

BRIEF REPORT

Bayesian latent variable models for the analysis of experimental psychology data

Edgar C. Merkle¹ · Ting Wang¹

Published online: 18 March 2016
© Psychonomic Society, Inc. 2016

Abstract In this paper, we address the use of Bayesian factor analysis and structural equation models to draw inferences from experimental psychology data. While such application is non-standard, the models are generally useful for the unified analysis of multivariate data that stem from, e.g., subjects' responses to multiple experimental stimuli. We first review the models and the parameter identification issues inherent in the models. We then provide details on model estimation via JAGS and on Bayes factor estimation. Finally, we use the models to re-analyze experimental data on risky choice, comparing the approach to simpler, alternative methods.

Keywords Bayesian statistics · Decision making

Traditional applications of factor analysis and related latent variable models include psychometric scale development, analysis of observational data, and possibly data reduction (though the related, but distinct, principal components

analysis is more relevant here). Because experimental psychologists do not typically embark on such applications, they may regard latent variable models as being irrelevant to their statistical toolboxes. Furthermore, applications of latent variable models to experimental data are relatively uncommon in the literature (though see Bagozzi & Yi, 1989; Donaldson, 2003; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000; Russell, Kahn, Spoth, & Altmaier, 1998; Wicherts et al., 2005). In this paper, we illustrate that Bayesian latent variable models can be advantageous for the analysis of multivariate data collected in many laboratory experiments. The models flexibly allow us to handle data from multiple trials, conditions, or scenarios, without the need for aggregation across trials or subjects. This allows us to pool data across multiple trials in order to draw general conclusions about the research question of interest, while preserving variability inherent in the raw data. We focus on Bayesian versions of the models because they are flexible and advantageous for model extension, comparison, and interpretation.

Bayesian approaches to estimating latent variable models have been considered for at least forty years, with increased attention to the subject following the computational advances in Markov chain Monte Carlo (MCMC). Early treatments include Press (1972), Martin and McDonald (1975), Bartholomew (1981), and Lee (1981), while later treatments that incorporate MCMC include Scheines et al. (1999), Lee (2007), Song and Lee (2012a), and Muthén and Asparouhov (2012). Some authors of the recent papers refer to a need for fully automated software that fits latent variable models via MCMC, with Mplus (Muthén & Muthén, 1998) providing SEM functionality and BUGS (Lunn et al., 2012), JAGS (Plummer, 2003), and Stan (Stan Development Team, 2014) providing general functionality. This automation has the reward of

This work was supported by National Science Foundation grant SES-1061334. The authors thank Carel Peeters, two anonymous reviewers, and the action editor for comments that improved the paper. Remaining errors are the authors' responsibility. The authors also thank Ellen Peters and Irwin Levin for making the data from their 2008 article freely available. Correspondence to Edgar C. Merkle, Department of Psychological Sciences, University of Missouri, Columbia, MO 65211. Email: merkle@missouri.edu.

✉ Edgar C. Merkle
merkle@missouri.edu

¹ University of Missouri, Columbia, MO, USA

easily estimating complex models coupled with the risk of estimating inappropriate models or drawing inappropriate conclusions (also see MacCallum, Edwards, & Cai, 2012; Steiger, 2001; Stromeier, Miller, Sriramachandramurthy, & DeMartino, 2014).

In the pages below, we first present an overview of factor analysis and structural equation models. We also discuss necessary topics for Bayesian model estimation, including prior distribution specification and model comparison methods. The Bayesian methods are trickier than they are for other models because of the inherent parameter identification issues. We then apply the models to data from an experiment on risky choice, highlighting the unified and unique results that the models can provide. We estimate all models in JAGS and carry out related computations in R; code to replicate all results is provided at the URL referenced at the end of the paper.

Models

To describe the models, we first set up a framework that is relevant to experimental research. We assume that n individuals respond to a series of stimuli that are all intended to measure the same concept. Across the stimuli, each individual contributes p observed variables. For example, in the illustration later, the stimuli are five binary choices. For each choice stimulus, individuals rate (i) the extent to which each individual choice option is attractive and (ii) the extent to which they prefer one stimulus over the other. Here, the stimuli are all intended to measure an individual's attraction to risk, and each individual provides $p = 15$ observed variables: three ratings for each of 5 stimuli. We generally assume that y_{ij} ($i = 1, \dots, n$; $j = 1, \dots, p$) is individual i 's response to observed variable j , and we also sometimes assume that individuals are randomly assigned to a between-subjects condition. This mimics many psychology experiments, where researchers are interested in studying whether experimental conditions impacted individuals' responses across multiple trials.

In the discussion below, we assume that y_{ij} is approximately continuous (so that our models fully rely on normal distributions). However, variants of these models can also be applied to discrete data such as binary choices or ordinal ratings. These variants are highly related to the item response models.

Factor model

The traditional factor model (e.g., Lawley & Maxwell, 1971; Bartholomew, Knott, & Moustaki, 2011) assumes that each individual has a stable underlying trait (e.g., attraction to risky choice) across all experimental stimuli. An

individual's response to each stimulus is driven by this trait. The stimuli are unique, however, so that stimulus characteristics also impact the individual's responses. Further, like many other models, noise impacts the responses.

The above ideas can be formally represented by:

$$y_{ij}|\theta_i, \gamma_j, \lambda_j, \psi_j \sim N(\mu_{ij}, \psi_j) \quad \text{for all } i, j \quad (1)$$

$$\mu_{ij} = \gamma_j + \lambda_j \theta_i \quad (2)$$

$$\theta_i \sim N(0, \phi). \quad (3)$$

Within this model, θ_i is the value of individual i 's stable trait. This value arises from a normal distribution with mean 0 and variance ϕ , reflecting the distribution of the trait across the population of interest. The observed response y_{ij} is perturbed from the trait by the sources mentioned in the previous paragraph: stimulus characteristics γ_j and λ_j , and random noise with variance ψ_j .

The stimulus characteristics include the intercept γ_j , reflecting stimulus j 's mean response in the population of interest, and slope λ_j , reflecting the extent to which stimulus j "taps into" the trait of interest. If λ_j equals 0, individual i 's standing on the trait has no impact on his/her response y_{ij} . Conversely, if λ_j is large (in the positive or negative direction), individual i 's standing on the trait has a large impact on his/her response. In traditional applications, the λ_j are called *factor loadings* and the trait of interest is called the *factor*.

Multiple factors The above model assumed that a single latent trait was associated with the observed responses y_{ij} . In some studies, multiple traits will have an association. For example, undergraduate participants' responses to a recognition memory task may be influenced both by their working memory capacity and by their attentional functioning during the experiment. If a participant has both high working memory and high attention, he/she should exhibit good test performance. Conversely, low values of either trait will result in lower performance.

Assuming that m traits impact observed responses, we can modify the previous model to arrive at an " m -factor model:"

$$y_{ij}|\theta_i, \gamma_j, \Lambda_j, \psi_j \sim N(\mu_{ij}, \psi_j) \quad \text{for all } i, j \quad (4)$$

$$\mu_{ij} = \gamma_j + \sum_{k=1}^m \lambda_{jk} \theta_{ik} \quad (5)$$

$$\theta_i \sim N_m(\mathbf{0}, \Phi), \quad (6)$$

where θ_i is a vector describing individual i 's standing on each of the m latent traits (with this vector now arising from a multivariate normal distribution) and Λ_j contains the m factor loadings for stimulus j . The matrix Φ contains information about the variance of each trait along

with covariances between pairs of traits. For the next section, it is helpful to combine the Λ_j into a $p \times m$ matrix Λ .

Exploratory versus confirmatory models From inspection of the above models, it is evident that there are some unidentified parameters. For example, Eq. 2 involves the multiplication of two unknown parameters $\lambda_j \times \theta_i$. If we multiply each λ by a constant and divide each θ by the same constant, our model predictions remain unchanged. This lack of identification is compounded by the fact that the Φ matrix typically contains additional free parameters. To address problems such as this, we need to fix some model parameters to constants (typically to zero or one). Additionally, when we consider the m -factor model from Eqs. 4 to 6, it is possible to use matrix algebra to illustrate a *rotational indeterminacy* problem. This implies that specific linear transformations (more complex than multiplication by a constant) of the λ parameters, coupled with inverse transformations of the θ parameters, lead to unchanged model predictions.

