \sim \mathrm{N}(\boldsymbol{H}\boldsymbol{\xi}, \boldsymbol{I} \otimes \boldsymbol{V}), \tag{21}$$

where $\boldsymbol{z}$ includes the $\boldsymbol{y}_i$ and $\boldsymbol{\eta}_i$ in sequence, and the identity matrix $\boldsymbol{I}$ is of dimension n. The Kronecker product $\boldsymbol{I} \otimes \boldsymbol{V}$ leads to a block diagonal matrix, where there is a $\boldsymbol{V}$ for each case i and zeros elsewhere (with the zeros representing independence of cases).

Now we are finally prepared to use the regression results to sample the free parameters in the $\boldsymbol{\xi}$ vector. Applying the result from (7) and assuming flat priors on $\boldsymbol{\xi}$ (of the same form as in (2)), we have

$$\boldsymbol{\xi} \mid \boldsymbol{H}, \boldsymbol{V} \sim \mathrm{N}\left((\boldsymbol{H}'(\boldsymbol{I} \otimes \boldsymbol{V})^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}'(\boldsymbol{I} \otimes \boldsymbol{V})^{-1}\boldsymbol{z}, (\boldsymbol{H}'(\boldsymbol{I} \otimes \boldsymbol{V})^{-1}\boldsymbol{H})^{-1} \right). \tag{22}$$

And if we wish to include normal priors $\boldsymbol{\xi} \sim \mathrm{N}(\boldsymbol{\xi}_0, \boldsymbol{\Sigma}_\xi)$, we can use the same trick that was used in the regression case: append $\boldsymbol{\xi}_0$ to the end of $\boldsymbol{z}$, add rows to $\boldsymbol{H}$ that select specific model parameters, and add a block to the end of the covariance matrix that holds $\boldsymbol{\Sigma}_\xi$.

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

The distribution in (22) (under flat priors) involves inversion of the matrix $\boldsymbol{I} \otimes \boldsymbol{V}$, which is of dimension $n(p+m) \times n(p+m)$. This will be very large in most applications and will lead to slow MCMC sampling. But the block-diagonal structure of this matrix allows us to handle each case i separately by inverting the smaller $(p+m) \times (p+m)$ matrix $\boldsymbol{V}$. These arguments lead us to simplifications that are related to the results of, for example, Asparouhov and Muthén (2010), which include normal prior distributions:

$$\boldsymbol{\xi} \mid \boldsymbol{H}, \boldsymbol{V} \sim \mathrm{N}(\boldsymbol{D}\boldsymbol{d}, \boldsymbol{D}) \tag{23}$$

$$\boldsymbol{D} = \left(\boldsymbol{\Sigma}_{\xi}^{-1} + \sum_{i=1}^{n} \boldsymbol{H}_i' \boldsymbol{V}^{-1} \boldsymbol{H}_i \right)^{-1} \tag{24}$$

$$\boldsymbol{d} = \boldsymbol{\Sigma}_{\xi}^{-1} \boldsymbol{\xi}_0 + \sum_{i=1}^{n} \boldsymbol{H}_i' \boldsymbol{V}^{-1} \boldsymbol{z}_i. \tag{25}$$

4.2.3. Sampling (co)variance parameters

With the latent variables, intercepts, loadings, and regression weights sampled, the only parameters that remain are the variances and covariances in $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$. For now, we assume that both these matrices consist of unrestricted blocks of free parameters. We consider relaxations of these restrictions in the next section.

In sampling the covariance parameters, we continue to follow regression results. The only difference is that we now have two sets of residuals: residuals for the $\boldsymbol{\eta}_i$ and residuals for the $\boldsymbol{y}_i$. We use these residuals to separately sample blocks of $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$, using the results from Equations (8) and (9). Specifically, for a standalone variance ψ_{jj} in $\boldsymbol{\Psi}$ with an Inverse Gamma(a_j, b_j) prior, we have

$$\psi_{jj} \mid \boldsymbol{\xi}, \boldsymbol{\eta} \sim \text{Inv-Gamma}\left(a_j + \frac{n}{2}, b_j + \frac{1}{2} \sum_{i=1}^{n} (\eta_{ij} - (\alpha_j + \sum_{k=1}^{m} B_{jk} \eta_{ik}))^2 \right). \tag{26}$$

For an unrestricted block $\boldsymbol{\Psi}_k$ with an Inverse Wishart(ν_k, $\boldsymbol{S}_k$) prior, we have

$$\boldsymbol{\Psi}_k \mid \boldsymbol{\xi}, \boldsymbol{\eta} \sim \text{Inv-Wishart}(\nu_k + n, \boldsymbol{S}_k + \boldsymbol{E}_k), \tag{27}$$

where $\boldsymbol{E}_k$ is the block of a matrix $\boldsymbol{E}$ that corresponds to $\boldsymbol{\Psi}_k$. The full $\boldsymbol{E}$ matrix is defined as

$$\boldsymbol{E} = \sum_{i=1}^{n} (\boldsymbol{\eta}_i - (\boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta}_i))(\boldsymbol{\eta}_i - (\boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta}_i))'.$$

These same ideas apply to the blocks of $\boldsymbol{\Theta}$. The only difference is that the definition of $\boldsymbol{E}$ now involves residuals of the $\boldsymbol{y}_i$:

$$\boldsymbol{E} = \sum_{i=1}^{n} (\boldsymbol{y}_i - (\boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i))(\boldsymbol{y}_i - (\boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i))'.$$

The above posteriors on (co)variance parameters condition on the intercepts, loadings, and other parameters in $\boldsymbol{\xi}$. It may be worthwhile to consider marginalizing over $\boldsymbol{\xi}$, similar to how we marginalized over the regression weights $\boldsymbol{\beta}$ in (5). This may lead to more efficient Gibbs sampling.

4.3. Summary

We have presented conditional posterior distributions for all parameters in a traditional SEM. There are a variety of additional situations we could consider, including missing data, fixed exogenous variables, and equality constraints. We outline these situations in the Appendix, which involve some modifications to the basic procedure described here. The results described above can be viewed as the core of other Gibbs sampling procedures.

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

5. Metropolis–Hastings sampling

The Gibbs sampler gets us far and is sufficient for many applications, but it cannot immediately handle all models in the SEM family. An important example involves residual correlation parameters. In this case, our $\boldsymbol{\Theta}$ matrix has some off-diagonal entries fixed to 0 and others free, not necessarily in a block diagonal structure. The full matrix is required to be positive definite, which influences the possible values that the free entries can take. When this matrix is not block diagonal, it is difficult to construct a coherent prior distribution (e.g., Merkle et al., 2023), which makes it difficult to sample the free entries. More generally, the Gibbs sampler requires us to use the specific prior distributions that we considered: Normal for intercepts and loadings, Inverse Gamma for variances, and Inverse Wishart for covariance matrices.

For situations where researchers want to use alternative prior distributions (such as some regularizing priors), we do not have nice conditional posterior distributions for our Gibbs sampler. In these situations, it is common to use the Metropolis–Hastings algorithm to sample the problematic parameters. This algorithm does not require us to know the form of the conditional posterior distribution. Instead, based on our current posterior sample, we *propose* a new sample from an easy-to-sample posterior distribution (such as the Normal). We then accept this proposal with a specific probability. After a burn-in period, and under "mild" regularity conditions, the accepted samples will follow the posterior distribution of interest. We omit the technical details here; further detail can be found in Patz and Junker (1999a, 1999b), who made extensive use of Metropolis–Hastings sampling in their IRT applications.

6. Hamiltonian sampling of Bayesian SEM

As illustrated by Mplus (Muthén & Muthén, 1998–2017), Blimp (Keller & Enders, 2023a), and other software, the Gibbs and Metropolis–Hastings samplers are sufficient for many applications of Bayesian SEM. But in some applications, the samplers yield a large degree of autocorrelation, which requires us to draw a large number of posterior samples (often in the tens of thousands or more). Additionally, because these samplers involve the $\boldsymbol{\eta}_i$, the number of model parameters increases with N. Both these issues can lead us to require large amounts of memory and/or time, which may be especially problematic when fitting a model to a large dataset. Hamiltonian sampling is an alternative method that can lead to more efficient sampling and scalability to large datasets. A key difference from Gibbs and Metropolis–Hastings is that Hamiltonian sampling uses the model gradient to efficiently explore the posterior distribution.

To gain intuition about HMC, it is helpful to think back to our simple posterior distribution from Figure 1. We said before that this posterior distribution was a mountain, and we now turn the mountain upside down (which is related to taking the negative of the model log-likelihood). Our posterior distribution becomes a bowl, with the mode being located at the bottom. HMC is similar to rolling a marble around the bowl, and sampling parameter values along the path that the marble travels. Gravity will tend to pull us toward the bottom of the bowl, so that plausible parameter values are sampled more often than implausible parameter values.

To carry out this sort of sampling, we need to formalize our intuition about the marble traveling through the posterior distribution. We describe key points below, with Thomas and Tu (2021) providing a related introduction. This discussion should prepare readers for expert texts on HMC with many more details, including Betancourt (2018) and Neal (2011).

