

Prior Sensitivity Analysis in Default Bayesian Structural Equation Modeling

Sara van Erp, Joris Mulder & Daniel L. Oberski

Tilburg University, The Netherlands

Abstract

Bayesian structural equation modeling (BSEM) has recently gained popularity because it enables researchers to fit complex models while solving some of the issues often encountered in classical 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 may harm the accuracy of the estimates.

In this paper, 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 an extensive simulation study, we find that these three default BSEM methods 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 (B)SEM literature and all code for conducting the prior sensitivity analysis is made available in the online appendix.

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

Prior Sensitivity Analysis in Default Bayesian Structural Equation Modeling

Author notes

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

This research was supported by an NWO Research Talent Grant.

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

Introduction

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—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), while 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 BSEM 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 with such accurately chosen priors has superior performance to frequentist SEM – wrongly chosen priors, 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). Therefore, instead of relying fully on expert judgements, researchers employing Bayesian analysis often attempt to choose the priors such that they are informative enough to yield BSEM's advantages, while not being so informative as to bias the results. "Noninformative" or "weakly informative" priors aim to achieve this goal without needing elicitation, although no general method of choosing a prior is guaranteed to accomplish this goal. For this reason B. O. Muthén and Asparouhov (2012, p. 320) recommended that researchers "vary the prior variance to study sensitivity in the results", a recommendation that echoes textbooks on Bayesian analysis (e.g. Gelman, Carlin, Stern, & Rubin, 2004, p. 189).

Previous research has investigated the performance of several weakly informative and noninformative 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 (Casella, 1985, 1992; Mulder, Hoijtink, & Klugkist, 2010), remain, to our knowledge, uninvestigated. Moreover, while 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, in press).

This article aims to further the practice and utility of BSEM by accomplishing two goals. First, we investigate the performance of several alternative default priors and compare them with the priors studied thus far, thereby investigating sensitivity of BSEM estimates to default prior choices in a simulation setting. Second, since we find that BSEM results can be highly sensitive to the choice of the default prior, we demonstrate an in-depth sensitivity analysis in an empirical example and provide practical guidelines on performing such an analysis in general BSEM

applications.

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, an extensive 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 paper we will consider a linear structural equation model with latent variables from the literature. The model (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, since 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 through pseudo-latent variables D , following Dunson, Palomo, and Bollen (2005).

[FIGURE 1]

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

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

The measurement model is given by:

$$\begin{aligned} \mathbf{y}_i &= \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta}_i + \mathbf{D}_i + \boldsymbol{\epsilon}_i^y && \text{with } \mathbf{D}_i \sim N(\mathbf{0}, \boldsymbol{\Omega}_D), \\ & && \text{and } \boldsymbol{\epsilon}_i^y \sim N(\mathbf{0}, \boldsymbol{\Sigma}_y) \\ \mathbf{x}_i &= \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi}_i + \boldsymbol{\delta}_i^x && \text{with } \boldsymbol{\delta}_i^x \sim N(\mathbf{0}, \boldsymbol{\Sigma}_x) \end{aligned}$$

Here, the structural mean and intercepts μ_ξ and α reflect the mean structure in the structural part of the model, while 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 \mathbf{B} and $\boldsymbol{\Gamma}$ represent the relations between the latent variables. The residual variances $\boldsymbol{\Sigma}_y$ and $\boldsymbol{\Sigma}_x$ 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. Appendix A 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; R Core Team, 2015), 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 due to a non-positive 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 due to the small sample size, whereas the second warning indicates an inadmissible parameter estimate; in this case the estimated variance of the pseudo-latent variable representing the relation between ϵ_4^y and ϵ_8^y , i.e., $\hat{\omega}_{D48}^2$, is negative. These warnings clearly illustrate that when using classical ML estimation, researchers may encounter certain problems which may be overcome by adopting a Bayesian approach. In order to apply the Bayesian

methodology, priors must be specified for each parameter vector.

Default priors for Bayesian SEM

We will focus on “default priors”, which are constructed in an automatic fashion and do not rely on substantive information. 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. 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. However, many different default priors exist and, consequently, different software packages use different default priors based on various heuristic arguments. For example, the commercial SEM software Mplus (L. K. Muthén & Muthén, 1998-2012) specifies a uniform improper prior for variance parameters by default, while the Bayesian modeling software WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) recommends vague proper inverse Gamma priors for the variances.

Three types of default priors are commonly used in the Bayesian literature: noninformative¹ improper priors, vague proper priors, and empirical Bayes priors. The first two have been used extensively in the BSEM literature, while the latter has, to our knowledge, not been applied to BSEM yet, but it is popular in the general literature on Bayesian modeling (Casella, 1985, 1992). 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. 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 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 above 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 paper 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. This prior can be written as the conjugate inverse Gamma prior with shape parameter $\alpha = 0$ and scale parameter $\beta = 0$ (see Appendix B).
- $p(\sigma^2) \propto \sigma^{-1}$. 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 unknown variance to obtain a proper posterior. It can be written as the conjugate inverse Gamma prior with shape parameter $\alpha = -\frac{1}{2}$ and scale parameter $\beta = 0$ (see Appendix B).

- $p(\sigma^2) \propto 1$. 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 unknown variance, at least three observations are needed to obtain a proper posterior for the variance. It can be written as the conjugate inverse Gamma prior with shape parameter $\alpha = -1$ and scale parameter $\beta = 0$ (see Appendix B).

Table 1 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 $+\infty$. Because this prior is not currently available to users of standard SEM software allowing for Bayesian estimation, such as Mplus, we will not consider it further in this paper. Thus, for the intercepts, means, loadings and regression coefficients, we will only investigate vague proper and empirical Bayes priors, which are discussed next.

[TABLE 1]

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, often conjugate forms, but the hyperparameters are chosen in a manner 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. For example, the latter option is used as default in WinBUGS. 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, specifically the normal prior $N(0, 10^{10})$, which is the default in Mplus. The vague proper priors that will be considered throughout this paper are summarized in Table 1.

A potential problem of vague proper priors is that the exact hyperparameters are arbitrarily chosen, while 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 8 schools on the second level, the posterior for the between-school variance was completely dominated by the inverse Gamma prior with small hyperparameters. It is yet unclear how this prior performs in structural equation models, which are considerably more complex than the 8 schools example studied by Gelman, in terms of the number of parameters and the relations between them.

Empirical Bayes priors

The third category of default priors that we consider consists of empirical Bayes priors (EB priors; Casella, 1985, 1992). The central idea behind the empirical Bayes methodology is that the hyperparameters are chosen based on the data at hand. The general idea of an EB prior is that it contains minimal information and has clear positive support in the region where the likelihood is concentrated. For example, a commonly used EB prior is centered around the ML estimate with relatively large prior variance. The main advantage of this approach arises in multiparameter problems where the information of all the data is combined to estimate each parameter, thereby resulting in more precise estimates (Casella, 1992).

Throughout this paper we will consider the following EB priors: ²

- EB prior for variances. This prior is specified as inverse Gamma with shape parameter $\alpha = \frac{1}{2}$, which implies that the prior carries the information equivalent to one data point (Gelman et al., 2004, p. 50). The scale parameter β is chosen in such a way that the prior median equals the ML estimate.
- EB prior for intercepts/means/factor loadings/regression coefficients. This prior is specified as normal with mean $\mu = 0$ and variance σ^2 equal to the squared ML estimate of that parameter plus the ML estimate of the residual variance of the model equation for that parameter. Thus, the prior is centered around zero, and two ML estimates are used to determine the prior variance. For example, for the measurement intercept of y_2 , ν_2^y , the

prior variance is equal to the squared ML estimate, i.e. $(\hat{\nu}_2^y)^2$, plus the ML estimate of the variance of the error δ_2^y , i.e. $\hat{\sigma}_{y2}^2$.

Since the prior variance is based on the observed effect, it will be large (small) in the case of large (small) observed effects. This ensures that the prior has clear positive support where the likelihood is concentrated. In addition, by including the error variance in the prior variance, the resulting prior contains the information of approximately one observation in the case of intercept parameters. For example, the standard error of the intercept ν_2^y equals $\frac{\hat{\sigma}_{y2}}{\sqrt{n}}$, and thus $\hat{\sigma}_{y2}$ would correspond to the standard error of one observation. A similar default prior was proposed by Mulder, Hoijtink, and Klugkist (2010) in Bayesian model selection. To our knowledge, this empirical Bayes prior has not been used for estimation in BSEM.

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 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 look at their frequentist properties (Bayarri & Berger, 2004). In particular, we were interested in (1) convergence of the Bayesian estimation procedure, (2) mean squared error (MSE), (3) quantiles, and (4) the percentage of significant direct and indirect effects. With a slight abuse of wording we shall refer to a “significant” result when the Bayesian credible interval does not contain the null value.

For the data generation, we considered four different sample sizes, ranging from 35 to 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 where the sample size is small (Heerwegh, 2014; Hox et al., 2012; Lee & Song, 2004), and the original sample size was only 75. We expect the influence of the prior to decline as sample size increases. We manipulated the population

values for the direct effect γ_{65} and the indirect effect $\gamma_{60} \cdot b_{21}$, since these are the parameters of substantive interest in the model. We also manipulated two loadings of \mathbf{y} (λ_4^y and λ_8^y) and the random variances of the measurement error correlations Ω_D . The latter 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). Generating 500 replications per condition, data was simulated under all combinations of population values for the direct and indirect effect and under all combinations of population values for the loadings and variances of error correlations, resulting in 15 different populations. Table 2 presents an overview of all 60 data-generating conditions.

Each condition was analyzed using different default priors with the same type of prior being specified for all parameters in the model at once. For every condition, we investigated the noninformative improper prior $\pi(\sigma^2) \propto 1$, the vague proper prior $\text{IG}(0.001, 0.001)$, and the EB prior. Note that for the first two specifications only the priors on the variances change, while the priors on the mean and regression parameters are specified as the normal prior $\text{N}(0, 10^{10})$. The noninformative improper priors $\pi(\sigma^2) \propto \sigma^{-1}$ and $\pi(\sigma^2) \propto \sigma^{-2}$, and the vague proper priors $\text{IG}(0.01, 0.01)$ and $\text{IG}(0.1, 0.1)$ were investigated only in those conditions where the population values of the variances of the error correlations differ, since those are the conditions in which we expect to see the largest differences due to the influence of the prior on the variances. In total this led to 276 conditions, as shown in Table 2.

The EB prior is 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 equal to 0.001. Preliminary analyses showed that the precise choice had little effect on the conclusions. 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 results, as long as the estimate was fixed to a small number, e.g. 0.001.

[TABLE 2]

Convergence was assessed using the potential scale reduction (PSR), taking $PSR < 1.05$ as a criterion, with a maximum of 75000 iterations. Estimation error was assessed using the mean squared error (MSE): $\frac{1}{S} \sum_{s=1}^S (\hat{\theta}_s - \theta)^2$ with S being the number of converged replications per cell and $\hat{\theta}_s$ being the posterior median for that parameter in a specific replication s . In addition, to check how well the posteriors reproduced the sampling distributions, we investigated the 2.5% and 97.5% quantiles for every parameter, i.e. the percentage of lower and upper bounds of credible intervals per cell that were higher than the true value. Finally, we looked at the percentage of significant effects for the direct effect γ_{65} and the indirect effect $b_{21} \cdot \gamma_{60}$. 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 3 shows the percentage of nonconverging replications for each prior and sample size, averaged across the population values. For all priors convergence increased with sample size and there was almost no convergence for the improper prior $\pi(\sigma^2) \propto \sigma^{-2}$, within the specified maximum of 75000 iterations. Because of the severe nonconvergence under this improper prior we shall not consider it further in this paper and for the other priors we will only consider the converged replications in those conditions with at least 50% convergence. 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.

[TABLE 3]

Mean squared error (MSE)

Given that the pattern of results for $N = 150$ and $N = 500$ did not differ substantially from that of $N = 75$, we will only present the results for $N = 35$ and $N = 75$. The results for $N = 150$ and $N = 500$ can be found in the online appendix. Figure 2 shows for each prior and type of parameter the MSE relative to the MSE of ML estimation per population value and parameter on

the logarithmic scale, $\ln(\text{MSE}_{\text{Bayes}}/\text{MSE}_{\text{ML}})$. Note that the vertical axis is truncated at $\ln(\text{MSE}_{\text{Bayes}}/\text{MSE}_{\text{ML}}) = 4$, 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 maximum-likelihood estimate. The labels on the vertical axes show the different priors for the variance parameters. For the noninformative improper and vague proper priors the mean and regression parameters have a normal $N(0, 10^{10})$ prior.