The m -factor model requires us to systematically constrain m^2 parameters in order to achieve identification (this does not count the mean of θ_i , which is fixed to 0). If we constrain only the minimum m^2 parameters, we obtain an *exploratory* factor model. This model allows for the possibility that all latent traits (contained in θ) are associated with all observed measures (the y_{ij}). We can obtain unique parameter estimates based on the specific m^2 parameter constraints employed, but there are an infinite number of other, equally-good sets of parameter values coupled with alternative parameter constraints (these arise from the rotational indeterminacy issue described in the previous paragraph). Because of this, the λ (and possibly Φ) parameters are typically transformed following model specification so that they are easily interpretable. This “interpretability” step is called *rotation*, and we provide some further information on it in the General Discussion (also see Browne, 2001). Regardless of the rotation step, however, the exploratory model can be useful because it allows us to compare models with different values of m . This can provide information about the number of latent traits associated with the observed measures, as well as about which latent traits are associated with which measures.

If we fix more than m^2 parameters, we obtain a *confirmatory* factor model. In this model, we assume that some latent traits have no association with some observed measures (consequently, some of the parameters in Λ are fixed to zero) and also assume fixed, nonzero values for some parameters. Rotational indeterminacy is no longer a problem here, though alternative parameter constraints can still impact parameter estimates. These are further described below.

Identification constraints for factor models Regardless of whether we have an exploratory or confirmatory model, the first m^2 parameter constraints must be set in a systematic fashion (see Jöreskog, 1969, 1979; Peeters, 2012a, b) to identify the likelihood. Popular strategies are described below; all strategies assume that the resulting Λ matrix is of full rank (e.g., that we cannot obtain one column of Λ through a linear combination of other columns) and that the resulting Φ matrix is positive definite.

Often, we start by setting $\Phi = I$, which means that each latent trait has a variance of 1 with each pair of traits having a 0 correlation. If we employ this constraint, we then systematically fix $m(m - 1)/2$ parameters in Λ : one column can have no zeros, one column must have one zero, one column must have two zeros, ..., and one column must have $(m - 1)$ zeros.

Alternatively to above, we can fix the diagonal of Φ (reflecting latent trait variances) to equal 1 but allow other parameters in Φ to be unconstrained. This allows for the possibility that latent traits are correlated. If we opt for these constraints, we must fix more parameters in Λ to zero. In particular, each column of Λ must now contain $(m - 1)$ zeros, for a total of $m(m - 1)$ zeroes in Λ . These restrictions, proposed by Jöreskog (1979), achieve “local” identification: they identify the model up to sign changes in columns of Λ . Peeters (2012b) describes how to make these constraints globally identified: in addition to the above constraints, one must force a nonzero parameter in each column of Λ to assume only positive or negative values.

Finally, another strategy involves placing all the restrictions in Λ . Now Φ is completely unconstrained, with each column of Λ containing $(m - 1)$ zeros. Each column of Λ must also contain a nonzero constant, typically one (with each one appearing in different rows), compensating for the unconstrained Φ . These restrictions, also proposed by Jöreskog (1979), achieve global identification.

The above constraints might make some readers uncomfortable, leading them to wonder whether estimated parameters can be interpreted given all the identification issues involved. We suggest first comparing exploratory models via Bayes factors (or other global model statistics), which should be approximately invariant to the specific identification constraints that are chosen. The Bayes factors will provide information about the best value of m or about which latent traits impact which observed measures. Interpretation of a single model may then proceed by studying the pattern of λ estimates across latent traits and observed variables; see Hoyle and Duvall (2004); Peeters et al. (2014), and Preacher and MacCallum (2003) for further discussion of related strategies. In addition to these strategies, the models can be extended so that we obtain estimates of experiment-specific parameters. This motivates the use of structural equation models, described later.

Exploratory models and priors In ML contexts, one often fits the model to the observed covariances between the p variables and integrates out the latent variables θ . In Bayesian contexts, one typically fits the model to the raw data, and θ is usually sampled simultaneously with other model parameters via MCMC. This sampling of the latent variables introduces some identification issues that are specific to Bayesian factor models. If these issues are not addressed, the sampled parameters will never converge to a stationary distribution.

To address the identification issues, we adopt a parameter expansion approach (Gelman 2004, 2006, Ghosh & Dunson, 2009) to fit exploratory factor models. Parameter expansion involves sampling from a “working model” with unidentified parameters. At each iteration of the sampling, we transform the sampled parameters to correspond to the “inferential model” with uniquely identified parameters. This improves the speed at which the sampled chains converge to the posterior distribution, while also handling some of the parameter identification issues inherent to the model.

The working model is intended to help us sample the factor loadings λ and the latent trait values θ . We allow the entire covariance matrix Φ^* to be unconstrained, and we also fix $m(m-1)$ parameters in Λ^* to zero. We then place the following prior distributions on the rest of the (unidentified) parameters:

$$\lambda_{jk}^* \sim N(0, 1) \text{ for } \lambda_{jk}^* \text{ free} \quad (7)$$

$$\theta_i^* \sim N_m(\mathbf{0}, \Phi) \quad i = 1, \dots, n \quad (8)$$

$$\Phi^{*-1} \sim \text{Wishart}(\mathbf{V}, m), \quad (9)$$

where \mathbf{V} is an $m \times m$ covariance matrix, and the asterisks on the θ s, λ s, and ϕ s imply that these parameters are unidentified. At each iteration, these unidentified working parameters are transformed to identified versions in the inferential model via

$$\lambda_{jk} = \text{sign}(\lambda_{kk}^*) \lambda_{jk}^* \phi_{kk}^{*1/2} \quad (10)$$

$$\theta_{ik} = \text{sign}(\lambda_{kk}^*) \phi_{kk}^{*-1/2} \theta_{ik}^* \quad (11)$$

$$\phi_{k\ell} = \frac{\text{sign}(\lambda_{kk}^*) \text{sign}(\lambda_{\ell\ell}^*) \phi_{k\ell}^*}{\sqrt{\phi_{kk}^* \phi_{\ell\ell}^*}}, \quad k, \ell = 1, \dots, m, \quad (12)$$

where $\text{sign}(\lambda_{kk}^*)$ equals either 1 or -1 , depending on whether λ_{kk}^* is positive or negative. These transformations automatically yield the remaining constraints necessary to globally identify the model parameters: in addition to the $m(m-1)$ parameters of Λ fixed to zero, we now have that parameters along the diagonal of Φ equal 1. Further, we have restricted one parameter in each column of Λ to be positive, which is one way to fulfill the sign restriction described by Peeters (2012b). This procedure results in

chains of sampled parameters that quickly converge to their posterior distributions for small values of m (say, 5 or less).

For large values of m , the specific λ parameters that are fixed to zero can impact convergence. If these constraints cause the posterior associated with one of the λ_{kk} to overlap with zero, then the chains cannot converge (because, under the parameter expansion approach, λ_{kk} can only be positive). This is only problematic if many of the observed variables are not associated with the latent traits of interest.

Structural equation models

The factor model can be extended in a variety of ways, many of which are relevant to experimental psychologists. For example, suppose that each participant i is assigned to one of two experimental conditions. We can allow unique parameters for each condition, which can provide information about the extent to which the parameters differ across conditions (e.g., Millsap, 2011; Verhagen & Fox, 2013). These types of models are often called *multiple-group models*.

We can also allow specific latent traits to influence other latent traits. For example, in situations where subjects choose between risky and riskless options, subjects' latent attractions to each type of option should influence their choice. Models that include directional influences of latent traits on other latent traits are no longer factor models; they are instead *structural equation models*. Structural equation models subsume factor models and can be written as

$$y_{ij} | \theta_i, \gamma_j, \Lambda_j, \psi_j \sim N(\mu_{ij}, \psi_j) \quad \text{for all } i, j \quad (13)$$

$$\mu_{ij} = \gamma_j + \sum_{k=1}^m \lambda_{jk} \theta_{ik} \quad (14)$$

$$\theta_{ik} = \alpha_k + \sum_{\ell=1}^m B_{k\ell} \theta_{i\ell} + \zeta_{ik} \quad \text{for all } i, k, \quad (15)$$

where α_k is the mean of latent trait k (often fixed to zero) and ζ_{ik} is a normally-distributed, zero-centered residual term associated with latent trait k . Equation 15 may look confusing to some because the θ parameters appear on both sides of the equation. This is the part of the equation that allows some latent traits to have directional influences on others. We require $B_{kk} = 0$, $k = 1, \dots, m$, which simply means that no latent trait exerts a directional influence on itself. Many of the other B parameters are also fixed to zero; the only nonzero B parameters correspond to latent variables that have a directional influence on another latent variable.