6.1. The Hamiltonian

Continuing with Figure 1 and the idea of a marble rolling around a bowl, we can conceptualize the marble as having a momentum with respect to the x-axis and a separate momentum with respect to the y-axis. The magnitude of each momentum is related to the speed of the marble along the corresponding axis, and the sign of each momentum tells us the direction. Taken together, the momentum variables determine the overall direction in which the marble travels.

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

For example, a momentum of $(10,0)$ means that we are moving from left to right along the x-axis (the Intercept axis), but we are not moving at all along the y-axis (the Slope axis). This corresponds to horizontal motion in Figure 1, where the Intercept values increase and the Slope remains fixed. Similarly, a momentum of $(-10,5)$ means that the Intercept values are decreasing and that the Slope values are increasing at a slower rate. If you are familiar with Etch-A-Sketch toys, the idea is similar: the speed at which you turn the left knob determines how you travel along the x-axis, and the speed at which you turn the right knob determines how you travel along the y-axis.

For HMC, we now assign each model parameter a corresponding momentum variable, in the manner outlined in the previous paragraph. Call this momentum vector $\boldsymbol{m}$ (not to be confused with the scalar m that signifies the number of latent variables in an SEM). If we have k total model parameters to sample, then $\boldsymbol{m}$ is a vector of length k. These variables signify how our sampler (e.g., how our marble) moves through the posterior distribution, which in turn influences the parameter values that we draw from the posterior distribution.

Continuing with our marble analogy, the *Hamiltonian* describes the total amount of energy in the marble. The total energy consists of potential energy, which involves the location of our marble in the posterior distribution (and is related to the height of the bowl), as well as kinetic energy, which is related to the momentum of our marble. This formalizes our intuition that the marble travels in a particular direction at a certain speed, with the shape of our bowl (i.e., the curvature of the bowl) influencing the marble's future directions and speeds.

It is convenient to assume that the marble's starting momentum is drawn from a multivariate normal:

$$\boldsymbol{m} \sim \mathrm{N}(0,\boldsymbol{M}),$$

where $\boldsymbol{M}$ is a tuning parameter that is often diagonal (further details on this later). Our Hamiltonian is then

$$H(\boldsymbol{\theta},\boldsymbol{m}) = -\log p(\boldsymbol{\theta} \mid \boldsymbol{Y}) + \frac{1}{2}\boldsymbol{m}'\boldsymbol{M}^{-1}\boldsymbol{m}, \tag{28}$$

where $\boldsymbol{\theta}$ is the model parameter vector and $\boldsymbol{Y}$ is the observed data. The term $p(\boldsymbol{\theta} \mid \boldsymbol{Y})$ is the model posterior distribution, which we often take to be the product of the model likelihood and prior, ignoring the normalizing constant. The Hamiltonian represents the total amount of kinetic energy and potential energy in the system, a total that is conserved as the marble rolls around the bowl.

With this setup, *Hamilton's equations* describe how $\boldsymbol{\theta}$ and $\boldsymbol{m}$ jointly change over time. Letting t be time, they are given as

$$\frac{\partial \boldsymbol{m}}{\partial t} = -\frac{\partial H(\boldsymbol{\theta},\boldsymbol{m})}{\partial \boldsymbol{\theta}} = \frac{\partial \log p(\boldsymbol{\theta} \mid \boldsymbol{Y})}{\partial \boldsymbol{\theta}} \tag{29}$$

$$\frac{\partial \boldsymbol{\theta}}{\partial t} = \frac{\partial H(\boldsymbol{\theta},\boldsymbol{m})}{\partial \boldsymbol{m}} = \boldsymbol{M}^{-1}\boldsymbol{m}, \tag{30}$$

that is, momentum changes based on the "incline" at our current location, and our location in the bowl simultaneously changes based on the velocity $\boldsymbol{M}^{-1}\boldsymbol{m}$. We can recognize Equation (29) as the model gradient: the log-posterior is the sum of the model's log-likelihood and log-priors, so (29) is the sum of the frequentist model gradient and first derivatives of log-priors. Solutions to these equations involve finding sequences of $\boldsymbol{\theta}$ and of $\boldsymbol{m}$ such that both equations are satisfied. We consider this difficult task in the next section.

6.2. Solving Hamilton's equations

Solutions to Equations (29) and (30) describe a path by which our marble travels through the posterior distribution. Bringing this back to the problem of model estimation via MCMC, we start our marble somewhere in the bowl, follows its path through the bowl for "awhile," then record the location of the

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

```r
leapfrog <- function(theta, momentum, grfun, data, eps, Minv) {
  newmom <- momentum + (eps/2) * grfun(theta, data)
  newtheta <- theta + eps * Minv %*% newmom
  newmom <- newmom + (eps/2) * grfun(newtheta, data)

  list(theta = newtheta, momentum = newmom)
}
```

Figure 2. R code for a leapfrog step. The grfun argument is a function computing the gradient of the log-posterior distribution, the eps argument is step size ε, and the Minv argument is $\boldsymbol{M}^{-1}$.

marble when "awhile" is over. That final location is our new posterior sample. Once we record the new posterior sample, we then repeat the process to draw another posterior sample.

In general, we cannot solve Hamilton's equations analytically. We instead use a *leapfrog integrator* to approximate the solution, which Betancourt (2018) calls "deceptively simple": the algorithm itself is simple, but the reasons why it works are not so simple. We now write our momentum vector and parameter vector as $\boldsymbol{m}(t)$ and $\boldsymbol{\theta}(t)$, respectively, signifying that these vectors change over time. We can then describe one leapfrog step as

$$\begin{aligned}
\boldsymbol{m}(t+\varepsilon/2) &= \boldsymbol{m}(t) + (\varepsilon/2)\frac{\partial \log p(\boldsymbol{\theta}(t) \mid \boldsymbol{Y})}{\partial \boldsymbol{\theta}(t)} \\
\boldsymbol{\theta}(t+\varepsilon) &= \boldsymbol{\theta}(t) + \varepsilon \boldsymbol{M}^{-1}\boldsymbol{m}(t+\varepsilon/2) \\
\boldsymbol{m}(t+\varepsilon) &= \boldsymbol{m}(t+\varepsilon/2) + (\varepsilon/2)\frac{\partial \log p(\boldsymbol{\theta}(t+\varepsilon) \mid \boldsymbol{Y})}{\partial \boldsymbol{\theta}(t+\varepsilon)},
\end{aligned}$$

where ε is the step size. In words, this leapfrog algorithm describes how we update the momentum and parameter vectors from time t to time $(t+\varepsilon)$ while (approximately) fulfilling Equations (29) and (30).

The above equations bear similarity to traditional optimization algorithms. The middle line, which updates the model parameters, looks similar to a parameter update step in other optimization algorithms. The surrounding lines, which update the momentum vector, look similar to a gradient descent update with step size $\varepsilon/2$. Of course, the terms involved in these updates are different from what we would use in other optimization algorithms. Figure 2 displays an R function for carrying out a leapfrog step.

6.3. Running the leapfrog integrator

The simple algorithm from Figure 2 gets us close to a full MCMC method, as outlined in the algorithm below.

1. Set tuning parameters $\boldsymbol{M}$ and ε, set initial values for $\boldsymbol{\theta}$, set the desired number of posterior samples (i.e., of iterations), and set the number of leapfrog steps per iteration n_ε.
2. For each iteration:
   (a) Draw $\boldsymbol{m}(0) \sim \mathrm{N}(0, \boldsymbol{M})$.
   (b) Take n_ε leapfrog steps, updating $\boldsymbol{m}$ and $\boldsymbol{\theta}$ at each step. This leads to $\boldsymbol{m}(n_\varepsilon \times \varepsilon)$ and $\boldsymbol{\theta}(n_\varepsilon \times \varepsilon)$.
   (c) Accept $\boldsymbol{\theta}(n_\varepsilon \times \varepsilon)$ with a Metropolis–Hastings probability (see Neal, 2011, Equation (3.6)).
   (d) Repeat, using the current value of $\boldsymbol{\theta}$ as the starting value of the next iteration.

The Metropolis–Hastings probability in (2c) is needed because the leapfrog integrator *approximates* a solution to Hamilton's equations.