For larger N , the MSEs of the Bayesian estimates relative to the MSE of ML estimation decreased for all priors. For both sample sizes, the EB prior performed best across parameters, generally better than or comparable to ML estimation. Note, however, the outliers for the EB prior for several parameters (σ_{y5}^2 , ω_{D15}^2 , λ_6^y , λ_7^y , and λ_8^y) and population values (medium to large direct or indirect effect). The noninformative improper priors also performed good for all parameters, showing only for the intercepts and loadings slightly higher MSEs than ML estimation. The vague proper priors performed worst, especially for lower hyperparameters. Interestingly, for $N = 35$ the vague proper priors performed worst for the intercept and loading parameters, with the measurement intercepts of \mathbf{y} being most affected. However, examination of the traceplots for selected replications showed nonconvergence for several parameters, indicating that the PSR criterion may not have filtered out all nonconverging analyses. Nevertheless, it is interesting to see how vague proper priors on variance parameters can indirectly influence location parameters in the model.

[FIGURE 2]

Quantiles

Coverage rates were examined by checking how often the lower 2.5% and upper 97.5% quantile estimates were above the true population value. Figure 3 shows the quantiles for $N = 35$ with the dashed lines indicating 2.5% and 97.5%. For the lower quantile, ML estimation and the EB prior performed best, with quantiles closest to the 2.5% for all parameters, except the intercepts, for which the EB prior resulted in lower quantiles above 2.5%. The vague proper

priors $IG(0.1, 0.1)$ and $IG(0.01, 0.01)$, and the noninformative improper prior $\pi(\sigma^2) \propto 1$ resulted in lower quantiles slightly above 2.5%, while the noninformative improper prior $\pi(\sigma^2) \propto \sigma^{-1}$ and the vague proper prior $IG(0.001, 0.001)$ performed worst with quantiles around 20-30%. One possible explanation for these extreme results is that the PSR criterion may not have filtered out all nonconverging analyses. There were several outliers for all priors at 100%. These corresponded to the lower quantiles for Ω_D when the population value for these parameters was equal to zero, a situation in which we always expect the lower bound of the credible interval to exceed the true value, since the priors only support values greater than zero. For the upper quantile, ML estimation and the EB prior generally performed worst with quantiles lower than 97.5% for most parameters, except the intercepts, for which they were closest to the desired 97.5%. The vague proper priors performed best for the regression coefficients and variances, but resulted in slightly lower quantiles for the intercepts and higher quantiles for the loadings. The noninformative improper priors resulted in slightly higher quantiles for all parameters.

Figure 4 shows the quantiles for $N = 75$, which were all closer to the desired quantiles compared to $N = 35$. For the lower quantile, the EB prior no longer performed best, showing more outliers compared to ML estimation, the vague proper priors $IG(0.1, 0.1)$ and $IG(0.01, 0.01)$, and the noninformative improper prior $\pi(\sigma^2) \propto \sigma^{-1}$. The noninformative improper prior $\pi(\sigma^2) \propto 1$ performed good as well, but showed slightly more outliers than $\pi(\sigma^2) \propto \sigma^{-1}$. The vague proper prior $IG(0.001, 0.001)$ generally resulted in quantiles closer to 2.5%, but still showed outliers for all parameters. Again, there were outliers for all priors at 100% which corresponded to Ω_D when the population value for these parameters was equal to zero. For the upper quantile, the results were similar to $N = 35$.

[FIGURE 3]

[FIGURE 4]

Direct and indirect effect

In practice, researchers often are not interested in all parameters of the model but only 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 5 shows the MSE of the direct effect for the different priors and ML estimation and the different sample sizes. It is clear that, despite the sensitivity of the results across parameters, the specific choice of the prior did not have a great influence on the estimation of the direct effect. For the indirect effect there were differences, especially when $N = 35$. In this case, the EB prior resulted in the lowest MSE, which was more than half as low as the MSE for ML estimation.

[FIGURE 5]

Finally, we consider the influence of the prior on 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. Figure 6 shows the percentage of significant direct effects for the different priors and for different population values of the direct and indirect effect. A dashed line at 5% is included for reference. Note that some points are missing when $N = 35$ due to nonconvergence.

When the true direct effect was zero, the noninformative improper and vague proper priors had a percentage of significant direct effects close to the nominal rate of 5%, as had ML estimation for $N = 75$. For $N = 35$, ML estimation resulted in badly controlled Type 1 error rates, as did the EB prior for both sample sizes. Consequently, when the true direct effect was not zero, the EB prior and ML estimation performed better than the noninformative improper and vague proper priors for $N = 35$, and for $N = 75$ only the vague proper prior resulted in lower percentages of significant direct effects.

Figure 7 shows the results for the conditions in which only the loadings λ_4^y and λ_8^y and the variances of the measurement error correlations Ω_D were manipulated. Here we see that ML estimation performed slightly better than the Bayesian methods, although the percentages significant direct effects were consistently too low across population values and across methods for $N = 35$. For $N = 75$, all methods resulted again in percentages that were too low except

when the population values for the loadings were equal to 2 and the variances of the error correlations equal to 0. For both sample sizes the percentage significant direct effects was generally higher when the correlations between the measurement errors were zero and when the two loadings were large.

Figure 8 shows the percentage significant indirect effects. When the true indirect effect was zero, all priors and ML estimation performed similarly with percentages close to the nominal 5%. However, as the true indirect effect increased the EB prior performed best compared to the other two priors and ML estimation. When only the loadings and measurement error correlations were manipulated, the EB prior showed the highest percentages significant indirect effects for $N = 35$, while there were no substantial differences between the priors or ML estimation for $N = 75$ (Figure 9).

[FIGURE 6]

[FIGURE 7]

[FIGURE 8]

[FIGURE 9]

Based on these results, we recommend against the use of the noninformative improper prior $\pi(\sigma^2) \propto \sigma^{-2}$ and the vague proper priors, due to convergence issues and large MSE's, respectively. When interested in the direct and indirect effect, we recommend to use the EB prior. However, there was not one prior that performed consistently better than the other priors or than ML estimation across all parameters. Despite the fact that all priors are commonly used default priors, the results differed substantially across situations. The inconsistencies in results across parameters, conditions, and outcome measures highlight the importance of conducting a prior sensitivity analysis. Therefore, the next section will provide 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 Muthén and Asparouhov (2012, p. 320), prior sensitivity analysis is an important step in BSEM. A prior sensitivity analysis can be conducted by rerunning the analysis with different choices for the prior. Due to 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. Due to 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, a specific prior can have a large effect on some parameters (e.g. variances of latent variables), and no or little effect on others (e.g. residual variances). Depaoli and Van de Schoot (in press) provided guidelines on conducting prior sensitivity analyses for general 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 paper). Therefore, we recommend to focus 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 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. Instead, we will change the priors in batches corresponding to the type of parameter, while keeping all other priors constant.

Based on the results of the simulation study, we recommend to include the following default priors in the prior sensitivity analysis: the noninformative improper priors $\pi(\sigma^2) \propto 1$ and $\pi(\sigma^2) \propto \sigma^{-1}$, and the EB prior for variance parameters. For mean and regression parameters, we recommend to include the vague proper prior $N(0, 10^{10})$ and the EB prior. As baseline we use the Mplus default prior setting, i.e. when the prior for one type of parameter is changed all other mean, intercept, loading, and regression parameters have the prior $N(0, 10^{10})$ and all other variance parameters have the prior $\pi(\sigma^2) \propto 1$. Note, however, that any prior setting can be used as baseline.

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, e.g. 95%. To assess prior sensitivity in the case of informative priors, we therefore recommend to vary the hyperparameters such that wider, more extreme regions are considered that contain 95% prior probability.

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 appendix 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 markov chain monte carlo (MCMC) sampling, it is important to ensure that the chains converge to the posterior distribution. Mplus provides an automatic criterion based on the potential scale reduction (PSR). We recommend to set the `BCONVERGENCE` option equal to 0.025 so that sampling stops once $PSR < 1.05$ or before that if the maximum number of iterations is reached. The maximum number of iterations can be specified through the `BITERATIONS` options and should depend on the model under consideration, with more complex models requiring a larger number of iterations. Preliminary analyses can be conducted to get an indication of the required number of iterations. More information on the PSR can be found in Gelman and Rubin (1992) or the Mplus User guide (L. K. Muthén & Muthén, 1998-2012). In addition, it is highly recommended to check the traceplots 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 posterior median (or other posterior summaries) across the priors; if a change in a parameter exceeds this boundary, the parameter can be classified as sensitive. 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. If a parameter is found to be sensitive to the choice of the prior, the results of the simulation study (Section 4) can be used to determine which priors can best be used to draw the most accurate conclusions. On the other hand, if the results are consistent for all different prior choices, we can conclude that the results are robust to the choice of the default prior.

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 where the results are quite sensitive to the choice of the prior. For comparison, we also include the standardized 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.

Step 2. In our model, we distinguish between the following types of parameters: the intercepts, ν_y , of the observed dependent variable-indicators \mathbf{y} ; the loadings Λ_y ; the intercepts, ν_x , of the observed independent variable-indicators \mathbf{x} ; the loadings Λ_x ; the structural intercepts, α ; the regression coefficient, b_{21} ; the regression coefficients, Γ ; the latent mean, μ_ξ ; the observed-variable residual variance matrices Σ_y and Σ_x ; the latent independent variable variance ω_ξ^2 ; the latent dependent variance matrix Ω_ζ ; and the variances of the pseudo-latent variables representing error covariances, Ω_D . We change the priors on each of these sets of parameters simultaneously and as default choices we consider the noninformative improper priors $\pi(\sigma^2) \propto 1$ and $\pi(\sigma^2) \propto \sigma^{-1}$, and the EB prior for variance parameters. For mean and regression parameters,

we consider the vague proper prior $N(0, 10^{10})$ and the EB prior.

In addition, informative priors are available in Dunson, Palomo, and Bollen (2005) for this model based on expert knowledge, which we included in our prior sensitivity analysis. We used the original specification, a specification in which the prior variance was half the original prior variance, and one in which it was twice the original prior variance. For the inverse Gamma prior this was done by multiplying or halving the shape parameter α , to influence the number of observations contained in the prior, and subsequently choosing the scale parameter β in such a way that the prior mean was equal to the original prior mean. For the normal prior, we kept the prior mean the same but doubled or halved the prior variance. The resulting hyperparameters and 95% prior regions are presented in Appendix C.

Step 3. Step 1 and 2 resulted in a total of 75 analyses, which we ran using MplusAutomation. All files for our analyses are available in the online appendix. Setting `BCONVERGENCE = 0.025` and `BITERATIONS = 75000(1000)`, all analyses converged. Figure 10 gives an example of a traceplot illustrating convergence and the corresponding estimated posterior densities of both chains, for one parameter in one analysis.

[FIGURE 10]

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 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, $IG(-1, 0)$. 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 this default specification, 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 $N = 35$, both of which are presented in Figure 11.

[FIGURE 11]

Results original data ($N = 75$). The second column of Figure 11 shows the deviations in standardized estimates for each Bayesian and ML estimate relative to the default setting in Mplus for every loading, regression, mean, and intercept parameter. The dotted line indicates the boundary of 0.1. The two parameters of interest in this application are the direct effect γ_{65} and the indirect effect $\gamma_{60} \cdot b_{21}$. The results for both parameters were insensitive to the default priors and the informative priors, even though the other (nuisance) parameters were sensitive to the choice of the prior. Specifically, the mean and intercept parameters crossed the boundary of 0.1 in some analyses in which the EB prior was specified, but the loadings and regression coefficients appeared robust for all default priors. The nuisance parameters showed more sensitivity to the informative priors, in particular to the informative priors for the loadings and regression coefficients, and to all informative priors for the mean and intercept parameters. Thus, we can conclude that the results for the parameters of interest are robust to the choice of the prior, despite the fact that the other parameters are sensitive to the informative priors and, in the case of the mean and intercepts, also to the EB prior.

