Psychological Methods

Prior Sensitivity Analysis in Default Bayesian Structural Equation Modeling

Sara van Erp, Joris Mulder, and Daniel L. Oberski

Online First Publication, November 27, 2017. http://dx.doi.org/10.1037/met0000162

Psychological Methods

1082-989X/17/$12.00 http://dx.doi.org/10.1037/met0000162

Prior Sensitivity Analysis in Default Bayesian Structural Equation Modeling

Sara van Erp and Joris Mulder
Tilburg University

Daniel L. Oberski
Utrecht University

Abstract

Bayesian structural equation modeling (BSEM) has recently gained popularity because it enables researchers to fit complex models and solve some of the issues often encountered in classical maximum likelihood estimation, such as nonconvergence and inadmissible solutions. An important component of any Bayesian analysis is the prior distribution of the unknown model parameters. Often, researchers rely on default priors, which are constructed in an automatic fashion without requiring substantive prior information. However, the prior can have a serious influence on the estimation of the model parameters, which affects the mean squared error, bias, coverage rates, and quantiles of the estimates. In this article, we investigate the performance of three different default priors: noninformative improper priors, vague proper priors, and empirical Bayes priors—with the latter being novel in the BSEM literature. Based on a simulation study, we find that these three default BSEM methods may perform very differently, especially with small samples. A careful prior sensitivity analysis is therefore needed when performing a default BSEM analysis. For this purpose, we provide a practical step-by-step guide for practitioners to conducting a prior sensitivity analysis in default BSEM. Our recommendations are illustrated using a well-known case study from the structural equation modeling literature, and all code for conducting the prior sensitivity analysis is available in the online supplemental materials.

Translational Abstract

Psychologists and social scientists often ask complex questions regarding group- and individual differences and how these change over time. To answer these questions, researchers generally use structural equation modeling (SEM), a general framework to fit complex models. Traditionally, structural equation models are estimated using maximum likelihood estimation, which uses only the data at hand. An alternative approach is Bayesian SEM, which is becoming increasingly popular because it can solve several problems of maximum likelihood estimation. Bayesian SEM combines the data at hand with a prior distribution of the parameters in the model. This prior distribution can be based on subjective beliefs or previous research. In practice, however, researchers tend to use noninformative "default" priors which are constructed in an automatic fashion without including any substantive prior information. Different default priors can be used for this purpose. Through a simulation study, we show that the exact choice of the default prior can have a serious influence on the estimation of the model parameters, especially when the sample size is small. Because of this finding, we recommend researchers who use default Bayesian SEM to always perform a prior sensitivity analysis to determine how robust the conclusions are across the various analyses. We provide a practical step-by-step guide on how to conduct a prior sensitivity analysis and we illustrate our recommendations with a practical application.

Keywords: Bayesian, structural equation models, default priors, sensitivity analysis

Supplemental materials: http://dx.doi.org/10.1037/met0000162.supp

Psychologists and social scientists often ask complex questions regarding group and individual differences and how these change over time. These complex questions necessitate complex methods such as structural equation modeling (SEM); its Bayesian version (Bayesian structural equation modeling; BSEM), in particular, has recently gained popularity (e.g., Kaplan, 2014) because it potentially resolves some of the difficulties with traditional frequentist SEM. For example, frequentist estimation of multilevel SEMs—

Sara van Erp and Joris Mulder, Department of Methodology and Statistics, Tilburg University; Daniel L. Oberski, Department of Methodology and Statistics, Utrecht University.

This research was supported by a Research Talent Grant from the Netherlands Organisation for Scientific Research. This work was presented at the Modern Modeling Methods conference in 2016 and the Interuniversity Graduate School of Psychometrics and Sociometrics summer conference in 2016. A preprint of this article has been made available on the Open Science Framework.

Correspondence concerning this article should be addressed to Sara van Erp, Department of Methodology and Statistics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, the Netherlands. E-mail: s.vanerp_1@tilburguniversity.edu

often employed when studying multiple classrooms, schools, or countries—has been found to perform badly in terms of bias and power with a small number of groups (Lüdtke, Marsh, Robitzsch, & Trautwein, 2011; Maas & Hox, 2005; Meuleman & Billiet, 2009; Ryu & West, 2009), whereas BSEM performed well even with small samples (Depaoli & Clifton, 2015; Hox, van de Schoot, & Matthijsse, 2012). BSEM may also reduce issues with nonconvergence (Kohli, Hughes, Wang, Zopluoglu, & Davison, 2015) and inadmissible estimates (Can, van de Schoot, & Hox, 2014; Dagne, Howe, Brown, & Muthén, 2002), it is computationally convenient for models with many latent variables (Harring, Weiss, & Hsu, 2012; Lüdtke, Robitzsch, Kenny, & Trautwein, 2013; Oravecz, Tuerlinckx, & Vandekerckhove, 2011), and it easily yields credible intervals (i.e., the Bayesian version of a confidence interval) on functions of parameters such as reliabilities (Geldhof, Preacher, & Zyphur, 2014) or indirect effects (Yuan & MacKinnon, 2009). Furthermore, BSEM allows researchers to assume that traditionally restricted parameters, such as cross-loadings, direct effects, and error covariances, are approximately, rather than exactly, zero by incorporating prior information (MacCallum, Edwards, & Cai, 2012; B. O. Muthén & Asparouhov, 2012).

However, to take advantage of BSEM, one challenge must be overcome (MacCallum et al., 2012): the specification of the prior distributions. Prior specification is an important but difficult part of any Bayesian analysis. Ideally, the priors should accurately reflect preexisting knowledge about the world, both in terms of the facts and the uncertainty about those facts. Previous research has shown that BSEM has superior performance to frequentist SEM from a subjective Bayesian perspective when priors reflect researchers' beliefs exactly; and from a frequentist perspective, BSEM outperforms frequentist SEM when priors reflect reality. Priors that do not reflect prior beliefs to infinite accuracy (Bayesian perspective), or that do not correspond to reality (frequentist perspective), however, can lead to severe bias (Baldwin & Fellingham, 2013; Depaoli, 2012, 2013, 2014; Depaoli & Clifton, 2015). Moreover, eliciting priors is a time-consuming task, and even experts are often mistaken and prone to overstating their certainty (e.g., Garthwaite, Kadane, & O'Hagan, 2005; Tversky, 1974). Additionally, in BSEM, it is generally more difficult to specify subjective priors because of the many parameters, some of which are not easily interpretable (e.g., latent variable variances). Therefore, instead of relying fully on expert judgments, researchers employing Bayesian analysis often use "default" priors. Default priors can be viewed as a class of priors that (a) do not contain any external substantive information, (b) are completely dominated by the information in the data, and (c) can be used in an automatic fashion for a Bayesian data analysis (Berger, 2006). For this reason, default priors seem particularly useful for SEM because they allow us to use the flexible Bayesian approach without needing to translate prior knowledge into informative priors.

Previous research has investigated the performance of several default priors for BSEM. Thus far, the BSEM priors studied have been limited to proper priors chosen to equal the true population values in expectation, or chosen purposefully to be biased in expectation by a certain percentage (Depaoli, 2012, 2013, 2014; Depaoli & Clifton, 2015). These studies yielded important insights into the consequences of prior choice. However, commonly suggested alternative default priors in the Bayesian literature, such as noninformative improper priors and empirical Bayes (EB) priors (e.g., Carlin & Louis, 2000a; Casella, 1985; Natarajan & Kass, 2000), remain, to our knowledge, uninvestigated. Moreover, although several authors agree that any BSEM analysis should be accompanied by a sensitivity analysis, the available practical guidelines to do so focus on the situation in which substantive information was used to specify the prior (Depaoli & van de Schoot, 2015). In addition, Depaoli and van de Schoot (2015) provide specific guidelines on checking the sensitivity of the results to different inverse Wishart priors for the covariance matrix. Our contribution is that we specifically focus on prior specification in BSEM, we focus on default prior specification when prior information is weak or completely unavailable, and we specifically focus on univariate priors (i.e., univariate normal, and inverse Gamma), which are easiest to interpret by applied researchers.

In addition to the lack of knowledge regarding priors in BSEM, there appears to be a lack of awareness of the importance of the prior as well. A recent review by van de Schoot, Winter, Ryan, Zondervan-Zwijnenburg, and Depaoli (2017) identified trends in and uses of Bayesian methods based on 1,579 articles published in psychology between 1990 and 2015. Of the 167 empirical articles employing regression techniques (including SEM), only 45% provided information about the prior that was used. Thirty-one percent of the articles did not discuss which priors were used at all, and 24% did not provide enough information to reconstruct the priors. In terms of the type of prior, 26.7% of the empirical articles used informative priors, of which only 4.5% (two articles) employed EB methods to choose the hyperparameters.

The literature review of van de Schoot et al. (2017) showed that a substantial part of Bayesian analyses in psychology relies on default priors. Now the problem is that the exact choice of the default prior may affect the conclusions substantially, as has been shown in the general Bayesian literature (e.g., Gelman, 2006; Lambert, Sutton, Burton, Abrams, & Jones, 2005) and will be shown in the context of BSEM in this article. Different software packages have implemented different default or weakly informative priors as their default software settings (see Table 1). With the development of more user-friendly Bayesian software, more nonexpert users are trying out Bayesian analysis, in general, and BSEM, in particular, and rely on the default software settings, without being fully aware of the influence and importance of the prior distributions. In the 167 empirical articles identified by van de Schoot et al. (2017), WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) was the most popular software program until 2012, but since 2013, this position has been taken over by the commercial SEM software Mplus (which has Bayesian methods implemented since 2010; L. K. Muthén & Muthén, 1998–2012; van de Schoot et al., 2017).

This article aims to further develop the practice and utility of default BSEM and to raise awareness that the exact choice of default prior in BSEM matters. Specifically, this article has the following three goals. First, we propose two novel EB prior settings that adapt to the observed data and are easy to implement. Second, we investigate the performance of several default priors, including the novel EB priors, and compare them with the priors studied thus far, thereby investigating prior sensitivity in default BSEM. Third, because the choice of the default prior can have a large effect on the estimates in small samples, we provide a step-by-step guide on how to perform a default prior sensitivity

Table 1
Overview of Default Priors in Software Packages for Bayesian Structural Equation Modeling

Software	Type of parameter	Default prior form	Default prior hyperparameters
Mplus (L. K. Muthén & Muthén, 1998–2012)	Intercepts/loadings/slopes	Normal	$N(0, 10^{10})$
	Thresholds	Normal	$N(0, 5)$
	Variance parameters	Inverse Gamma	$IG(-1, 0)$
	Variance covariance matrices continuous variables	Inverse Wishart	$IW(0, -p - 1)$
	Variance covariance matrices categorical variables	Inverse Wishart	$IW(I, p + 1)$
	Class proportions mixture models	Dirichlet	$D(10, 10, \ldots, 10)$
Blavaan (Merkle & Rosseel, 2015)	Measurement intercepts	Normal	$N(0, 1{,}000)$
	Structural intercepts/loadings/regression coefficients	Normal	$N(0, 100)$
	Precision residuals	Gamma	$G(1, .5)$
	Blocks of precision parameters	Wishart	$W(I, p + 1)$
	Correlations	Beta[a]	$B(1, 1)$
Stan (Carpenter et al., 2017)	All parameters	Uniform	
Amos (Arbuckle, 2013)	All parameters	Uniform	
WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000)	All parameters	No defaults; manually specify proper priors	
JAGS (Plummer, 2003)	All parameters	No defaults; manually specify proper priors	

Note. p = number of variables; I = identity matrix; JAGS = Just Another Gibbs Sampler.
[a] The beta prior for correlations in blavaan has support $(-1, 1)$ instead of $(0, 1)$.

analysis. Note that we focus on frequentist properties of the different default priors, such as bias, mean squared errors (*MSE*s), and coverage rates. We take this perspective because it is common to focus on frequentist properties when assessing the performance of default priors (Bayarri & Berger, 2004). A different and popular perspective on Bayesian statistics is that of updating one's prior beliefs with the data. In this perspective, instead of default priors, informative priors are used that contain external subjective information about the magnitude of the parameters before observing the data. Specification of such subjective priors will not be explored in the current article.

The rest of this article is organized as follows. We first introduce the BSEM model using a running example from the SEM literature. In the subsequent section, we discuss possible priors that have been suggested both in the BSEM and in the wider Bayesian analysis literature. Subsequently, a simulation study investigates the effect these prior choices have on BSEM estimates. We then provide practical guidelines based on the results of the simulation for practitioners who wish to perform their own sensitivity analysis. Finally, we apply these guidelines to empirical data from the running example, providing a demonstration of sensitivity analysis in BSEM.

A Structural Equation Model

Throughout this article, we will consider a linear structural equation model with latent variables from the literature. We have selected this model because it is one of the most popular example models in SEM. Furthermore, the model includes a mediation effect, which is of interest in substantive research. As a result, investigation of this model will not only result in general insights regarding default priors in BSEM but will also provide specific information about default priors for mediation analysis. The model (see Figure 1) describes the influence of the level of industrialization in 1960 (ξ) on the level of political democracy in 1960 (η^{60}) and 1965 (η^{65}) in 75 countries. Industrialization is measured by three indicators and the level of democracy by four indicators at each time point. The indicators for level of democracy consist of expert ratings, and, because some of the ratings come from the same expert at both time points or the same source in the same year, several measurement errors correlate, which we model

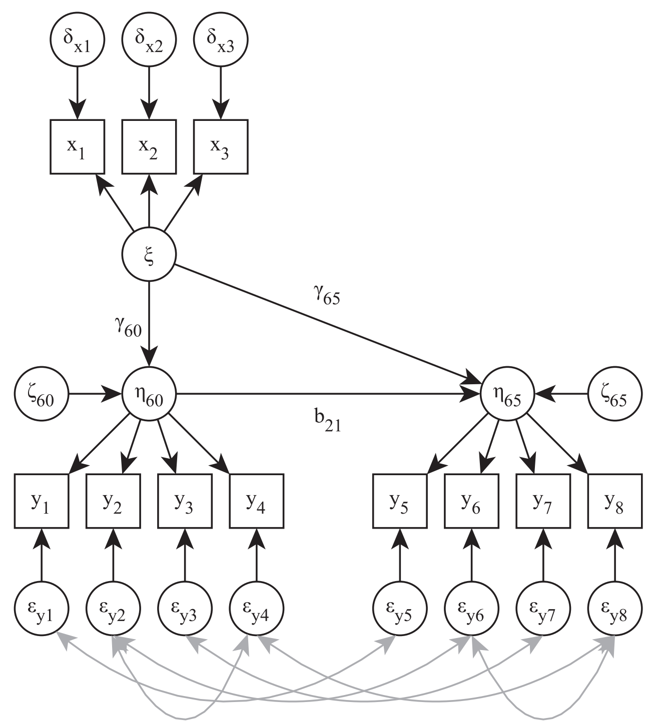

Figure 1. Structural equation model describing the influence of industrialization in 1960 (ξ) on political democracy in 1960 (η^{60}) and 1965 (η^{65}).

through pseudolatent variables $\boldsymbol{D}$, following Dunson, Palomo, and Bollen (2005).

The structural model (for $i = 1, \ldots, n$) is given by

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\xi_i + \boldsymbol{\zeta}_i \quad \text{with} \quad \xi_i \sim N(\mu_\xi, \omega_\xi^2), \quad \text{and} \quad \boldsymbol{\zeta}_i \sim N(\mathbf{0}, \boldsymbol{\Omega}_\zeta)$$

The measurement model is given by

$$\mathbf{y}_i = \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta}_i + \mathbf{D}_i + \boldsymbol{\epsilon}_i^y \quad \text{with} \quad \mathbf{D}_i \sim N(\mathbf{0}, \boldsymbol{\Omega}_D), \quad \text{and} \quad \boldsymbol{\epsilon}_i^y \sim N(\mathbf{0}, \Sigma_y)$$

$$\mathbf{x}_i = \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x\xi_i + \boldsymbol{\delta}_i^x \quad \text{with} \quad \boldsymbol{\delta}_i^x \sim N(\mathbf{0}, \Sigma_x)$$

Here, the structural mean and intercepts μ_ξ and $\boldsymbol{\alpha}$ reflect the mean structure in the structural part of the model, whereas the measurement intercepts $\boldsymbol{\nu}_y$ and $\boldsymbol{\nu}_x$ reflect the mean structure in the measurement part of the model. The loadings $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ represent the relations between the latent variables and their indicators, and the structural regression coefficients $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ represent the relations between the latent variables. The residual variances Σ_y and Σ_x reflect the variation in the measurement errors, and the random variances ω_ξ^2, $\boldsymbol{\Omega}_\zeta$, and $\boldsymbol{\Omega}_D$ reflect the variation in the latent variables. In practice, researchers are often most interested in the relations between the latent variables, in this case, the direct effect γ_{65} and the indirect effect $\gamma_{60} \cdot b_{21}$ of industrialization in 1960 on political democracy in 1965. Note that this specification is more restrictive than necessary because, following Dunson et al. (2005), we fix the nonzero entries in the weight matrix for $\boldsymbol{D}$ to be 1. As a result negative covariances are not allowed. A solution to this problem is implemented in the R package blavaan (Merkle & Rosseel, 2015; R Core Team, 2015). The model is identified by restricting the first intercept and loading of each latent variable, that is, $\lambda_y^1 = \lambda_y^5 = \lambda_x^1 = 1$ and $\nu_y^1 = \nu_y^5 = \nu_x^1 = 0$. The Appendix provides the full model in matrix form, and a more detailed description of the data and the model can be found in Bollen (1980, 1989).