Figure 3 shows a single iteration of this process in action, with cherrypicked data, cherrypicked starting values of $\boldsymbol{m}$ and $\boldsymbol{\theta}$, and a cherrypicked value of ε. The top left panel shows the first 10 leapfrog

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

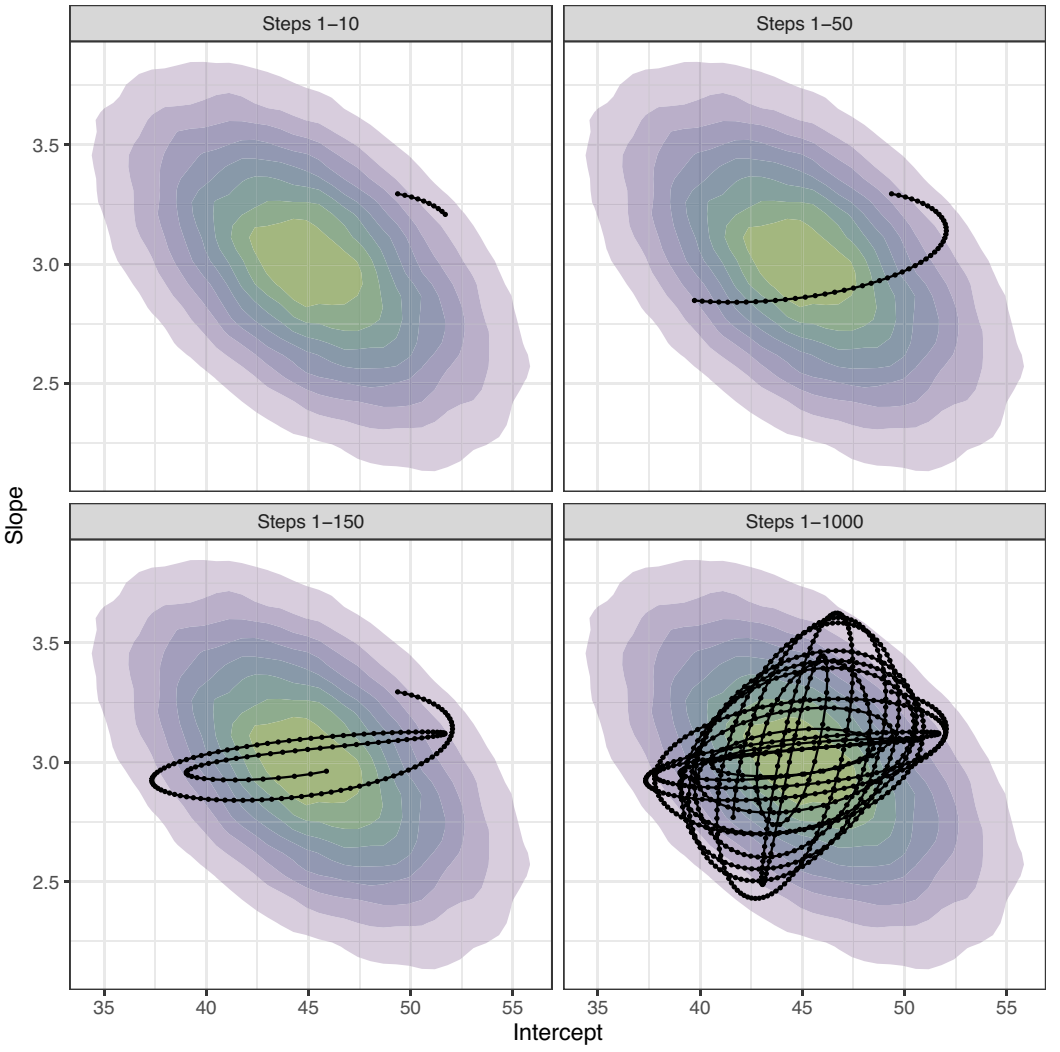


Figure 3. Example of a leapfrog algorithm. Each panel shows a different number of steps in the same path.

steps, with increasing numbers of steps shown as we move toward the bottom right panel. We see that, by Step 50 (top right panel), we have moved most of the way across the posterior distribution. At Step 150 (bottom left), we have circled around the mode multiple times. At Step 1,000 (bottom right), we retrace our steps repeatedly. This figure helps show why HMC requires smooth posterior distributions and excludes discrete parameters: roughly, the sampler cannot travel up a wall or off a cliff. So it is desirable to transform all model parameters to have unbounded support. For example, when we have a variance parameter that is bounded from below by 0, we would want to sample the unbounded log of the variance to avoid a wall at 0. This requires some bookkeeping of the reparameterized log-likelihood, gradient, and prior, which Stan mostly handles automatically and behind the scenes.

Returning to Figure 3, we see that the bottom right panel is wasteful from a computational standpoint: we are repeatedly visiting the same parts of the posterior distribution, when it would be better to stop at an earlier point and begin a new iteration (i.e., send the marble on a new trajectory). The "No U-Turn"

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

sampler (NUTS) of Hoffman and Gelman (2014) helps us automatically determine a good stopping point. The idea is to stop the sampler as soon as it turns back on itself. In Figure 3, this would occur a few steps after Step 50, which can be seen by comparing the top right panel to the bottom left panel. Mathematically, this happens when

$$(\boldsymbol{\theta}(t) - \boldsymbol{\theta}(0))' \boldsymbol{m}(t) < 0, \tag{31}$$

where $\boldsymbol{\theta}(t)$ and $\boldsymbol{m}(t)$ are the parameter values and momenta at step t, and $\boldsymbol{\theta}(0)$ is the starting parameter value at the current iteration.

While the above provides some general intuition about HMC and NUTS, the actual NUTS algorithm is more complex than this. The NUTS algorithm moves forward and backward from $t = 0$, comparing parameter and momentum values at the points it visits, determining when to stop via a rule that is related to (31), and accepting a new value of $\boldsymbol{\theta}$ via Metropolis–Hastings-like probability. It also adaptively determines ε during the warmup period. Though we leave out full details here, our intuition is strong enough that we can start to make sense of the warning messages that may appear in Stan and related software:

- *Divergent transition*: After taking some number of leapfrog steps, the value of $H()$ has changed significantly. This happens because the leapfrog integrator is a numerical approximation, and it is problematic because $H()$ is supposed to be conserved. It occurs when the geometry of the posterior distribution has varying amounts of curvature, and/or when the step size ε is excessively large.
- *Maximum treedepth*: The leapfrog sampler took the maximum number of steps (which is set in advance) without encountering a U-turn. This can happen when the step size is excessively small.

These warnings serve as diagnostics that something is wrong with the model specification and/or with the sampling process. The warnings can sometimes be addressed by reparameterizing the model, by marginalizing parameters out of the likelihood, or by changing some sampler settings. Based on past experience, though, it is more likely that the warnings occur due to user error in model specification. This is a distinct advantage of HMC over Gibbs and Metropolis–Hastings samplers: a Gibbs sampler will often continue to run even when the specified model is problematic, while the Hamiltonian sampler will fail "loudly" (Betancourt, 2020).

6.4. Running and tuning HMC

At this point, it is helpful to consider the things that are needed to run an HMC algorithm. We divide these into pieces that the user specifies, and other pieces required by the software.

The user will specify the data $\boldsymbol{Y}$, the model likelihood $p(\boldsymbol{Y} \mid \boldsymbol{\theta})$, and the prior distribution $p(\boldsymbol{\theta})$. The product of the latter two functions gives us the posterior (up to the normalizing constant), $p(\boldsymbol{\theta} \mid \boldsymbol{Y})$. We additionally need the ability to evaluate the posterior gradient, which the user does not supply. For specific families of models like the SEM family from Equation (15), the gradient can be specified analytically. For the vast world of models that Stan can handle, analytic derivatives are not supplied. Stan and other software use automatic differentiation, which is fast and which can be flexibly applied to a large variety of models. Cudeck (2005) provides a nice overview of automatic differentiation for psychometric models that was ahead of its time, considering today's widespread use of automatic derivatives in Bayesian models, neural networks, and related methods.

Additional tuning parameters required by HMC involve the "mass matrix" $\boldsymbol{M}$ that influences momentum values, the step size ε, and the number of leapfrog steps n_ε. Hoffman and Gelman (2014) and related work develop methods that can adaptively set these values based on one's model and data. For example, the "No U-Turn" idea eliminates the need to set n_ε; we instead take additional steps until

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

the sampler turns on itself (or until we hit the maximum number of steps). The step size ε represents the distance between the points in Figure 3. From the figure, we can see that ε needs to be in a "Goldilocks" zone for optimal sampling: if it is too small, the sampler will not be able to explore the full posterior, and if it is too large, the sampler will stray from the high-density part of the posterior. Finally, the mass matrix $\boldsymbol{M}$ determines the distribution of momentum values $\boldsymbol{m}$ and is also involved in the leapfrog update of $\boldsymbol{\theta}$. We seek a value such that $\boldsymbol{M}^{-1}$ approximates the posterior covariance matrix, with $\boldsymbol{M}$ often being diagonal so that its inverse is easy to compute.

During the Stan "warmup" period, the posterior variances are approximated so that a suitable value of $\boldsymbol{M}$ is chosen. Further, the step size ε is tuned to achieve an optimal acceptance rate of $\boldsymbol{\theta}$ proposals (Stan Development Team, 2024). This warmup period aids Stan in flexible application to a wide variety of models. Newer variations of NUTS seek to improve the tuning in various manners. For example, Bou-Rabee et al. (2025) recently described a WALNUTS sampler that dynamically sets ε based on the local geometry of the posterior distribution. This appears very promising for models with complicated geometries.