Recently, Depaoli and van de Schoot (in press) suggested the use of the PSR as measure of prior sensitivity. Normally, the PSR is used to assess convergence by computing the variance between and within multiple chains with different starting values. These variances are used to estimate the PSR which will be closer to 1 as the two chains are more similar, which implies that both chains are converged to the same posterior distribution (Asparouhov & Muthén, n.d.; Gelman & Rubin, 1992). To assess prior sensitivity one could consider two chains from analyses with different priors to assess whether these chains move to the same posterior distribution. By default, Mplus runs two MCMC chains with different starting values for one analysis. We computed the PSR separately for these two chains by comparing one chain from each analysis with the corresponding chain from the analysis in which all parameters have the Mplus default prior setting. Thus, each prior comparison results in two PSR values for each parameter, one for chain 1 and one for chain 2. The results are shown in Figure 12 with the dotted line at 1.1 indicating a cut-off point often used when employing the PSR to assess convergence. When using

the PSR as prior sensitivity check with this cut-off, more parameters are flagged as being prior sensitive. Specifically, the parameters of substantive interest showed sensitivity to the EB prior on Σ_y , as well as sensitivity to the informative priors. The other parameters also showed sensitivity to the EB prior and all informative priors. However, the results of the PSR were not always consistent across chains. For example, the direct effect γ_{65} was only flagged as sensitive to the EB prior on Σ_y in the first chain.

[FIGURE 12]

Results subset 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 due to the small sample size and thus these results are excluded. The first column of Figure 11 shows the deviations from the baseline priors. The parameters of substantive interest γ_{65} and $\gamma_{60} \cdot b_{21}$ were robust to the default priors. However, they were sensitive to the informative priors, with the analyses that crossed the boundary corresponding to an informative prior on α . This illustrates how priors on specific parameters can indirectly influence estimates of other parameters as well. The other loadings and regression coefficients were only sensitive to the informative priors, even the vague specifications. The mean and intercept parameters were sensitive to both the informative priors and some default priors, specifically the EB prior and $\pi(\sigma^2) \propto \sigma^{-1}$. Thus, the parameters of interest are only sensitive to the informative priors, as are the other loadings and regression coefficients, but the mean and intercept parameters are sensitive to both the default and the informative priors.

Figure 13 shows the prior sensitivity according to the PSR, with the dashed line indicating the cut-off of 1.1. The parameters of interest were sensitive to two default priors: the EB prior and $\pi(\sigma^2) \propto \sigma^{-1}$. In addition, they were sensitive to the informative priors. The other priors were also sensitive to these two default priors and the informative priors. Although the PSR may be an interesting measure for assessing prior sensitivity, we generally recommend to use differences in posterior summaries to assess sensitivity since these cut-offs can be tailored to substantive

considerations. The cut-off of the PSR method at 1.1 is arbitrarily chosen.

[FIGURE 13]

Discussion

BSEM is 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 paper, 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 use of the EB prior if researchers are interested in the direct and indirect effect. We recommend against the use of vague proper priors for variances, since the results obtained under these priors depend on the specific choice for the hyperparameters in an arbitrary manner. Further, we recommend against the use of the noninformative improper prior $\pi(\sigma^2) \propto \sigma^{-2}$, since it suffers from major convergence problems. Of the other two noninformative improper priors, $\pi(\sigma^2) \propto 1$ performed better than $\pi(\sigma^2) \propto \sigma^{-1}$, especially in terms of the coverage rates of the interval estimates. In general, there was not one default prior that outperformed the others in all settings and across all parameters. 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.

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. For both sample sizes, we saw that the results were sensitive to the informative priors, especially the informative priors on α proved influential on the parameters of interest. This is not surprising given the fact that the original informative priors were substantially different from the ML estimates. For example, the ML estimate for α_{60} in the original data was -2.031 whereas the informative priors were centered around 1 and did not even include this value in the 95% prior

region in the informative case. A difference between prior and ML estimate also occurred for the measurement intercepts ν_x and ν_y although the effects of these priors were less influential on the parameters of substantive interest. When using informative priors, the estimates can thus be substantially different than when using default priors. If this is the case, caution is needed when interpreting the results. Possibly, the user needs to reevaluate the specified informative priors, for instance by setting larger prior variances, especially in the case of small datasets.

We investigated only conditionally conjugate priors since these are available in Mplus. However, many non-conjugate 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). For intercept, mean, and regression parameters a robust alternative is the t -distribution, which has been proposed for logistic models by Gelman, Jakulin, Pittau, and Su (2008) and which includes the Cauchy prior 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.

To conclude, our simulation study showed that BSEM is still in a too early stage to rely on one single default prior. Therefore, default BSEM should be used with care and a prior sensitivity analysis should always be conducted, especially when the sample size is small. Our online appendix is intended to provide practitioners with a template for doing so.

Footnotes

¹Here, 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.

²Initially, we also considered an informative EB prior. For variance parameters this prior was specified the same as the EB prior considered in this paper, but with the shape parameter equal to $\frac{n}{8}$, so that the prior holds the equivalent of 25% of the information in the data. For the mean and regression parameters, the informative EB prior had a prior mean equal to the ML estimate and a prior variance of 1. However, the results for this prior were extremely bad, possibly because the prior was too informative. This indicates that it can be difficult to determine the right variance for EB priors with hyperparameters based on the ML estimates and that it is better to automatically set the information in the EB prior equal to one observation, as we did in the current specifications.

Table 1

Overview of the default prior specifications per parameter considered throughout this paper.

Parameter type	Parameter	Default	Prior		
			Noninformative improper	Vague proper	Empirical Bayes (EB)
Latent variable variances	ω_ξ^2	$\pi(\omega_\xi^2) \propto 1$	$\pi(\omega_\xi^2) \propto 1$	IG(0.001, 0.001)	IG $\left(\frac{1}{2}, \hat{\omega}_\xi^2 \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)\right)$
			$\pi(\omega_\xi^2) \propto \omega_\xi^{-1}$	IG(0.01, 0.01)	
			$\pi(\omega_\xi^2) \propto \omega_\xi^{-2}$	IG(0.1, 0.1)	
	ω_ζ^2	$\pi(\omega_\zeta^2) \propto 1$	$\pi(\omega_\zeta^2) \propto 1$	IG(0.001, 0.001)	IG $\left(\frac{1}{2}, \hat{\omega}_\zeta^2 \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)\right)$
			$\pi(\omega_\zeta^2) \propto \omega_\zeta^{-1}$	IG(0.01, 0.01)	
			$\pi(\omega_\zeta^2) \propto \omega_\zeta^{-2}$	IG(0.1, 0.1)	
	ω_D^2	$\pi(\omega_D^2) \propto 1$	$\pi(\omega_D^2) \propto 1$	IG(0.001, 0.001)	IG $\left(\frac{1}{2}, \hat{\omega}_D^2 \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)\right)$
			$\pi(\omega_D^2) \propto \omega_D^{-1}$	IG(0.01, 0.01)	
			$\pi(\omega_D^2) \propto \omega_D^{-2}$	IG(0.1, 0.1)	
Residual variances	σ_y^2	$\pi(\sigma_y^2) \propto 1$	$\pi(\sigma_y^2) \propto \sigma_y^{-2}$	IG(0.001, 0.001)	IG $\left(\frac{1}{2}, \hat{\sigma}_y^2 \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)\right)$
			$\pi(\sigma_y^2) \propto \sigma_y^{-1}$	IG(0.01, 0.01)	
			$\pi(\sigma_y^2) \propto \sigma_y^{-2}$	IG(0.1, 0.1)	
	σ_x^2	$\pi(\sigma_x^2) \propto 1$	$\pi(\sigma_x^2) \propto \sigma_x^{-2}$	IG(0.001, 0.001)	IG $\left(\frac{1}{2}, \hat{\sigma}_x^2 \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)\right)$
			$\pi(\sigma_x^2) \propto \sigma_x^{-1}$	IG(0.01, 0.01)	
			$\pi(\sigma_x^2) \propto \sigma_x^{-2}$	IG(0.1, 0.1)	
Structural intercepts	α	N(0, 10 ¹⁰)	-	N(0, 10 ¹⁰)	N(0, $\hat{\alpha}^2 + \hat{\omega}_\zeta^2$)
Structural regression coefficients	b	N(0, 10 ¹⁰)	-	N(0, 10 ¹⁰)	N(0, $\hat{b}^2 + \hat{\omega}_\zeta^2$)
	γ	N(0, 10 ¹⁰)	-	N(0, 10 ¹⁰)	N(0, $\hat{\gamma}^2 + \hat{\omega}_\zeta^2$)
Latent variable mean	μ_ξ	N(0, 10 ¹⁰)	-	N(0, 10 ¹⁰)	N(0, $\hat{\mu}_\xi^2 + \hat{\omega}_\xi^2$)
Measurement intercepts	ν_y	N(0, 10 ¹⁰)	-	N(0, 10 ¹⁰)	N(0, $\hat{\nu}_y^2 + \hat{\sigma}_y^2$)
	ν_x	N(0, 10 ¹⁰)	-	N(0, 10 ¹⁰)	N(0, $\hat{\nu}_x^2 + \hat{\sigma}_x^2$)
Loadings	λ_y	N(0, 10 ¹⁰)	-	N(0, 10 ¹⁰)	N(0, $\hat{\lambda}_y^2 + \hat{\sigma}_y^2$)
	λ_x	N(0, 10 ¹⁰)	-	N(0, 10 ¹⁰)	N(0, $\hat{\lambda}_x^2 + \hat{\sigma}_x^2$)

Q^{-1} : denotes the regularized inverse Gamma function.

Table 2

Overview of the data generating and analysis conditions included in the simulation study.

Variable	# Levels	Levels	Values		
<i>Data generating conditions</i>					
Sample size	4		$N \in \{35, 75, 150, 500\}$		
Direct effect	3	Zero:	$\gamma_{65} = 0$		
		Medium:	$\gamma_{65} = 1$		
		Large:	$\gamma_{65} = 2$		
Indirect effect	3	Zero:	$\gamma_{60} = 0$		
		Medium:	$\gamma_{60} = 1$		
		Large:	$\gamma_{60} = 2$		
Loadings	3	Zero:	$\lambda_4^y, \lambda_8^y = 0$		
		Medium:	$\lambda_4^y, \lambda_8^y = 1$		
		Large:	$\lambda_4^y, \lambda_8^y = 2$		
Error covariances	2	Zero:	$\Omega_D = \mathbf{0}$		
		Medium:	$\Omega_D = \mathbf{1}$		
<i>Analysis conditions</i>					
Priors	10	Noninformative improper:	$\pi(\sigma^2) \propto 1$ and $N(0, 10^{10})$		
			$\pi(\sigma^2) \propto \sigma^{-1}$ and $N(0, 10^{10})$		
			$\pi(\sigma^2) \propto \sigma^{-2}$ and $N(0, 10^{10})$		
			Vague proper:	$IG(0.001, 0.001)$ and $N(0, 10^{10})$	
				$IG(0.01, 0.01)$ and $N(0, 10^{10})$	
				$IG(0.1, 0.1)$ and $N(0, 10^{10})$	
			Empirical Bayes (EB):		$IG\left(\frac{1}{2}, \hat{\sigma}^2 \cdot Q^{-1}\left(\frac{1}{2}, \frac{1}{2}\right)\right)$ and $N(0, \hat{\mu}^2 + \hat{\sigma}^2)$

Table 3

Percentage nonconverging replications for the default priors in the simulation study, averaged across population values.

N	$\pi(\sigma^2) \propto 1$	$\pi(\sigma^2) \propto \sigma^{-1}$	$\pi(\sigma^2) \propto \sigma^{-2}$	IG(0.001, 0.001)	IG(0.01, 0.01)	IG(0.1, 0.1)	EB
35	9.1	27.1	99.7	34.9	4.3	3.3	0.7
75	0.4	0.3	99.0	2.8	0	0	0.4
150	0	0	98.2	0	0	0	0.05
500	0	0.2	54.4	0.2	0	0	0.2

Note. N = sample size. EB = Empirical Bayes prior. $\pi(\sigma^2) \propto 1$, $\pi(\sigma^2) \propto \sigma^{-1}$, $\pi(\sigma^2) \propto \sigma^{-2}$ = 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. Location parameters have the normal $N(0, 10^{10})$ prior, except for the EB prior.