Finally, the structural equation model presented above can be further extended to allow for interactions between latent variables (e.g., Klein & Moosbrugger, 2000). To do

so, we maintain most of the previous model, replacing only Eq. 15 with

$$\theta_{ik} = \alpha_k + \sum_{\ell=1}^m B_{k\ell} \theta_{i\ell} + \sum_{q_1=1}^m \sum_{q_2=1}^m \xi_{kq_1q_2} \theta_{iq_1} \theta_{iq_2} + \zeta_{ik} \quad \text{for all } i, k, \quad (16)$$

where ζ_{ik} is still a zero-centered residual and many of the ξ parameters are fixed to zero (as are many of the B parameters). This model is difficult to fit via ML methods and simpler to fit via Bayesian methods (Lee et al., 2007). Bayesian methods allow us to sample the latent variables θ_i , which in turn allow us to deal with the conditional distribution of y_i given θ_i . These conditional distributions are normal, regardless of whether or not we employ latent variable interactions. In contrast, ML methods work on the marginal distribution of y_i , integrating out the latent variable θ_i . When there are latent variable interactions in the model, this marginal distribution is no longer normal. Consequently, special ML estimation methods must be developed specifically for models that include latent variable interactions. For more discussion of these issues and specific ML model estimation methods, see Cudeck et al. (2009), Klein and Moosbrugger (2000), Klein and Muthén (2007), and Marsh et al. (2013).

Identification constraints for SEMs Structural equation models' likelihoods are identified when both the measurement part (14) and the structural part (15) are identified (Lee, 2007). Identification of the measurement part is typically handled in a manner similar to the factor analysis models. There exist no general rules for identifying the structural part, however (Bollen and Davis, 2009). Researchers approach this problem by satisfying some (not unique) sufficient conditions that are relatively easy to satisfy in practice (Lee, 2007). These conditions often focus on *nonrecursive* relationships between latent variables, which include reciprocal directional effects, feedback loops, and correlated residuals. The idea is that, to achieve identification, we should minimize the number of directional influences and covariances between latent variables (i.e., we want to avoid cases where most of the B parameters are estimated and where the covariance matrix associated with ζ is nearly unrestricted). See Rigdon (1995) for further discussion of these issues.

SEMs and priors To specify the priors for structural equation models, we adopt an approach related to that described by Merkle (2011), Curtis (2010), and others. We fix the latent variable variances ϕ to 1 and then, for each factor, restrict at least one loading λ to have a prior distribution that is truncated from below at 0. This truncation solves the

“sign change” issue that was described by Peeters (2012b). We could also use a parameter expansion approach here, but it adds extra complexity to the already-complex JAGS code. Further details on this issue, along with specific prior distributions used, appears in the Application section. For the JAGS code used to fit the models, see the Computational Details section at the end of the paper.

Model comparison: Bayes factor computation

In a latent variable context, Bayes factors are useful for comparing models with differing numbers of latent traits (i.e., models with different values of m) or for comparing structural equation models with different structures. We compute log-Bayes factors to compare pairs of models in this paper, with positive values generally conveying support for more complex models (which contain extra parameters or factors). Kass and Raftery (1995) provide some rules of thumb for their interpretation, suggesting that log-Bayes factors below 1 are “not worth more than a bare mention,” values from 1 to 3 are “positive,” values from 3 to 5 are “strong,” and values greater than 5 are “very strong.” Like most rules of thumb, these values are subject to debate.

To compare models with different values of m , factor analysts have traditionally used the method of path sampling for Bayes factor computation (Lee & Song, 2002; Ghosh & Dunson, 2009). This method was originally described by Meng (1998), but recent work suggests that it can be problematic when it is applied to factor analysis (Dutta and Ghosh, 2013). Hence, we rely on the Laplace approximation to the Bayes factor (Kass & Raftery, 1995; Lewis & Raftery, 1997), and we also use the Savage-Dickey method (Dickey & Lientz, 1970; Wagenmakers et al., 2010) to check our results. We now briefly describe the Laplace approximation and the Savage-Dickey density ratio. Alternative methods for computing Bayes factors are described by, e.g., Kass and Raftery (1995), Lopes and West (2004), Rouder and Morey (2012), and Rouder et al. (2012).

Laplace approximation

The Laplace approximation to the Bayes factor (which we use throughout the paper) is used to approximate the integrals involved in the candidate models' marginal likelihoods (the ratio of which yields the Bayes factor). Given a model, the Laplace approximation to the marginal log-likelihood is (Lewis & Raftery, 1997)

$$\log(f(Y)) \approx \frac{q}{2} \log(2\pi) + \frac{1}{2} \log(|J^{-1*}|) + \log(f(\Omega^*)) + \log(f(Y|\Omega^*)), \quad (17)$$

where Ω is the model parameter vector of length q , Ω^* is a posterior central tendency measure of Ω , J^{-1*} is the inverse of the information matrix evaluated at Ω^* , and $f()$ is the probability distribution of the terms within its parentheses. This approximation has been found to be accurate when the posterior distribution is unimodal without long tails (Lewis and Raftery, 1997; Tierney & Kadane, 1986), and its error in approximating the Bayes factor is of order $O(n^{-1})$ (Kass & Raftery, 1995). Raftery (1993) has previously considered the approximation's application to the types of models described here.

The beauty of the Laplace approximation involves the fact that all four terms on the right of Eq. 17 can often be obtained directly from MCMC output. The first term is trivial. The second term can be estimated via the covariance matrix of sampled parameters. The third term requires Ω^* , which is a posterior central tendency measure of each sampled parameter (and we are evaluating the prior distributions at Ω^*). Finally, the fourth term requires us to evaluate the likelihood at Ω^* .

The fourth term can be somewhat difficult to evaluate when there are random effects in the model that are not officially counted as model parameters; this includes the θ parameters in the factor models described here. As Lewis and Raftery (1997) describe, we must integrate out the random effects in order to calculate the fourth term. This is easy to do for factor models involving normal distributions because we can solve the integral analytically. The solution is:

$$y_i | \gamma, \Lambda, \Phi, \Psi \sim N_p(\gamma + \Lambda \Phi \Lambda' + \Psi), \quad (18)$$

where y_i is individual i 's data vector, γ is a vector of intercepts, Ψ is a diagonal matrix of residual variances, and Λ (the collection of factor loadings) and Φ (the latent trait covariance matrix) were defined previously. For SEMs involving normal distributions the solution is

$$y_i | \gamma, \Lambda, \Phi, \Psi, \alpha, B, \Delta \sim N_p(\gamma + \Lambda \alpha, \Lambda B \Phi B' \Lambda' + \Lambda \Delta \Lambda' + \Psi), \quad (19)$$

where y_i , γ , Λ and Ψ are defined in the same way as Eq. 18; α is a vector of latent trait means; B is a matrix containing the B parameters from Eq. 15; and Δ is a matrix containing the variances of the residual terms ζ_{ik} from Eq. 15. The matrix Φ is still the latent trait covariance matrix, but the vital difference between SEM and factor analysis involves the entries of Φ . In factor analysis, each entry of Φ is either a single parameter or a constraint (0 or 1). In SEMs, we need to derive many entries of Φ via conditional distributions. This is one reason why the Bayesian SEM framework

can be extended more flexibly. For example, for latent variable interactions, we can separate observed variables that “receive” a latent interaction from those that do not, then rely on the conditional distribution of the former observed variables (conditioned on the other latent variables) to arrive at the full marginal distribution. If our models involve distributions other than the normal (as is the case for, say, logistic regressions and many item response models), then integration of the random effects is less straightforward. Further MCMC steps could be used here; see Lewis and Raftery (1997) for more detail.