6.5. Application to psychometric models

So far, the HMC discussion has not involved much detail about psychometric models; we have only mentioned a generic posterior distribution and model gradient. Considering psychometric models specifically, the conditional/marginal distinction in Equations (14) and (15) becomes important for estimation speed and efficiency. Under the conditional specification, the latent variables $\boldsymbol{\eta}_i$ all become model parameters, leading to a high-dimensional space that HMC must navigate. This also leads to more overhead in gradient calculations, because automatic derivatives are applied to each $\boldsymbol{\eta}_i$ separately. In essence, the conditional approach entails adding hundreds or thousands of extra model parameters to the model.

Merkle et al. (2021) compared conditional and marginal estimation approaches for traditional SEMs in Stan. They showed that, at relatively small values of p (up to 11), sampling speed and efficiency are considerably improved when we use the marginal likelihood from (15). Information criteria like WAIC (Watanabe, 2010) and PSIS-LOO (Vehtari et al., 2017) are also more stable and theoretically appropriate under the marginal approach (Merkle et al., 2019). And if users wish to sample the latent variables under this approach, they can rely on the analytic result from (18) and (19) to draw the latent variables after model estimation, conditioned on other parameters.

A limitation of the marginal approach is that analytic results for the latent variables do not always exist for novel psychometric models. Relatedly, sampling the $\boldsymbol{\eta}_i$ can be advantageous because it often leads to conditional independence between observed variables. We sometimes need this conditional independence to make progress with model estimation. For Stan and related software, we may encounter tradeoffs between fast and efficient estimation of traditional models under a marginal approach, versus slow and inefficient estimation of novel models under a conditional approach.

6.6. HMC summary

We now have some intuition about how HMC works theoretically and about how the NUTS algorithm works practically. Our two-dimensional example in Figure 3 is helpful for intuition, and Feng (2024) provides an interactive web application that is recommended for building further intuition. We can imagine how the HMC procedure becomes more complicated for highly parameterized models: each parameter becomes a new dimension of the posterior distribution, potentially making the high-density part of the distribution more difficult to locate and to explore. In the next section, we consider extensions of our MCMC samplers to item response models and to multilevel SEM.

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

7. Item response models

The concepts described thus far provide a foundation for developing MCMC methods for item response models. In fact, from a Bayesian standpoint, many distinctions between item response models and factor analysis models disappear. This is especially true of the graded response model with probit link function, which can be conceptualized as a traditional factor analysis model placed on latent responses that are chopped to yield ordinal data. Because the graded response model contains the Rasch model and two-parameter model as special cases, the results from previous sections of this paper are applicable to many traditional IRT models. We provide further detail below. We then consider the estimation of other traditional IRT models, including models with logit link functions, the three- and four-parameter models, and the generalized partial credit model (GPCM). These all require further machinery, as compared to the graded response model with probit link function.

7.1. Graded response model

When using MCMC for estimation, the graded response model with probit link function and ordinal factor analysis are the same model (e.g., Takane & de Leeuw, 1987). Similar to frequentist modeling, we can assume continuous latent responses that give rise to the ordinal observed variables. In particular, for an ordinal variable y with K categories, we assume

$$\begin{aligned} y = 1 \text{ if } y^* &< \tau_1 \\ y = 2 \text{ if } \tau_1 < y^* &< \tau_2 \\ y = 3 \text{ if } \tau_2 < y^* &< \tau_3 \\ &\vdots \\ y = K \text{ if } y^* &> \tau_{K-1}, \end{aligned}$$

where $\tau_1 < \tau_2 < \tau_3 < \cdots < \tau_{K-1}$. Our SEM from Equations (10) and (11) is then placed on the $\boldsymbol{y}^*$ instead of the $\boldsymbol{y}$.

This setting provides us at least two options for MCMC estimation. Under the first option, we sample the $\boldsymbol{\eta}_i$ to achieve conditional independence between observed variables. This allows us to evaluate the (multinomial) likelihood of each observed variable separately, which can be done with the help of the univariate normal cumulative distribution function. Under the second option, we follow the data augmentation approach of Chib and Greenberg (1998). This involves sampling the $\boldsymbol{y}^*$ variates from truncated normal distributions, then pretending that the $\boldsymbol{y}^*$ are observed, continuous data. Both approaches can be handled by Gibbs samplers as well as Hamiltonian samplers.

Regardless of the specific approach chosen, we must sample extra parameters, as compared to the setting of continuous observed variables. Sampling the $\boldsymbol{\eta}_i$ involves Nm additional parameters, and sampling the $\boldsymbol{y}_i^*$ involves Np additional parameters. It is worth noting that we cannot immediately use a model likelihood that involves response patterns, which is often done in frequentist applications of IRT (e.g., Baker & Kim, 2004). This is because, to use response patterns, we require that individuals are identically distributed. The identically distributed requirement does not typically hold for MCMC because it would require us to marginalize over both the $\boldsymbol{\eta}_i$ and $\boldsymbol{y}_i^*$. So MCMC estimation of ordinal SEMs is less efficient than estimation of SEMs with continuous data, especially for large values of N, m, and p.

7.2. Logit link functions

IRT models typically use logit instead of probit link functions. The logit and probit link functions are generally equivalent to one another, and authors like McDonald (1999) describe simple transformations for approximating a logit model via a probit model or vice versa. We could embed such transformations in an MCMC method for probit link functions to approximately estimate a model with a logit link

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

function. Alternatively, for models with one latent variable per person, we can make progress with a similar Gibbs sampler by sampling the latent variables from a logistic distribution instead of a normal distribution (as was done around Equations (18) and (19); see Fox, 2010, p. 77). This approach still requires a Metropolis–Hastings step for the item parameters, and it can be difficult to extend to multidimensional latent variables because the multivariate logistic distribution does not have as many nice properties as the multivariate normal distribution. Finally, a latent response approach based on the Pólya–Gamma distribution can also be used to estimate a model with a logit link function. This approach was considered by Jiang and Templin (2019) for the two-parameter logistic model and has been applied recently to more complex psychometric models (e.g., Balamuta & Culpepper, 2022; Zens et al., 2023).

7.3. Three- and four-parameter models

Researchers often add "guessing" and "slipping" item parameters to the two-parameter model of binary data, leading to the three- and four-parameter IRT models (Barton & Lord, 1981). Culpepper (2016) describes a Gibbs sampler for these models involving analytic conditional posterior distributions, which is expected to be more efficient than other samplers that involve Metropolis–Hastings steps. He follows Béguin and Glas (2001) in sampling a latent binary response from the two-parameter model, which facilitates derivation of conditional posteriors of the four-parameter model. Sampling from the two-parameter model is similar to what we have already seen, with additional steps being required to sample the guessing and slipping parameters of the four-parameter model. This procedure poses a problem for HMC because it involves the latent binary responses, which are discrete parameters. We further consider discrete parameters in the General Discussion.

7.4. Generalized partial credit model

As compared to the graded response model, MCMC methods for the GPCM have received less attention. Patz and Junker (1999b) develop a Gibbs sampler with Metropolis Hastings steps, with analytic conditional posteriors being potentially complicated by the fact that the latent response interpretation of the GPCM is more complicated than that of the graded response model. This latent response interpretation is not often discussed in a GPCM context, but it arises often in the literature on ordinal regression model (e.g., Bürkner & Vuorre, 2019; Fullerton, 2009). Fox (2010, p. 104) outlines part of a Gibbs sampler for the partial credit model (not generalized) that involves model-predicted cumulative probabilities, and some further consideration of this approach may be worthwhile.

8. Multilevel models

Along with item response models, the methods discussed here can be extended to multilevel structural equation models. Some of these extensions are related to the GLLAMM framework of Skrondal and Rabe-Hesketh (2004), while Fox and colleagues have considered multilevel IRT models (e.g., Fox, 2005, 2007; Fox & Glas, 2001). The most common multilevel SEM describes students nested within schools, involving latent variables at both the student level and at the school level. We again face a tradeoff in using a likelihood that conditions on latent variables, versus using a likelihood that marginalizes over the latent variables. Use of the conditional likelihood is simpler, allows for more flexibility, and has been the focus of much of the previous work. But it requires sampling all the school and student latent variables, which can lead to sampling inefficiency. Use of the marginal likelihood can often capitalize on analytic results that were originally developed for SEMs with random intercepts (e.g., McDonald & Goldstein, 1989; Rosseel, 2021). These results simplify the evaluation of high-dimensional multivariate normal likelihoods, leading to HMC samplers that are especially fast and efficient. This functionality is included in recent versions of *blavaan* (Merkle et al., 2021; Merkle & Rosseel, 2018). But the story

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

is different for SEMs with random slopes: at the time of writing, use of the conditional likelihood is required. It could be worthwhile to apply an approach related to that of Rockwood (2020), where we sample only the subset of latent variables that leads to likelihood simplifications and/or to computational speedups.

9. General discussion

In this paper, we reviewed MCMC approaches to psychometric model estimation, using a regression-based framework to handle many popular models under one approach. We provided intuition about how the samplers work, highlighted the tradeoffs between flexibility and efficiency that exist for many models, and provided discussion of psychometric models not directly covered under our framework. In this final section, we discuss some needs and ideas about future Bayesian estimation methods for psychometric modeling. These include developments related to prior distributions, HMC, and copulas.