In the application, we will use the original data containing observations from 75 countries (available in the lavaan package in R; Rosseel, 2012) and a subset of the data containing only the first 35 observations. Maximum likelihood (ML) estimation for this subset gives two warnings: (1) The standard errors of the parameter estimates may not be reliable because of a nonpositive definite first-order derivative product matrix, and (2) the latent variable covariance matrix is not positive definite. The first warning can be an indication of weak empirical identification because of the small sample size, whereas the second warning indicates an inadmissible parameter estimate; in this case, the estimated variance of the pseudolatent variable representing the relation between ϵ_4^y and ϵ_8^y, that is, $\hat{\omega}_{D48}^2$, is negative. These warnings clearly illustrate that when using classical ML estimation, researchers may encounter certain problems that may be overcome by adopting a Bayesian approach, in which prior distributions are specified in the subspace of admissible solutions.

Default Priors for Bayesian SEM

Default priors have the following three key properties. First, they do not contain external subjective information. Second, they are completely dominated by the data. Third, they can be used in an automatic fashion for a Bayesian data analysis. Because of the second property, these priors are also referred to as "noninformative" or "weakly informative" priors. Such priors are often used in Bayesian analysis, including BSEM, when no substantive information is available or when the researcher does not wish to incorporate any substantive information through the prior distribution. Default priors allow researchers to use the powerful and flexible Bayesian approach without needing to specify an informative prior based on one's prior knowledge, which can be a difficult and time-consuming task. Different software packages use different default priors in their automatic settings based on various heuristic arguments (see Table 1). For example, the commercial SEM software Mplus (L. K. Muthén & Muthén, 1998–2012) specifies a uniform improper prior for variance parameters by default, whereas the Bayesian modeling software WinBUGS (Lunn et al., 2000) recommends vague proper inverse Gamma priors for the variances.

Default priors can roughly be divided in the following three categories: noninformative[1] improper priors, vague proper priors, and EB priors. The first two have been used extensively in the BSEM literature, whereas the latter has, to our knowledge, not been applied to BSEM yet, but it is popular in the general literature on Bayesian modeling (Carlin & Louis, 2000a; Casella, 1985; Natarajan & Kass, 2000). In this study, we will focus on different priors from each of these three commonly used types of default priors. Some priors are chosen because they are the default setting in popular software programs (specifically Mplus and blavaan), whereas other priors are chosen because they are widely used or because they have been shown to perform well in certain situations. The EB priors are novel in BSEM and have been included to investigate whether an EB approach can be advantageous in BSEM.

For all three types, we focus on priors that have a conditionally conjugate form. Conditionally conjugate priors have the advantage that they result in fast computation because the resulting conditional posteriors have known distributional forms (i.e., they have the same distribution as the prior) from which we can easily sample. Specifically, the conditionally conjugate prior for a location parameter (e.g., intercepts, loadings, and regression coefficients) is the normal distribution, and for a variance parameter, it is the inverse Gamma distribution. Note that these are univariate priors. Generally, it is possible to instead specify a multivariate normal distribution for the vector of location parameters and an inverse Wishart prior for the covariance matrix. However, in this article, we focus specifically on univariate priors for separate parameters. We will now discuss these different default priors in more detail.

Noninformative Improper Priors

Noninformative improper priors are most commonly used in "objective Bayesian analysis" (Berger, 2006). In a simple normal distribution with unknown mean μ and unknown variance σ^2, for example, the standard noninformative improper prior $p(\mu, \sigma^2) \propto \sigma^{-2}$ (known as Jeffreys prior) yields exactly the same point and

[1] Here, "noninformative" refers to the ultimate goal of the prior rather than its actual behavior. In fact, many different noninformative priors exist (Kass & Wasserman, 1996), which often result in slightly different estimates.

interval estimates for the population mean as does classical ML estimation; hence the name "objective Bayes." An improper prior is not a formal probability distribution because it does not integrate to unity. A potential problem of noninformative improper priors is that the resulting posteriors may also be improper, which occurs when there is too little information in the data (Hobert & Casella, 1996). In the example of a normal distribution with unknown mean and variance, we need at least two distinct observations in order to obtain a proper posterior for μ and σ^2 when starting with the improper Jeffreys prior. Currently, little is known about the performance of these types of priors in BSEM. Throughout this article, we will therefore consider the following noninformative improper priors for variance parameters σ^2:

- $p(\sigma^2) \propto \sigma^{-2}$. This prior is most commonly used in objective Bayesian analysis for variance components. It is equivalent to a uniform prior on $\log(\sigma^2)$. There have been reports, however, that this prior results in improper posteriors for variances of random effects in multilevel analysis (e.g., Gelman, 2006). In a simple normal model with known mean and unknown variance, at least one observation is needed for this prior to result in a proper posterior for the variance.
- $p(\sigma^2) \propto \sigma^{-1}$. This prior was recommended by Berger (2006) and Berger and Strawderman (1996) for variance components in multilevel models. For this prior, at least two observations are needed in a normal model with known mean and unknown variance to obtain a proper posterior.
- $p(\sigma^2) \propto 1$. This prior is the default choice in Mplus (L. K. Muthén & Muthén, 1998–2012). Gelman (2006) noted that it may result in overestimation of the variance. When using this prior in a normal model with known mean and unknown variance, at least three observations are needed to obtain a proper posterior for the variance.

Each of these noninformative improper priors can be written as the conjugate inverse Gamma prior. The inverse Gamma distribution is given by

$$p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-(\alpha+1)}\exp\left(-\frac{\beta}{\sigma^2}\right) \text{ with shape } \alpha > 0 \text{ and scale } \beta > 0.$$

When the shape parameter $\alpha = 0$ and the scale parameter $\beta = 0$, we obtain $p(\sigma^2) \propto \sigma^{-2}$. When the shape parameter $\alpha = -\frac{1}{2}$ and the scale parameter $\beta = 0$, we obtain $p(\sigma^2) \propto \sigma^{-1}$. When the shape parameter $\alpha = -1$ and the scale parameter $\beta = 0$, we obtain $p(\sigma^2) \propto 1$.

Table 2 presents these priors for all variance components in our model. For the intercepts, means, loadings, and regression coefficients, the standard noninformative improper prior is the uniform prior from $-\infty$ to $+\infty$. The vague proper prior $N(0, 10^{10})$ approximates this uniform prior. Thus, for the intercepts, means, loadings, and regression coefficients, we will only investigate vague proper and EB priors, which are discussed next.

Vague Proper Priors

A common solution to avoid improper posteriors while keeping the idea of noninformativeness in the prior is to specify vague proper priors. These priors are formal probability distributions, in which the hyperparameters are chosen such that the information in the prior is minimal. In the case of variance parameters, vague proper priors can be specified as conjugate inverse Gamma priors with hyperparameters close to zero, typically 0.1, 0.01, or 0.001. These priors approximate the improper prior $p(\sigma^2) \propto \sigma^{-2}$ (Berger, 2006). The latter option, $IG(\epsilon, \epsilon)$, with $\epsilon = 0.001$ is used as example throughout the WinBUGS manual. We will consider these three typical prior specifications for the variance parameters in our model. Note that smaller hyperparameters lead to a prior that is more peaked around zero. For means and regression parameters, we will investigate a normal prior with a large variance. This vague proper prior approximates a flat prior. Specifically, we shall use the normal prior $N(0, 10^{10})$, which is the default in Mplus. In addition, we will consider the blavaan default setting for location parameters, which is the normal prior $N(0, 1000)$ for the measurement intercepts and the normal prior $N(0, 100)$ for the loadings, structural intercepts, and structural regression coefficients. We will label this prior coupled with $\pi(\sigma^2) \propto 1$ for variances "vague normal." The vague proper priors that will be considered throughout this article are summarized in Table 2.

A potential problem of vague proper priors is that the exact hyperparameters are arbitrarily chosen, and this choice can greatly affect the final estimates. For example, there is no clear rule stating how to specify the shape and scale parameter of the inverse Gamma prior, namely, 0.1, 0.01, 0.001, or perhaps even smaller. Gelman (2006) showed that in a multilevel model with eight schools on the second level, the posterior for the between-school variance was completely dominated by the inverse Gamma prior with small hyperparameters. In addition, the inverse Gamma prior depends on the scale of the data. Specifically, an $IG(0.1, 0.1)$ prior might be a noninformative choice when the data are standardized, but can be very informative when the data are unstandardized. It is yet unclear how this prior performs in structural equation models, which are considerably more complex than the eight schools example studied by Gelman, in terms of the number of parameters and the relations between them.

Empirical Bayes Priors

The third type of default priors we consider is EB priors. The central idea behind the EB methodology is that the hyperparameters are chosen based on the data at hand (see, e.g., Carlin & Louis, 2000a, Chapter 3). This results in a prior with substantial probability mass in the region in which the likelihood is concentrated. EB methodology can be seen as a compromise between classical and Bayesian approaches (Casella, 1992). Because all of the data are used to inform the prior distribution, EB methods are useful for combining evidence (e.g., across neighborhoods, see Carter & Rolph, 1974; or across law schools, see Rubin, 1980). In our application, we expect an EB prior informed by the data of all countries to provide better estimates because it adds more information to the analysis than improper or vague priors, or ML estimation. There is also a computational advantage of EB priors, as noted by Carlin and Louis (2000b), who state that the Markov Chain Monte Carlo (MCMC) sampler based on EB priors can be more stable.

We focus on the parametric EB approach, in which a specific distributional form of the prior is assumed, typically conjugate, with only the hyperparameters unknown. Different methods have been proposed to obtain the hyperparameters in this set-

Table 2
Overview of the Default Prior Specifications per Parameter Considered Throughout This Article

		Prior			
Parameter type	Parameter	Default	Noninformative improper	Vague proper	Empirical Bayes
Latent variable variances	ω^2_ξ	$\pi(\omega^2_\xi) \propto 1$	$\pi(\omega^2_\xi) \propto 1$	IG(.001, .001)	$\text{IG}\left(\frac{1}{2}, \hat{\omega}^2_\xi \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)\right)$
			$\pi(\omega^2_\xi) \propto \omega_\xi^{-1}$	IG(.01, .01)	
			$\pi(\omega^2_\xi) \propto \omega_\xi^{-2}$	IG(.1, .1)	
	ω^2_ζ	$\pi(\omega^2_\zeta) \propto 1$	$\pi(\omega^2_\zeta) \propto 1$	IG(.001, .001)	$\text{IG}\left(\frac{1}{2}, \hat{\omega}^2_\zeta \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)\right)$
			$\pi(\omega^2_\zeta) \propto \omega_\zeta^{-1}$	IG(.01, .01)	
			$\pi(\omega^2_\zeta) \propto \omega_\zeta^{-2}$	IG(.1, .1)	
	ω^2_D	$\pi(\omega^2_D) \propto 1$	$\pi(\omega^2_D) \propto 1$	IG(.001, .001)	$\text{IG}\left(\frac{1}{2}, \hat{\omega}^2_D \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)\right)$
			$\pi(\omega^2_D) \propto \omega_D^{-1}$	IG(.01, .01)	
			$\pi(\omega^2_D) \propto \omega_D^{-2}$	IG(.1, .1)	
Residual variances	σ^2_y	$\pi(\sigma^2_y) \propto 1$	$\pi(\sigma^2_y) \propto 1$	IG(.001, .001)	$\text{IG}\left(\frac{1}{2}, \hat{\sigma}^2_y \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)\right)$
			$\pi(\sigma^2_y) \propto \sigma_y^{-1}$	IG(.01, .01)	
			$\pi(\sigma^2_y) \propto \sigma_y^{-2}$	IG(.1, .1)	
	σ^2_x	$\pi(\sigma^2_x) \propto 1$	$\pi(\sigma^2_x) \propto 1$	IG(.001, .001)	$\text{IG}\left(\frac{1}{2}, \hat{\sigma}^2_x \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)\right)$
			$\pi(\sigma^2_x) \propto \sigma_x^{-1}$	IG(.01, .01)	
			$\pi(\sigma^2_x) \propto \sigma_x^{-2}$	IG(.1, .1)	
Structural intercepts	α	N(0, 10^{10})	—	N(0, 10^{10})	N(0, $\hat{\alpha}^2 + \hat{\omega}^2_\zeta$)
			—	N(0, 100)	
Structural regression coefficients	b	N(0, 10^{10})	—	N(0, 10^{10})	N(0, $\hat{b}^2 + \hat{\omega}^2_\zeta$)
			—	N(0, 100)	
	γ	N(0, 10^{10})	—	N(0, 10^{10})	N(0, $\hat{\gamma}^2 + \hat{\omega}^2_\zeta$)
			—	N(0, 100)	
Latent variable mean	μ_ξ	N(0, 10^{10})	—	N(0, 10^{10})	N(0, $\hat{\mu}^2_\xi + \hat{\omega}^2_\xi$)
			—	N(0, 100)	
Measurement intercepts	υ_y	N(0, 10^{10})	—	N(0, 10^{10})	N(0, $\hat{\nu}^2_y + \hat{\sigma}^2_y$)
			—	N(0, 1000)	
	υ_x	N(0, 10^{10})	—	N(0, 10^{10})	N(0, $\hat{\nu}^2_x + \hat{\sigma}^2_x$)
			—	N(0, 1000)	
Loadings	λ_y	N(0, 10^{10})	—	N(0, 10^{10})	N(0, $\hat{\lambda}^2_y + \hat{\sigma}^2_y$)
			—	N(0, 100)	
	λ_x	N(0, 10^{10})	—	N(0, 10^{10})	N(0, $\hat{\lambda}^2_x + \hat{\sigma}^2_x$)
			—	N(0, 100)	

Note. Q^{-1} = regularized inverse Gamma function; IG = Inverse Gamma; N = normal.

ting. First, the hyperparameters can be estimated using the marginal distribution of the data (i.e., the product of the likelihood and the prior integrated over the model parameters: $p(Data) = \int f(Data \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the vector with parameters, with prior $p(\boldsymbol{\theta})$, and f denotes the likelihood of the data given the unknown model parameters; see, e.g., Carlin & Louis, 2000a; Casella, 1985; Laird & Ware, 1982). An example is the EB version of the well-known g-prior of Zellner (1986). The g-prior is centered around a reference (or null) value in which g controls the prior variance. In EB methodology, g is estimated from the marginal distribution (e.g., see Liang et al., 2008, and the references therein). The resulting EB prior can be seen as the best predictor of the observed data. The difficulty with this approach, however, is that an analytic expression of the marginal distribution of the data may not be available for large complex models. This is also the case in structural equation models, and therefore we will not use the marginal distribution of the data to construct an EB prior in this article.

A second EB approach, which is simpler to implement, is to specify weakly informative priors centered around the estimates (e.g., ML) of the model parameters. These priors contain minimal information so that the problem of double use of the data is negligible. This type of EB prior has been investigated in generalized multilevel modeling (Kass & Natarajan, 2006; Natarajan & Kass, 2000) and SEM (Dunson et al., 2005). This approach, however, may perform badly when the estimates that are used for centering the EB priors are unstable, which is the case in SEM with small samples.

For this reason, an alternative EB prior is proposed, which is novel in the BSEM literature. The idea is to first center the prior around a reference value to minimize its dependence on unstable estimates. Subsequently, the other hyperparameters are estimated from the data from a simplified model to keep the solution tractable. The idea of this EB prior was inspired by the constrained posterior prior approach of Mulder et al. (2009; Mulder, Hoijtink, & Klugkist, 2010) for Bayesian model selection. As will be shown, the proposed EB prior generally contains less information than the information of one observation. For this reason, the double use of the data and the resulting underestimation of the posterior variance, a known problem of

EB methodology (e.g., Carlin & Louis, 2000a; Darnieder, 2011; Efron, 1996), is expected to be negligible.

EB priors for intercepts, means, factor loadings, and regression coefficients. Here, we discuss the EB prior for the intercepts, means, factor loadings, and regression coefficients for which the conditionally conjugate prior is normally distributed. To estimate the hyperparameters, that is, the prior mean and the prior variance, we simplify the endeavor by constructing a normal prior for α, denoted by $N(\mu_\alpha,\tau_\alpha^2)$, for the following model,

$$y_i = \alpha + \epsilon_i \text{ with } \epsilon_i \sim N(0, \sigma^2), \quad (1)$$

for $i = 1, \ldots, n$ observations, with unknown error variance σ^2. Note that α in Model 1 denotes a location parameter, that is, an intercept, mean, factor loading, or regression coefficient in the model considered in this article. The prior mean μ_α is set equal to a reference (or null) point value to avoid the heavy dependence on the data. In the current setting, we set the prior mean equal to zero because of its special meaning of "no effect". The prior variance τ_α^2 is then estimated as the variance in a restricted model in which the prior mean, $\mu_\alpha = 0$, is plugged in for α, that is, $y_i \sim N(0, \tau_\alpha^2)$, for $i = 1, \ldots, n$. The variance in this model can be estimated as $\hat{\tau}_\alpha^2 = n^{-1}\sum_{i=1}^{n} y_i^2 \approx E[Y^2]$. Subsequently, an expression for the estimate $\hat{\tau}_\alpha^2$ can be obtained by deriving $E[Y^2]$ from the original Model 1 according to

$$E[Y^2] = \text{Var}(Y) + (E[Y])^2 = \sigma^2 + \alpha^2. \quad (2)$$

Thus, the prior variance is chosen to be equal to the sum of the estimated error variance and the square of the estimated effect, that is, to $\hat{\tau}_\alpha^2 = \hat{\sigma}^2 + \hat{\alpha}^2$.