Savage-dickey density ratio

The Savage-Dickey density ratio is useful for computing Bayes factors associated with nested models; such models often mimic null hypotheses. In such situations, we often wish to test a hypothesis that a model parameter (or some function of model parameters) equals zero. We can gain information related to this hypothesis by comparing two models: Model A, where the parameter in question is freely estimated, and Model B, where the parameter in question is constrained to zero. In computing a Bayes factor to compare these models, we gain information related to the chance that the model parameter actually equals zero.

The Savage-Dickey density ratio allows us to compute these types of Bayes factors by estimating only Model A (where the focal parameter is free). Assume that we wish to calculate the Bayes factor associated with $\omega = 0$ versus $\omega \neq 0$, where ω is a specific entry of the parameter vector Ω . Then we have that

$$BF_{BA} = \frac{p(\omega = 0 | Y)}{p(\omega \neq 0)}; \quad (20)$$

that is, the Bayes factor is the ratio between the posterior density evaluated at $\omega = 0$ and the prior density evaluated at $\omega \neq 0$. The prior density is known, and the posterior density becomes normal for large n (e.g., Gelman, Carlin, Stern, & Rubin, 2004). For smaller n , the posterior density can be approximated via splines or kernel-based methods; see Wagenmakers et al. (2010) for more detail. Regardless, we can evaluate the posterior density using the MCMC output and a minimal amount of extra computation.

Throughout the application below, we report Bayes factors calculated via the Laplace approximation. We also used the Savage-Dickey method to ensure that both methods agreed; while the Savage-Dickey methods are not reported here, the code used to calculate all Bayes factors is included in the replication code. We now move to the application, which can be viewed as the start of a serious data analysis.

Application: Risky choice

Peters and Levin (2008) studied framing effects of risky and riskless choice alternatives in variants of the Asian Disease Problem (Tversky & Kahneman, 1981). This problem requires subjects to choose between two treatments for a fictional disease that impacts 600 people. In a “positive frame” version of the problem, a riskless treatment saves 200 lives, while a risky treatment saves all 600 lives with probability 1/3 and no lives with probability 2/3. In a “negative frame” version of the problem, a riskless treatment causes 400 people to die, while a risky treatment causes 0 deaths with probability 1/3 and 600 deaths with probability 2/3. The question framing (number dead versus number alive) is known to have an impact on choice: “death” phrasing causes people to choose the risky option, while “alive” phrasing causes people to choose the riskless option.

Peters and Levin (2008) were interested in the extent to which each option’s attractiveness influenced choice, along with the impact of numeracy on the relationship between attractiveness and choice. They created five variants of the Asian Disease Problem involving human deaths due to disease, animal deaths due to wildfire, and crop loss due to drought. Subjects were assigned to one of two “frame” conditions, where the problems were phrased as described above. Subjects reported a preference between options on a seven-point scale (from “much prefer the riskless option” to “much prefer the risky option”), and they also rated the attractiveness of each option. Thus, for each problem, subjects gave three ratings on seven-point scales: attraction to the risky option, attraction to the riskless option, and preference. For each subject, the data therefore include 10 attractiveness ratings (of the risky option and the riskless option for each of the five problems) and 5 preference ratings (for each of the five problems). These ratings are all treated as continuous below, though similar models could be specified if, e.g., choice were binary.

The original authors’ focal hypothesis was that subjects’ numeracy would moderate the relationship between attractiveness ratings and choice: as numeracy increases, attractiveness ratings should be more predictive of choice. To study this hypothesis, the authors employed separate ANOVAs on each of the five problem variants and on data that were averaged over the problems. They generally found statistically significant interaction effects in agreement with their focal hypothesis. In the analysis described here, we employ factor models and structural equation models to simultaneously study the experimental data across all five problems and to handle potential differences between problems.

Methods

We propose the use of factor analysis and SEM to study the authors’ original hypotheses. This allows us to draw unified inferences about the full experimental data from a single model, removing the need to do separate analyses of each problem. While the modeling here can be viewed as the start of a full analysis, we omit some steps and issues due to space. We further describe these omissions in the Discussion section.

Our modeling proceeds in two steps. First, we fit 1- and 2-factor models (i.e., Eqs. 4 to 6, with $m = 1$ and $m = 2$) to the ten attractiveness ratings given by each subject (one risky rating and one riskless rating, for each of the five questions). The factors (latent variables) represent participants’ attractions to risky and riskless alternatives in the context of this experimental study; they are used here to pool information across all five problems. In comparing a one-factor model to a two-factor model, we obtain information about the nature of attraction to risky and riskless alternatives: are these two extremes of a continuum, or are they unique dimensions? The former implies that a person who is attracted to risky alternatives is also repelled by riskless alternatives, while the latter implies that a person can be simultaneously attracted to (or repelled by) both risky and riskless alternatives. This comparison provides information about the best way to model the effect of attractiveness ratings on choice. To preview our results, we find that the 2-factor model is better than the 1-factor model.

Next, we build off the results of the first step to examine the effects of risky/riskless attraction on choice, as well as to draw inferences related to the experimental manipulations and to numeracy. The model is most easily conceptualized via the path diagram in Fig. 1. This diagram illustrates a series of regression-like relationships (directed arrows) between observed variables (squares) and latent traits (circles). The observed variables AR1 to AR5 represent riskless attraction ratings for items 1–5; the observed variables AR6 to AR10 represent risky attraction ratings for items 6–10; and the observed variables C1 to C5 represent choice ratings for items 1–5. Double-headed arrows imply variance parameters.

The base portion of the model involves three main latent variables: a riskless variable representing the riskless attraction ratings, a risky variable representing the risky attraction ratings, and a choice variable representing the choice ratings. The experimental manipulation (frame), numeracy, and a frame-by-numeracy interaction predict all three of the main latent variables. Finally, there appear two latent interaction variables: one between riskless attraction and numeracy, and one between risky attraction and numeracy. These are notable because (i) they help us study the original

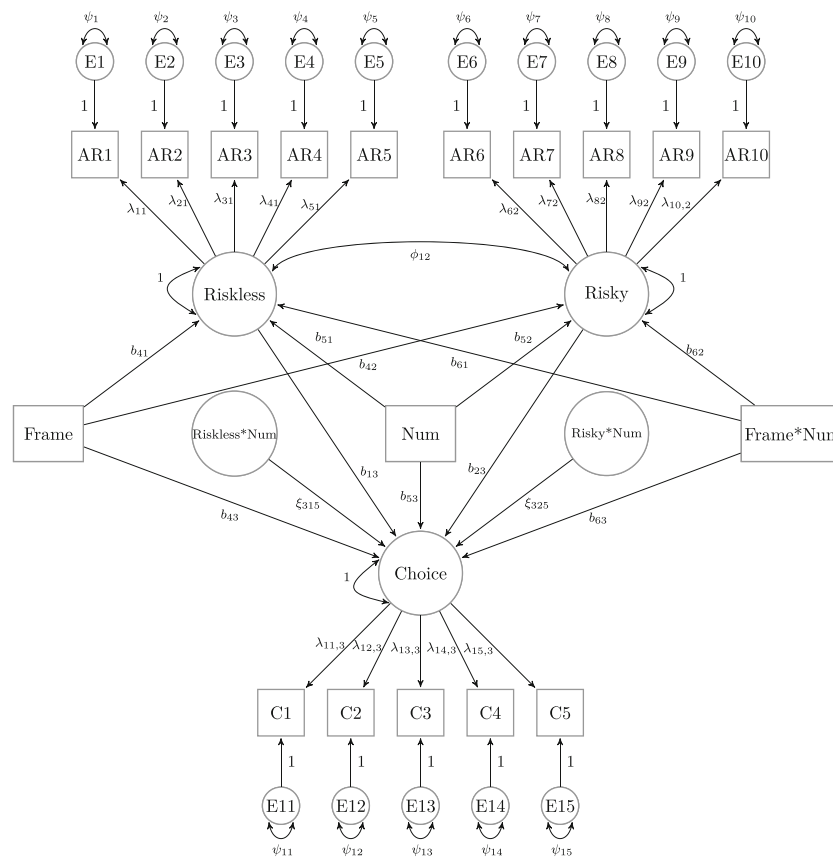


Fig. 1 Path diagram of the structural equation model used in the application

authors' focal hypotheses, and (ii) interactions involving latent variables are difficult to estimate via ML methods.