9.1. Prior distributions

A wide variety of prior distributions have been used for psychometric modeling, with conjugate priors and regularizing priors being especially popular. As we mentioned in the Metropolis–Hastings section, well-defined prior distributions are sometimes unavailable or unimplemented for some model parameters. This includes model covariance matrices with equality constraints or other fixed entries, as well as priors on threshold parameters (where parameters need to be ordered in a specific way) and priors on factor loadings with indeterminate signs. Merkle et al. (2023) reviewed these situations and provided some recommendations, but more work could be done to explictly define prior distributions for all parameters that a psychometric modeler could encounter. The goal should be to define prior distributions from which we can draw samples, which aids in prior predictive checks (e.g., Vanpaemel, 2020) and in sampler verification methods such as simulation-based calibration (e.g., Kim et al., 2022; Modrák et al., 2025). Pinkney (2024) made recent progress on priors for restricted correlation matrices, and we encourage more research along these lines.

9.2. Extensions to HMC

Stan is a great and flexible piece of software for HMC, and it has been very beneficial for psychometrics research. But the field of psychometrics might also consider HMC samplers that are specifically tailored to psychometric models, outside of Stan. For example, we have analytic results for psychometric model gradients and Hessians, and these results might be used to improve the efficiency of HMC. Additionally, we might make progress with complex models by sampling a subset of latent variables that leads to conditional independence (related to Rockwood, 2020), or by combining HMC samplers with Gibbs samplers. Expanding on the latter idea, it is possible to use our analytic result from Equation (18) to sample a model's latent variables from their conditional posterior distributions, then use HMC to sample other parameters. This might allow us to benefit from the efficiency of HMC when sampling measurement and structural parameters, and to employ HMC in models that involve discrete parameters. Martinez and Templin (2023) have recently considered related ideas for cognitive diagnosis models that involve many discrete parameters. Further work appears fruitful, with care being required to ensure that the HMC tuning parameters interact well with the other parts of the Gibbs sampler.

9.3. Alternative dependency structures

Finally, in addition to the traditional SEM-like models of covariances between observed variables, we see copulas as providing potential for flexible Bayesian model estimation. Copulas have already been applied to some problems in psychometrics, including generating non-normal data (Mair et al., 2012), handling violations of conditional independence (Braeken, 2011; Braeken et al., 2013; Braeken et al., 2007), and

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

obtaining flexible latent variable distributions (Nikoloulopoulos & Joe, 2015). Further work could be done to handle observed variables of different types (continuous, ordinal, count, etc.) within a marginal SEM framework. That is, the latent variables that generate a Gaussian copula could be parameterized via the covariance matrix from (15), while simultaneously maintaining the necessary marginal distributions of observed variables. Related ideas have been discussed in the context of the *brms* package (Bürkner, 2017), especially see https://github.com/paul-buerkner/brms/issues/304.

9.4. Summary

In discussing standard MCMC approaches to Bayesian estimation of psychometric models, we hope that the intuition paired with technical results inspires others to make continued progress with the methods. We should also mention related approaches that were neglected here. One involves the idea of specifying a multivariate model through a series of univariate, conditional distributions (Enders et al., 2018; Keller, 2021; Keller & Enders, 2023b; Kim et al., 2025), which is implemented in the Blimp software (Keller & Enders, 2023a). This idea is related to the methods in BUGS and JAGS, and the approach can be used to estimate complex psychometric models.

Other approaches involve sophisticated approximations of posterior distributions, including variational inference (Blei et al., 2017), amortized inference (Radev et al., 2023), and integrated nested Laplace approximations (INLAs; Rue et al., 2009). As applied to Bayesian models, variational inference seeks to approximate the posterior distribution via a normal (or other) distribution, with the methods being recently applied to psychometric models (e.g., Cho et al., 2024). Amortized inference uses neural networks and related methods to predict a model's posterior distribution given data. INLA uses Laplace approximations of the model posterior distribution, with the *INLAvaan* package (Jamil, 2024) containing an implementation for some psychometric models. All three of these approximations can lead to faster inferences as compared to MCMC, and further applications to Bayesian psychometric models are encouraged.

These approaches, as well as the ones discussed in the paper, benefit from the major computational advances witnessed in recent decades. As we continue to accrue computing resources, psychometric and statistical theory perhaps becomes less necessary in all situations. For example, it is possible to code an inefficient MCMC sampler and then rely on fast computing and large amounts of memory to obtain reliable results. But we think that the theoretical results discussed here provide a foundation that will continue to be useful for pushing the boundaries of models and datasets that can be considered. Maximally efficient methods will require incorporation of theoretical results alongside computing resources.

Data availability statement. All results were obtained using the R system for statistical computing (R Core Team, 2024), version 4.5.3. R is freely available under the General Public License 2 from the Comprehensive R Archive Network at http://CRAN.R-project.org/.

Funding statement. This work was supported by the Institute of Education Sciences, U.S. Department of Education, Grant R305D210044.

Competing interests. The author declares none.

References

Arminger, G., & Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika*, *63*(3), 271–300. https://doi.org/10.1007/BF02294856

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 359–388. https://doi.org/10.1080/10705511.2017.1406803

Asparouhov, T., & Muthén, B. (2010). Bayesian analysis using Mplus: Technical implementation. http://www.statmodel.com/download/Bayes3.pdf

Bainter, S. A. (2017). Bayesian estimation for item factor analysis models with sparse categorical indicators. *Multivariate Behavioral Research*, *52*(5), 593–615. https://doi.org/10.1080/00273171.2017.1342203

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

Baker, F. B., & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques*. (2nd ed.) CRC Press.
Balamuta, J. J., & Culpepper, S. A. (2022). Exploratory restricted latent class models with monotonicity requirements under Pólya-gamma data augmentation. *Psychometrika*, *87*(3), 903–945. https://doi.org/10.1007/s11336-021-09815-9
Bartholomew, D. J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology*, *34*, 93–99.
Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model. *ETS Research Report Series*, *1981*. https://doi.org/10.1002/j.2333-8504.1981.tb01255.x
Béguin, A. A., & Glas, C. A. W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, *66*(4), 541–561. https://doi.org/10.1007/BF02296195
Betancourt, M. (2018). A conceptual introduction to Hamiltonian Monte Carlo. Preprint, arXiv:1701.02434.
Betancourt, M . (2020). Identity crisis. https://betanalpha.github.io/assets/case_studies/identifiability.html
Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773
Bou-Rabee, N., Carpenter, B., Kleppe, T. S., & Liu, S. (2025). The within-orbit adaptive leapfrog No-U-Turn sampler. Preprint, arXiv:2506.18746.
Braeken, J. (2011). A boundary mixture approach to violations of conditional independence. *Psychometrika*, *76*(1), 57–76. https://doi.org/10.1007/s11336-010-9190-4
Braeken, J., Kuppens, P., De Boeck, P., & Tuerlinckx, F. (2013). Contextualized personality questionnaires: A case for copulas in structural equation models for categorical data. *Multivariate Behavioral Research*, *48*(6), 845–870. https://doi.org/10.1080/00273171.2013.827965
Braeken, J., Tuerlinckx, F., & De Boeck, P. (2007). Copula functions for residual dependency. *Psychometrika*, *72*(3), 393–411. https://doi.org/10.1007/s11336-007-9005-4
Breiman, L. (1997). No Bayesians in foxholes. *IEEE Expert: Intelligent Systems and their Applications*, *12*, 21–24.
Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01
Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, *100*(5), 1–54. https://doi.org/10.18637/jss.v100.i05
Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101. https://doi.org/10.1177/2515245918823199
Cai, J.-H., Lee, S.-Y., & Song, X.-Y. (2008). Bayesian analysis of nonlinear structural equation models with mixed continuous, ordered and unordered categorical, and nonignorable missing data. *Statistics and Its Interface*, *1*(1), 99–114. https://doi.org/10.4310/sii.2008.v1.n1.a9
Cain, M. K., & Zhang, Z. (2018). Fit for a Bayesian: An evaluation of PPP and DIC for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(1), 39–50. https://doi.org/10.1080/10705511.2018.1490648
Chen, M., Chow, S.-M., Oravecz, Z., & Ferrer, E. (2023). Fitting Bayesian stochastic differential equation models with mixed effects through a filtering approach. *Multivariate Behavioral Research*, *58*(5), 1014–1038. https://doi.org/10.1080/00273171.2023.2171354
Chib, S., & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, *85*, 347–361.
Cho, A. E., Xiao, J., Wang, C., & Xu, G. (2024). Regularized variational estimation for exploratory item factor analysis. *Psychometrika*, *89*(1), 347–375. https://doi.org/10.1007/s11336-022-09874-6
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*(2), 136–157. https://doi.org/10.3102/1076998615570945
Cudeck, R. (2005). Fitting psychometric models with methods based on automatic differentiation. *Psychometrika*, *70*(4), 599–617. https://doi.org/10.1007/s11336-005-1404-9
Culpepper, S. A. (2016). Revisiting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika*, *81*(4), 1142–1163. https://doi.org/10.1007/s11336-015-9477-6
Curtis, S. M. (2010). BUGS code for item response theory. *Journal of Statistical Software*, *36*, 1–34.
Depaoli, S. (2021). *Bayesian structural equation modeling*. Guilford Publications.
Depaoli, S., Lai, K., & Yang, Y. (2021). Bayesian model averaging as an alternative to model selection for multilevel models. *Multivariate Behavioral Research*, *56*(6), 920–940. https://doi.org/10.1080/00273171.2020.1778439
Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software*, *77*(5), 1–35. https://doi.org/10.18637/jss.v077.i05
Edwards, M. C. (2010). A Markov chain Monte Carlo approach to confirmatory item factor analysis. *Psychometrika*, *75*(3), 474–497. https://doi.org/10.1007/s11336-010-9161-9
Enders, C. K., Keller, B. T., & Levy, R. (2018). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological Methods*, *23*, 298–317.
Fazio, L., & Bürkner, P.-C. (2025). Gaussian distributional structural equation models: A framework for modeling latent heteroscedasticity. *Multivariate Behavioral Research*, *60*(4), 840–858. https://doi.org/10.1080/00273171.2025.2483252