This prior has two important properties. First, this prior has clear positive support where the likelihood is concentrated, which is a key property of an EB prior. This can be seen from the fact that the prior variance will be large (small) when the difference between the observed effect and the prior mean, $\hat{\alpha}$, is large (small). Note, however, that by centering the prior around a reference value instead of the observed effect, the prior will be less sensitive to the instability of the ML estimates. Second, this EB prior does not contain more information than the information of a single observation. To see this note that the standard error of α is equal to $\sigma/\sqrt{n}$. Furthermore, the EB prior standard deviation is smallest when $\hat{\alpha} = 0$, in which case $\hat{\tau}_\alpha = \hat{\sigma}$, which corresponds to the standard error based on a single observation. If $\hat{\alpha} \neq 0$, then $\hat{\tau}_\alpha > \hat{\sigma}$, which implies less information than the information in a single observation. For this reason, we expect the problem of using the data twice (i.e., for prior specification and estimation) to be negligible. This behavior is illustrated in Figure 2. In the case of no effect, that is, $\hat{\alpha} = 0$ (Data 1), the prior variance is equal to the error variance. When $\hat{\alpha} \neq 0$ (Data 2), the prior variance becomes larger, that is, the error variance plus the squared estimated effect, to ensure positive support around $\hat{\alpha}$. Note that the EB prior behaves similarly to the constrained posterior prior (Mulder et al., 2010), with the difference that the EB prior is simpler to compute.

This methodology will be used to construct EB priors for the intercepts, means, factor loadings, and regression coefficients. For example, for the measurement intercept of y_2, denoted by ν_2^y, the EB prior variance is equal to the squared ML estimate, that is, $(\hat{\nu}_2^y)^2$, plus the ML estimate of the variance of the error δ_2^y, that is, $\hat{\sigma}_{y2}^2$, and thus, the EB prior is distributed as, $\nu_2^y \sim N(0,(\hat{\nu}_2^y)^2 + \hat{\sigma}_{y2}^2)$. An overview of all EB priors for location parameters is given in Table 2.

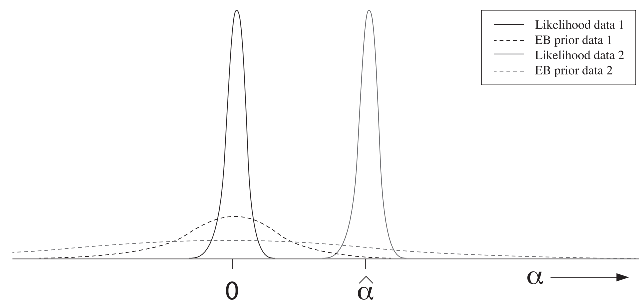

Figure 2. Illustration EB priors for location parameters. The EB prior for α is $N(0,\tau_\alpha^2)$, with τ_α^2 equal to $\hat{\sigma}^2$ for Data 1, and equal to $\hat{\sigma}^2 + \hat{\alpha}^2$ for Data 2. See the online article for the color version of this figure.

EB priors for variance components. The methodology to construct EB priors for location parameters cannot be used for variance components because a clear reference (or null) value is generally unavailable for these types of parameters. Therefore, we will consider two alternative approaches for the priors for the variances, which will be combined with the EB prior for intercepts, means, factor loadings, and regression coefficients. In the first combination, which we will label EB1, the priors for the variance components are centered around the ML estimate, as was also considered by Natarajan and Kass (2000). Again, we consider the conditionally conjugate prior, which has an inverse Gamma distribution with a shape parameter and a scale parameter. The shape parameter controls the amount of prior information. By setting the shape parameter to $\frac{1}{2}$, the prior carries the information that is equivalent to one data point (Gelman, Carlin, Stern, & Rubin, 2004, p. 50). For this reason, the double use of the data is also not a serious concern in this case. The scale parameter is chosen such that the prior median equals the ML estimate, that is, $\hat{\sigma}^2 \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)$, with $\hat{\sigma}^2$ denoting the ML estimate of the variance parameter and Q^{-1} denoting the regularized inverse Gamma function (see Table 2). A potential issue of this prior is its heavy dependence on the ML estimates of the variances. Therefore, in the second combination, labeled EB2, noninformative uniform priors are specified for the variances. Thus, the EB2 prior can be seen as a hybrid EB prior, in which the priors for location parameters depend on the data and the priors for variance parameters are completely independent of the data.[2]

A Simulation Study of Default BSEM Analyses

Even though all the discussed priors reflect some form of default noninformative BSEM analysis, each choice may result in different conclusions. A simulation study was set up to investigate the performance of the different default priors in the industrialization and political democracy model, a classical SEM application. A common method to check the performance of objective priors is to

[2] We also considered a third combination: an EB prior for the variances (i.e., an inverse Gamma prior centered around the ML estimate) combined with the $N(0, 10^{10})$ prior for the location parameters. Thus, this combination can also be seen as a hybrid EB prior, in which the priors for the variance parameters depend on the data and the priors for the location parameters are completely independent of the data. However, this combination performed similarly to the uniform prior for variances combined with the $N(0, 10^{10})$ prior for location parameters, and we have therefore not considered it further.

look at their frequentist properties (Bayarri & Berger, 2004). In particular, we were interested in (a) convergence of the Bayesian estimation procedure, (b) (relative) bias, (c) *MSE*, (d) coverage rates, (e) quantiles, and (f) Type I error rates of the direct and indirect effects. We will end the section with a general conclusion regarding the performance of the different default priors.

For the data generation, we considered four different sample sizes: 35, 75, 150, and 500. For such a complex SEM, a sample size of 35 or 75 might seem extremely small. However, BSEM is often recommended especially in situations in which the sample size is small (Heerwegh, 2014; Hox et al., 2012; Lee & Song, 2004). Furthermore, the original sample size was only 75. We expect the influence of the prior to decline as sample size increases. The population values were equal to the ML estimates of the original data (see the online supplemental materials for an overview). We manipulated the population values for the direct effect γ_{65} and the indirect effect $\gamma_{60} \cdot b_{21}$, as these are the parameters of substantive interest in the model. We also manipulated two loadings of $\boldsymbol{y}$ (λ_4^y and λ_8^y) and the variances of the pseudolatent variables $\boldsymbol{\Omega_D}$, which represent the measurement error correlations. These variances were manipulated because previous research indicates that the vague proper priors for variances are especially influential when the variance parameter is estimated to be close to zero (Gelman, 2006). The manipulations of the population values are shown in Table 3 and were selected in such a way that we obtained a wide range of values. We used a fractional design in which we simulated data under two combinations of population values: (1) combinations of population values for the direct and indirect effect ($3 \times 3 = 9$ conditions), and (2) combinations of population values for the loadings and variances of error correlations ($3 \times 2 = 6$ conditions). In total, this resulted in 15 different populations. From each population, we generated 500 data sets per sample size.[3] Table 3 presents an overview of all 60 data-generating conditions.

Each condition was analyzed using the nine different default prior combinations with the same type of prior being specified for all parameters in the model at once, that is, three noninformative improper priors, three vague proper priors, two EB priors, and the vague normal setting. Note that in the case of the noninformative improper and vague proper priors, only the priors on the variance parameters change, whereas the priors on the mean and regression parameters are specified as the normal prior $N(0, 10^{10})$. In addition, ML estimation was included for each condition, leading to a total of $10 \times 15 \times 4 = 600$ conditions, as shown in Table 3.

The EB priors are based on the ML estimates, which sometimes included Heywood cases (i.e., an estimated negative variance). In the case of negative ML estimates for the variance parameters, we set the prior median for the EB prior on variance parameters equal to 0.001. Preliminary analyses showed that the precise choice had little effect on the posterior. For the mean, intercept, loading, and regression parameters, the residual variances of the model equations were sometimes estimated to be negative, in which case we fixed them to zero for computation of the prior variance. Again, preliminary analyses indicated that the exact choice did not have any clear influence on the posterior, as long as the estimate was fixed to a small number, for example, 0.001.

For each analysis, we ran two MCMC chains and discarded the first half of each chain as burn-in. Convergence was originally assessed using the potential scale reduction (PSR), taking PSR $<$ 1.05 as a criterion, with a maximum of 75,000 iterations. Based on reviewers' comments, we reran the conditions for $N = 35$ with a fixed number of iterations, first 50,000 for each chain and then 100,000 for those replications that did not obtain a PSR < 1.05 (using the population values as starting values). We then assessed convergence by selecting those replications with a PSR < 1.1[4] and manually checked a part of the trace plots. Given that the results did not differ substantially from the original results, we only took this approach for $N = 35$.[5]

Estimation error was assessed using (relative) bias and *MSE*. Bias was computed as $\frac{1}{S}\sum_{s=1}^{S}(\hat{\theta}_s - \theta)$, with S being the number of converged replications per cell, θ being the population value of that parameter, and $\hat{\theta}_s$ being the posterior median[6] for that parameter in a specific replication s. Relative bias was computed as $\frac{1}{S}\sum_{s=1}^{S}\left(\frac{\hat{\theta}_s - \theta}{\theta}\right)$ and is only defined in those population conditions for which $\theta \neq 0$. *MSE* was computed based on the true population value as $\frac{1}{S}\sum_{s=1}^{S}(\hat{\theta}_s - \theta)^2$. We obtained the 95% coverage interval by computing how often the population value was contained in the 95% credible or confidence interval. In addition, to check how well the posteriors reproduced the sampling distributions, we investigated the 2.5% and 97.5% quantiles for every parameter. The 2.5% (97.5%) quantile was computed as the proportion of times that the lower (upper) bound of the 95% confidence/credible interval was higher (lower) than the population value. Ideally, these should equal 2.5% and 97.5%, respectively. Finally, we looked at the Type I error rates for the direct effect γ_{65} and the indirect effect $\gamma_{60} \cdot b_{21}$. The ML results are included for comparison. All analyses were done in Mplus (Version 7.2) and R, using the package MplusAutomation (Hallquist & Wiley, 2014).

Convergence

Table 4 shows the percentage of converged replications for each prior and sample size, averaged across the population values. For $N = 35$, the EB2 prior resulted in the highest convergence (98.9%), followed by the vague normal prior (94.1%). For all priors, convergence generally increased with sample size and there was almost no convergence for the improper prior $\pi(\sigma^2) \propto \sigma^{-2}$ when $N \leq 150$. This is not surprising because this prior is known to result in improper posteriors for variances of random effects in multilevel analysis (e.g., Gelman, 2006). Because of the severe nonconvergence under this improper prior, we shall not consider it

[3] We created cumulative average plots to assess whether 500 replications were enough to attain Monte Carlo convergence, which was the case.

[4] Although less conservative, we used this cutoff because after 100,000 iterations, some parameters had a PSR slightly above 1.05, whereas the trace plots indicated convergence based on eyeballing. An example is given in the online supplemental materials.

[5] Some priors and conditions were added later during the review process. Specifically, for $N = 150$ and $N = 500$, two vague proper, and two noninformative improper, priors were added later in those population conditions for which the direct and indirect effect were manipulated, as well as the vague normal and EB2 prior in all population conditions. For these conditions, we simply ran all replications with 100,000 iterations and assessed convergence by selecting those replications with a PSR $<$1.1, given that this strategy had proven correct for $N = 35$.

[6] We compared the posterior medians with the posterior means. Overall, these two posterior summaries did not differ substantially. However, in some replications, the posterior mean for a specific parameter was extremely high. This can happen when the MCMC sampler samples some extreme values that have a large influence on the posterior mean, but not on the posterior median. Therefore, we used the posterior median as point estimate.

Table 3
Overview of the Data Generating and Analysis Conditions Included in the Simulation Study

Variable	# Levels	Values	
Data generating conditions			
Sample size	4	$N \in \{35, 75, 150, 500\}$	
Direct effect	3	$\gamma_{65} = 0$	
		$\gamma_{65} = 1$	
		$\gamma_{65} = 2$	
Indirect effect	3	$\gamma_{60} \cdot b_{21} = 0 \times .837$	
		$\gamma_{60} \cdot b_{21} = 1 \times .837$	
		$\gamma_{60} \cdot b_{21} = 2 \times .837$	
Loadings	3	$\lambda_4^y, \lambda_8^y = 0$	
		$\lambda_4^y, \lambda_8^y = 1$	
		$\lambda_4^y, \lambda_8^y = 2$	
Error covariances	2	$\omega_{D_{15}}^2, \omega_{D_{24}}^2, \omega_{D_{26}}^2, \omega_{D_{37}}^2, \omega_{D_{48}}^2, \omega_{D_{68}}^2 = 0$	
		$\omega_{D_{15}}^2, \omega_{D_{24}}^2, \omega_{D_{26}}^2, \omega_{D_{37}}^2, \omega_{D_{48}}^2, \omega_{D_{68}}^2 = 1$	
Analysis conditions			
Priors	9	Noninformative improper:	$\pi(\sigma^2) \propto 1$ and $N(0, 10^{10}$; Mplus default)
			$\pi(\sigma^2) \propto \sigma^{-1}$ and $N(0, 10^{10})$
			$\pi(\sigma^2) \propto \sigma^{-2}$ and $N(0, 10^{10})$
		Vague proper:	IG(.001, .001) and $N(0, 10^{10})$
			IG(.01, .01) and $N(0, 10^{10})$
			IG(.1, .1) and $N(0, 10^{10})$
		Vague normal:	N(0, 1000) and N(0, 100) and $\pi(\sigma^2) \propto 1$
		Empirical Bayes (EB):	
		EB1:	$IG\left(\frac{1}{2}, \hat{\sigma}^2 \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)\right)$ and $N(0, \hat{\mu}^2 + \hat{\sigma}^2)$
		EB2:	$\pi(\sigma^2) \propto 1$ and $N(0, \hat{\mu}^2 + \hat{\sigma}^2)$
Maximum likelihood	1		

Note. The maximum likelihood estimates from the original data were $\hat{\gamma}_{65} = 0.572$; $\hat{\gamma}_{60} \cdot \hat{b}_{21} = 1.483 \cdot 0.837$; $\hat{\lambda}_4^y = 1.265$; $\hat{\lambda}_8^y = 1.266$; $\hat{\omega}_{D_{15}}^2 = 0.624$; $\hat{\omega}_{D_{24}}^2 = 1.313$; $\hat{\omega}_{D_{26}}^2 = 2.153$; $\hat{\omega}_{D_{37}}^2 = 0.795$; $\hat{\omega}_{D_{48}}^2 = 0.348$; and $\hat{\omega}_{D_{68}}^2 = 1.356$; N = normal; IG = Inverse Gamma; Q^{-1} = regularized inverse Gamma function.

further in this article. Note that this may imply that the vague proper priors $IG(\epsilon, \epsilon)$, with $\epsilon = 0.1$, 0.01, or 0.001 can perform badly, as they approximate the improper prior $\pi(\sigma^2) \propto \sigma^{-2}$. Specifically, trace plots of converged replications for these vague proper priors showed occasional peaks, resulting in relatively high posterior medians for those replications. We will only present the results in those population conditions with at least 50% convergence, and we only consider the converged replications. Convergence percentages for each separate population condition are available in the online supplemental materials. The ML analysis always converged but often resulted in estimated negative variances. Specifically, in 53.9% of the replications, at least one Heywood case occurred.

Bias

Table 5 presents the relative bias, with the bias in brackets, for selected parameters in the model, for $N = 35$ and $N = 75$ averaged across population values. Results for all parameters in the model are available in the online supplemental materials. Following L. K. Muthén and Muthén (2002), relative biases exceeding 10% are regarded as substantial and shown in bold. Given that the influence of the prior is greatest for small samples, we will focus on $N = 35$ and $N = 75$ in presenting the results and only mention the results for $N = 150$ and $N = 500$ briefly. The results for $N = 150$ and $N = 500$ can be found in the online supplemental materials.

Table 4
Percentage Converged Replications for the Default Priors in the Simulation Study, Averaged Across Population Values

N	Mplus default	$\pi(\sigma^2) \propto \sigma^{-1}$	$\pi(\sigma^2) \propto \sigma^{-2}$	IG(.001, .001)	IG(.01, .01)	IG(.1, .1)	Vague normal	EB1	EB2
35	73.7	53.3	0	51.4	68.9	88.7	94.1	83.3	98.9
75	99.6	99.5	1.03	97.2	99.9	99.9	99.0	99.6	100
150	100	100	1.77	100	100	100	95.0	99.9	99.6
500	100	99.8	45.6	99.8	100	100	98.7	99.8	98.8

Note. Location parameters have the normal $N(0, 10^{10})$ prior, except for the vague normal and EB priors. N = sample size. Mplus default = $\pi(\sigma^2) \propto 1$ combined with the normal $N(0, 10^{10})$ prior; $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters; IG(.001, .001), IG(.01, .01), IG(.1, .1) = vague proper priors variance parameters; Vague normal = $\pi(\sigma^2) \propto 1$ prior for variance parameters combined with the normal $N(0, 1000)$ prior for measurement intercepts and the normal $N(0, 100)$ prior for the other location parameters; EB1 = Empirical Bayes prior location and variance parameters; EB2 = EB prior location parameters combined with $\pi(\sigma^2) \propto \sigma^{-1}$ prior variance parameters.