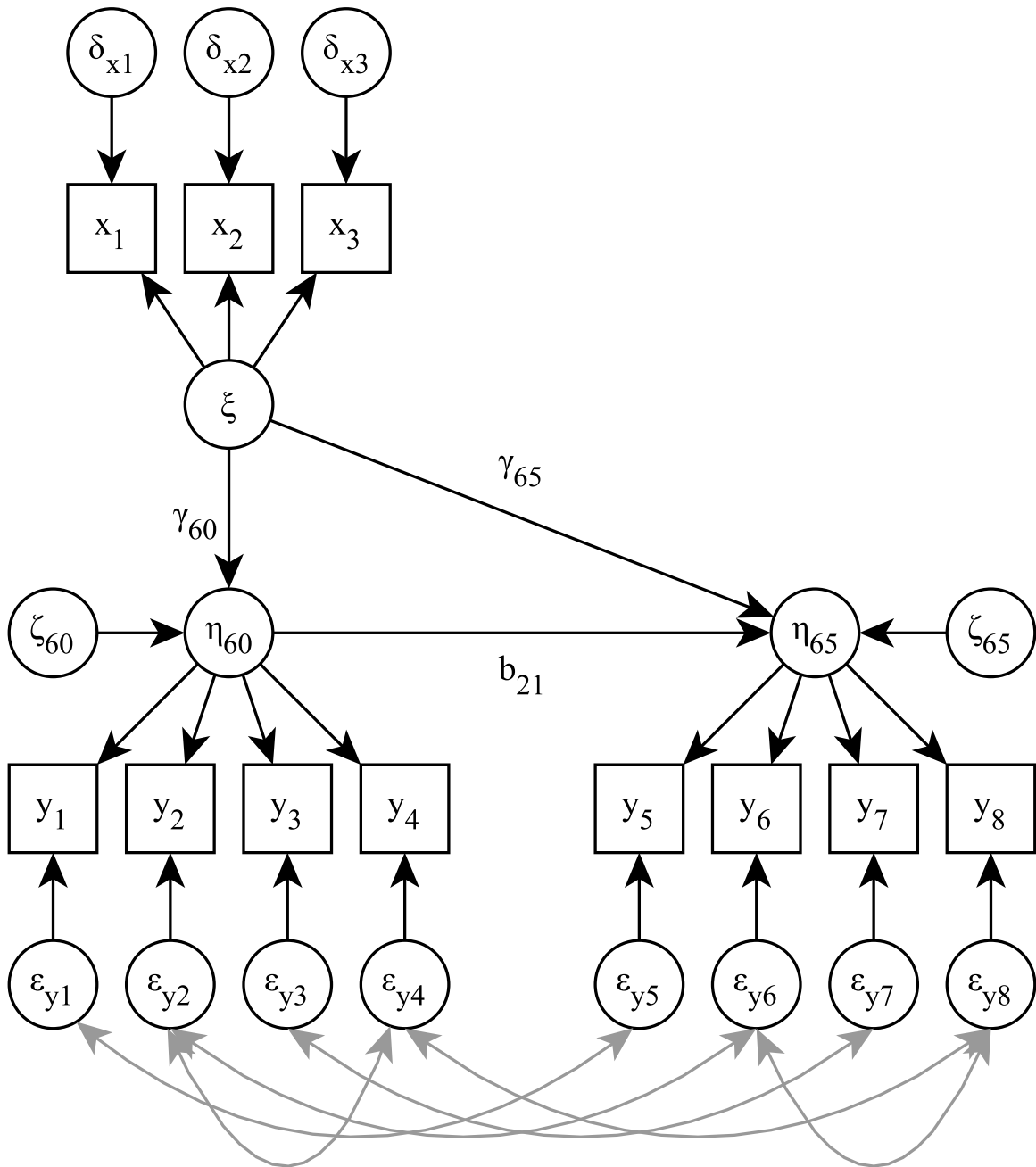


Figure 1. Structural equation model describing the influence of industrialization in 1960 (ξ) on political democracy in 1960 (η^{60}) and 1965 (η^{65}).

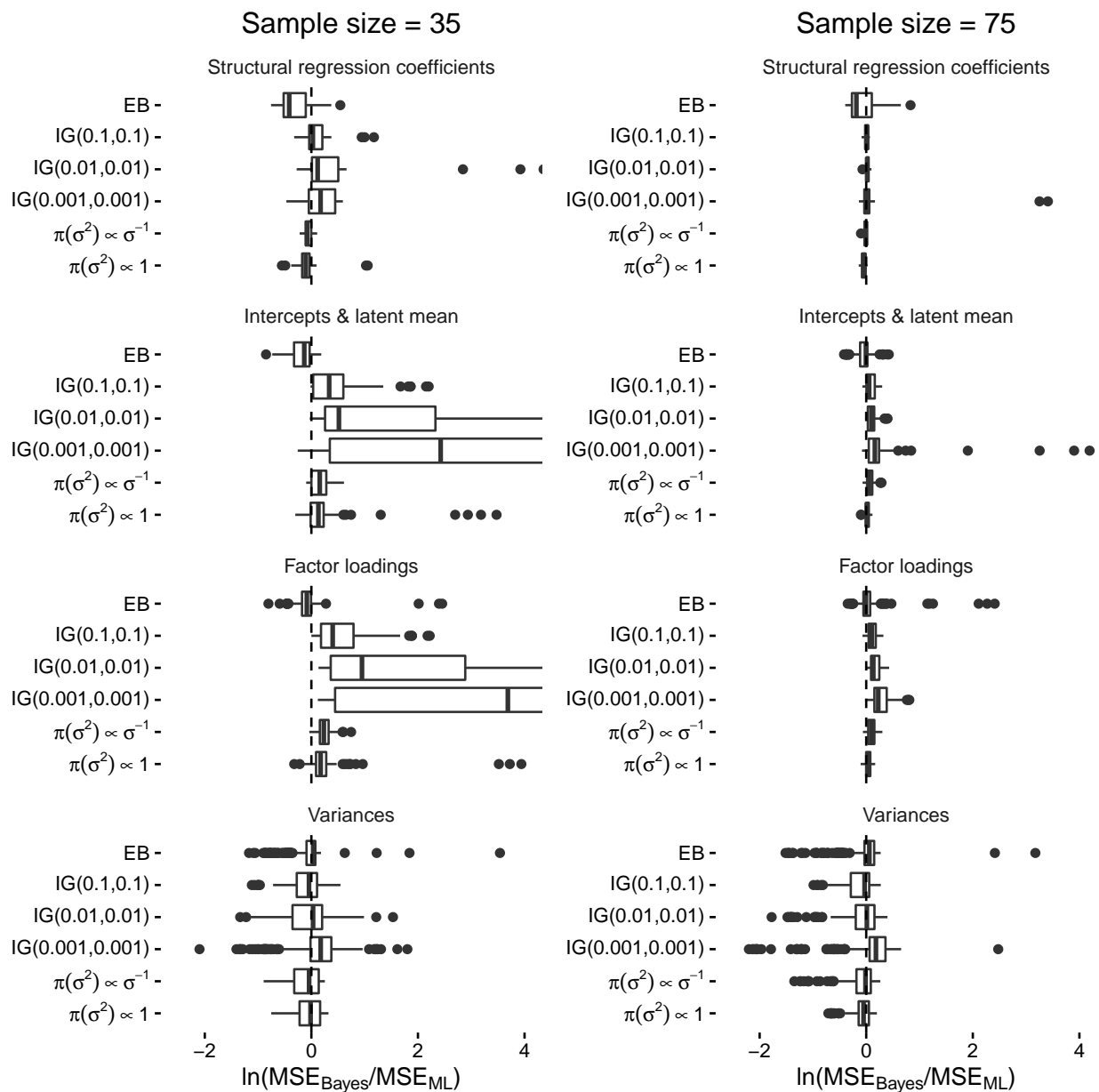


Figure 2. Mean squared error (MSE) for each default prior divided by the MSE for maximum likelihood (ML) estimation. EB = Empirical Bayes prior. $\pi(\sigma^2) \propto 1$, $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters. $IG(0.001, 0.001)$, $IG(0.01, 0.01)$, $IG(0.1, 0.1)$ = vague proper priors variance parameters. Location parameters have the normal $N(0, 10^{10})$ prior, except for the EB prior. Vertical dashed lines indicate where the MSE for the Bayesian estimates equals the MSE for the ML estimates.

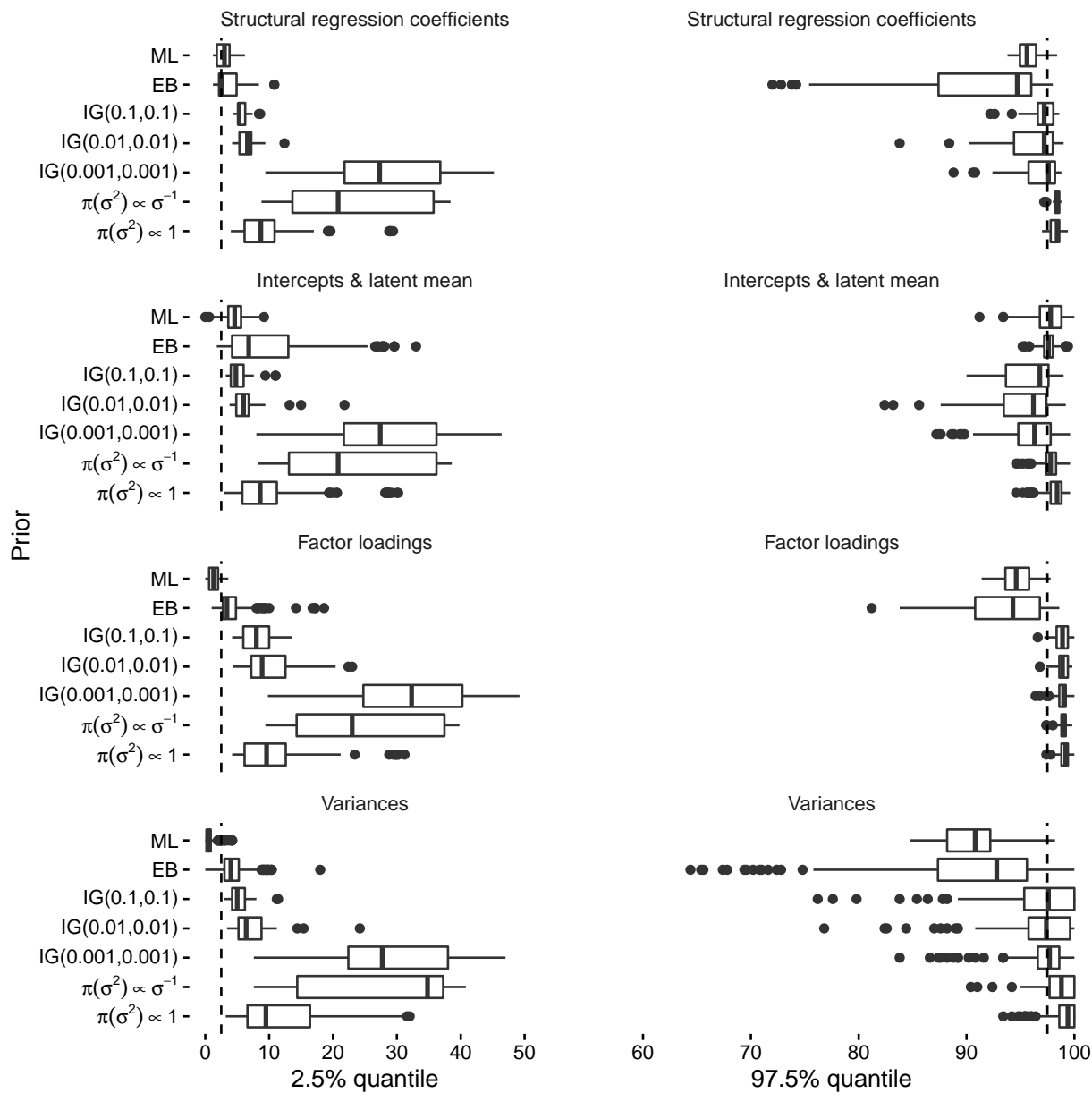


Figure 3. 2.5% and 97.5% quantiles for each default prior and maximum likelihood (ML) estimation for $N = 35$. EB = Empirical Bayes prior. $\pi(\sigma^2) \propto 1$, $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters. IG(0.001, 0.001), IG(0.01, 0.01), IG(0.1, 0.1) = vague proper priors variance parameters. Location parameters have the normal $N(0, 10^{10})$ prior, except for the EB prior. Vertical dashed lines indicate the desired 2.5% and 97.5%