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

Feng, C . (2024). The Markov chain Monte Carlo gallery. https://chi-feng.github.io/mcmc-demo/app.html?algorithm=HamiltonianMC&target=banana

Fox, J.-P. (2005). Multilevel IRT using dichotomous and polytomous response data. *British Journal of Mathematical and Statistical Psychology*, *58*(1), 145–172. https://doi.org/10.1348/000711005x38951

Fox, J.-P. (2007). Multilevel IRT modeling in practice with the package mlirt. *Journal of Statistical Software*, *20*(5). https://doi.org/10.18637/jss.v020.i05

Fox, J.-P. (2010). *Bayesian item response modeling: Theory and applications*. Springer.

Fox, J.-P., & Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*, 271–288.

Fox, J.-P., Mulder, J., & Sinharay, S. (2017). Bayes factor covariance testing in item response models. *Psychometrika*, *82*(4), 979–1006. https://doi.org/10.1007/s11336-017-9577-6

Fullerton, A. S. (2009). A conceptual framework for ordered logistic regression models. *Sociological Methods & Research*, *38*, 306–347.

Garnier-Villarreal, M., & Jorgensen, T. D. (2020). Adapting fit indices for Bayesian structural equation modeling: Comparison to maximum likelihood. *Psychological Methods*, *25*(1), 46–70.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. (3rd ed.) Chapman & Hall/CRC.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

Glas, C. A. W., & Meijer, R. R. (2003). A Bayesian approach to person fit analysis in item response theory models. *Applied Psychological Measurement*, *27*(3), 217–233. https://doi.org/10.1177/0146621603027003003

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(47), 1593–1623. http://jmlr.org/papers/v15/hoffman14a.html

Hoijtink, H., & van de Schoot, R. (2018). Testing small variance priors using prior-posterior predictive p values. *Psychological Methods*, *23*(3), 561–569. https://doi.org/10.1037/met0000131

Jamil, H. (2024). *INLAvaan: Bayesian structural equation modeling with INLA*. R package version 0.1.0.9013. https://github.com/haziqj/INLAvaan

Ji, F., Rabe-Hesketh, S., & Skrondal, A. (2023). Diagnosing and handling common violations of missing at random. *Psychometrika*, *88*(4), 1123–1143. https://doi.org/10.1007/s11336-022-09896-0

Jiang, Z., & Templin, J. (2019). Gibbs samplers for logistic item response models via the Pólya–gamma distribution: A computationally efficient data-augmentation strategy. *Psychometrika*, *84*(2), 358–374. https://doi.org/10.1007/s11336-018-9641-x

Jöreskog, K. G., & Sörbom, D. (1997). *LISREL 8 user's reference guide*. Scientific Software International.

Jorgensen, T. D., Forney, K. J., Hall, J. A., & Giles, S. M. (2018). Using modern methods for missing data analysis with the social relations model: A bridge to social network analysis. *Social Networks*, *54*, 26–40. https://doi.org/10.1016/j.socnet.2017.11.002

Kaplan, D. (2023). *Bayesian statistics for the social sciences*. (2nd ed.) Guilford.

Kaplan, D., Chen, J., Yavuz, S., & Lyu, W. (2023). Bayesian dynamic borrowing of historical information with applications to the analysis of large-scale assessments. *Psychometrika*, *88*, 1–30. https://doi.org/10.1007/s11336-022-09869-3

Keller, B. T. (2021). An introduction to factored regression models with blimp. *The Psychiatrist*, *4*(1), 10–37. https://doi.org/10.3390/psych4010002

Keller, B. T., & Enders, C. K. (2023a). *Blimp user's guide* (3rd ed.) https://www.appliedmissingdata.com/blimp

Keller, B. T., & Enders, C. K. (2023b). An investigation of factored regression missing data methods for multilevel models with cross-level interactions. *Multivariate Behavioral Research*, *58*, 938–963. https://doi.org/10.1080/00273171.2022.2147049

Kim, J., Lee, S., Kim, S., & Keller, B. T. (2025). Tutorial for Bayesian multilevel structural equation modeling using blimp. *Structural Equation Modeling*, 1–13. https://doi.org/10.1080/10705511.2025.2526035

Kim, S., Moon, H., Modrák, M., & Säilynoja, T. (2022). *SBC: Simulation based calibration for rstan/cmdstanr models*. https://hyunjimoon.github.io/SBC/

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, *95*, 391–413.

Lee, S.-Y. (1981). A Bayesian approach to confirmatory factor analysis. *Psychometrika*, *46*, 153–160.

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. John Wiley & Sons.

Lee, S.-Y., & Song, X.-Y. (2002). Bayesian selection on the number of factors in a factor analysis model. *Behaviormetrika*, *29*, 23–39.

Lee, S.-Y., Song, X.-Y., & Tang, N.-S. (2007). Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling*, *14*, 404–434.

Lee, S.-Y., & Xia, Y.-M. (2008). A robust Bayesian approach for structural equation models with missing data. *Psychometrika*, *73*(3), 343–364. https://doi.org/10.1007/s11336-008-9060-5

Levy, R., & McNeish, D. (2023). Perspectives on Bayesian inference and their implications for data analysis. *Psychological Methods*, *28*(3), 719–739. https://doi.org/10.1037/met0000443

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

Levy, R., & Mislevy, R. J. (2017). *Bayesian psychometric modeling*. https://doi.org/10.1201/9781315374604
Liu, H., Depaoli, S., & Marvin, L. (2022). Understanding the deviance information criterion for SEM: Cautions in prior specification. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(2), 278–294.
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*, 353–367.
Lu, Z.-H., Chow, S.-M., & Loken, E. (2016). Bayesian factor analysis as a variable-selection problem: Alternative priors and consequences. *Multivariate Behavioral Research*, *51*, 519–539.
Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Chapman & Hall/CRC.
MacCallum, R. C., Edwards, M. C., & Cai, L. (2012). Hopes and cautions in implementing Bayesian structural equation modeling. *Psychological Methods*, *17*, 340–345.
Mair, P., Satorra, A., & Bentler, P. M. (2012). Generating nonnormal multivariate data using copulas: Applications to SEM. *Multivariate Behavioral Research*, *47*(4), 547–565. https://doi.org/10.1080/00273171.2012.692629
Martin, J. K., & McDonald, R. P. (1975). Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika*, *40*, 505–517.
Martinez, A. J., & Templin, J. (2023). Estimating Bayesian diagnostic models with attribute hierarchies with the Hamiltonian-Gibbs hybrid sampler. *Multivariate Behavioral Research*, *58*(1), 141–142. https://doi.org/10.1080/00273171.2022.2160950
McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum.
McDonald, R. P., & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two level data. *British Journal of Mathematical and Statistical Psychology*, *42*, 215–232.
McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. (2nd ed.) Chapman and Hall/CRC.
Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, *9*, 538–558.
Merkle, E. C. (2011). A comparison of imputation methods for Bayesian factor analysis models. *Journal of Educational and Behavioral Statistics*, *36*, 257–276. https://doi.org/10.3102/1076998610375833
Merkle, E. C., Ariyo, O., Winter, S. D., & Garnier-Villarreal, M. (2023). Opaque prior distributions in Bayesian latent variable models. *Methodology*, *19*(3), 228–255. https://doi.org/10.5964/meth.11167
Merkle, E. C., Fitzsimmons, E., Uanhoro, J., & Goodrich, B. (2021). Efficient Bayesian structural equation modeling in Stan. *Journal of Statistical Software*, *100*(6), 1–22.
Merkle, E. C., Furr, D., & Rabe-Hesketh, S. (2019). Bayesian model assessment: Use of conditional vs marginal likelihoods. *Psychometrika*, *84*, 802–829. https://doi.org/10.1007/s11336-019-09679-0
Merkle, E. C., & Rosseel, Y. (2018). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, *85*(4), 1–30.
Modrák, M., Moon, A. H., Kim, S., Bürkner, P., Huurre, N., Faltejsková, K., Gelman, A., & Vehtari, A. (2025). Simulation-based calibration checking for Bayesian computation: The choice of test quantities shapes sensitivity. *Bayesian Analysis*, *20*(2), 461–488. https://doi.org/10.1214/23-BA1404
Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335.
Muthén, L. K., & Muthén, B. (1998–2017). *Mplus user's guide*. (8th ed.) Muthén & Muthén.
Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. L. Jones & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*. London: Chapman and Hall/CRC. https://doi.org/10.1201/b10905-6
Nikoloulopoulos, A. K., & Joe, H. (2015). Factor copula models for item response data. *Psychometrika*, *80*(1), 126–150. https://doi.org/10.1007/s11336-013-9387-4
O'Muircheartaigh, C., & Moustaki, I. (1999). Symmetric pattern models: A latent variable approach to item non-response in attitude scales. *Journal of the Royal Statistical Society A*, *162*, 177–194.
Padgett, R. N. (2022). *Bayesian psychometric modeling*. https://noah-padgett.github.io/Bayesian-Psychometric-Modeling/
Padgett, R. N., Morgan, G. B., & Lomas, T. (2024). Regularizing threshold priors with sparse response patterns in Bayesian factor analysis with categorical indicators. *Psychological Methods*. https://doi.org/10.1037/met0000682
Padgett, R. N., & Winter, S. D. (2025). Threshold priors in item factor analysis under DELTA and THETA parameterizations. https://doi.org/10.31234/osf.io/8mkv7_v1
Patz, R. J., & Junker, B. W. (1999a). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, *24*, 146–178.
Patz, R. J., & Junker, B. W. (1999b). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics*, *24*, 342–366.
Pinkney, S. (2024). A short note on a flexible Cholesky parameterization of correlation matrices. Preprint, arXiv:2405.07286.
Pokropek, A., Schmidt, P., & Davidov, E. (2020). Choosing priors in Bayesian measurement invariance modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(5), 750–764. https://doi.org/10.1080/10705511.2019.1703708
Press, S. J. (1972). *Applied multivariate analysis: Using Bayesian and frequentist methods of inference*. Holt, Rinehart, and Winston.