Table 5

Relative Bias and Bias (in Parentheses) for Each Default Prior and ML Estimation, Averaged Across Population Values, for Selected Parameters

Prior	γ_{60}	γ_{65}	b_{21}	$\gamma_{60} \cdot b_{21}$	α_{60}	α_{65}	λ_2^y	ν_2^y	$\omega^2_{\zeta 60}$	$\omega^2_{\zeta 65}$	ω^2_{D15}	σ^2_{y6}	σ^2_{x2}
Sample size = 35													
Mplus default	.044 (.053)	.026 (.024)	−.016 (−.013)	.016 (.017)	**.13** (−.264)	.05 (−.118)	.038 (.047)	.069 (−.179)	.067 (.266)	**1.207** (.208)	**.332** (.239)	.099 (.142)	**.427** (.051)
$\pi(\sigma^2) \propto \sigma^{-1}$	.051 (.068)	−.009 (−.002)	−.003 (−.003)	.035 (.045)	**.185** (−.336)	.006 (−.034)	.026 (.053)	.078 (−.181)	.014 (.159)	**.373** (.082)	.011 (.133)	**−.227** (−.095)	**.134** (.017)
IG(.001, .001)	.037 (.056)	0 (.005)	**.271** (.227)	.035 (.043)	**.137** (−.278)	**.611** (−1.426)	**.468** (.588)	**1.411** (−3.683)	−.019 (−.074)	.005 (.001)	−.053 (.017)	**−.117** (−.17)	**−.111** (−.013)
IG(.01, .01)	−.05 (−.048)	−.047 (−.021)	**.191** (.274)	−.042 (−.027)	**−.185** (.251)	**.364** (−.388)	**.752** (1.133)	**2.015** (−2.444)	**−.134** (−.631)	**.161** (.033)	**.378** (.238)	−.208 (−.097)	.068 (.006)
IG(.1, .1)	−.04 (−.046)	.01 (.004)	.034 (.046)	−.042 (−.038)	**−.15** (.244)	**.119** (−.163)	**.155** (.255)	**.417** (−.715)	**−.157** (−.728)	**.667** (.128)	**.269** (.25)	**−.238** (−.111)	**.339** (.04)
Vague normal	−.055 (−.062)	.002 (.001)	.007 (.006)	−.069 (−.067)	**−.162** (.328)	.012 (−.028)	.068 (.086)	**.101** (−.264)	−.042 (−.167)	**1.186** (.204)	**.424** (.294)	.099 (.143)	**.436** (.052)
EB1	−.077 (−.095)	**−.149** (−.103)	.011 (.009)	−.079 (−.08)	**−.23** (.468)	**−.2** (.467)	−.019 (−.023)	**−.162** (.422)	−.04 (−.159)	**.212** (.036)	−.037 (−.002)	−.089 (−.128)	.036 (.004)
EB2	−.079 (−.100)	**−.155** (−.109)	−.011 (−.009)	−.099 (−.102)	**−.236** (.479)	**−.251** (.585)	−.06 (−.075)	**−.229** (.599)	.084 (.334)	**1.491** (.256)	**.268** (.200)	.096 (.139)	**.499** (.060)
ML	.021 (.030)	.002 (.006)	.017 (.015)	.032 (.035)	.074 (−.142)	.036 (−.088)	.018 (.024)	.038 (−.101)	−.061 (−.235)	**−.246** (−.043)	−.078 (−.053)	**−.107** (−.159)	−.033 (−.004)
Sample size = 75													
Mplus default	−.016 (−.018)	.036 (.029)	−.013 (−.01)	−.036 (−.035)	−.041 (.084)	.039 (−.092)	.026 (.033)	.035 (−.09)	−.003 (−.01)	**.688** (.118)	**.189** (.142)	.016 (.023)	**.211** (.025)
$\pi(\sigma^2) \propto \sigma^{-1}$	−.018 (−.034)	.038 (.017)	−.017 (−.003)	−.041 (−.04)	−.058 (.166)	.01 (−.045)	.019 (.041)	.043 (−.123)	−.016 (−.107)	**.232** (.037)	**.118** (.075)	−.094 (−.002)	.034 (.001)
IG(.001, .001)	−.034 (−.04)	−.002 (.002)	.007 (.006)	−.038 (−.037)	−.094 (.192)	.004 (−.009)	.06 (.075)	**.104** (−.273)	−.055 (−.216)	**−.117** (−.02)	.027 (.03)	−.062 (−.089)	**−.148** (−.018)
IG(.01, .01)	−.022 (−.039)	0 (.008)	.004 (.007)	−.022 (−.03)	−.078 (.196)	.004 (−.053)	.032 (.056)	.089 (−.194)	−.052 (−.243)	−.022 (.002)	.078 (.036)	−.097 (−.085)	−.066 (−.005)
IG(.1, .1)	−.005 (−.027)	.023 (.019)	−.006 (0)	−.016 (−.031)	−.018 (.143)	.01 (−.075)	.026 (.058)	.075 (−.194)	−.051 (−.253)	**.334** (.079)	.063 (.088)	**−.153** (−.091)	**.144** (.018)
Vague normal	−.036 (−.044)	.025 (.019)	−.009 (−.008)	−.051 (−.052)	**−.113** (.23)	.028 (−.065)	.022 (.028)	.036 (−.093)	−.004 (−.017)	**.632** (.109)	**.18** (.135)	.032 (.046)	**.169** (.02)
EB1	−.068 (−.087)	**−.102** (−.076)	.012 (.01)	−.061 (−.067)	**−.205** (.416)	**−.141** (.33)	−.012 (−.016)	**−.104** (.272)	−.023 (−.09)	**−.146** (−.025)	−.046 (−.015)	−.069 (−.099)	−.092 (−.011)
EB2	−.068 (−.088)	−.074 (−.051)	−.01 (−.008)	−.083 (−.088)	**−.211** (.429)	**−.131** (.307)	−.027 (−.034)	**−.128** (.333)	.044 (.174)	**.761** (.131)	**.133** (.108)	.009 (.013)	**.259** (.031)
ML	−.009 (−.011)	.020 (.018)	.008 (.007)	−.003 (−.003)	−.029 (.000)	.045 (−.021)	.010 (.005)	.024 (−.017)	−.029 (−.087)	**−.135** (−.043)	−.041 (.000)	−.061 (−.010)	−.005 (−.038)

Note. Location parameters have the normal $N(0, 10^{10})$ prior, except for the vague normal and EB priors. Values for the relative bias greater than .10 are shown in bold. ML = maximum likelihood estimation; Mplus default = $\pi(\sigma^2) \propto 1$ combined with the normal $N(0, 10^{10})$ prior; $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters; IG(.001, .001), IG(.01, .01), IG(.1, .1) = vague proper priors variance parameters; Vague normal = $\pi(\sigma^2) \propto 1$ prior for variance parameters combined with the normal $N(0, 1000)$ prior for measurement intercepts and the normal $N(0, 100)$ prior for the other location parameters; EB1 = Empirical Bayes prior location and variance parameters; EB2 = EB prior location parameters combined with $\pi(\sigma^2) \propto \sigma^{-1}$ prior variance parameters.

For $N = 35$, the vague proper priors resulted in relative biases greater than 0.10 for most location parameters, followed by the EB priors. The noninformative priors (Mplus default, $\pi[\sigma^2] \propto \sigma^{-1}$) and the vague normal prior resulted in only a few parameters with substantial bias. ML performed best, with relative bias close to zero for all location parameters. For all priors, some parameters showed more bias than others. Specifically, the measurement intercepts ν_y and the structural intercepts α often resulted in large relative bias.

For $N = 75$, the bias decreased for all priors, resulting in only a few location parameters with relative bias exceeding 10% for the noninformative improper, vague proper, and vague normal priors. The EB priors performed worst, with six and seven location parameters showing relative biases greater than 0.10, whereas ML again resulted in relative bias close to zero for all parameters. Again, the measurement intercepts ν_y showed most bias. For $N = 150$, only the vague proper and EB priors showed relative biases greater than 0.10 for some location parameters, whereas for $N = 500$, the relative bias was close to zero for all priors and location parameters.

For the variance parameters in the model, when $N = 35$, the vague proper prior IG(0.1, 0.1), the vague normal prior, and the EB2 prior resulted in most cases with substantial bias, followed by the improper priors, and the vague proper prior IG(0.01, 0.01). The vague proper prior IG(0.001, 0.001) and the EB1 prior performed good, with only three variance parameters having relative biases greater than 0.10, and ML performed best with only two parameters with relevant bias. Generally, the estimated latent variable variances showed more bias then the estimated error variances, especially $\omega^2_{\zeta 65}$, which had relative bias greater than 0.10 for ML and all priors, except the IG(0.001, 0.001) prior.

For $N = 75$, the vague normal and EB2 prior resulted in most biased variance parameters, followed by the Mplus default setting (i.e., the improper prior $\pi(\sigma^2) \propto 1$ combined with the N(0, 10^{10}) prior), then the vague proper priors IG(0.001, 0.001) and IG(0.1, 0.1), and next the EB1 prior. The improper prior $\pi(\sigma^2) \propto \sigma^{-1}$ and the vague proper prior IG(0.01, 0.01) performed well, with only two variances exceeding the relative bias of 10%, and ML performed best. Again, the estimates for the latent variable variances were generally more biased than the estimates for the error variances. For $N = 150$, all methods resulted in some biased variance parameters, except the improper prior $\pi(\sigma^2) \propto \sigma^{-1}$ and ML estimation. For $N = 500$, only IG(0.001, 0.001), IG(0.01, 0.01), the Mplus default, and the EB1 prior resulted in some biased variance parameters.

Overall, ML estimation performed best in terms of relative bias for both the variance and location parameters. This can be explained by the fact that ML estimation does not force separate variance components to be positive. For the location parameters, the vague proper priors and the EB priors performed worst and for the variance parameters, the vague normal and EB2 prior performed worst. Of the Bayesian methods, the Mplus default setting performed best for the location parameters and the EB1 prior performed best for the variance parameters.

Mean Squared Error

Figure 3 shows for each prior and type of parameter the *MSE*s relative to the *MSE* of ML estimation per population value and parameter on the logarithmic scale, $\ln(MSE_{\text{Bayes}}/MSE_{\text{ML}})$. The results are categorized by structural regression coefficients, intercepts and latent mean, factor loadings, and variance parameters. Note that the vertical axis is truncated at $\ln(MSE_{\text{Bayes}}/MSE_{\text{ML}}) = 4$, excluding the extreme situations in which the *MSE* of the prior is more than exp(4) $\approx$ 55 times higher than the *MSE* of the ML estimate, which occurred for the vague proper priors. Tables with the numerical values for the *MSE* are available in the online supplemental materials. For $N = 35$, the vague proper priors performed worst, especially for the intercepts and factor loadings. This can be explained by the occasional extreme values drawn for these priors, as noted previously, and this unstability in the MCMC sampler is related to the fact that the vague proper priors approximate the problematic improper prior $\pi(\sigma^2) \propto \sigma^{-2}$. All other Bayesian methods resulted in smaller or approximately equal *MSE*s in comparison with ML estimation. In particular, the EB priors and the vague normal prior performed best for the structural regression coefficients. For $N = 75$, $N = 150$, and $N = 500$, there were hardly any clear differences between the *MSE*s using the different methods.

We can thus conclude that all methods perform similarly in terms of *MSE*s, except for the vague proper priors, which performed considerably worse.

Coverage Rates

Table 6 shows the coverage rates of the 95% confidence intervals and 95% Bayesian credible intervals for selected parameters. Coverage rates higher than 97.5% or lower than 92.5% are considered as substantially deviating from the desired 95%, and are marked in bold. Coverage rates for all parameters in the model are available in the online supplemental materials. For $N = 35$, the EB1 and EB2 priors performed worst, with low coverage for 13 and 12 location parameters, respectively, followed by the vague proper prior IG(0.01, 0.01), which showed low coverage for nine parameters. The vague normal prior showed coverage rates for eight parameters that were too high, as did the Mplus default setting for five parameters. ML estimation resulted in coverage rates that were too low for four parameters. The improper prior $\pi(\sigma^2) \propto \sigma^{-1}$ and the vague proper priors IG(0.001, 0.001) and IG(0.1, 0.1) performed best.

For $N = 75$, the vague proper prior IG(0.001, 0.001) performed worst with low coverage rates for 16 location parameters, followed by the EB1 prior with low coverage rates for 14 parameters. The EB2 prior also showed low coverage for eight parameters. The vague normal prior showed coverage rates for two parameters that were slightly too high. The other priors and ML estimation resulted in coverage rates between 92.5% and 97.5% for all parameters. For $N = 150$ and $N = 500$, all methods resulted in coverage rates for the location parameters close to 95%, except for the EB priors when $N = 150$.

We now discuss the coverage rates for the variance parameters.[7] For $N = 35$, ML estimation performed worst, with low coverage

[7] We deleted some values for the coverage rates for the vague normal prior, the vague proper priors, the Mplus default, and the EB2 prior that were equal to zero. These values occurred only for Ω_D when the population value was zero. This happened because the lower bound of the credible interval was always greater than zero for these priors, and thus a population value of zero is, by definition, never contained in the credible interval. For the EB1 prior and the improper prior $\pi(\sigma^2) \propto \sigma^{-1}$, the lower bound did equal zero in several replications, thereby resulting in a coverage not equal to zero in this situation. Note that because the lower bound did not equal zero in all replications, the resulting coverage for these priors was low when the population values were equal to zero. For ML, the lower bound of the confidence interval can be negative.

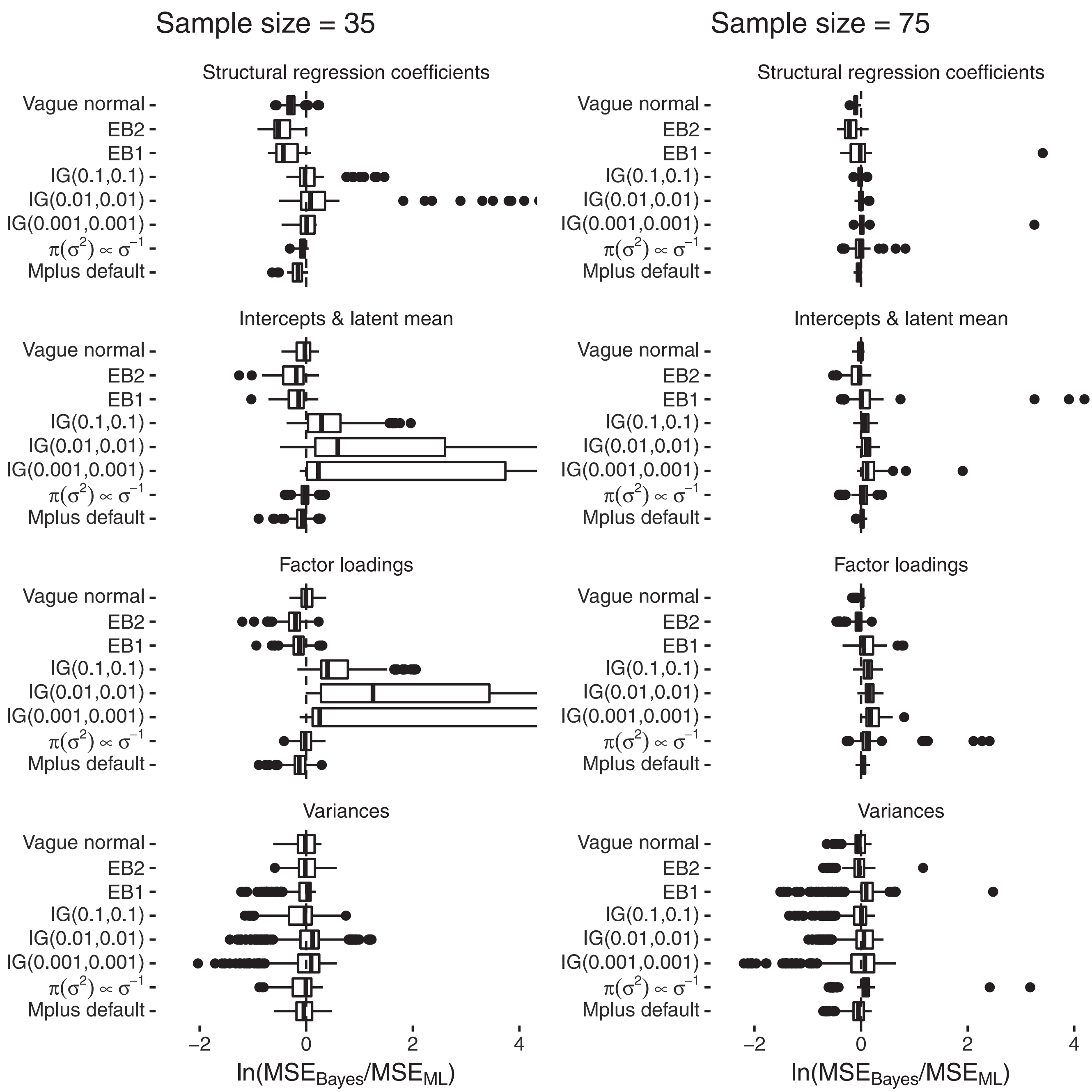

Figure 3. Mean squared error (*MSE*) for Bayesian estimation using different default priors divided by the *MSE* for maximum likelihood (ML) estimation on the natural logarithmic scale. Location parameters have the normal $N(0, 10^{10})$ prior, except for the vague normal and EB priors. Vertical dashed lines indicate where the *MSE* for the Bayesian estimates equals the *MSE* for the ML estimates. EB1 = Empirical Bayes prior location and variance parameters; EB2 = EB prior location parameters combined with $\pi(\sigma^2) \propto \sigma^{-1}$ prior variance parameters; Vague normal = $\pi(\sigma^2) \propto 1$ prior for variance parameters combined with the normal $N(0, 1000)$ prior for measurement intercepts and the normal $N(0, 100)$ prior for the other location parameters; Mplus default = $\pi(\sigma^2) \propto 1$ combined with the normal $N(0, 10^{10})$ prior; $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters; IG(0.001, 0.001), IG(0.01, 0.01), IG(0.1, 0.1) = vague proper priors variance parameters.