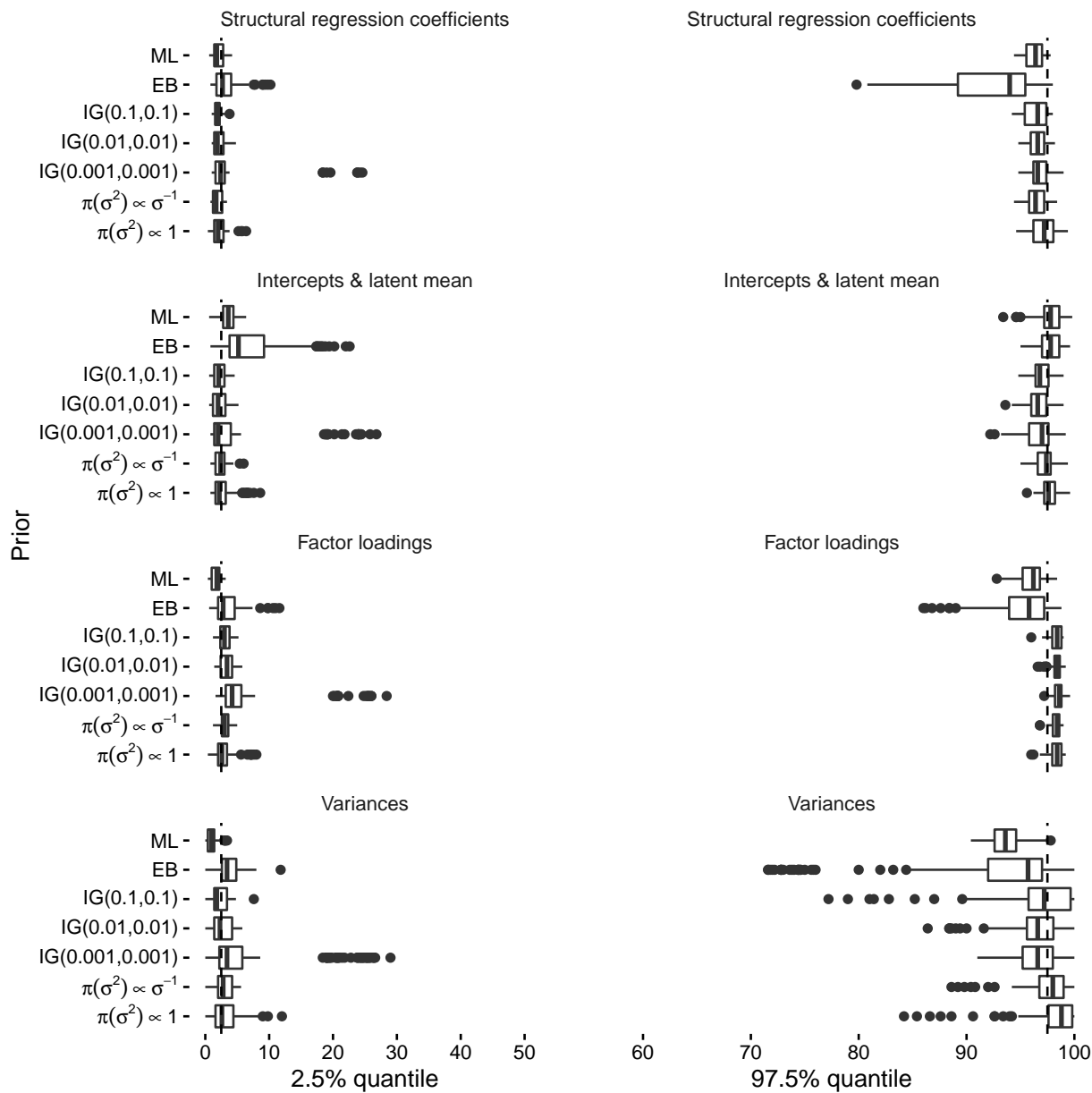


Figure 4. 2.5% and 97.5% quantiles for each default prior and maximum likelihood (ML) estimation for $N = 75$. EB = Empirical Bayes prior. $\pi(\sigma^2) \propto 1$, $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters. $IG(0.001, 0.001)$, $IG(0.01, 0.01)$, $IG(0.1, 0.1)$ = vague proper priors variance parameters. Location parameters have the normal $N(0, 10^{10})$ prior, except for the EB prior. Vertical dashed lines indicate the desired 2.5% and 97.5%

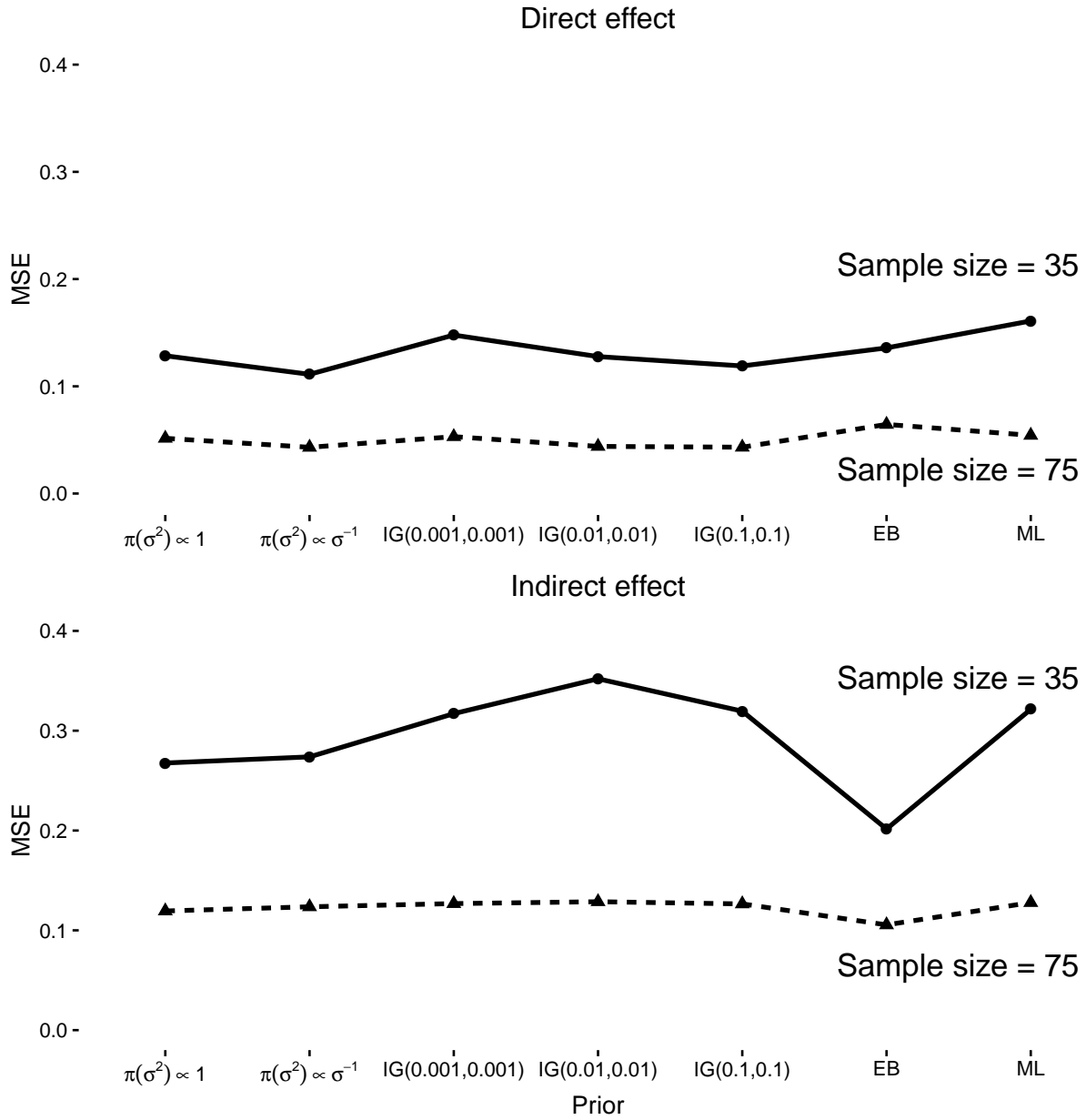


Figure 5. Mean squared error (MSE) for each default prior and maximum likelihood (ML) estimation for the direct effect γ_{65} and the indirect effect $\gamma_{60} \cdot b_{21}$. EB = Empirical Bayes prior. $\pi(\sigma^2) \propto 1$, $\pi(\sigma^2) \propto \sigma^{-1}$ = noninformative improper priors variance parameters. IG(0.001, 0.001), IG(0.01, 0.01), IG(0.1, 0.1) = vague proper priors variance parameters. Location parameters have the normal $N(0, 10^{10})$ prior, except for the EB prior.

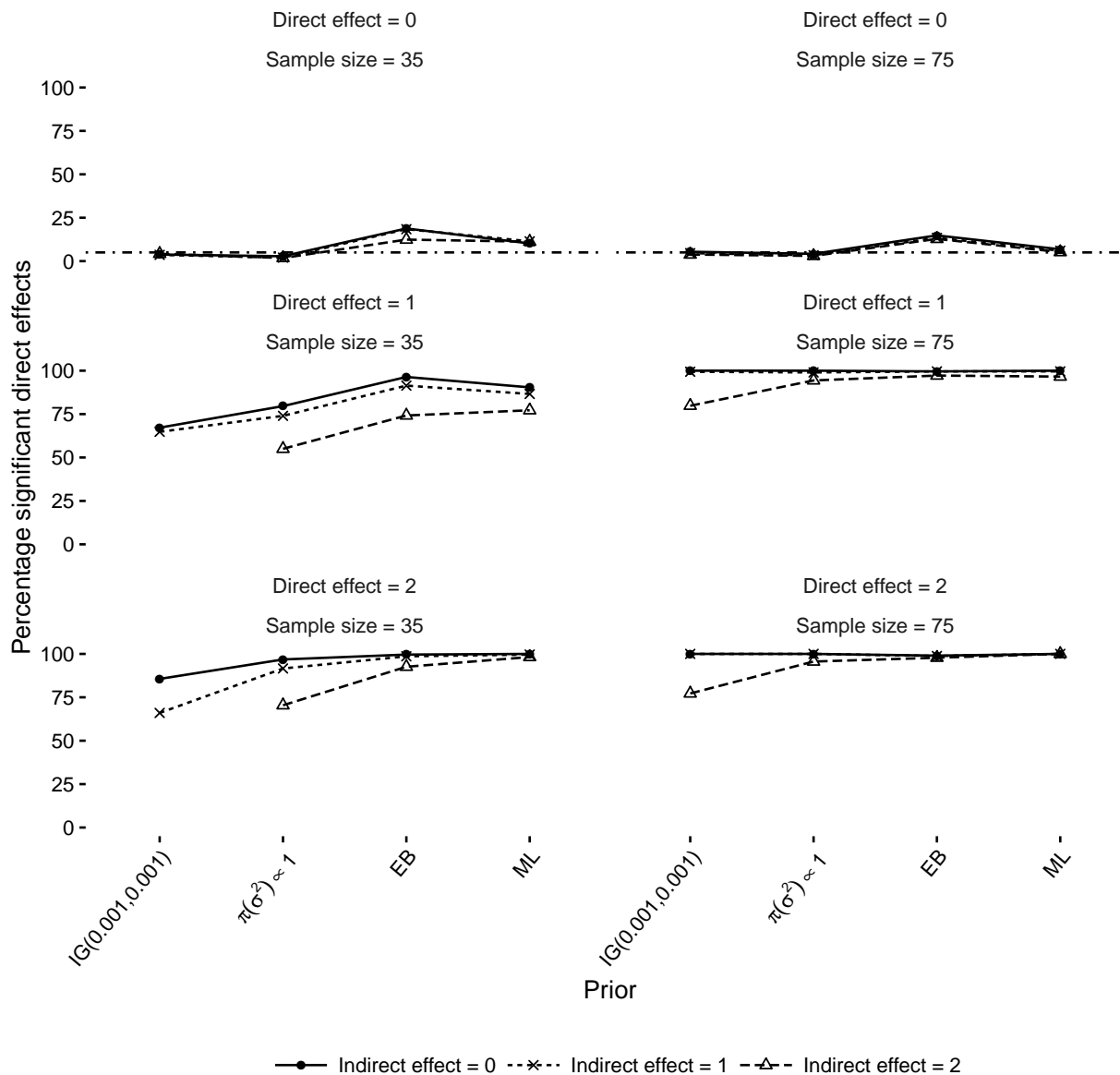


Figure 6. Percentage significant direct effects γ_{65} for each default prior and maximum likelihood (ML) estimation for different population values. EB = Empirical Bayes prior. $IG(0.001, 0.001)$ = vague proper prior variance parameters. $\pi(\sigma^2) \propto 1$ = noninformative improper prior variance parameters. Location parameters have the normal $N(0, 10^{10})$ prior, except for the EB prior. Horizontal dashed lines indicate the nominal error rate when the true effect equals zero.