All observed variables are assumed to be conditionally normal, so that the path diagram can also be written as a combination of Eqs. 13, 14, and 16. The parameter subscripts in Fig. 1 match the subscripts that would be used in the equation version of the model (not shown). Additionally, to obtain the interactions from Eq. 16, we implicitly create a latent numeracy variable that is perfectly related to the observed numeracy variable.

Prior distributions on model parameters are generally taken to be non-informative. For the factor analysis models, many of the prior distributions are listed in Eqs. 7 to 9. The remaining parameters include the intercepts associated with observed variables (γ_j) and residual variances (ψ_j); these classes of parameters are assigned priors of

$$\gamma_j \sim N(0, 10^3) \quad (21)$$

$$\psi_j \sim \text{Inv-Gamma}(0.1, 0.1), \quad (22)$$

where $j = 1, 2, \dots, 15$. The γ_j parameters are given exceptionally-large variances because they are essentially nuisance parameters here. They represent the mean of each observed variable and play little role in the parameters of

interest, which include the λ_{jk} , ϕ_{12} , and ψ_j parameters, where $j = 1, 2, \dots, 15$; $k = 1, \dots, 6$. For completeness, note that the variances of the three observed covariates (labeled “Frame,” “Num,” and “Frame by Num”) are fixed to their sample estimates.

For the structural equation model in Fig. 1, the $b_{k\ell}$ and $\xi_{kq_1q_2}$ parameters (paths between latent variables) can be somewhat difficult to sample. This is because, for these parameters, the chains sampled via JAGS (and related software) exhibit high autocorrelation. When the prior variances are very large, the MCMC algorithm can accept a “bad” posterior value, and the high autocorrelation implies that it will take a long time for the chain to recover from that bad value. As a result, we use “weakly informative” prior distributions on the parameters (Gelman, 2006): prior distributions with variances constrained to reflect plausible parameter values. For example, because the latent variables in Fig. 1 are constrained to have means of 0 and variances of 1 (and because we know that, for the data considered here, no latent variable will be perfectly predictive of another), we can be nearly certain that the λ_{jk} , $b_{k\ell}$, and $\xi_{kq_1q_2}$ parameters are between -10 and 10 . Because 99 % of the normal distribution lies within 3 standard deviations of the mean (and because we expect the parameters to be closer to zero than to -10 or

10), a normal distribution with mean 0 and variance 10 is mildly informative for these parameters. Note that we are supplying no information about the parameters' signs; the prior distributions are always symmetric around zero. The only mild information we provide is related to the expected variability of the parameters.

In Fig. 1, there are five main types of parameters: λ_{jk} , $b_{k\ell}$, $\xi_{kq_1q_2}$, ψ_j , and ϕ_{12} . There also exist observed-variable intercepts γ_j that are not displayed. The prior distributions for these classes of parameters are

$$\gamma_j \sim N(0, 10^3) \quad (23)$$

$$\lambda_{jk} \sim \text{TN}_{\mathbb{R}^+}(0, 10) \quad (24)$$

$$b_{k\ell} \sim N(0, 10) \quad (25)$$

$$\xi_{kq_1q_2} \sim N(0, 10) \quad (26)$$

$$\psi_j \sim \text{Inv-Gamma}(0.1, 0.1) \quad (27)$$

$$\phi_{12} \sim \text{Unif}(-1, 1), \quad (28)$$

where $\text{TN}_{\mathbb{R}^+}$ is a normal distribution truncated from below at 0. The truncated normals are used here as a shortcut to ensure that the sign of the loadings λ_{jk} and the signs of the latent variables θ_i do not switch on one another. The shortcut can be safely used here because we know the loadings all have the same sign and are far from zero. If this were not the case, we could instead estimate each latent variable's variance, fixing a single λ parameter to 1 for each latent variable. We could also adopt a parameter expansion approach, similar to that described for the factor analysis model.

All models are sampled for 5,000 iterations following an adaptation/burn-in of 6,000 iterations. Model convergence is assessed via time series plots and the Gelman-Rubin statistic (Gelman & Rubin, 1992); for the prior distributions outlined above, all parameters achieve values of the Gelman-Rubin statistic less than 1.1.

Results

Correlations between the 15 observed ratings (riskless attractiveness, risky attractiveness, choice) are presented in Fig. 2 (other descriptive statistics can be found in the original authors' paper). Correlations for the negative frame condition are displayed above the main diagonal, and correlations for the positive frame condition are displayed below the main diagonal. Blue shadings imply negative correlations, and red shadings imply positive correlations. It is seen that the riskless ratings (AR1–AR5) generally display large correlations with one another, as do the risky ratings (AR6–AR10). Correlations between the riskless ratings and risky ratings are also generally positive: for a given scenario, subjects generally found the risky and riskless options to be more or less appealing. This may be because the

options had the same expected value, and some subjects found the expected value to be more appealing than others. Correlations between the choice ratings (C1–C5) and the attractiveness ratings are generally smaller, and they generally differ in sign by type of option: the riskless ratings are generally negatively related to choice ratings, while the risky ratings are generally positively related. The original authors' focal hypotheses involve moderation of these latter correlations by numeracy.

Model estimation was on the order of minutes (on our computers, two minutes or less), as opposed to seconds or hours. We build up to the focal hypotheses by first comparing two exploratory factor models: a one-factor model and a two-factor model (i.e., a model with $m = 1$ versus a model with $m = 2$). As stated previously, this comparison provides information about the nature of attraction to risky and riskless alternatives: are these attractions two extremes of a single dimension, or are they unique dimensions? In employing the Laplace approximation to compare models, we obtain an estimated log-Bayes factor of 29.02 in favor of the two-factor model. This implies that attraction to risky options and attraction to riskless options should be treated as separate dimensions. Table 1 displays the posterior mean factor loadings (i.e., the estimated λ s) and reinforces this result. The table shows that the riskless alternatives (labeled AR1 to AR5) all have large, positive loadings on the first factor and near-zero loadings on the second factor. The risky alternatives (labeled AR6 to AR10) exhibit the converse result, with the estimated correlation between factors being 0.42. The loadings that equal exactly zero (i.e., that equal 0 instead of 0.00) reflect identification constraints for the exploratory model; these constraints led to factor loadings that were easy to interpret, without the need for further rotation. Based on these results, we maintain a 2-factor confirmatory model for the remainder of the analyses. This model fixes to zero the loadings that are already near zero in Table 1: the risky alternatives on the "Riskless" factor, and the riskless alternatives on the "Risky" factor.

We are now prepared to additionally model the impact of numeracy on attraction, along with the impacts of frame, numeracy, and attraction on choice via the model displayed in Fig. 1. The results are displayed in the rightmost columns of Table 2, including 95 % posterior intervals and log-Bayes factors associated with different parameters.