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Rabe-Hesketh, S., & Skrondal, A. (2023). Ignoring non-ignorable missingness. *Psychometrika*, *88*(1), 31–50. https://doi.org/10.1007/s11336-022-09895-1

Radev, S. T., Schmitt, M., Schumacher, L., Elsemüller, L., Pratz, V., Schälte, Y., Köthe, U., & Bürkner, P.-C. (2023). Bayesflow: Amortized Bayesian workflows with neural net-works. *Journal of Open Source Software*, *8*(89), 5702. https://doi.org/10.21105/joss.05702

Rast, P., Martin, S., Liu, S., & Williams, D. (2022). A new frontier for studying within-person variability: Bayesian multivariate generalized autoregressive conditional het-eroskedasticity models. *Psychological Methods*, *27*, 856–873. https://doi.org/10.1037/met0000357.

Rockwood, N. J. (2020). Maximum likelihood estimation of multilevel structural equation models with random slopes for latent covariates. *Psychometrika*, *85*(2), 275–300. https://doi.org/10.1007/s11336-020-09702-9

Roman, Z. J., & Brandt, H. (2021). A latent auto-regressive approach for Bayesian structural equation modeling of spatially or socially dependent data. *Multivariate Behavioral Research*. https://doi.org/10.1080/00273171.2021.1957663

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. http://www.jstatsoft.org/v48/i02/

Rosseel, Y. (2021). Evaluating the observed log-likelihood function in two-level structural equation modeling with missing data: From formulas to R code. *The Psychiatrist*, *3*(2), 197–232. https://doi.org/10.3390/psych3020017

Rouder, J. N., Morey, R. D., Verhagen, J., Province, J. M., & Wagenmakers, E.-J. (2016). Is there a free lunch in inference? *Topics in Cognitive Science*, *8*, 520–547.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, *64*, 37–52.

Seaman, J. W., Seaman, J. W., & Stamey, J. D. (2012). Hidden dangers of specifying non-informative priors. *The American Statistician*, *66*(2), 77–84. https://doi.org/10.1080/00031305.2012.695938

Shi, J.-Q., & Lee, S.-Y. (1997). A Bayesian estimation of factor score in confirmatory factor model with polytomous, censored or truncated data. *Psychometrika*, *62*, 29–50.

Shi, J.-Q., & Lee, S.-Y. (1998). Bayesian sampling-based approach for factor analysis models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology*, *51*, 233–252.

Sinharay, S. (2004). Experiences with Markov chain Monte Carlo convergence assessment in two psychometric examples. *Journal of Educational and Behavioral Statistics*, *29*(4), 461–488. https://doi.org/10.3102/10769986029004461

Sinharay, S. (2005). Assessing fit of unidimensional item response theory models using a Bayesian approach. *Journal of Educational Measurement*, *42*(4), 375–394. https://doi.org/10.1111/j.1745-3984.2005.00021.x

Sinharay, S., Johnson, M. S., & Stern, H. S. (2006). Posterior predictive assessment of item response theory models. *Applied Psychological Measurement*, *30*(4), 298–321. https://doi.org/10.1177/0146621605285517

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation modeling*. Chapman & Hall.

Song, X.-Y., & Lee, S.-Y. (2001). Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. *British Journal of Mathematical and Statistical Psychology*, *54*, 237–263.

Song, X.-Y., & Lee, S.-Y. (2012). A tutorial on the Bayesian approach for analyzing structural equation models. *Journal of Mathematical Psychology*, *56*, 135–148.

Stan Development Team. (2024). *Stan reference manual, version 2.36*. http://mc-stan.org/

Stromeyer, W. R., Miller, J. W., Sriramachandramurthy, R., & DeMartino, R. (2015). The prowess and pitfalls of Bayesian structural equation modeling: Important considerations for management research. *Journal of Management*, *41*, 491–520.

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393–408.

Thomas, S., & Tu, W. (2021). Learning Hamiltonian Monte Carlo in R. *The American Statistician*, *75*, 403–413. https://doi.org/10.1080/00031305.2020.1865198

Tong, X., & Zhang, Z. (2020). Robust Bayesian approaches in growth curve modeling: Using student's t distributions versus a semiparametric method. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(4), 544–560. https://doi.org/10.1080/10705511.2019.1683014

Uanhoro, J. O. (2023). Modeling misspecification as a parameter in Bayesian structural equation models. *Educational and Psychological Measurement*. https://doi.org/10.1177/00131644231165306

Van Erp, S., & Browne, W. J. (2021). Bayesian multilevel structural equation modeling: An investigation into robust prior distributions for the doubly latent categorical model. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(6), 875–893. https://doi.org/10.1080/10705511.2021.1915146

Van Erp, S., Mulder, J., & Oberski, D. L. (2018). Prior sensitivity analysis in default Bayesian structural equation modeling. *Psychological Methods*, *23*, 363–388.

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

Van Zundert, C., Somer, E., & Miočević, M. (2022). Prior predictive checks for the method of covariances in Bayesian mediation analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(3), 428–437. https://doi.org/10.1080/10705511.2021.1977648

Vanpaemel, W. (2020). Strong theory testing using the prior predictive and the data prior. *Psychological Review*, *127*, 136–145.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432.

Verhagen, A. J., & Fox, J.-P. (2013). Bayesian tests of measurement invariance. *British Journal of Mathematical and Statistical Psychology*, *66*, 383–401.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Winter, S. D., & Depaoli, S. (2022). Performance of model fit and selection indices for Bayesian structural equation modeling with missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(4), 531–549. https://doi.org/10.1080/10705511.2021.2018656

Winter, S. D., & Depaoli, S. (2023). Illustrating the value of prior predictive checking for Bayesian structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*(6), 1000–1021. https://doi.org/10.1080/10705511.2022.2164286

Wothke, W. (2000). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 205–224). Lawrence Erlbaum Associates.

Woźniak, T. (2024). *bsvars: Bayesian estimation of structural vector autoregressive models*. R package version 3.1. https://doi.org/10.32614/CRAN.package.bsvars

Zens, G., Frühwirth-Schnatter, S., & Wagner, H. (2023). Ultimate Pólya gamma samplers—Efficient MCMC for possibly imbalanced binary and categorical data. *Journal of the American Statistical Association*, *119*(548), 2548–2559. https://doi.org/10.1080/01621459.2023.2259030

Zhang, J., Li, J., & Liu, C. (2014). Robust factor analysis using the multivariate t-distribution. *Statistica Sinica*, *24*, 291–312.

Zhang, J., Templin, J., & Mintz, C. E. (2022). A model comparison approach to posterior predictive model checks in Bayesian confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(3), 339–349. https://doi.org/10.1080/10705511.2021.2012682

Zhang, X., Tao, J., Wang, C., & Shi, N.-Z. (2019). Bayesian model selection methods for multilevel IRT models: A comparison of five DIC-based indices. *Journal of Educational Measurement*, *56*, 3–27.