Table 6
Coverage Rates of 95% Bayesian Credible Intervals and 95% Confidence Intervals for Each Default Prior and ML Estimation, Averaged Across Population Values, for Selected Parameters

Prior	γ_{60}	γ_{65}	b_{21}	$\gamma_{60} \cdot b_{21}$	α_{60}	α_{65}	λ_2^y	ν_2^y	$\omega^2_{\zeta 60}$	$\omega^2_{\zeta 65}$	ω^2_{D26}	σ^2_{y4}	σ^2_{x2}
						Sample size = 35							
Mplus default	96.5	97.2	**98.2**	97	96.9	97.3	**98.2**	**97.7**	97.1	96.6	**98**	**98**	**97.7**
$\pi(\sigma^2) \propto \sigma^{-1}$	96.2	96.3	96.7	96.4	96.2	96	97.2	97	96.4	**98.2**	**83.3**	95.3	**98.8**
IG(.001, .001)	94.8	95	95.4	96	94.9	95	95.4	95.8	94.9	**98.2**	**91.2**	**92.5**	**98.1**
IG(.01, .01)	**91.8**	94.5	94	93.6	**91.9**	94.3	**90.8**	**91.3**	**86.2**	**98.8**	94	94	**99**
IG(.1, .1)	93.4	96.3	96	94.9	93.6	95.8	94.5	94.7	**88.2**	**98.1**	95.7	95.3	**98.9**
Vague normal	96.7	**97.8**	**98.1**	97	97	**97.5**	**98.1**	**97.7**	95.1	96.9	**97.9**	**98.1**	**97.6**
EB1	94.9	**77.6**	**91.6**	93.4	95	**72.4**	**88.5**	**87.1**	**91.7**	**64.6**	**79.3**	**88.8**	**84.7**
EB2	96.3	**81.4**	96.0	95.0	96.9	**74.2**	**91.1**	**89.2**	95.0	95.3	**77.8**	**98.0**	96.4
ML	93	**90.2**	93.7	93.8	92.7	**90.9**	92.9	93.1	**87.4**	**90.7**	**90.4**	**92**	94
						Sample size = 75							
Mplus default	94.3	95.8	95.6	94.1	94.5	94.8	95.3	95.4	94.7	94.5	95.8	94.5	95.6
$\pi(\sigma^2) \propto \sigma^{-1}$	94.4	95.2	95.8	94.4	94.7	94.9	95	95.4	94.3	**97.6**	92.8	93.8	96.3
IG(.001, .001)	**91.6**	**92.2**	92.7	**91.6**	**91.8**	**91.8**	**91.5**	**91.8**	**90.5**	94.7	**72.5**	**87.9**	**89.9**
IG(.01, .01)	94.4	95	95.4	94.2	94.5	94.5	94.4	94.7	93.3	**97.7**	93	**91.5**	96.5
IG(.1, .1)	94.4	95.8	95.3	94.3	94.4	95.2	94.4	94.8	93.5	97.1	95.1	**92**	**97.9**
Vague normal	95.3	96.6	97	95.1	95.6	96.3	96.2	95.9	95.2	96.3	96.4	94.9	96.3
EB1	92.8	**82.5**	**89.1**	**90.8**	93.3	**80.6**	**91.3**	**90.3**	92.6	**73.2**	**81.5**	**89.1**	**86.3**
EB2	94.2	**87.4**	95.1	92.5	94.8	**83.3**	92.9	**92.2**	94.6	94.1	**76.8**	95.1	95.2
ML	93.9	93.6	95.3	93.8	94.1	93.3	94.5	94.7	**92.0**	92.7	93.0	92.8	95.2

Note. Location parameters have the normal $N(0, 10^{10})$ prior, except for the vague normal and EB priors. Coverage rates lower than 92.5% and higher than 97.5% are shown in bold. ML = maximum likelihood estimation; Mplus default = $\pi(\sigma^2) \propto 1$ combined with the normal $N(0, 10^{10})$ prior; $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters; IG(.001, .001), IG(.01, .01), IG(.1, .1) = vague proper priors variance parameters; Vague normal = $\pi(\sigma^2) \propto 1$ prior for variance parameters combined with the normal $N(0, 1000)$ prior for measurement intercepts and the normal $N(0, 100)$ prior for the other location parameters; EB1 = Empirical Bayes prior location and variance parameters; EB2 = EB prior location parameters combined with $\pi(\sigma^2) \propto \sigma^{-1}$ prior variance parameters.

rates for 19 parameters. The EB1 prior also showed low coverage for 14 parameters, as did the EB2 prior for nine parameters, whereas the Mplus default setting showed coverage rates for 11 parameters that were too high, as did the vague normal prior for nine parameters. The improper prior $\pi(\sigma^2) \propto \sigma^{-1}$ and the vague proper priors showed coverage rates for six or eight parameters outside the range of 92.5% and 97.5%. For $N = 75$, coverage rates improved for most priors and ML estimation, except for the vague proper prior IG(0.001, 0.001) and the EB1 prior, which resulted in extreme values for 16 and 15 variance parameters, respectively. For $N = 150$, coverage rates for all variance parameters were close to 95% for the vague normal and EB2 prior. The EB1 prior performed worst, followed by the IG(0.001, 0.001) and $\pi(\sigma^2) \propto \sigma^{-1}$ priors. For $N = 500$, ML estimation performed best and the EB1 prior performed worst, followed by the $\pi(\sigma^2) \propto \sigma^{-1}$ and IG(0.001, 0.001) priors.

In summary, for location parameters, the EB priors performed worst in terms of coverage rates, whereas for the variance parameters, ML estimation, the vague proper prior IG(0.001, 0.001), and the EB1 prior performed worst. Across all parameters, the improper prior $\pi(\sigma^2) \propto \sigma^{-1}$ and the vague proper prior IG(0.1, 0.1) performed best.

Quantiles

Even in the case of perfect coverage rates of 95%, it may be that the underestimation of the interval estimate occurs much more often than overestimation (or the other way around). To assess this, we investigated the lower 2.5% and upper 97.5% quantiles. The quantiles were obtained by computing how often the lower 2.5% and upper 97.5% bounds of the credible/confidence intervals were above or below the true population value. Figure 4 shows the quantiles for $N = 35$, with the dashed lines indicating 2.5% and 97.5%. The results are categorized by structural regression coefficients, intercepts and latent mean, factor loadings, and variance parameters. The numerical values on which Figure 4 is based are available in the online supplemental materials. For the lower quantile, all priors resulted in quantiles close to the desired 2.5%, except for the two EB prior settings in the case of intercepts and the vague proper priors in the case of factor loadings. For the upper quantile, the noninformative improper priors and the vague normal prior generally performed best, with upper quantiles close to or slightly higher than 97.5%. For the structural regression coefficients and factor loadings, the EB priors performed worst, followed by ML estimation. For the variance parameters, the EB1 prior performed badly, as did ML estimation. For the intercepts, all priors and ML estimation resulted in quantiles close to the desired 97.5%.

Figure 5 shows the quantiles for $N = 75$, which were all closer to the desired quantiles compared with $N = 35$. For the lower quantile, the EB priors resulted in too-high quantiles for the intercepts. Note also the outliers for the vague proper prior IG(0.001, 0.001) across parameters. For the upper quantile, the EB priors again performed worst for the structural regression coefficients and loadings, followed by ML estimation. For the variance parameters, the EB1 prior and ML estimation also

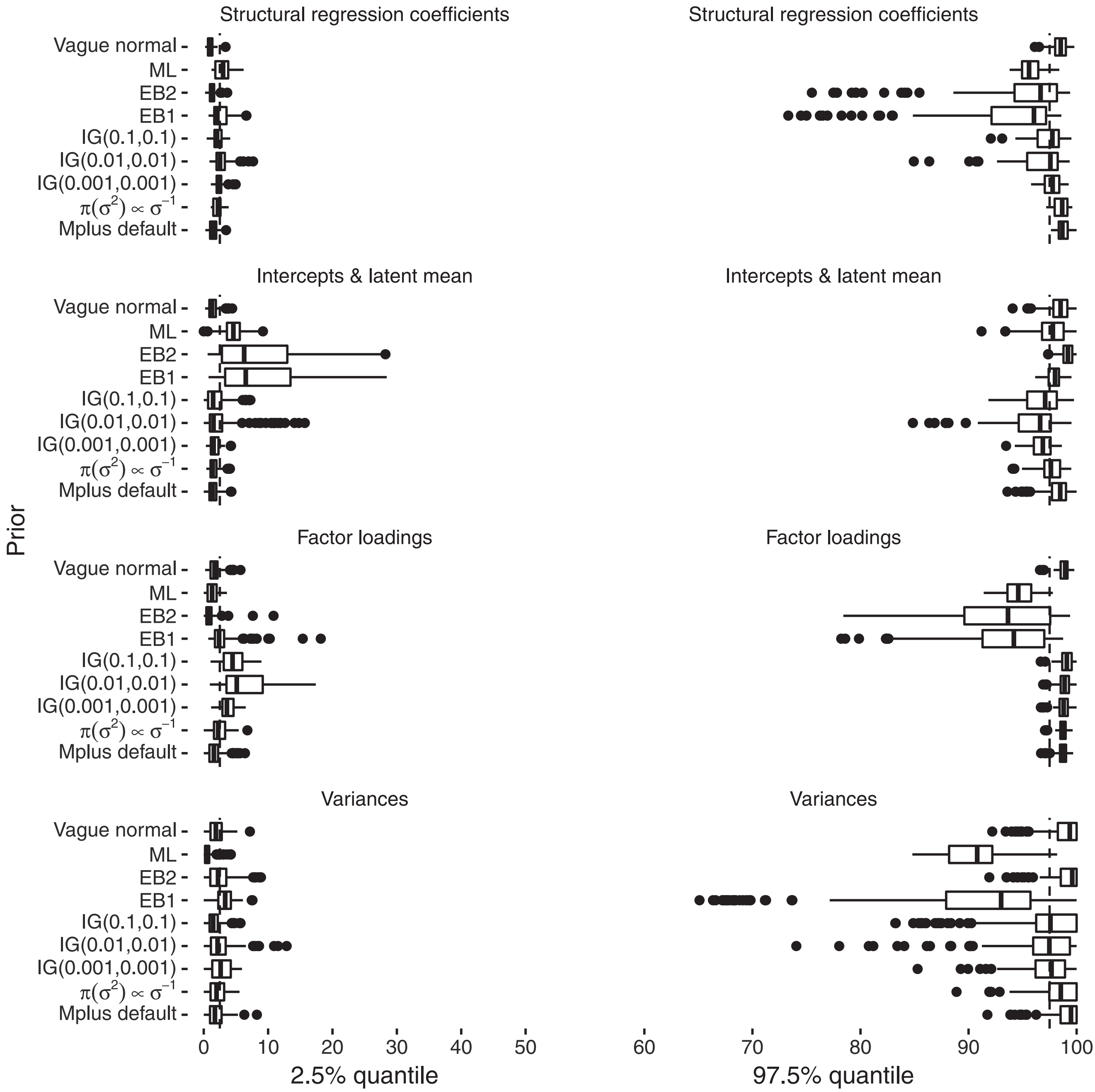

Figure 4. 2.5% and 97.5% quantiles for each default prior and maximum likelihood (ML) estimation for N = 35. Location parameters have the normal $N(0, 10^{10})$ prior, except for the vague normal and EB priors. Vertical dashed lines indicate the desired 2.5% and 97.5%. ML = maximum likelihood estimation; EB1 = Empirical Bayes prior location and variance parameters; EB2 = EB prior location parameters combined with $\pi(\sigma^2) \propto \sigma^{-1}$ prior variance parameters; Vague normal = $\pi(\sigma^2) \propto 1$ prior for variance parameters combined with the normal $N(0, 1{,}000)$ prior for measurement intercepts and the normal $N(0, 100)$ prior for the other location parameters; Mplus default = $\pi(\sigma^2) \propto 1$ combined with the normal $N(0, 10^{10})$ prior; $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters; IG(0.001, 0.001), IG(0.01, 0.01), IG(0.1, 0.1) = vague proper priors variance parameters.

performed badly, whereas for the intercepts, all priors and ML estimation resulted in quantiles close to the desired 97.5%. For N = 150, most priors resulted in quantiles close to the desired 2.5% and 97.5%, except for the vague normal and EB2 priors, which were slightly higher or lower for some parameters. For N = 500, all methods generally resulted in correct quantiles.

To conclude, overall the EB priors performed worst in terms of quantiles, whereas the noninformative improper priors and the vague normal prior performed best.

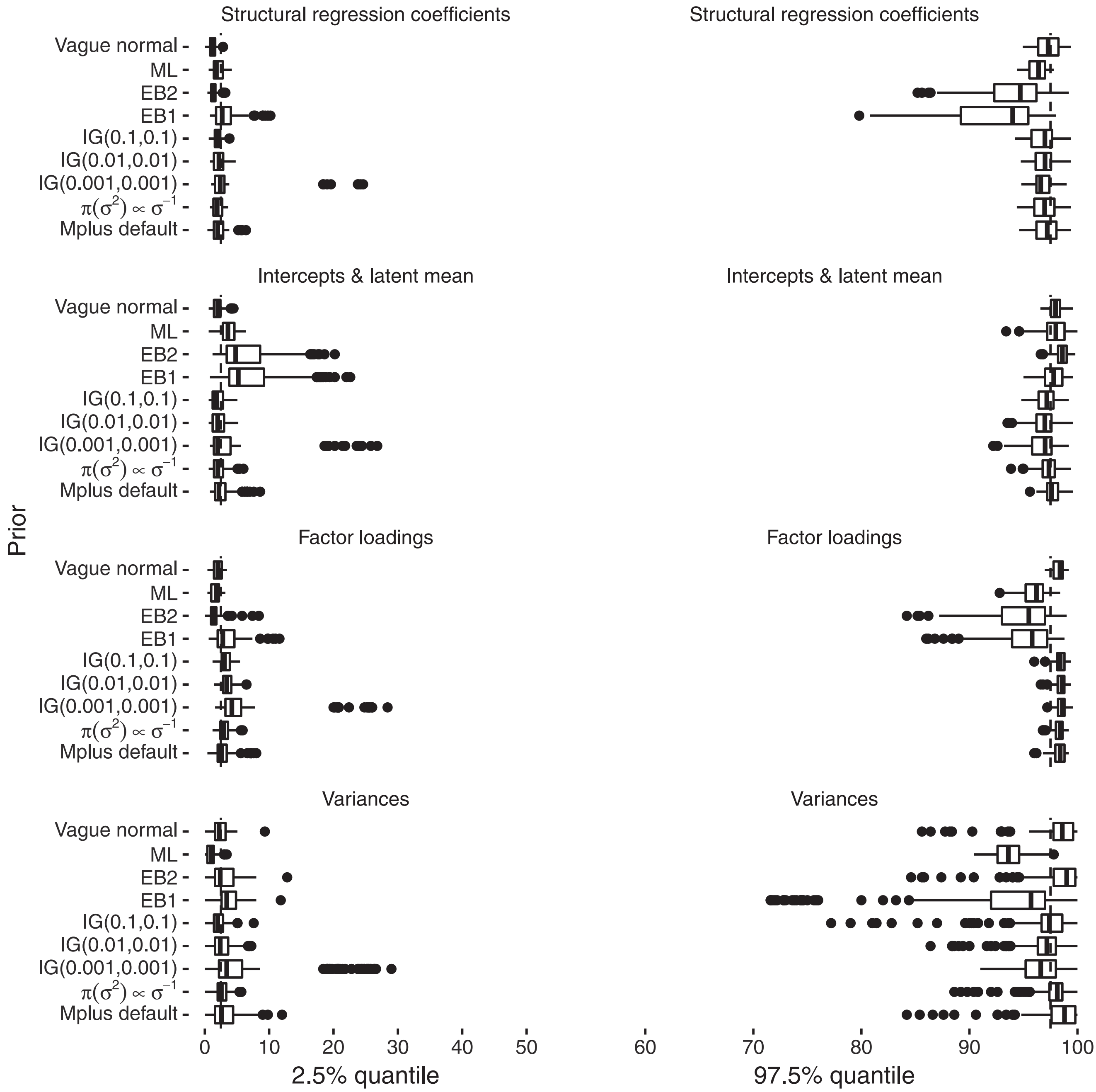

Figure 5. 2.5% and 97.5% quantiles for each default prior and maximum likelihood (ML) estimation for N = 75. Location parameters have the normal $N(0, 10^{10})$ prior, except for the vague normal and EB priors. Vertical dashed lines indicate the desired 2.5% and 97.5%. ML = maximum likelihood estimation; EB1 = Empirical Bayes prior location and variance parameters; EB2 = EB prior location parameters combined with $\pi(\sigma^2) \propto \sigma^{-1}$ prior variance parameters; Vague normal = $\pi(\sigma^2) \propto 1$ prior for variance parameters combined with the normal $N(0, 1000)$ prior for measurement intercepts and the normal $N(0, 100)$ prior for the other location parameters; Mplus default = $\pi(\sigma^2) \propto 1$ combined with the normal $N(0, 10^{10})$ prior; $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters; IG(0.001, 0.001), IG(0.01, 0.01), IG(0.1, 0.1) = vague proper priors variance parameters.