There is a small effect of frame on riskless attraction, whereby people were more attracted to the riskless option in the positive frame condition (the condition where options were phrased in terms of "number alive"). The posterior interval associated with this effect hovered around zero, and the log-Bayes factor favored a model where this effect equaled zero. We also modeled effects of numeracy

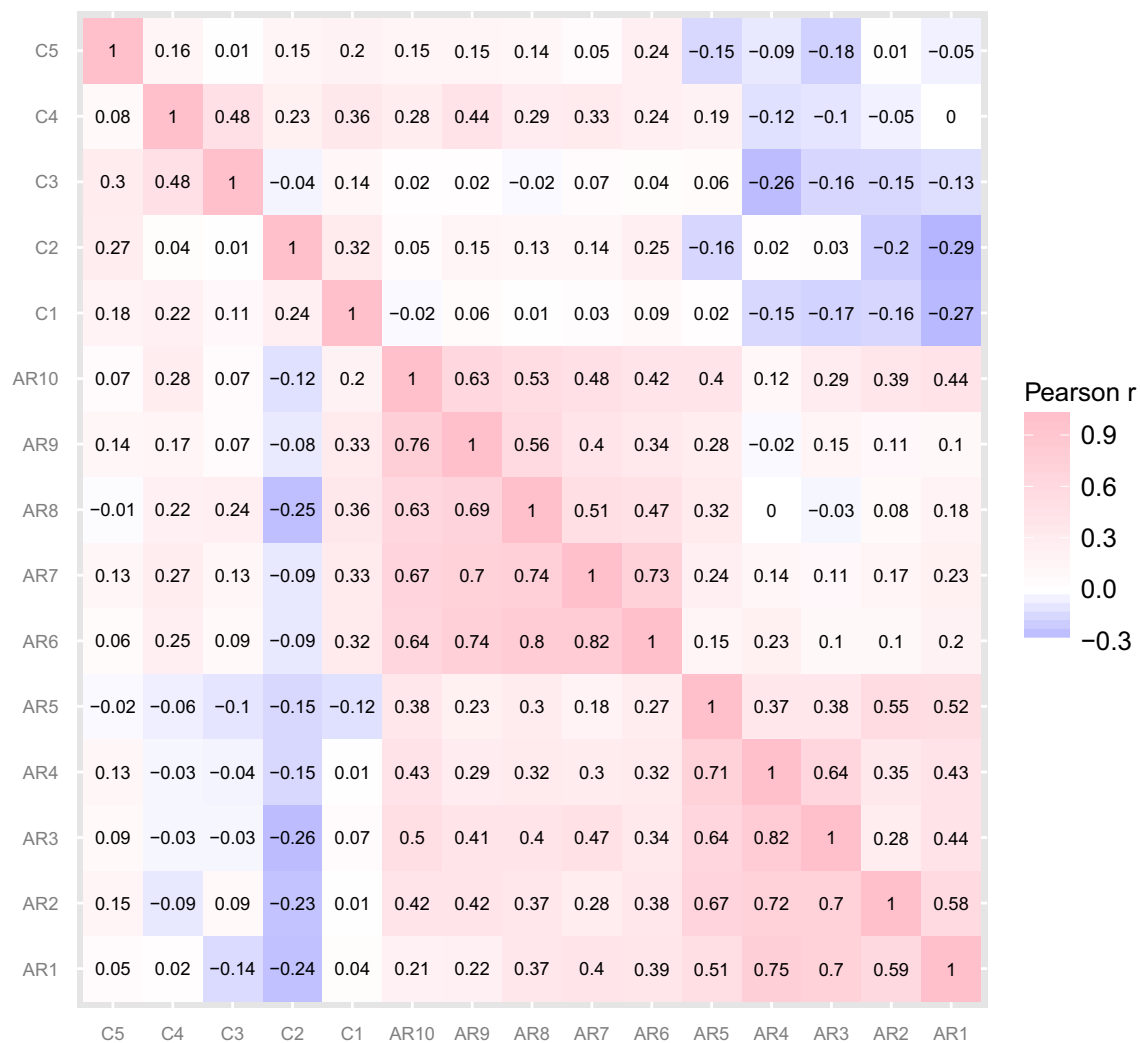


Fig. 2 Correlation matrix of data from Peters & Levin (2008). Correlations for the positive frame condition are below the main diagonal; correlations for the negative frame condition are above the main diagonal. Variables include riskless attraction rating on items 1–5

(AR1–AR5); risky attraction rating on items 1–5 (AR6–AR10); and choice rating on items 1–5 (C1–C5). Red values signify correlations closer to 1, and blue values signify negative correlations

and numeracy-by-frame interactions on the attraction latent variables, but these effects were unsupported by both the posterior intervals and Bayes factors.

We now consider effects of frame, numeracy, and the latent attraction variables on choice. We found the standard framing effect on choice, whereby the positive frame condition led people to prefer the riskless option. The log-Bayes

factor was around 5, with the posterior interval excluding zero. There was also a small numeracy \times frame interaction: numerate subjects were less impacted by question framing as compared to less-numerate subjects. This interaction had a posterior interval that hovered around zero, with the log-Bayes factor mildly preferring a model without the interaction.

Table 1 Estimated factor loadings from the exploratory model with $m = 2$

	AR1	AR2	AR3	AR4	AR5	AR6	AR7	AR8	AR9	AR10
Riskless	1.19	1.15	1.19	1.28	1.13	0	0.05	−0.03	−0.02	0.30
Risky	0	−0.05	0.00	−0.17	−0.06	1.20	1.14	1.12	1.11	0.98

AR1 to AR5 represent attraction to the riskless options for items 1 to 5, and AR6 to AR10 represent attraction to the risky options. Rows correspond to the two factors, labeled “Riskless” and “Risky”

Table 2 Comparison of original (general linear model) results, 95 % posterior intervals from the factor model, and log(Bayes factors) from the structural equation model

Effect	Dependent Variable	Original	Posterior interval	Log(BF)
Frame	Attraction (riskless)	$F(1, 106) = 5.2, p < .05$	(0.02, 0.81)	−0.87
Frame	Attraction (risky)	Unreported (not significant)	(−0.58, 0.19)	−2.07
Numeracy	Attraction (riskless)	$F(1, 104) = 10.8, p < .01$	(−0.07, 0.32)	−3.41
Numeracy	Attraction (risky)	Unreported (not significant)	(−0.16, 0.23)	−4.31
Numeracy × Frame	Attraction (riskless)	$F(1, 104) = 3.9, p = .05$	(−0.5, 0.07)	−2.45
Numeracy × Frame	Attraction (risky)	Unreported (not significant)	(−0.33, 0.22)	−4.29
Frame	Choice	$F(1, 104) = 18.9, p < .001$	(−3.04, −0.68)	5.08
Numeracy	Choice	$F(1, 104) = .04, p = .85$	(−0.53, 0.22)	−2.69
Numeracy × Frame	Choice	$F(1, 104) = 1.8, p = .28$	(0, 1.23)	−0.8
Attraction (riskless)	Choice	$F(1, 100) = 6.76, p < .01$	(−1.46, −0.2)	1.97
Attraction (risky)	Choice	$F(1, 100) = 7.29, p < .01$	(0.36, 1.85)	5.15
Numeracy × Attraction (riskless)	Choice	$F(1, 100) = 9.6, p = .002$	(−0.92, −0.12)	1.51
Numeracy × Attraction (risky)	Choice	$F(1, 100) = 4.4, p = .04$	(−0.09, 0.59)	−2.79

Negative log(Bayes factor) implies that the “no-effect” model is preferred. Some of the original results have been converted from *t* statistics to *F* statistics for uniformity

Finally, to address the focal hypothesis, attraction ratings were positively related to choice ratings, more so for numerate subjects. First, increases in riskless attraction were associated with preferences for the riskless option, and increases in risky attraction were associated with preferences for the risky option. The riskless attraction latent variable had a weaker association with choice than did risky attraction, with log-Bayes factors of 1.97 and 5.15, respectively. As Fig. 1 shows, we also included numeracy × riskless attraction and numeracy × risky attraction interactions. The interaction term involving riskless attraction had a posterior interval that did not overlap with zero, while the interaction term involving risky attraction had a posterior interval that did overlap with zero. The log-Bayes factor favors the model that includes the riskless interaction (1.51) but not the risky interaction (−2.79). This suggests a small numeracy × riskless attraction association with choice, similar to the original authors’ hypotheses.

Discussion

We now discuss some general issues associated with the results in Table 2, along with some limitations of the current analyses.

Prior sensitivity and comparison Focusing on results in Table 2, the magnitude of the Bayes factors is impacted by the prior distributions used (for the model here, prior distributions with smaller variances generally produce larger Bayes factors). This property of Bayes factors is known, and it has been the subject of both praise and criticism (Liu & Aitkin, 2008; Vanpaemel, 2010). The fact that Bayes factors are sensitive to choice of prior distribution is helpful in that they summarize the data while also accounting for our previous knowledge of the phenomenon under study. Others point out that, if one needs to surpass a Bayes factor threshold for a journal publication (similar to $p < .05$),

then it is disconcerting that the threshold can be met by manipulating the prior distribution. To address this issue, one can fit the model under multiple prior distributions to examine the results' sensitivities.