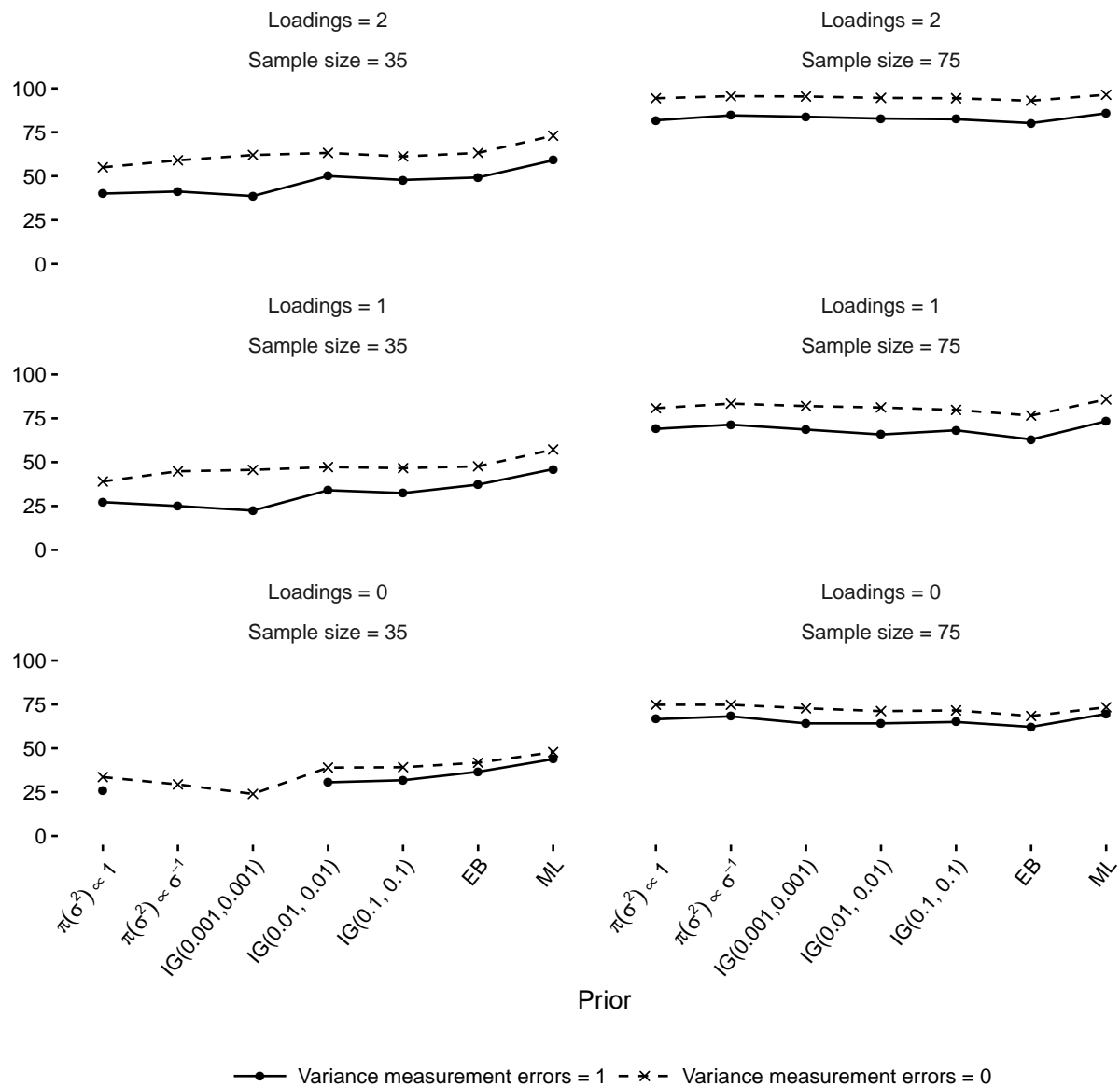


Figure 7. Results simulation: Percentage significant direct effects γ_{65} for each default prior and maximum likelihood (ML) estimation when the population value for the direct effect $\gamma_{65} = 0.572$. EB = Empirical Bayes prior. $\pi(\sigma^2) \propto 1$, $\pi(\sigma^2) \propto \sigma^{-1}$ = 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. Location parameters have the normal $N(0, 10^{10})$ prior, except for the EB prior.

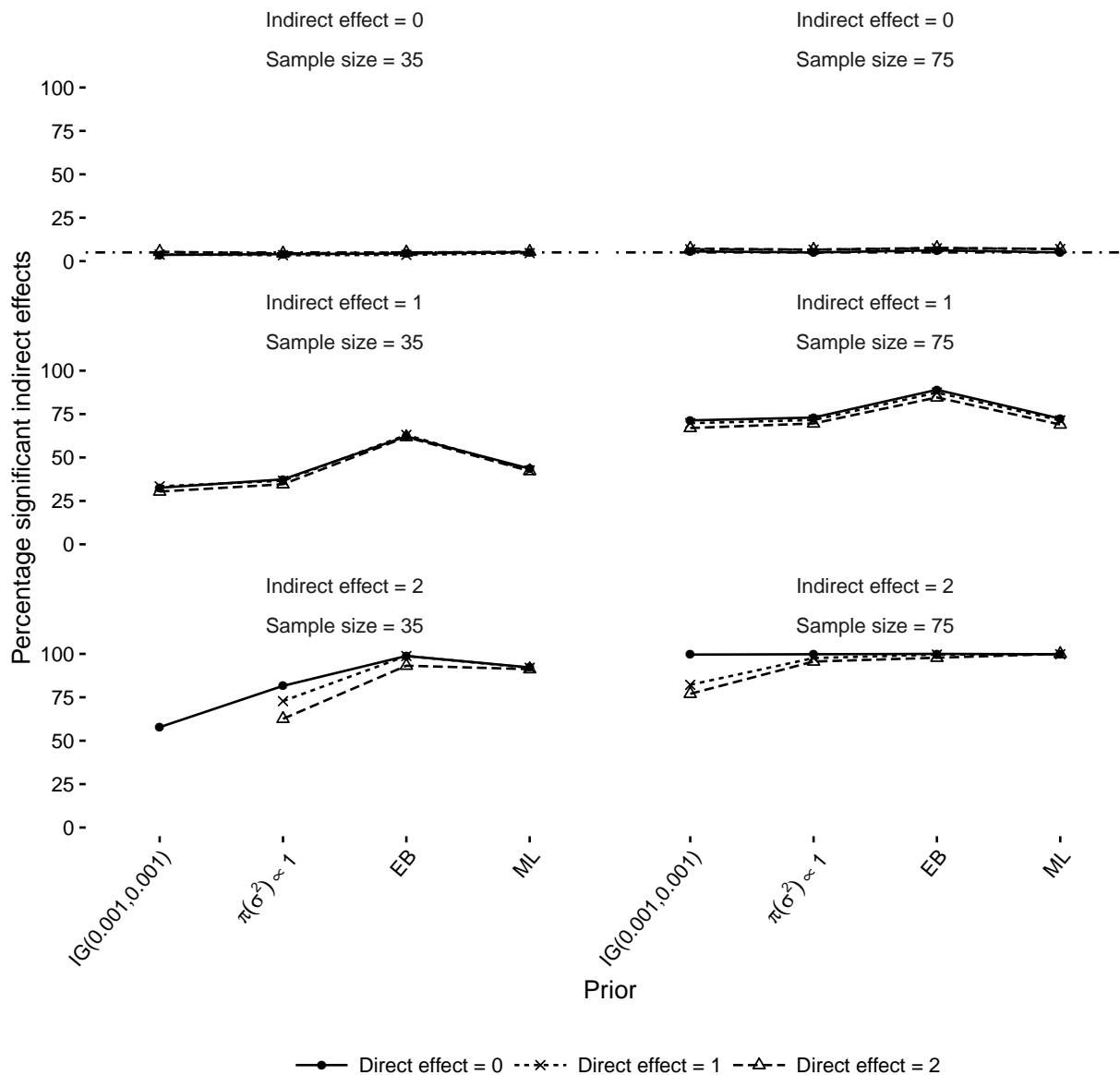


Figure 8. Percentage significant indirect effects $\gamma_{60} \cdot b_{21}$ for each default prior and maximum likelihood (ML) estimation for different population values. EB = Empirical Bayes prior. IG(0.001, 0.001) = vague proper prior variance parameters. $\pi(\sigma^2) \propto 1$ = noninformative improper prior variance parameters. Location parameters have the normal $N(0, 10^{10})$ prior, except for the EB prior. Horizontal dashed lines indicate the nominal error rate when the true effect equals zero.

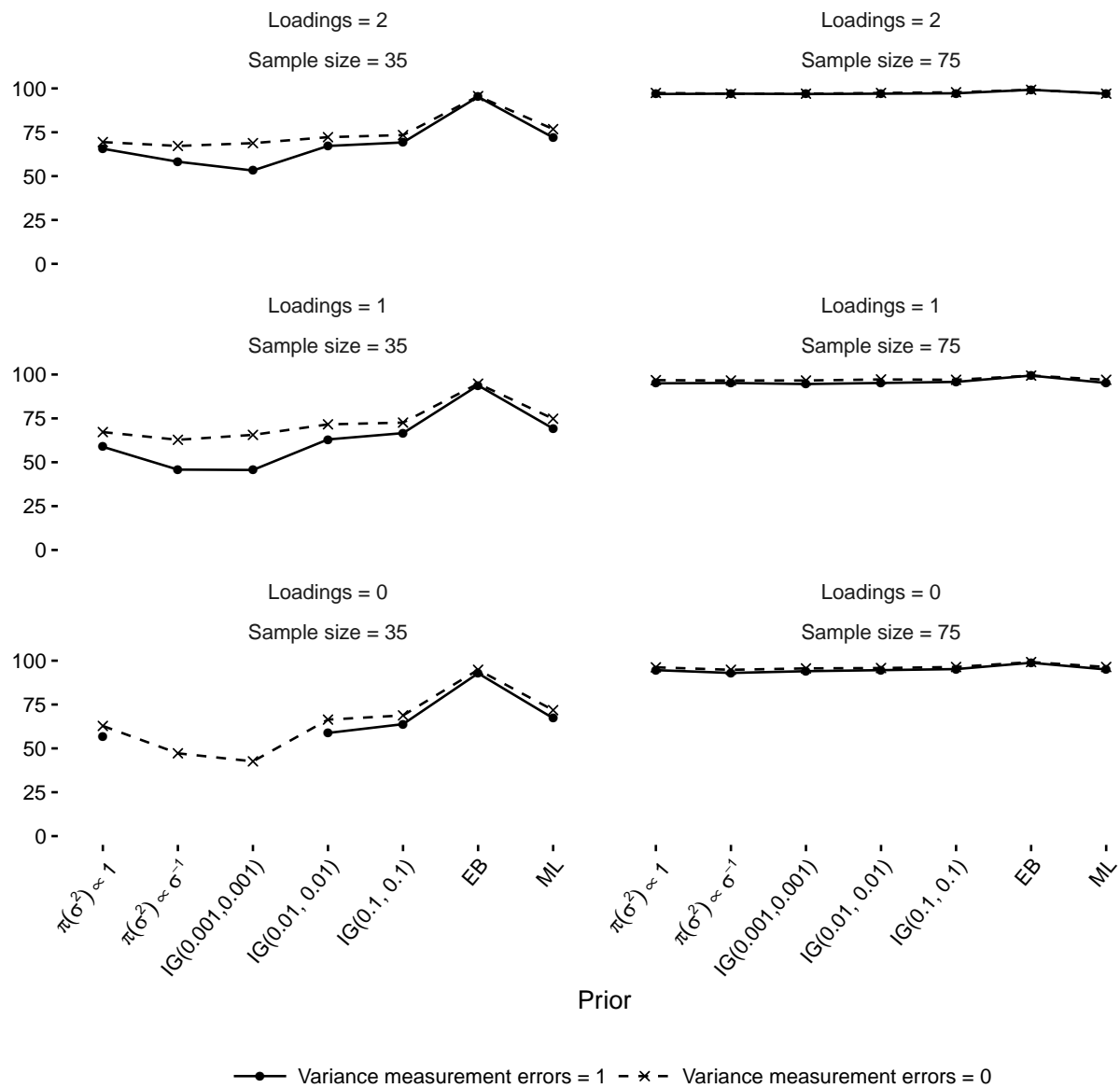


Figure 9. Percentage significant indirect effects $\gamma_{60} \cdot b_{21}$ for each default prior and maximum likelihood (ML) estimation when the population value for the indirect effect $\gamma_{60} \cdot b_{21} = 1.241$. EB = Empirical Bayes prior. $\pi(\sigma^2) \propto 1$, $\pi(\sigma^2) \propto \sigma^{-1}$ = 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. Location parameters have the normal $N(0, 10^{10})$ prior, except for the EB prior.