Direct and Indirect Effect

In practice, researchers are often mainly interested in those parameters related to the research question. In this model, the parameters of substantive interest are the direct effect γ_{65} and the indirect effect $\gamma_{60} \cdot b_{21}$. Figure 6 shows the *MSE*s of the direct and indirect effect for the different priors and ML estimation for N = 35 and N = 75. The figure shows that for N = 35, the vague

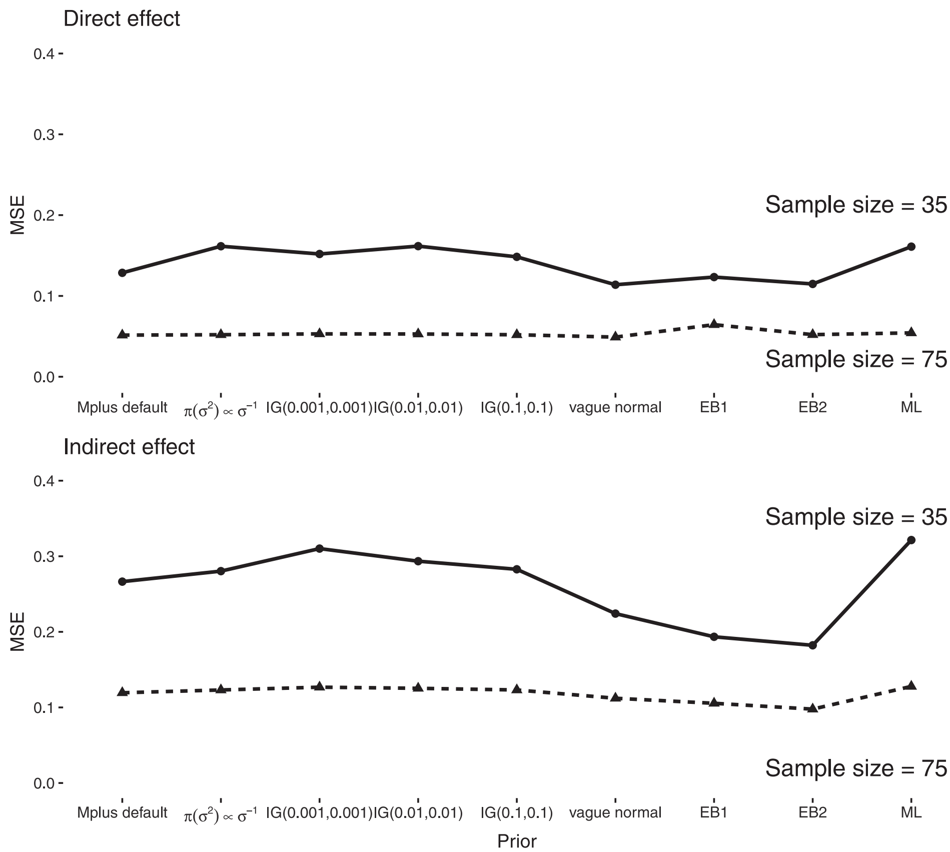

Figure 6. Location parameters have the normal $N(0, 10^{10})$ prior, except for the vague normal and EB priors. Mean squared error (*MSE*) for each default prior and maximum likelihood (ML) estimation for the direct effect γ_{65} and the indirect effect $\gamma_{60} \cdot b_{21}$. ML = maximum likelihood estimation; EB1 = Empirical Bayes prior location and variance parameters; EB2 = EB prior location parameters combined with $\pi(\sigma^2) \propto \sigma^{-1}$ prior variance parameters; Vague normal = $\pi(\sigma^2) \propto 1$ prior for variance parameters combined with the normal $N(0, 1000)$ prior for measurement intercepts and the normal $N(0, 100)$ prior for the other location parameters; Mplus default = $\pi(\sigma^2) \propto 1$ combined with the normal $N(0, 10^{10})$ prior; $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters; IG(0.001, 0.001), IG(0.01, 0.01), IG(0.1, 0.1) = vague proper priors variance parameters.

normal and EB priors resulted in the smallest *MSE*s for the direct effect; the other methods showed slightly worse results. For the indirect effect and $N = 35$, the smallest *MSE*s were obtained using the EB2 prior, followed by the EB1 prior, and the vague normal prior. ML estimation performed worst.

Furthermore, we consider the Type I error rates for the direct and indirect effect, that is, the percentage of replications for which zero is not included in the 95% credible interval, or in the 95% confidence interval for the ML estimates, when the population value is zero. Table 7 shows the Type I error rates for the different priors and different population conditions. For $N = 35$, the error rates for the direct effect were closest to the nominal 5% for the vague proper priors. For the Mplus default setting and the vague normal prior, the rates were too low. For the EB priors and for ML estimation, the Type I error rates were too high. For the indirect effect, the differences between priors were smaller and, in general, most priors resulted in error rates close to 5%, except for the vague normal and EB2 priors, which resulted in error rates that were too low. For $N = 75$, the EB priors again resulted in error rates that were too high for the direct effect, whereas the rates for ML estimation were closer to 5%. For the indirect effect, all priors and ML estimation generally resulted in error rates slightly higher than 5%. For $N = 150$ and $N = 500$, all Type I error rates were generally close to 5%, except for the direct effect for the vague proper priors and EB priors when $N = 150$.

Based on these results, we can conclude that there is not one default prior that performed consistently better than the other priors or than ML estimation across all parameters or outcomes, especially in small samples. When looking across all parameters for $N = 35$, ML estimation performed best in terms of bias; in terms of *MSE*, the EB priors performed best. However, both the EB priors and ML estimation performed badly in terms of coverage and quantiles. The vague proper priors performed worst in case of bias and *MSE* but best in terms of coverage and Type I error rates for the direct effect and indirect effect. Overall, the noninformative improper priors $\pi(\sigma^2) \propto 1$ and $\pi(\sigma^2) \propto \sigma^{-1}$ performed best across all parameters and outcomes, with good coverage rates and quantiles and average bias and *MSE*. One disadvantage of the improper priors in small samples is their high nonconvergence percentage, especially for $\pi(\sigma^2) \propto \sigma^{-1}$.

Table 7
Type 1 Error Rates for the Direct and Indirect Effect, for the Different Default Priors in Different Population Conditions

Parameter	Mplus default	$\pi(\sigma^2) \propto \sigma^{-1}$	IG(.001, .001)	IG(.01, .01)	IG(.1, .1)	Vague normal	EB1	EB2	ML
				Sample size = 35					
Direct effect	2.6	NA	5.9	4.1	3.5	2.0	15.8	12.1	10.9
Indirect effect	4.2	4.0	NA	5.1	5.2	2.9	4.4	2.8	5.1
				Sample size = 75					
Direct effect	3.5	4.0	4.6	4.3	3.4	3.0	13.7	9.7	6.0
Indirect effect	6.1	6.1	6.5	6.4	6.1	5.4	7.1	6.1	6.3

Note. Location parameters have the normal $N(0, 10^{10})$ prior, except for the vague normal and EB priors. Some results are not available (NA) for conditions that did not have at least 50% convergence. Mplus default = $\pi(\sigma^2) \propto 1$ combined with the normal $N(0, 10^{10})$ prior; $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters; IG(.001, .001), IG(.01, .01), IG(.1, .1) = vague proper priors variance parameters; Vague normal = $\pi(\sigma^2) \propto 1$ prior for variance parameters combined with the normal $N(0, 1000)$ prior for measurement intercepts and the normal $N(0, 100)$ prior for the other location parameters; EB1 = Empirical Bayes prior location and variance parameters; EB2 = EB prior location parameters combined with $\pi(\sigma^2) \propto \sigma^{-1}$ prior variance parameters; ML = maximum likelihood estimation.

Of the noninformative improper priors, the Mplus default setting (i.e., $\pi(\sigma^2) \propto 1$ for variances coupled with $N(0, 10^{10})$ for location parameters) performed best in terms of bias for the location parameters but worse than $\pi(\sigma^2) \propto \sigma^{-1}$ for the variance parameters. In terms of coverage rates, the Mplus default setting outperformed the $\pi(\sigma^2) \propto \sigma^{-1}$ prior, especially for the variance parameters. For the *MSE* and quantiles, they did not differ substantially. When considering the direct and indirect effect, the parameters of practical interest, both noninformative improper priors performed well in terms of bias and coverage. Finally, although the Type I error rate for the direct effect was slightly too low for the Mplus default setting when $N = 35$, this result was not available for the $\pi(\sigma^2) \propto \sigma^{-1}$ prior, because of high nonconvergence. Based on these differences in performance between the noninformative improper priors, we thus recommend the Mplus default priors as general choice for BSEM, with the important observation that this setting does not perform perfectly.

Even though the different default priors that were investigated in this section are routinely used in practice, their performance varies greatly across conditions. Therefore, a prior sensitivity analysis is highly recommendable, especially in small samples when clear prior information is absent. The next section provides guidelines on how to perform such an analysis in default BSEM.

A Practical Guide to Prior Sensitivity Analysis

Given the results of the simulation study, and in line with recommendations by B. O. Muthén and Asparouhov (2012, p. 320), prior sensitivity analysis is an important step in BSEM. The goal of a prior sensitivity analysis is to assess whether the results of the BSEM analysis are influenced by the specific default prior that is used. When the conclusions are similar using the different default priors, we can be confident that the results are reliable and robust to default prior specification. On the other hand, if the different default approaches result in substantially different conclusions, some care must be taken regarding the reliability of the results.

A prior sensitivity analysis can be conducted by rerunning the analysis with different choices for the prior. Because of the large number of parameters, possible prior choices, and possible settings for each choice, conducting a sensitivity analysis can become quite involved. Nevertheless, prior sensitivity analysis is particularly relevant in the context of SEM. As a result of the complex relationships inherent in structural equation models, a prior on a specific parameter can indirectly influence other parts of the model as well. Consequently, the effects of the prior for different parameters may cancel out but can also accumulate. Moreover, some parameters are more influenced by their prior distribution (e.g., variances of latent variables) compared with other parameters (e.g., residual variances). If a parameter is highly influenced by its prior distribution, this influence can carry through the model and affect other parameters as well. Depaoli and van de Schoot (2015) provided guidelines on conducting prior sensitivity analyses for Bayesian analyses with informative priors. The goal of this section is to provide a step-by-step guide on how to conduct a prior sensitivity analysis in BSEM using default priors. This analysis is recommended when prior information is weak, or when a researcher prefers to exclude external information in the statistical analysis. We will illustrate the guidelines on the democracy and industrialization model from Section 2.

Step 1: Decide Which Parameters to Investigate

The first step in conducting a sensitivity analysis for structural equation models is to decide which parameters to focus on. Although it is important to change the prior on each parameter, there are generally only a few parameters of substantive interest (e.g., the direct and indirect effect in the model considered throughout this article). Therefore, we recommend focusing primarily on the parameters of substantive interest in determining the sensitivity to the prior. Which parameters are of interest will, of course, vary across different applications. In addition to the parameters of substantive interest, we recommend focusing on the latent variable variances in the model as well, as these are generally most sensitive to the choice of the prior. These variance parameters can therefore unduly influence other parameters as well. In this first step, it is also helpful to consider which magnitudes of changes in the parameter values would constitute meaningful differences in the parameters. These magnitudes will be used in Step 4 to determine when a parameter is sensitive to the choice of the prior.

Step 2: Decide Which Priors to Include

The second step consists of deciding which priors to include. Software such as Mplus limits the choice of possible priors by allowing a limited set of prior choices, such as normal priors for location parameters (e.g., intercepts, regression coefficients) and inverse Gamma priors for variance parameters. Given the large number of parameters in the model, it is infeasible to alter the prior for each parameter one at a time. In addition, it is more realistic to change the prior for all parameters in the model simultaneously because, in general, researchers will specify the same type of default prior for all parameters in the model.

Based on the results of the simulation study, we recommend including the following default priors in the prior sensitivity analysis: the noninformative improper priors $\pi(\sigma^2) \propto \sigma^{-1}$ and $\pi(\sigma^2) \propto 1$, and the vague proper priors $IG(\epsilon, \epsilon)$ with $\epsilon = 0.001$, 0.01, and 0.1 for variance parameters, combined with the vague proper prior $N(0, 10^{10})$ for location parameters; the vague normal prior; and the EB priors. Note that it is important to consider multiple values for ϵ when considering the vague proper prior $IG(\epsilon, \epsilon)$, as the choice of ϵ can have a large influence on the results. When the results are not robust to the exact choice of ϵ, we do not recommend using these priors for drawing substantive conclusions. On the other hand, when the results are robust to the choice of ϵ, the results are reliable for drawing substantive conclusions. We do not recommend including the improper prior $\pi(\sigma^2) \propto \sigma^{-2}$, because of its severe nonconvergence.

Furthermore, when prior knowledge is available, a researcher can use an informative prior. This prior can be specified by choosing the hyperparameters in such a way that the resulting prior has high probability on those parameter values deemed plausible by previous research or by an expert in the field. The challenge in specifying informative priors is to specify the hyperparameters such that the prior probability that the parameter falls in a plausible parameter region equals a certain percentage, for example, 95%. When informative priors are used, we follow the recommendation of Depaoli and van de Schoot (2015) to compare the results of the informative priors with results obtained using default priors.

Step 3: Technical Implementation (Mplus)

The R package MplusAutomation (Hallquist & Wiley, 2014) can be used to automatically create and run the Mplus input files for the analyses with different priors. Subsequently, the results of all analyses can be read into R simultaneously. In the online supplemental materials we provide the code for our sensitivity analysis, which can be used as a template for a prior sensitivity analysis using MplusAutomation.

One issue when conducting the analyses in an automatic way is how to assess convergence. When using MCMC sampling, it is important to ensure that the chains converge to the posterior distribution. Mplus provides an automatic criterion based on the PSR. Sampling will continue until the cutoff for the PSR defined in the BCONVERGENCE option is reached, or before that if the maximum number of iterations is reached (specified through the BITERATIONS option). The maximum number of iterations should depend on the model under consideration, with more complex models requiring a larger number of iterations, and preliminary analyses can be conducted to get an indication of the required number of iterations. In addition, when using the PSR, it is recommended to rerun the analysis with twice as many iterations to avoid preliminary fulfillment of the PSR criterion. More information on the PSR can be found in Gelman and Rubin (1992) or the Mplus user guide (L. K. Muthén & Muthén, 1998–2012). A second option is to not rely on the automatic criterion to determine the number of iterations, but to specify a fixed number of iterations (using the FBITERATIONS options) and subsequently assessing convergence diagnostics, such as the PSR. Again, preliminary analyses can be conducted to get an indication of the required number of iterations. Note that there are other methods to automatically assess convergence, for example, blavaan has a setting that relies on the PSR in combination with Raftery and Lewis's (1992) convergence diagnostic to determine the number of draws. Regardless of which automated criterion is used to assess convergence, it is highly recommended to check the trace plots of the posterior draws for all parameters.

Step 4: Interpretation of the Results

The marginal posterior distributions for each parameter in the model can be summarized in different ways. As Bayesian point estimates, the mean, median, or mode can be used. By default, Mplus provides the posterior median, which is also the summary we used. In addition, we considered the 95% credible interval. As noted in Step 1, in order to conclude whether the results are sensitive to the prior, the researcher must first decide what constitutes a meaningful difference in parameters of interest, based on the application at hand. In other words, boundaries must be specified for the changes in results across the priors; if a change in a parameter exceeds this boundary, the parameter can be classified as sensitive. Changes can be evaluated by comparing the results obtained with a specific prior with the results obtained with the original prior distribution. To define a meaningful boundary, it may be helpful to set the bounds on the standardized estimates, which are generally easier to interpret. In addition, because the standardized estimates automatically include the scale of the variables, only the sensitivity of the latent mean, intercept, loading, and regression parameters needs to be considered. However, other options are possible, for example, the threshold can be based on qualitative differences. One such option is to classify a parameter as sensitive if a different prior results in a different sign of the estimate, or if the expected parameter change (EPC; Saris, Satorra, & van der Veld, 2009), or EPC-interest (Oberski, 2014) exceeds a certain cutoff. The EPC estimates the change in a parameter when relaxing a constraint, while the EPC-interest estimates the expected change in the parameter of interest. Sensitivity with respect to other outcomes may also be evaluated, such as whether the credible interval includes zero or not; or whether model fit measures, such as the posterior predictive p value (Gelman, Meng, & Stern, 1996), exceed a threshold such as 0.05.