Zhang, Z., Hamagami, F., Wang, L., Nesselroade, J. R., & Grimm, K. J. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, *31*(4), 374–383. https://doi.org/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*(1), 47–78. https://doi.org/10.1080/10705511.2013.742382

Appendix

A. Gibbs sampling extensions

In this appendix, we describe extensions of the Gibbs sampler to situations that are common to SEM, including datasets with missing observations, models with standalone observed variables, and models with equality constraints. Part of this discussion is inspired by the way that *lavaan* (Rosseel, 2012) constructs models.

A.1. Missing data

If we are willing to assume that the data are missing at random, it is relatively easy to handle missing data. Equation (14) describes the joint distribution of the observed and missing variables for case *i*. It is then easy to obtain the conditional distribution of missing values given observed variables based on the analytic properties of the multivariate normal distribution (the same property that gave us the conditional distribution of latent variables given observed, in Equations (18) and (19)). At each iteration, we sample from this conditional distribution to fill in the missing data, then carry out the other steps of the Gibbs sampler as if we had complete data.

Another option involves use of a "full information" likelihood (e.g., Wothke, 2000), where we skip over missing entries in the data. This does not require us to impute the missing entries, but it requires more bookkeeping in the Gibbs sampler. To efficiently evaluate the multivariate normal likelihood here, it is helpful to first arrange the cases by missing data pattern. This allows us to compute the inverse and determinant of the model-implied covariance matrix once per missing data pattern, then reuse the results for all cases with the same pattern.

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

Yet another option is to impute the missing values before model estimation, using the analyst's choice of imputation model (often an unrestricted multivariate normal). This provides some flexibility for model estimation, but it can also lead to a conflict in one's analytic strategy: the researcher fits one model to their dataset but then decides that another model should be used for imputation. This conflict sometimes leads to model posterior distributions with increased uncertainty (Meng, 1994; Merkle, 2011).

If one is unwilling to assume that data are missing at random, then the situation becomes more complicated. The approach of O'Muircheartaigh and Moustaki (1999) is notable here, where the missingness indicators (0 for missing, 1 for observed) are incorporated into the model as binary observed variables. It is then possible to specify a latent "missingness" variable that influences other parts of the model. While this approach cannot handle all "missing not at random" mechanisms, it provides flexibility and usually includes the missing at random model as a special case. For further discussion of missingness mechanisms and other approaches, see Ji et al. (2023) and Rabe-Hesketh and Skrondal (2023).

A.2. Standalone observed variables

Many models in the SEM framework include observed variables that are not directly associated with a latent variable. We can usually handle these "standalone" observed variables by defining an extra latent variable. The loading from this extra latent variable to the observed variable is fixed to 1, and the residual variance of the observed variable (in $\boldsymbol{\Theta}$) is fixed to 0 (side note: fixing a variance to 0 can cause problems with JAGS and Stan model specifications!). Then the corresponding entries of $\boldsymbol{\eta}_i$ are fixed to their observed values and are not sampled.

Further modifications are required for exogenous, standalone observed variables. For these variables, the mean and variance of the latent variable are fixed to the sample mean and variance of the observed variable. We then partition $\boldsymbol{\eta}_i$ into $(\boldsymbol{\eta}_i^{n\prime}\ \boldsymbol{\eta}_i^{x\prime})'$, where $\boldsymbol{\eta}_i^x$ contains exogenous observed variables and $\boldsymbol{\eta}_i^n$ contains the remaining variables in $\boldsymbol{\eta}_i$. Then we need to sample $\boldsymbol{\eta}_i^n$ conditioned on $\boldsymbol{y}$ *and* on $\boldsymbol{\eta}_i^x$, which can be accomplished with the following modifications to Equations (18) and (19).

- In both equations, off-diagonal entries of $(\boldsymbol{I}-\boldsymbol{B})^{-1}$ in rows and columns corresponding to exogenous observed variables are set to 0. Diagonal entries of $(\boldsymbol{I}-\boldsymbol{B})^{-1}$ corresponding to exogenous observed variables are set to 1.
- Let $\boldsymbol{B}^*$ be the $m \times m_x$ submatrix of $\boldsymbol{B}$, whose columns correspond to the m_x exogenous observed variables. Let $\boldsymbol{y}^*$ be the $m_x \times 1$ subvector of $\boldsymbol{y}$, whose entries correspond to the exogenous observed variables. Then the final $(\boldsymbol{I}-\boldsymbol{B}')\boldsymbol{\Psi}^{-1}\boldsymbol{\alpha}$ term of Equation (18) becomes

$$(\boldsymbol{I}-\boldsymbol{B}')\boldsymbol{\Psi}^{-1}(\boldsymbol{\alpha}+\boldsymbol{B}^*\boldsymbol{y}^*).$$

We also require some modifications to handle the fixed factor loading of 1. These are covered in the next section, which is more generally about handling equality constraints.

A.3. Equality constraints

Two primary types of equality constraints in SEM are: (i) constraints that fix a specific parameter to a nonzero constant and (ii) constraints specifying that free parameters are equal to one another (to an unknown value). Below, we consider how each can be incorporated into the Gibbs sampler; also see Section 5 of Asparouhov and Muthén (2010).

A.3.1. Parameters fixed to a constant

For parameters that are fixed to a nonzero constant, many steps of the Gibbs sampler remain the same, just with some entries of the model matrices being fixed. An exception to this idea occurs in the "sampling location parameters" step. If we encounter a nonzero fixed entry of $\boldsymbol{\nu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\alpha}$, or $\boldsymbol{B}$, we residualize those terms from the $\boldsymbol{z}$ vector in Equation (21). Let $\boldsymbol{H}_{\text{fix}}$ be the same as $\boldsymbol{H}$, with columns corresponding to free parameters becoming all zero. Let $\boldsymbol{H}_{\text{free}}$ be the submatrix of $\boldsymbol{H}$ that includes only columns corresponding to free parameters. Similarly, let $\boldsymbol{\xi}_{\text{free}}$ contain only free parameters. Then Equation (21) becomes

$$(\boldsymbol{z}-\boldsymbol{H}_{\text{fix}}\boldsymbol{\xi}) \sim \text{N}(\boldsymbol{H}_{\text{free}}\boldsymbol{\xi}_{\text{free}}, \boldsymbol{I}\otimes\boldsymbol{V}).$$

A.3.2. Parameters constrained to be equal

For parameters constrained to be equal to one another, we must consider whether the constraints involve variance parameters or location parameters. For the Gibbs sampler outlined in this paper, we cannot generally constrain variance parameters to be equal to location parameters because these parameters are updated separately.

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press

For variance parameters, we modify Equation (26) to include all the variances that are constrained to be equal. For example, say that we constrain $\psi_{11} = \psi_{22} = \psi^*$, with ψ^* having an Inverse Gamma(a^*, b^*) prior. Then we have

$$\psi^* \mid \boldsymbol{\xi}, \boldsymbol{\eta} \sim \text{Inv-Gamma}\left(a^* + n, b^* + \frac{1}{2}\sum_{i=1}^{n}(\eta_{i1} - (\alpha_1 + \sum_{k=1}^{m} B_{1k}\eta_{ik}))^2 + \frac{1}{2}\sum_{i=1}^{n}(\eta_{i2} - (\alpha_2 + \sum_{k=1}^{m} B_{2k}\eta_{ik}))^2\right).$$

If we were to constrain covariance matrix blocks to be equal to one another, then a similar result applies to Equation (27).

For location parameters with equality constraints, we modify the results from Equation (23). The $\boldsymbol{\xi}$ vector now includes only nonredundant location parameters, so that we do not double-count parameters with equality constraints. But we then need to ensure that the $\boldsymbol{D}$ and $\boldsymbol{d}$ matrices have the correct, "nonredundant" dimensions. To do so, we

- Compute $\sum_{i=1}^{n} \boldsymbol{H}_i' \boldsymbol{V}^{-1} \boldsymbol{z}_i$ as if we had no equality constraints, then sum the terms corresponding to equality-constrained parameters. Then we obtain the appropriate dimension of $\boldsymbol{d}$ (where it is assumed that $\boldsymbol{\xi}_0$ and $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$ already exclude nonredundant parameters).
- Compute $\sum_{i=1}^{n} \boldsymbol{H}_i' \boldsymbol{V}^{-1} \boldsymbol{H}_i$ as if we had no equality constraints. Then sum the rows corresponding to equality-constrained parameters and collapse them. Then sum the columns corresponding to equality-constrained parameters and collapse them. Now we obtain the appropriate dimension of $\boldsymbol{D}$ (where it is assumed that $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$ already excludes nonredundant parameters).

Cite this article: Merkle, E. C. (2026). Bayesian Estimation of Normal and Probit Psychometric Models. *Psychometrika*, 1599–1625. https://doi.org/10.1017/psy.2026.10119

https://doi.org/10.1017/psy.2026.10119 Published online by Cambridge University Press