We can further use Table 2 to compare results arising from the Bayesian models to those of the original authors. The third column of the table displays the original authors' ANOVA-based results. Some of these results are not based on exactly the same data as our results, because the original authors sometimes focused on three of the five scenarios and sometimes aggregated data (whereas we consistently include all five scenarios). Nevertheless, we see that the posterior intervals and Bayes factors often agree with the ANOVA results, in that intervals fail to overlap with zero and log-Bayes factors are positive when p -values are small. When there are disagreements, the Bayes factors and posterior intervals appear to be more conservative than the other measures.

There are also a few situations where the posterior intervals fail to overlap with zero, yet the Bayes factors favor a model that excludes the effect. These different conclusions stemming from posterior intervals versus Bayes factors arise from their different intents: posterior intervals summarize the value of a single model parameter, while Bayes factors signify whether or not the associated parameter(s) generally improve our model. While the former are readily interpreted and may be more familiar to researchers (due to the similarity between posterior intervals and confidence intervals), the latter seems more useful for summarizing the extent to which a parameter is generally important to one's theory (somewhat similarly to a frequentist likelihood ratio test). Bayes factors additionally account for the complexity afforded to us by extra model parameters, which is relevant here because some types of SEM parameters may afford us more complexity than other types. For further discussion of Bayes factors and of model complexity, see Jefferys and Berger (1992) and Spiegelhalter and Smith (1982).

Limitations If the focus of this paper were risky choice (as opposed to Bayesian latent variable models), further model extension and study would be warranted. The model used here posits three separate latent variables for risky attraction, riskless attraction, and choice. Conditional on these latent variables, the observed variables are posited to be independent. This assumption is likely to be incorrect for observed variables stemming from the same problem. That is, even after conditioning on the latent variables, there may remain a correlation between the three observed variables associated with any one problem (e.g., the variables labeled AR1, AR6, and C1 in Figure 1). To account for these issues, we can allow the residual terms associated with these three variables (E1, E6, and E11) to be correlated. The introduction of residual correlations poses a difficult problem for

Bayesian SEM estimation (Barnard et al., 2000; Chib & Greenberg, 1998; Palomo et al., 2007), though model estimation is possible (e.g., Muthén & Asparouhov, 2012) and will become easier in the future (e.g., Merkle & Rosseel 2016).

Aside from residual correlations, several issues deserve more attention than they received here. First, it would be useful to include more descriptive statistics that supplement the model. Peters and Levin (2008) provided many tables that are useful along these lines. Second, it would be useful to study the absolute fits of the models in addition to the relative fits. This may involve posterior predictive checks (e.g., Gelman et al., 2004) or calculation of a Bayes factor comparing the proposed models to saturated models (i.e., multivariate normal models with a free parameter for each mean and covariance). Finally, the statistical issue of suppression may be considered in more detail. The SEM from Fig. 1 involves two positively-related latent variables predicting a third latent variable with opposite signs. This can lead to instability in the associated regression weights and standard errors. While the signs of the regression weights involving latent variables are the same as the majority of relevant, observed-variable correlations in Fig. 2, some instability may still exist.

In the General Discussion below, we provide detail about further model extensions, uses, and needs.

General discussion

In this paper, we have demonstrated the utility of applying Bayesian latent variable models to multivariate experimental psychology data. We first provided background on factor analysis models, structural equation models, and issues specifically related to Bayesian estimation of these models. We then considered some model extensions and applied them to data from a decision making experiment on risky choice. The models allowed us to draw conclusions across multiple stimuli in a unified manner, removing the need for data aggregation and for separate, stimulus-specific analyses. Bayesian methods further allowed us to easily estimate models with latent variable interactions, which are difficult to estimate via ML methods. In the paragraphs below, we address additional Bayesian model extensions and needs. These include general comparisons to ML models, considerations of exploratory factor analysis, and cognitive psychometric models.

Bayesian versus ML Advantages of the Bayesian approach to structural equation modeling include easy extension to complex situations, along with non-asymptotic estimates of the variability in parameter estimates. Expanding on the former point, the methods described here were easily

extended to handle latent variable interactions, and they are also easily extended to handle models with both continuous and ordinal observed variables. In particular, Song and Lee (2012b) provide a tutorial on Bayesian estimation of these types of models, with further details in their book (Song & Lee, 2012a). These methods are very powerful because they provide a single framework for estimating many structural equation models that a researcher may conceptualize. This framework is not necessarily the best for building an intuition of structural equation models, however, because it is so general. Researchers' desired models can often be written in a manner that is more concise than the general framework.

Disadvantages of Bayesian structural equation models, as compared to ML models, lie in model specification and estimation. Model specification in, e.g., JAGS is very different from model specification in traditional ML software like Mplus or LISREL, so that users coming from these traditions will find that their prior software experience is not terribly helpful. Relatedly, Bayesian methodology has more “moving parts,” including prior distribution specification and assessments of model convergence. These represent new topics for users of ML software, but we are optimistic that users can master the new topics.

Bayesian exploratory factor analysis Bayesian methods offer novel ways to estimate exploratory factor models. In this paper, we always fixed specific parameters to zero in order to identify model parameters. If the factor model is exploratory, we could then rotate model parameters' (the λ s) posterior means to have a more-interpretable structure. This is a two-step process, just like the ML case: we first estimate the model, and we then rotate parameters. Under the Bayesian approach, we could simultaneously rotate and sample the factor loading matrices at each iteration of the MCMC algorithm, which may lead to more-interpretable solutions than the ML algorithm (which can only use the single set of factor loadings arising from the ML estimates).

Related methods have been the focus of recent work. In particular, Peeters (2012a) describes a method by which factor models with multiple values of m are simultaneously estimated, rotated, and compared to one another via marginal likelihoods (the building blocks of Bayes factors). Conti et al. (2014) propose a method whereby the value of m is allowed to change from iteration to iteration, with the possibility that some latent traits are associated with no observed variables. Importantly, the latter authors assume that every observed variable has a nonzero loading associated with only one latent trait; this assumption is restrictive and results in a model that, using the definitions in this paper, would be called “confirmatory” instead of “exploratory.” Finally, Ročková and George (2014) propose an alternative method whereby m is set to a large value, and “spike-and-slab” priors are used to enforce simplicity

of the Λ matrix. These prior distributions fix weak loadings to zero, so that smaller values of m are obtained by fixing to zero all loadings associated with a latent trait. This method is computationally difficult but appears promising for future application.

Cognitive psychometric models The methods described here are also highly related to the development of *cognitive* latent variable models (e.g., Nunez, Srinivasan, & Vandekerckhove, 2015; Turner et al., 2013; Turner, van Maanen, & Forstmann, 2015; Vandekerckhove, 2014), where the latent variables are used to tie multiple types of data (response time, BOLD response, survey data, etc) together in a single model. For example, in Vandekerckhove's application, a diffusion model was used to describe response times from an executive functioning task. Participants also completed two depression scales, with the latent variables in the model predicting both the depression scales and diffusion model parameters. This allowed for novel information about specific aspects of performance on the response time task that are related to the depression scales. Cognitive psychometric models generally offer new avenues for combining psychometric knowledge with cognitive modeling (also see, e.g., Tuerlinckx & De Boeck, 2005; van der Maas, Molenaar, Maris, Kievit, & Borsboom, 2011). The model extensions mentioned here, coupled with computational advances in model estimation and increased communication between experimental psychologists and psychometricians, yield fruitful prospects for the use of Bayesian latent variable models in cognitive science. We encourage readers to explore these prospective models and the novel inferences that they can provide.

Computational details

All results were obtained using the R system for statistical computing (R Development Core Team, 2014) version 3.2.3 and JAGS software for Bayesian computation version 3.4.0, employing the helper package *rjags* 4-5 (Plummer, 2014). R and the package *rjags* are freely available under the General Public License 2 from the Comprehensive R Archive Network at <http://CRAN.R-project.org/>. JAGS is freely available under the General Public License 2 from Sourceforge at <http://mcmc-jags.sourceforge.net/>. R and JAGS code for replication of our results is available at <http://semttools.R-Forge.R-project.org/>.

References

- Bagozzi, R. P., & Yi, Y. (1989). On the use of structural equation models in experimental designs. *Journal of Marketing Research*, 26, 271–284.