The results of the prior sensitivity analysis can fall into one of three categories: (1) the results are not sensitive to the choice of the prior; (2) the results obtained using default priors do not vary, but these results differ from the result obtained using informative priors; and (3) the results vary across all priors, both default and informative. In the first scenario, we can conclude that the results are robust to the choice of the prior. In the second scenario, the information embedded in the informative prior influences the results, as would be expected. As noted by Berger (2006), subjective prior elicitation is difficult and can generally only provide certain characteristics of the prior (e.g., the location), whereas other features (e.g., the parametric form) are typically chosen in a convenient way. Thus, in the second scenario, the researcher should be certain that the chosen informative prior is an accurate reflection of one's prior belief, in terms of the prior guess (i.e., the prior mean) and the prior uncertainty (i.e., the prior variance and prior's distributional form). If the certainty about the informative prior cannot be warranted, the analysis based on default priors is recommended for substantive conclusions. If the informative prior is an accurate representation of the prior beliefs, the results from this prior can be used for final conclusions, and the results of the default analyses can be used as a reference.

In the third scenario, there is an additional difficulty that the results are also not robust to the different default priors. This may occur when the sample is relatively small and the prior is (possibly unintentionally) relatively informative. In this situation, one option is to collect more data. The advantage of Bayesian analysis is that we can simply collect additional data and combine them with the original data, whereas classical methods, such as confidence intervals and *p* values, require a fixed sampling plan before data analysis (e.g., Robert, 2007, p. 23). Thus, the researcher can continue to collect additional observations until the results are no longer sensitive to the priors. However, it is not always feasible or possible to collect more data. In that case, we recommend that the researcher reports all results or the range of results obtained using the different default priors. The range can be computed by first combining all posterior draws from the different priors, and then computing the median and the lower and upper bounds of the 95% credible interval of the combined set of posterior draws. In addition, differences between default priors can be examined graphically, for example, using boxplots such as those shown in Figure 7. The interpretation should then focus on how the substantive conclusions (e.g., the effect size of the indirect effect) vary across the default priors. This mirrors the recommendation of Leamer (1983) and relates to robust Bayesian analysis (Berger, 2006), in which the results from multiple prior distributions are combined to obtain a range of results.

Note that the three scenarios can occur for different parameters. If results between priors vary only for the nuisance parameters and not for the parameters of interest (i.e., the direct and indirect effect), reliable conclusions can be drawn about these parameters of interest. Only when the researcher wishes to draw conclusions about the complete fitted model, sensitivity of the nuisance parameters to the priors should be taken into account.

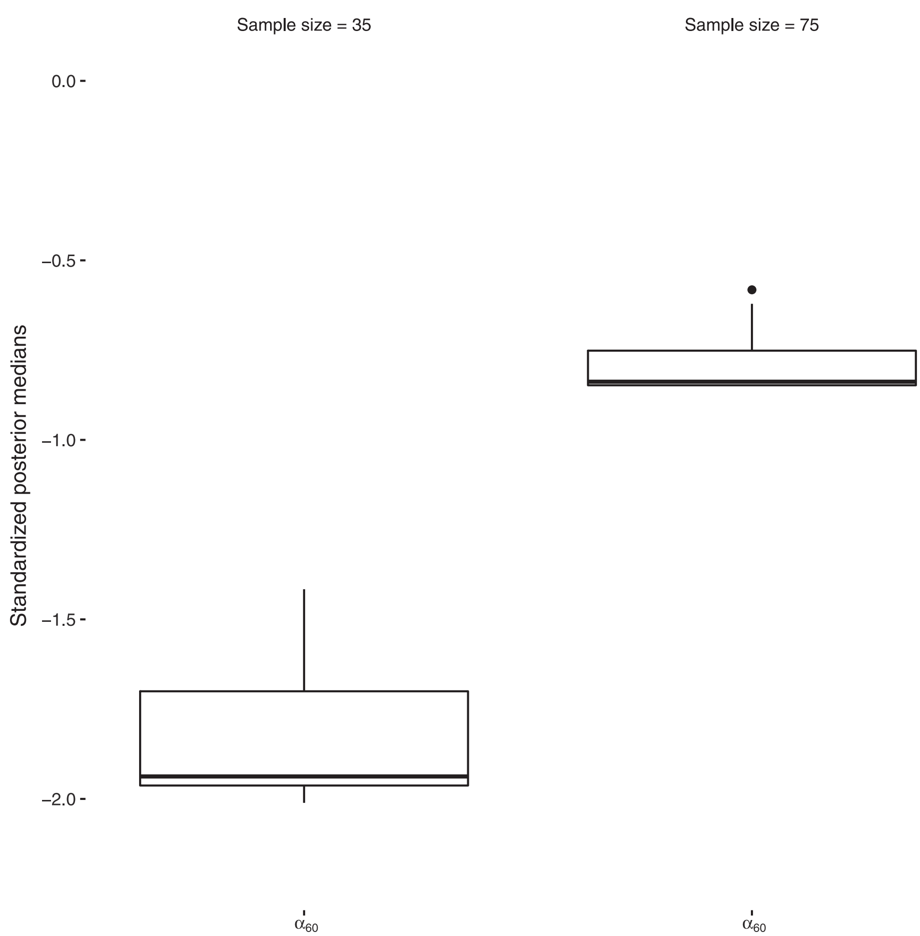

Figure 7. Standardized posterior medians for α_{60} in the prior sensitivity analysis.

Empirical Application: Democracy and Industrialization Data

We applied these steps to the original data from the democracy and industrialization application, which has a sample size of 75. In addition, we took the first 35 observations of the original data to illustrate a prior sensitivity analysis in a situation in which the results are quite sensitive to the choice of the prior. For comparison, we also include the ML estimates in the results.

Step 1

In this specific application, the parameters of substantive interest are the direct effect γ_{65} and the indirect effect $\gamma_{60} \cdot b_{21}$. We decided that a standardized change of 0.1 would constitute a meaningful difference in the parameters. Note that the variables have been standardized with respect to the variances of both $\boldsymbol{y}$ and $\boldsymbol{x}$.

Step 2

We include in our sensitivity analysis the noninformative improper prior $\pi(\sigma^2) \propto \sigma^{-1}$ (combined with the vague proper prior $N(0, 10^{10})$ for location parameters), the Mplus default setting (i.e., $\pi(\sigma^2) \propto 1$ for variance parameters and $N(0, 10^{10})$ for location parameters), the vague proper priors (with $\epsilon = 0.1$, 0.01, and 0.001, combined with the vague proper prior $N(0, 10^{10})$ for location parameters), the vague normal prior, and the EB priors. The Mplus default prior setting is used as baseline to which the other priors are compared. In general, however, the original prior distribution will serve as baseline. Although we do not recommend the use of the improper prior $\pi(\sigma^2) \propto \sigma^{-2}$, we did include this prior in the sensitivity analysis for illustrative purposes.

In addition, informative priors are available in Dunson et al. (2005) for this model based on expert knowledge, which we included in our prior sensitivity analysis (see the online supplemental materials). We assume that these priors reflect the available prior knowledge and the certainty about this knowledge correctly.

Step 3

We ran the analyses using MplusAutomation. All files for our analyses are available in the online supplemental materials. We did not rely on the PSR criterion to determine the amount of iterations, but instead specified a fixed number of 75,000 iterations (through the FBITERATIONS option). For each analysis, we checked whether the PSR < 1.1 for all parameters and we eyeballed each trace plot to ensure convergence. An example of a trace plot illustrating convergence and the corresponding estimated posterior densities is available in the online supplemental materials.

Step 4

To assess sensitivity, we compared the standardized median for each prior with the standardized median obtained when using the Mplus default prior settings (the baseline). The Mplus default settings correspond to a normal prior, $N(0, 10^{10})$, for means and regression parameters and an improper prior, $\pi(\sigma^2) \propto 1$, for variances, implemented as an inverse Gamma prior, $IG(-1, 0)$. Note that, in practice, the original prior distributions will serve as baseline. As noted in Step 1, a standardized change of 0.1 would constitute a meaningful difference. Consequently, if the standardized median of a prior deviated more than 0.1 from the standardized median obtained under the baseline, we concluded that the results are sensitive to the prior. We will first discuss the results for the original data ($N = 75$), followed by the results for $N = 35$.

Results: Original Data ($N = 75$)

Table 8 shows the standardized and unstandardized ML estimates and posterior medians for the direct effect for each analysis and sample size, as well as the 95% confidence and credible intervals. Standardized estimates that deviate more than 0.1 from the estimates obtained with the Mplus default setting are presented in bold. It is clear that for $N = 75$, none of the estimates exceed the cutoff, and thus we can conclude that the direct effect is not sensitive to the prior. There are differences between the priors in terms of credible intervals. Specifically, for the informative prior, the lower bound is negative, whereas it is positive for all other priors and ML estimation. Consequently, a test for the direct effect using the informative prior would result in the conclusion that the effect is not substantially different from zero, whereas the other priors and ML estimation would lead to this conclusion. Note that the informative prior results in the smallest credible interval because of the additional information added through the prior. In addition, the EB priors have slightly smaller credible intervals compared with the other default priors, because the EB priors include more prior information than the default priors.

Table 9 shows the point estimates and 95% confidence and credible intervals for the indirect effect for each analysis and sample size. None of the estimates exceed the cutoff of 0.1, and thus the indirect effect is not sensitive to the choice of the prior. The confidence and credible intervals for the indirect effect show less variation compared with the direct effect so that conclusions based on interval testing for the indirect effect would not differ across the Bayesian methods or ML estimation. Only the width of the credible intervals differ, with the informative prior showing the smallest credible intervals, followed by the EB priors.

Results: Subset of Original Data ($N = 35$)

For the subset of 35 observations from the original data, the ML analysis led to empirical weak identification and inadmissible estimates because of the small sample size, and thus these results are excluded. The standardized and unstandardized posterior medians, and 95% credible intervals for the direct effect, are presented in Table 8. The standardized median for the informative prior is shown in bold, indicating that this estimate differs more than 0.1 from the estimate obtained under the Mplus default setting. In addition, the informative prior is the only prior resulting in a negative lower bound of the 95% credible interval. Thus, for $N = 35$, the direct effect is sensitive to the choice of the prior. The results for the indirect effect are presented in Table 9. Again, the indirect effect is less sensitive to the choice of the prior. None of the standardized estimates exceed the cutoff, and the only clear difference between the priors is that the width of the credible interval is smallest for the informative prior, followed by the EB priors.

Table 8
Standardized and Unstandardized Point Estimates and 95% Confidence and Credible Intervals for the Direct Effect γ_{65} in the Prior Sensitivity Analysis

Prior	Standardized estimate	Unstandardized estimate	Lower bound 95% CI	Upper bound 95% CI	Width 95% CI
		Sample size = 35			
Mplus default	.299	1.137	.132	2.193	2.061
$\pi(\sigma^2) \propto \sigma^{-1}$	.284	1.052	.090	2.087	1.997
IG(.001, .001)	.270	.990	.074	2.041	1.967
IG(.01, .01)	.278	1.019	.059	2.029	1.970
IG(.1, .1)	.283	1.052	.088	2.053	1.965
Vague normal	.293	1.085	.114	2.086	1.972
EB1	.270	.975	.160	1.741	1.581
EB2	.274	.997	.137	1.812	1.675
Informative	**.090**	.225	−.427	.791	1.218
		Sample size = 75			
Mplus default	.183	.574	.098	1.092	.994
$\pi(\sigma^2) \propto \sigma^{-1}$	.177	.554	.082	1.046	.964
IG(.001, .001)	.174	.552	.089	1.023	.934
IG(.01, .01)	.175	.549	.078	1.031	.953
IG(.1, .1)	.177	.555	.084	1.052	.968
Vague normal	.182	.569	.096	1.080	.984
EB1	.153	.475	.055	.873	.818
EB2	.158	.488	.064	.910	.846
Informative	.109	.288	−.126	.678	.804
ML	.182	.572	.114	1.030	.916

Note. Location parameters have the normal $N(0, 10^{10})$ prior, except for the vague normal, EB, and informative priors. Standardized estimates deviating more than .1 from the estimate obtained under the Mplus default prior settings are shown in bold. Mplus default = $\pi(\sigma^2) \propto 1$ combined with the normal $N(0, 10^{10})$ prior; $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters; IG(.001, .001), IG(.01, .01), IG(.1, .1) = vague proper priors variance parameters; Vague normal = $\pi(\sigma^2) \propto 1$ prior for variance parameters combined with the normal $N(0, 1000)$ prior for measurement intercepts and the normal $N(0, 100)$ prior for the other location parameters; EB1 = Empirical Bayes prior location and variance parameters; EB2 = EB prior location parameters combined with $\pi(\sigma^2) \propto \sigma^{-1}$ prior variance parameters; ML = maximum likelihood estimation; CI = confidence/credible interval.

Conclusions From the Prior Sensitivity Analysis

Based on the scenarios described in Step 4 of the prior sensitivity guide, we can thus conclude that for $N = 75$, the estimates of the parameters of interest (i.e., the direct and indirect effect) are robust to the choice of the prior (the first scenario). However, the credible interval for the direct effect included zero for the informative prior, whereas it did not include zero for the default priors. Thus, when testing the direct effect, we find ourselves in the second scenario. The same holds for $N = 35$, in which both the estimate and credible interval of the direct effect were sensitive to the informative prior. Therefore, careful consideration of the informative prior for the direct effect is necessary. The informative prior for the direct effect was the normal prior $N(0.5, 2)$, which results in 95% prior probability on the interval (−3.43, 4.42; see the online supplemental materials). Compared with the default priors, which are more spread out, the informative prior shrinks the estimate for the direct effect toward the prior mean, resulting in a smaller estimate. If the informative prior has been specified with care and accurately reflects the prior beliefs (we assume this was the case), the results obtained with the informative prior can be used for substantive conclusions, which implies no significant direct effect. The default analysis, which suggests a significant direct effect, can be reported as a reference analysis to show that the information in the data implies a significant direct effect.

For both sample sizes, the nuisance parameters were sensitive to the default priors as well. Thus, if the goal of the analysis is to draw conclusions about the full model, the third scenario is applicable. If no informative priors were specified, and if it is not possible to collect more data, the researcher should consider and report the (range of) results from all default priors. By combining the posterior draws from all default priors and computing the median and bounds of the 95% credible interval, we can obtain a range for all parameters, which is reported in Table 10. Some of the credible intervals based on all posterior draws are very wide. This common behavior of a robust Bayesian analysis (e.g., Berger, 2006) can be explained by the fact that there is very little information in the data to fit the relatively complex SEM model.

Additionally, we can examine the differences between the default priors graphically, for example, by plotting the standardized posterior medians for each parameter, as is done in Figure 7 for the structural intercept α_{60}. From Figure 7, we can see that for $N = 35$, the estimated medians vary from −1.4 to −2 and the researcher should further examine these differences between the priors. For example, in this case, the smallest estimates are obtained using the EB priors, whereas the improper and vague proper priors generally result in estimates close to −2 and the vague normal prior lies in between. Of the default priors, the EB priors are most informative, as they include information regarding the ML estimates. The improper and vague proper priors are least informative, as they have the largest posterior variance, and the vague normal prior lies in between. Thus, for more informative default priors, the estimate for α_{60} becomes smaller. If informative priors were specified and the third scenario is applicable, the researcher should carefully consider each of the informative priors and, if in doubt concerning

Table 9
Standardized and Unstandardized Point Estimates and 95% Confidence and Credible Intervals for the Indirect Effect $\gamma_{60} \cdot b_{21}$ in the Prior Sensitivity Analysis

Prior	Standardized estimate	Unstandardized estimate	Lower bound 95% CI	Upper bound 95% CI	Width 95% CI
			Sample size = 35		
Mplus default	.430	1.572	.556	2.938	2.382
$\pi(\sigma^2) \propto \sigma^{-1}$	.446	1.599	.580	2.962	2.382
IG(.001, .001)	.459	1.627	.603	2.992	2.389
IG(.01, .01)	.456	1.622	.609	2.996	2.387
IG(.1, .1)	.452	1.622	.623	2.976	2.353
Vague normal	.422	1.505	.537	2.798	2.261
EB1	.405	1.417	.537	2.546	2.009
EB2	.387	1.353	.473	2.487	2.014
Informative	.461	1.125	.593	1.897	1.304
			Sample size = 75		
Mplus default	.385	1.191	.522	1.982	1.460
$\pi(\sigma^2) \propto \sigma^{-1}$	.394	1.221	.548	2.017	1.469
IG(.001, .001)	.399	1.241	.569	2.050	1.481
IG(.01, .01)	.397	1.229	.561	2.035	1.474
IG(.1, .1)	.393	1.208	.553	2.011	1.458
Vague normal	.383	1.177	.522	1.950	1.428
EB1	.376	1.150	.592	1.801	1.209
EB2	.364	1.108	.554	1.751	1.197
Informative	.377	.988	.582	1.482	.900
ML	.396	1.242	.542	1.941	1.399

Note. Location parameters have the normal $N(0, 10^{10})$ prior, except for the vague normal, EB, and informative priors. Standardized estimates deviating more than .1 from the estimate obtained under the Mplus default prior settings are shown in bold. Mplus default = $\pi(\sigma^2) \propto 1$ combined with the normal $N(0, 10^{10})$ prior; $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters; IG(.001, .001), IG(.01, .01), IG(.1, .1) = vague proper priors variance parameters; Vague normal = $\pi(\sigma^2) \propto 1$ prior for variance parameters combined with the normal $N(0, 1000)$ prior for measurement intercepts and the normal $N(0, 100)$ prior for the other location parameters; EB1 = Empirical Bayes prior location and variance parameters; EB2 = EB prior location parameters combined with $\pi(\sigma^2) \propto \sigma^{-1}$ prior variance parameters; ML = maximum likelihood estimation; CI = confidence/credible interval.

whether the prior accurately reflects the prior belief, the researcher should consider the results of all default priors.