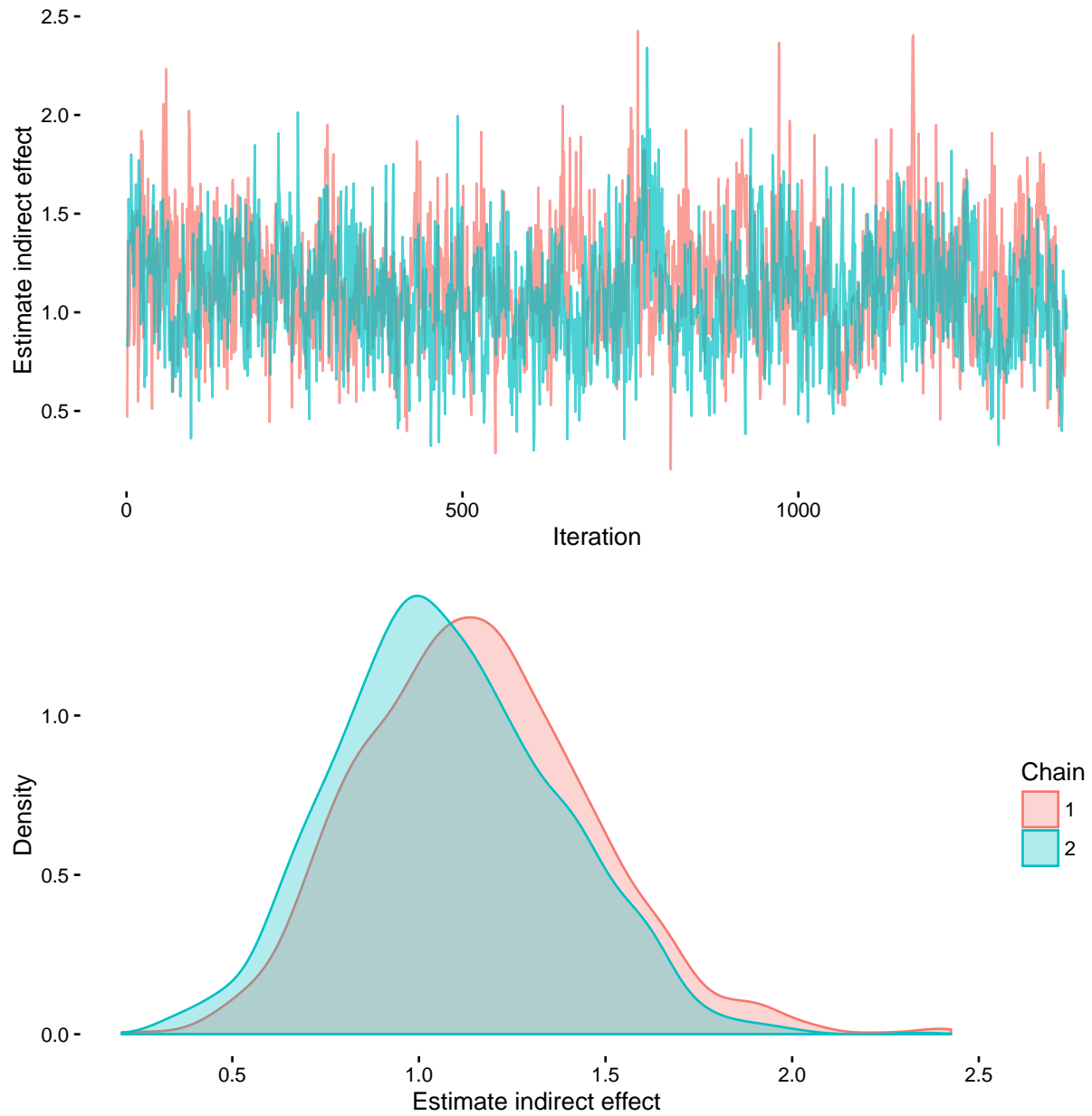


Figure 10. Traceplot indicating convergence (top) and estimated posterior densities (bottom) for the indirect effect $\gamma_{60} \cdot b_{21}$ in the analysis with the empirical Bayes prior on the structural intercepts α and the Mplus default priors on the other parameters.

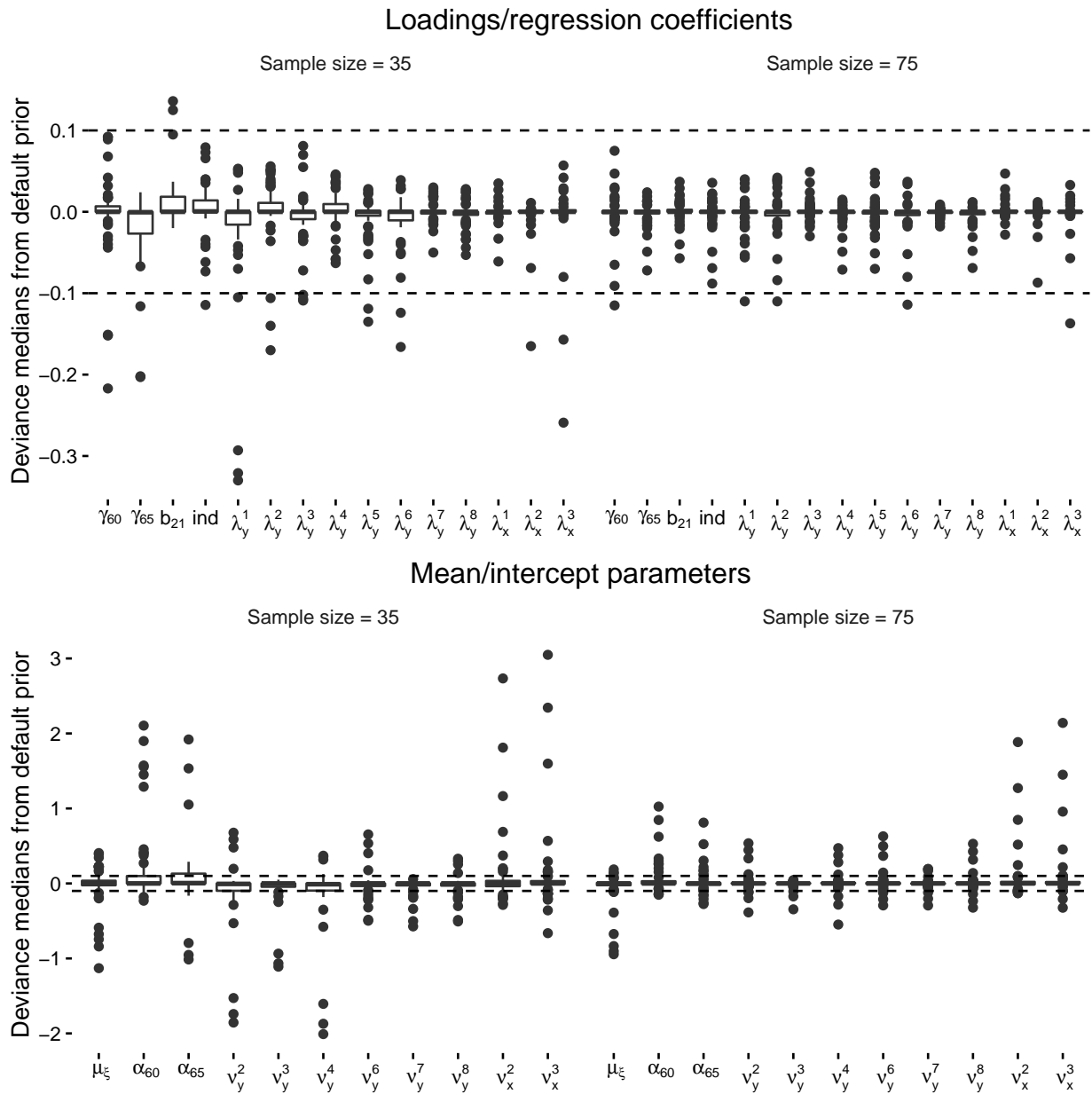


Figure 11. Deviance of the standardized posterior medians per parameter for each prior in the sensitivity analysis from the default prior setting in Mplus. Horizontal dashed lines indicate the cut-off used to determine prior sensitivity.

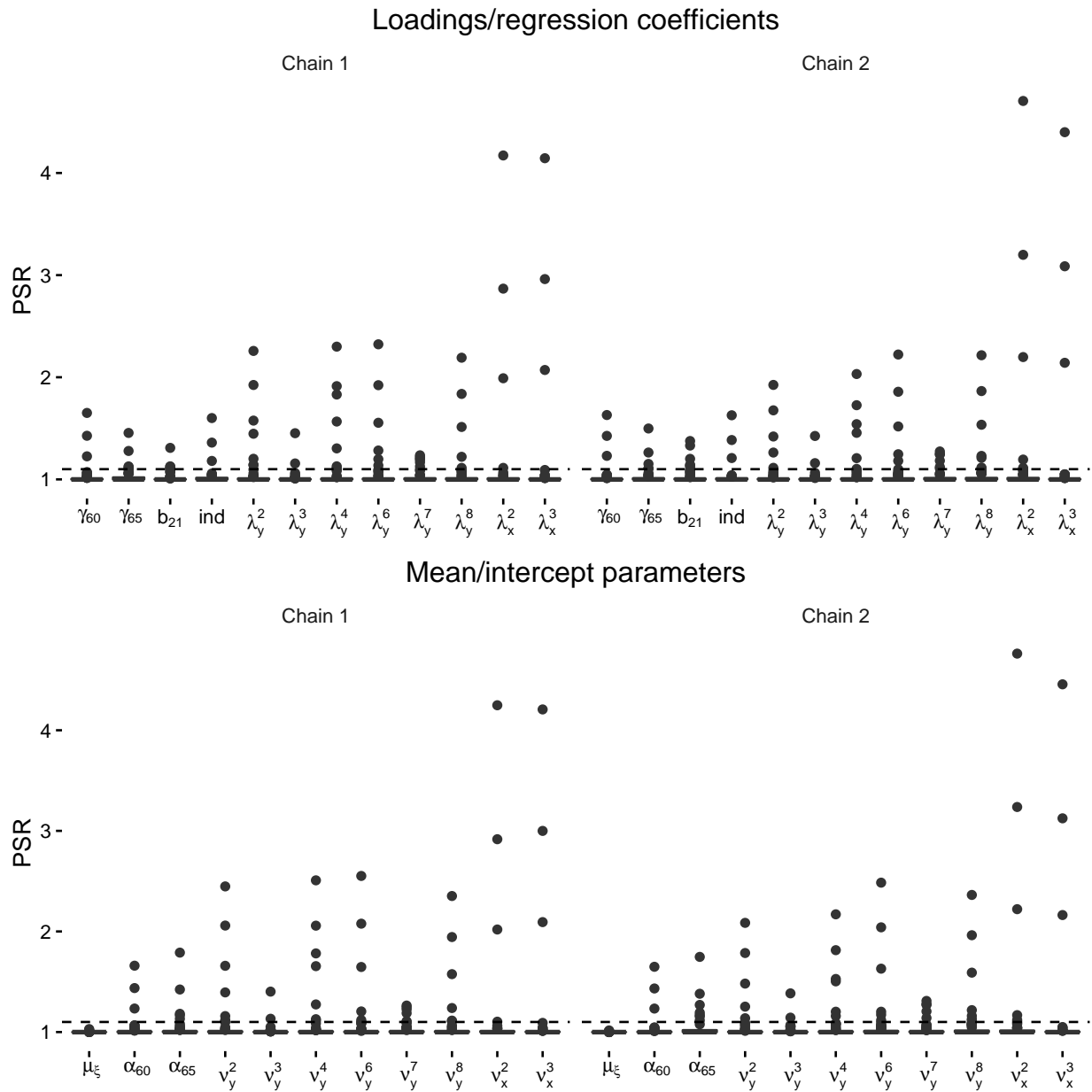


Figure 12. Potential scale reduction (PSR) per parameter for each prior in the sensitivity analysis for the original data ($N = 75$). Horizontal dashed lines indicate the cut-off used to determine prior sensitivity.