- Barnard, J., McCulloch, R., & Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10, 1281–1311.
- Bartholomew, D. J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology*, 34, 93–99.
- Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach*, 3rd. Wiley.
- Bollen, K. A., & Davis, W. R. (2009). Two rules of identification for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 523–536.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36, 111–150.
- Chib, S., & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, 85, 347–361.
- Conti, G., Frühwirth-Schnatter, S., Heckman, J. J., & Piatek, R. (2014). Bayesian exploratory factor analysis. *Journal of Econometrics*, 183, 31–57.
- Cudeck, R., Harring, J. R., & du Toit, S. H. C. (2009). Marginal maximum likelihood estimation of a latent variable model with interaction. *Journal of Educational and Behavioral Statistics*, 34, 131–144.
- Curtis, S. M. (2010). BUGS code for item response theory. *Journal of Statistical Software*, 36, 1–34.
- Dickey, J. M., & Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *The Annals of Mathematical Statistics*, 41, 214–226.
- Donaldson, G. W. (2003). General linear contrasts on latent variable means: Structural equation hypothesis tests for multivariate clinical trials. *Statistics in Medicine*, 22, 2893–2917.
- Dutta, R., & Ghosh, J. K. (2013). Bayes model selection with path sampling: Factor models and other examples. *Statistical Science*, 28, 95–115.
- Gelman, A. (2004). Parameterization and Bayesian modeling. *Journal of the American Statistical Association*, 99, 537–545.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 3, 515–534.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* 2nd edn. London: Chapman Hall.
- Meng, X. L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, 13, 163–185.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7, 457–511.
- Ghosh, J., & Dunson, D. B. (2009). Default prior distributions and efficient posterior computation in Bayesian factor analysis. *Journal of Computational and Graphical Statistics*, 18, 306–320.
- Hoyle, R. H., & Duvall, J. L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis. In Kaplan, D. (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 301–316). Thousand Oaks, CA: Sage.
- Jefferys, W., & Berger, J. O. (1992). Ockham's razor and Bayesian analysis. *American Scientist*, 80, 64–72.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202.
- Jöreskog, K. G. (1979). Author's addendum. In Magidson, J. (Ed.), *Advances in factor analysis and structural equation models* (pp. 40–43). Cambridge: Abt Books.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795.
- Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65, 457–474.
- Klein, A., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, 42, 647–673.
- Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. New York: American Elsevier Publishing Co.
- 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*. Chichester: Wiley.
- 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.
- Lewis, S. M., & Raftery, A. E. (1997). Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association*, 92, 648–655.
- Liu, C. C., & Aitkin, M. (2008). Bayes factors: Prior sensitivity and model generalizability. *Journal of Mathematical Psychology*, 52, 362–375.
- Lopes, H. F., & West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, 14, 41–67.
- 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.
- Marsh, H. W., Wen, Z., Hau, K. T., & Nagengast, B. (2013). Structural equation models of latent interaction and quadratic effects. In Hancock, G. R., & Mueller, R. O. (Eds.), *Structural equation modeling: A second course*, 2nd edn. (pp. 267–308). Information Age Publishing.
- Martin, J. K., & McDonald, R. P. (1975). Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika*, 40, 505–517.
- Merkle, E. C. (2011). A comparison of imputation methods for Bayesian factor analysis models. *Journal of Educational and Behavioral Statistics*, 36, 257–276.
- Merkle, E. C., & Rosseel, Y. (2016). *blavaan: Bayesian structural equation models via parameter expansion*. Manuscript under review.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howarter, A. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.
- 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). *Mplus user's guide*, 7th edn. Los Angeles, CA: Muthén & Muthén.
- Nunez, M. D., Srinivasan, R., & Vandekerckhove, J. (2015). Individual differences in attention influence perceptual decision making. *Frontiers in Psychology*, 8(18), 1–13.
- Palomo, J., Dunson, D. B., & Bollen, K. (2007). Bayesian structural equation modeling. In Lee, S. Y. (Ed.), *Handbook of latent variable and related models* (pp. 163–188). Elsevier.
- Peeters, C. F. W. (2012a). Bayesian exploratory and confirmatory factor analysis: Perspectives on constrained-model selection, Utrecht University.
- Peeters, C. F. W. (2012b). Rotational uniqueness conditions under oblique factor correlation metric. *Psychometrika*, 77, 288–292.

- Peeters, C. F. W., Dziura, J., & van Wesel, F. (2014). Pathophysiological domains underlying the metabolic syndrome: An alternative factor analytic strategy. *Annals of Epidemiology*, 24, 762–770.
- Peters, E., & Levin, I. P. (2008). Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. *Judgment and Decision Making*, 3, 435–448.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Hornik, K., Leisch, F., & Zeileis, A. (Eds.), *Proceedings of the 3rd international workshop on distributed statistical computing*.
- Plummer, M. (2014). rjags: Bayesian graphical models using MCMC. <http://CRAN.R-project.org/package=rjags>. R package version 3-13.
- Preacher, K. J., & MacCallum, R. C. (2003). Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2, 13–43.
- Press, S. J. (1972). *Applied multivariate analysis: Using Bayesian and frequentist methods of inference*. New York: Holt, Rinehart, and Winston.
- R Development Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: <http://www.R-project.org>. ISBN 3-900051-07-0.
- Raftery, A. E. (1993). Bayesian model selection in structural equation models. In Bollen, K. A., & Long, J. S. (Eds.), *Testing structural equation models* (pp. 163–180). Beverly Hills, CA: Sage.
- Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research*, 30(3), 359–383.
- Ročková, V., & George, E. I. (2014). Fast Bayesian factor analysis via automatic rotations to sparsity. Obtained Jan 8, 2015 from <http://www.ssc.upenn.edu/~fdiebold/papers/misc/RockovaAndGeorge2014FactorAnalysis.pdf>.
- Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, 47, 877–903.
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56, 356–374.
- Russell, D. W., Kahn, J. H., Spoth, R., & Altmaier, E. M. (1998). Analyzing data from experimental studies: A latent variable structural equation modeling approach. *Journal of Counseling Psychology*, 45, 18–29.
- Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, 64, 37–52.
- Song, X. Y., & Lee, S. Y. (2012a). *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences*. Chichester, England: Wiley.
- Song, X. Y., & Lee, S. Y. (2012b). A tutorial on the Bayesian approach for analyzing structural equation models. *Journal of Mathematical Psychology*, 56, 135–148.
- Spiegelhalter, D. J., & Smith, A. F. M. (1982). Bayes factors for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society B*, 44, 377–387. *Journal of the Royal Statistical Society B*, 44, 377–387.
- Stan Development Team (2014). *Stan Modeling Language Users Guide and Reference Manual*, Version 2.5.0. <http://mc-stan.org/>.
- Steiger, J. H. (2001). Driving fast in reverse. *Journal of the American Statistical Association*, 96, 331–338.
- Stromeyer, W. R., Miller, J. W., Sriramachandramurthy, R., & DeMartino, R. (2014). The prowess and pitfalls of Bayesian structural equation modeling: Important considerations for management research. *Journal of Management*.
- Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81, 82–86.
- Tuerlinckx, F., & De Boeck, P. (2005). Two interpretations of the discrimination parameter. *Psychometrika*, 70, 629–650.
- Turner, B. M., Forstmann, B. U., Wagenmakers, E. J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage*, 72, 193–206.
- Turner, B. M., van Maanen, L., & Forstmann, B. U. (2015). Informing cognitive abstractions through neuroimaging: The neural drift diffusion model. *Psychological Review*, 122, 312–336.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–458.
- van der Maas, H. L. J., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological Review*, 118, 339–356.
- Vandekerckhove, J. (2014). A cognitive latent variable model for the simultaneous analysis of behavioral and personality data. *Journal of Mathematical Psychology*, 60, 58–71.
- Vanpaemel, W. (2010). Prior sensitivity in theory testing: An apologia for the Bayes factor. *Journal of Mathematical Psychology*, 54, 491–498.
- Verhagen, A. J., & Fox, J. P. (2013). Bayesian tests of measurement invariance. *British Journal of Mathematical and Statistical Psychology*, 66, 383–401.
- Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*, 60, 158–189.
- Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology*, 89(5), 696–716.