Discussion

Bayesian methods are a useful alternative to ML estimation for structural equation models. In the case of small samples, ML estimation can result in empirical weak identification and inadmissible estimates, whereas BSEM analyses can prevent these problems. In order to use the BSEM framework, however, prior distributions must be specified for the model parameters. In this article, we focused on default priors that can be applied in an automatic fashion for a BSEM analysis when prior knowledge is absent or if a researcher does not wish to include external information. Based on the results, we recommend the Mplus default setting (i.e., the noninformative improper prior $\pi(\sigma^2) \propto 1$ for variance parameters, combined with the vague proper prior $N(0, 10^{10})$ for location parameters) as general default prior for BSEM. In general, we recommend against the use of the improper prior $\pi(\sigma^2) \propto \sigma^{-2}$, as it suffers from major convergence problems, and against the vague proper priors, which approximate this improper prior and consequently lead to instable MCMC estimation. The vague proper priors can be considered in the prior sensitivity analysis only when multiple values for the hyperparameters are included and the results do not vary across these choices. The performance of the different default priors varied greatly across conditions. For this reason, it is highly recommended to consider several default priors when performing a default BSEM analysis, to assess robustness of the results to the choice of the prior.

For $N = 35$, ML estimation performed better than the Bayesian methods in terms of bias. This can be explained by the fact that ML estimation does not force the separate variances to be positive, while the priors we have considered are only defined in the region in which the separate variances are positive. In the case of small samples, the likelihood has support for negative variances while the default priors give probability zero to negative variances to obtain interpretable estimates at the cost of introducing some bias. Given that variance parameters are often nuisance parameters, one might argue that minimizing bias is preferred over interpretable estimates. It would be interesting to adopt a Bayesian approach in which priors for the variances have support in the negative subspace of certain variance parameters and compare the bias with ML estimation (e.g., Mulder & Fox, 2013). Generally, however, ML estimation cannot be recommended for small samples ($N = 35$) because of the low coverage rates for the variance parameters and high Type I error rates for testing direct effects. For larger samples (i.e., $N = 75$, 150, 500), ML performed well in terms of all outcome measures and generally outperformed the Bayesian methods. Therefore, if ML estimation is feasible, it is recommended for large samples ($N \geq 75$). Note, however, that for more complex models than the SEM studied here, new simulations have to be performed to check whether the sample size of the data at hand is large enough to fit this more complex model using ML. Additionally, ML estimation has the disadvantage that confidence

Table 10
Posterior Medians and Lower and Upper Bounds of the 95% Credible Interval Based on All Posterior Draws from Sensitivity Analyses With Default Priors

Parameter	Lower bound 95% CI	Median	Upper bound 95% CI	Lower bound 95% CI	Median	Upper bound 95% CI
		Sample size = 35			Sample size = 75	
γ_{60}	.211	.538	.771	.211	.428	.617
γ_{65}	.032	.281	.530	.032	.172	.323
b_{21}	.541	.780	.953	.541	.885	.971
$\gamma_{60} \cdot b_{21}$	.114	.420	.735	.114	.379	.599
λ_1^y	.609	.845	.955	.609	.845	.919
λ_2^y	.492	.737	.886	.492	.692	.811
λ_3^y	.456	.721	.871	.456	.717	.826
λ_4^y	.637	.842	.951	.637	.831	.909
λ_5^y	.683	.850	.936	.683	.802	.882
λ_6^y	.528	.758	.887	.528	.726	.832
λ_7^y	.668	.839	.931	.668	.818	.894
λ_8^y	.483	.718	.861	.483	.812	.892
λ_1^x	.887	.950	.986	.887	.917	.957
λ_2^x	.935	.985	1.000	.935	.973	.998
λ_3^x	.742	.868	.935	.742	.867	.919
μ_ξ	5.781	7.775	10.120	5.781	7.415	9.038
α_{60}	−3.750	−1.694	.898	−3.750	−.723	.972
α_{65}	−3.458	−1.791	−.179	−3.458	−.999	.022
ν_2^y	−1.959	−.741	.015	−1.959	−.600	−.086
ν_3^y	−.637	.258	1.091	−.637	.213	.768
ν_4^y	−1.575	−.308	.395	−1.575	−.692	−.226
ν_6^y	−1.600	−.850	−.223	−1.600	−.857	−.370
ν_7^y	−.649	.038	.663	−.649	−.108	.384
ν_8^y	−1.076	−.328	.359	−1.076	−.704	−.238
ν_2^x	−5.936	−4.254	−3.001	−5.936	−4.060	−3.095
ν_3^x	−5.515	−3.969	−2.627	−5.515	−3.901	−2.955

intervals depend on the sampling plan. This implies that optional stopping or deciding to collect more data because of inaccurate results (i.e., wide confidence intervals) is not straightforward to incorporate in a classical analysis using confidence intervals (e.g., Robert, 2007, p. 23). Bayesian credible intervals, on the other hand, abide by the likelihood principle (Berger & Wolpert, 1984), and therefore the results are invariant of the sampling plan.

We have proposed two EB priors that are novel in BSEM. Although the EB priors performed best in terms of *MSE*, they did not perform well in terms of bias, coverage rates, quantiles, and Type I error rates. Several studies have found that EB priors can result in an underestimation of the posterior variance (Carlin & Louis, 2000a; Darnieder, 2011; Efron, 1996), which can partly explain the low coverage rates. Furthermore, the EB priors considered in this article were developed to be generally applicable. For example, the proposed EB prior for location parameters is centered around zero, with the prior variance chosen such that the prior has positive support where the likelihood is concentrated. However, for some parameters such as intercepts, the prior mean of zero might not be realistic and can lead to biased estimates. In addition, it may be that in certain extreme situations (e.g., when the error variances are approximately zero), data-dependent priors, such as our EB priors, should not be used. In general, we believe that the EB methodology offers interesting possibilities for BSEM; however, more research is needed for further development of good EB priors.

We provided guidelines on how to conduct a (default) prior sensitivity analysis in Mplus and illustrated these guidelines on a structural equation model from the literature. An important step is choosing which parameters are of substantive interest. If these parameters are insensitive to the default priors, the estimates can be readily interpreted even if the estimates of some nuisance parameters show prior sensitivity. If the estimated parameters of interest are sensitive to the default priors, that is, they differ more than the chosen threshold value, the (range of) results of all default priors should be reported. To obtain robust bounds for the interval estimates, we recommend combining the posterior draws of the different default Bayesian analyses, and subsequently reporting the upper and lower bounds of the 95% credible intervals based on the combined set of draws (e.g., Berger, 2006).

We investigated only conditionally conjugate priors because these are available in Mplus. However, many nonconjugate priors have been proposed in the Bayesian literature as more robust (i.e., less influential) alternatives. For example, Gelman (2006) and Polson and Scott (2012) proposed the half-Cauchy prior for random effects variances, which can be implemented in a Gibbs sampler relatively easy through parameter expansion. A second option for random effects variances is a Gamma prior in combination with posterior mode estimates, which has been proposed in the context of meta-analysis by Chung, Rabe-Hesketh, and Choi (2013). Note that the choice of the prior for residual variances is considerably less important than the prior for the variances of latent variables (e.g., Polson & Scott, 2012). For intercept, mean, and regression parameters a robust alternative is the *t*-distribution, which has been proposed as prior for logistic models by Gelman, Jakulin, Pittau, and Su (2008), and as error distribution to obtain robust models (e.g., robust growth curve models; Zhang, Lai, Lu, & Tong, 2013). The *t*-distribution includes the Cauchy distribution as special case when the number of degrees of freedom is set to 1. These priors should be investigated in the context of BSEM to assess their performance and determine whether they can be used as default priors.

References

Arbuckle, J. L. (2013). *Amos 22 user's guide*. Chicago, IL: IBM SPSS.

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

Bayarri, M. J., & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science, 19,* 58–80.

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

Berger, J. O., & Strawderman, W. E. (1996). Choice of hierarchical priors: Admissibility in estimation of normal means. *The Annals of Statistics, 24,* 931–951.

Berger, J. O., & Wolpert, R. L. (1984). *The likelihood principle*. Hayward, CA: Institute of Mathematical Statistics.

Bollen, K. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review, 45,* 370–390.

Bollen, K. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Can, S., van de Schoot, R., & Hox, J. (2014). Collinear latent variables in multilevel confirmatory factor analysis: A comparison of maximum likelihood and Bayesian estimations. *Educational and Psychological Measurement, 75,* 406–427.

Carlin, B. P., & Louis, T. A. (2000a). *Bayes and empirical Bayes methods for data analysis* (2nd ed.). New York, NY: Chapman and Hall/CRC.

Carlin, B. P., & Louis, T. A. (2000b). Empirical Bayes: Past, present and future. *Journal of the American Statistical Association, 95,* 1286–1289.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1), 1–32.

Carter, G. M., & Rolph, J. E. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities. *Journal of the American Statistical Association, 69,* 880–885.

Casella, G. (1985). An introduction to empirical Bayes data analysis. *The American Statistician, 39,* 83–87.

Casella, G. (1992). Illustrating empirical Bayes methods. *Chemometrics and Intelligent Laboratory Systems, 16,* 107–125.

Chung, Y., Rabe-Hesketh, S., & Choi, I.-H. (2013). Avoiding zero between-study variance estimates in random-effects meta-analysis. *Statistics in Medicine, 32,* 4071–4089.

Dagne, G. A., Howe, G. W., Brown, C. H., & Muthén, B. O. (2002). Hierarchical modeling of sequential behavioral data: An empirical Bayesian approach. *Psychological Methods, 7,* 262–280.

Darnieder, W. F. (2011). *Bayesian methods for data-dependent priors* (Unpublished doctoral dissertation). The Ohio State University, Columbus, OH.

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

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

Depaoli, S. (2014). The impact of inaccurate informative priors for growth parameters in Bayesian growth mixture modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 21,* 239–252.

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

Depaoli, S., & van de Schoot, R. (2015). Improving transparency and replication in Bayesian statistics: The WAMBS-checklist. *Psychological Methods, 22,* 240–261.

Dunson, D. B., Palomo, J., & Bollen, K. (2005). *Bayesian structural equation modeling* (Technical report). Retrieved from http://www.samsi.info/sites/default/files/tr2005-05.pdf

Efron, B. (1996). Empirical Bayes methods for combining likelihoods. *Journal of the American Statistical Association, 91,* 538–550.

Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association, 100,* 680–701.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19,* 72–91.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis, 1,* 515–534.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London, UK: Chapman and 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,* 1360–1383.

Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 6,* 733–807.

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

Hallquist, M., & Wiley, J. (2014). *MplusAutomation: Automating Mplus model estimation and interpretation [Computer software manual]*. Retrieved from http://CRAN.R-project.org/package=MplusAutomation (R package version 0.6–3)

Harring, J. R., Weiss, B. A., & Hsu, J.-C. (2012). A comparison of methods for estimating quadratic effects in nonlinear structural equation models. *Psychological Methods, 17,* 193–214.

Heerwegh, D. (2014). *Small sample Bayesian factor analysis*. Retrieved from http://www.lexjansen.com/phuse/2014/sp/SP03.pdf

Hobert, J. P., & Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association, 91,* 1461–1473.

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

Kaplan, D. (2014). *Bayesian statistics for the social sciences*. New York, NY: Guilford Press.

Kass, R. E., & Natarajan, R. (2006). A default conjugate prior for variance components in generalized linear mixed models (comment on article by Browne and Draper). *Bayesian Analysis, 1,* 535–542.

Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association, 91,* 1343–1370.

Kohli, N., Hughes, J., Wang, C., Zopluoglu, C., & Davison, M. L. (2015). Fitting a linear–linear piecewise growth mixture model with unknown knots: A comparison of two common approaches to inference. *Psychological Methods, 20,* 259–275.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38,* 963–974.

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

Leamer, E. E. (1983). Let's take the con out of econometrics. *The American Economic Review, 73,* 31–43.

Lee, S.-Y., & Song, X.-Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research, 39,* 653–686.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association, 103,* 410–423.

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2× 2 taxonomy of multilevel latent contextual models: Accuracy–bias trade-offs in full and partial error correction models. *Psychological Methods, 16,* 444–467.

Lüdtke, O., Robitzsch, A., Kenny, D. A., & Trautwein, U. (2013). A general and flexible approach to estimating the social relations model using Bayesian methods. *Psychological Methods, 18,* 101–119.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10,* 325–337.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1,* 86–92.

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

Merkle, E. C., & Rosseel, Y. (2015). *Blavaan: Bayesian structural equation models via parameter expansion*. Retrieved from https://arxiv.org/abs/1511.05604

Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods, 3,* 45–58.

Mulder, J., & Fox, J.-P. (2013). Bayesian tests on components of the compound symmetry covariance matrix. *Statistics and Computing, 23,* 109–122.

Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference, 140,* 887–906.

Mulder, J., Klugkist, I., van de Schoot, R., Meeus, W. H. J., Selfhout, M., & Hoijtink, H. (2009). Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology, 53,* 530–546.

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

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal, 9,* 599–620.

Natarajan, R., & Kass, R. E. (2000). Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association, 95,* 227–237.

Oberski, D. L. (2014). Evaluating sensitivity of parameters of interest to measurement invariance in latent variable models. *Political Analysis, 22,* 45–60.

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2011). A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological Methods, 16,* 468–490.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124). Vienna, Austria: Technische Universität Wien. Retrieved from https://www.r-project.org/conferences/DSC-2003/Proceedings/

Polson, N. G., & Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis, 7,* 887–902.

Raftery, A. E., & Lewis, S. (1992). How many iterations in the Gibbs sampler. *Bayesian statistics, 4,* 763–773.

R Core Team. (2015). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Robert, C. P. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementations*. New York, NY: Springer Science+Business Media.

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

Rubin, D. B. (1980). Using empirical Bayes techniques in the law school validity studies. *Journal of the American Statistical Association, 75,* 801–816.

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling, 16,* 583–601.

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling, 16,* 561–582.

Tversky, A. (1974). Assessing uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological), 36,* 148–159.

van de Schoot, R., Winter, S., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods, 22,* 217–239.

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods, 14,* 301–322.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In P. K. Goel & A. Zellner (Eds.), *Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti* (pp. 233–243). Amsterdam, the Netherlands: Elsevier.

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

(*Appendix follows*)

Appendix

Industrialization and Political Democracy Model in Matrix Form

The structural model (for $i = 1, \ldots, n$) is given in matrix form as

$$\begin{pmatrix} \eta_i^{60} \\ \eta_i^{65} \end{pmatrix} = \begin{pmatrix} \alpha^{60} \\ \alpha^{65} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ b_{21} & 0 \end{pmatrix} \begin{pmatrix} \eta_i^{60} \\ \eta_i^{65} \end{pmatrix} + \begin{pmatrix} \gamma^{60} \\ \gamma^{65} \end{pmatrix} \xi_i + \begin{pmatrix} \zeta_i^{60} \\ \zeta_i^{65} \end{pmatrix},$$

with ξ_i representing industrialization level in country i in 1960, and η_i^{60} and η_i^{65} representing political democracy in country i in 1960 and 1965, respectively. The parameters of interest in this model are the direct and indirect effect of industrialization in 1960 on political democracy in 1965, γ_{65} and $\gamma_{60} \cdot b_{21}$, respectively.

The measurement model for political democracy is given by

$$\begin{pmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \\ y_{4i} \\ y_{5i} \\ y_{6i} \\ y_{7i} \\ y_{8i} \end{pmatrix} = \begin{pmatrix} 0 \\ \nu_2^y \\ \nu_3^y \\ \nu_4^y \\ 0 \\ \nu_6^y \\ \nu_7^y \\ \nu_8^y \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ \lambda_2^y & 0 \\ \lambda_3^y & 0 \\ \lambda_4^y & 0 \\ 0 & 1 \\ 0 & \lambda_6^y \\ 0 & \lambda_7^y \\ 0 & \lambda_8^y \end{pmatrix} \begin{pmatrix} \eta_i^{60} \\ \eta_i^{65} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} D_i^{15} \\ D_i^{24} \\ D_i^{26} \\ D_i^{37} \\ D_i^{48} \\ D_i^{68} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i}^y \\ \epsilon_{2i}^y \\ \epsilon_{3i}^y \\ \epsilon_{4i}^y \\ \epsilon_{5i}^y \\ \epsilon_{6i}^y \\ \epsilon_{7i}^y \\ \epsilon_{8i}^y \end{pmatrix},$$

with $\boldsymbol{D}$ representing a vector of pseudo-latent variables used to model the correlations between measurement errors in such a way that the covariance matrix Σ_y remains a diagonal matrix.

The measurement model for industrialization level is given by

$$\begin{pmatrix} x_{1i} \\ x_{2i} \\ x_{3i} \end{pmatrix} = \begin{pmatrix} 0 \\ \nu_2^x \\ \nu_3^x \end{pmatrix} + \begin{pmatrix} 1 \\ \lambda_2^x \\ \lambda_3^x \end{pmatrix} \xi_i + \begin{pmatrix} \delta_{1i}^x \\ \delta_{2i}^x \\ \delta_{3i}^x \end{pmatrix}.$$

Received April 29, 2016
Revision received June 14, 2017
Accepted June 30, 2017 ■