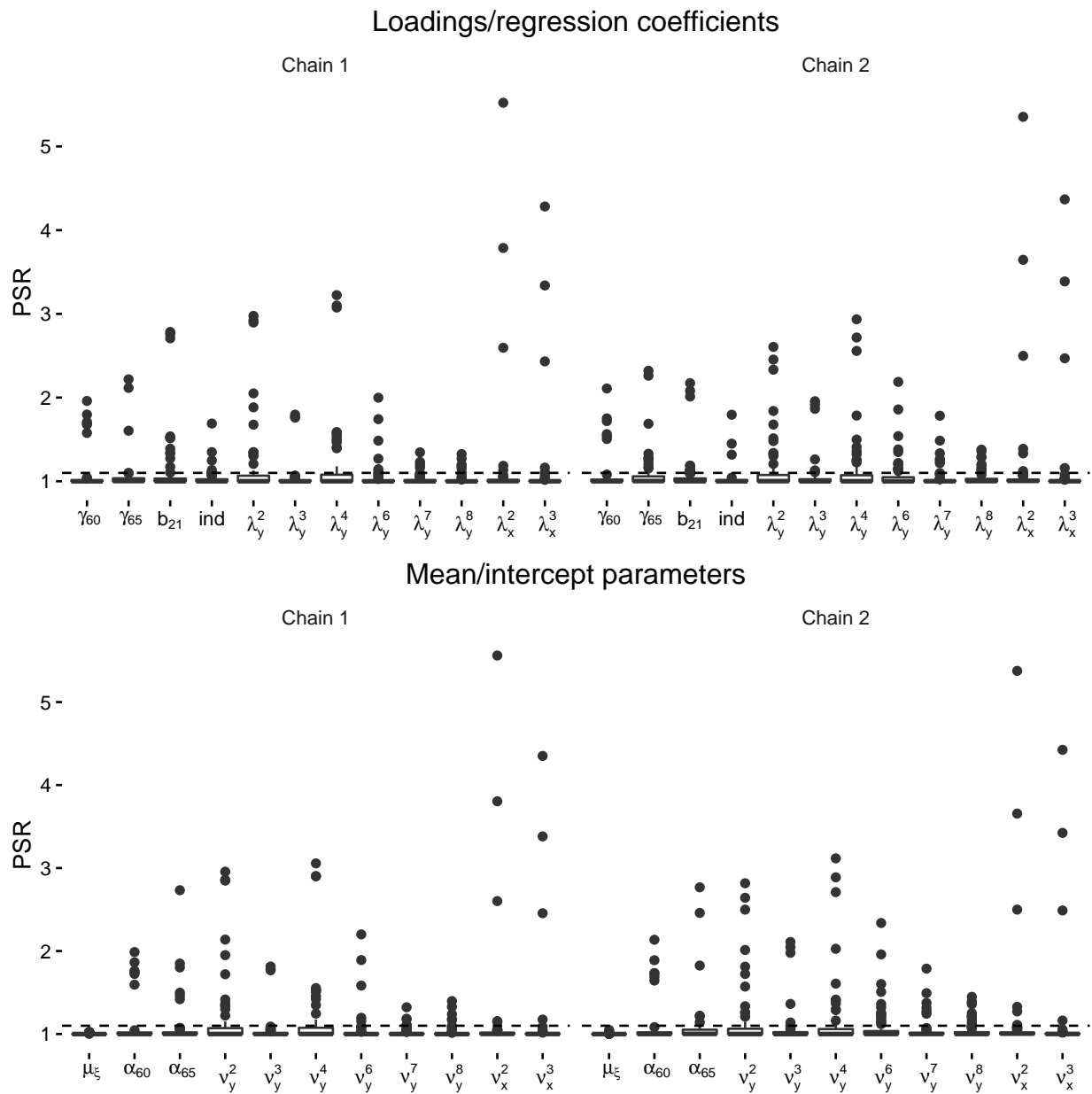


Figure 13. Potential scale reduction (PSR) per parameter for each prior in the sensitivity analysis for the subset of the original data ($N = 35$). Horizontal dashed lines indicate the cut-off used to determine prior sensitivity.

Appendix A

Industrialization and political democracy model in matrix form

The structural model (for $i = 1, \dots, n$) 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 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:

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

Appendix B

Implementation noninformative improper priors variance parameters in Mplus

Here, we show how the noninformative improper priors can be implemented as inverse Gamma priors with specific choices for the hyperparameters. The inverse Gamma distribution is given by:

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

We consider the following noninformative improper priors:

- $p(\sigma^2) \propto \sigma^{-2} = IG(0, 0)$. Inserting the hyperparameters into the kernel of the inverse Gamma distribution and simplifying gives:

$$p(\sigma^2) = (\sigma^2)^{-(0+1)} e\left(-\frac{0}{\sigma^2}\right) = (\sigma^2)^{-1} \cdot 1 = \sigma^{-2}$$

- $p(\sigma^2) \propto \sigma^{-1} = IG(-\frac{1}{2}, 0)$. Inserting the hyperparameters into the kernel of the inverse Gamma distribution and simplifying gives:

$$p(\sigma^2) = (\sigma^2)^{-(-\frac{1}{2}+1)} e\left(-\frac{0}{\sigma^2}\right) = (\sigma^2)^{-\frac{1}{2}} \cdot 1 = \sigma^{-1}$$

- $p(\sigma^2) \propto 1 = IG(-1, 0)$. Inserting the hyperparameters into the kernel of the inverse Gamma distribution and simplifying gives:

$$p(\sigma^2) = (\sigma^2)^{-(-1+1)} e\left(-\frac{0}{\sigma^2}\right) = (\sigma^2)^0 \cdot 1 = 1$$

Appendix C

informative priors considered in the prior sensitivity analysis

Table C1

Hyperparameters and 95% prior regions per parameter for the informative priors in the sensitivity analysis.

Parameter	Informativeness					
	Mild		Moderate (original)		Strong	
	Prior	95% interval	Prior	95% interval	Prior	95% interval
ω_{ξ}^2	IG(3, 2)	[0.28, 3.23]	IG(6, 5)	[0.43, 2.26]	IG(12, 11)	[0.56, 1.77]
ω_{ζ}^2	IG(3, 2)	[0.28, 3.23]	IG(6, 5)	[0.43, 2.26]	IG(12, 11)	[0.56, 1.77]
ω_D^2	IG(3, 2)	[0.28, 3.23]	IG(6, 5)	[0.43, 2.26]	IG(12, 11)	[0.56, 1.77]
σ_y^2	IG(5, 16)	[1.56, 9.91]	IG(10, 36)	[2.10, 7.55]	IG(20, 76)	[2.56, 6.20]
σ_x^2	IG(3, 0.4)	[0.055, 0.642]	IG(6, 1)	[0.086, 0.454]	IG(12, 2.2)	[0.11, 0.35]
α	N(1, 4)	[-6.86, 8.84]	N(1, 2)	[-2.91, 4.90]	N(1, 1)	[-0.97, 2.96]
b	N(1, 4)	[-6.86, 8.84]	N(1, 2)	[-2.91, 4.90]	N(1, 1)	[-0.97, 2.96]
γ_{60}	N(1.5, 4)	[-6.32, 9.33]	N(1.5, 2)	[-2.45, 5.41]	N(1.5, 1)	[-0.47, 3.46]
γ_{65}	N(0.5, 4)	[-7.31, 8.36]	N(0.5, 2)	[-3.43, 4.42]	N(0.5, 1)	[-1.45, 2.45]
μ_{ξ}	N(5, 2)	[1.09, 8.94]	N(5, 1)	[3.03, 6.96]	N(5, 0.5)	[4.02, 5.98]
ν_y	N(0, 2)	[-3.94, 3.93]	N(0, 1)	[-1.96, 1.96]	N(0, 0.5)	[-0.98, 0.98]
ν_x	N(0, 2)	[-3.94, 3.93]	N(0, 1)	[-1.96, 1.96]	N(0, 0.5)	[-0.98, 0.98]
λ_y	N(1, 4)	[-6.86, 8.84]	N(1, 2)	[-2.91, 4.90]	N(1, 1)	[-0.97, 2.96]
λ_x	N(1, 4)	[-6.86, 8.84]	N(1, 2)	[-2.91, 4.90]	N(1, 1)	[-0.97, 2.96]

References

- Asparouhov, T., & Muthén, B. (n.d.). *Bayesian analysis using Mplus: Technical implementation*. Retrieved 18-04-2015, from <https://www.statmodel.com/download/Bayes3.pdf>
- 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*(2), 151-164.
- Bayarri, M. J., & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science, 19*(1), 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*(3), 931-951.
- Bollen, K. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review, 45*(3), 370-390.
- Bollen, K. (1989). *Structural equations with latent variables*. New York: John Wiley and Sons.
- 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*(3), 406-427.
- Casella, G. (1985). An introduction to empirical Bayes data analysis. *The American Statistician, 39*(2), 83-87.
- Casella, G. (1992). Illustrating empirical Bayes methods. *Chemometrics and Intelligent Laboratory Systems, 16*(2), 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*(23), 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*(2), 262.
- Depaoli, S. (2012). Measurement and structural model class separation in mixture CFA: ML/EM

- versus MCMC. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(2), 178–203.
- Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological methods*, 18(2), 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(2), 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*, 1-25.
- Depaoli, S., & van de Schoot, R. (in press). Improving transparency and replication in Bayesian statistics: The WAMBS-checklist. *Psychological Methods*.
- Dunson, D. B., Palomo, J., & Bollen, K. (2005). *Bayesian structural equation modeling* (Tech. Rep.). SAMSI. Retrieved from <http://www.samsi.info/sites/default/files/tr2005-05.pdf>
- Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470), 680–701.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19(1), 72-91.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Analysis*, 1(3), 515-534.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London, England: 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(4), 1360-1383.

- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 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(2), 193-214.
- Heerwegh, D. (2014). Small sample Bayesian factor analysis. In *Phuse*. 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(436), 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(2), 87-93.
- Kaplan, D. (2014). *Bayesian statistics for the social sciences*. Guilford Publications.
- Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435), 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(2), 259-275.
- 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(4), 653-686.
- 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(4), 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(1), 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(3), 86-92.
- MacCallum, R. C., Edwards, M. C., & Cai, L. (2012). Hopes and cautions in implementing Bayesian structural equation modeling. *Psychological Methods*, 17(3), 340-345.
- Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods*, 3(1), 45-58.
- 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.
- Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313-335.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide. seventh edition*. Los Angeles, CA: Muthén and Muthén.
- Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2011). A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological methods*, 16(4), 468-490.
- Polson, N. G., & Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, 7(4), 887-902.
- R Core Team. (2015). R: A language and environment for statistical computing. [Computer software manual]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical*

- Software*, 48(2), 1-36. Retrieved from <http://www.jstatsoft.org/v48/i02/>
- Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling*, 16(4), 583–601.
- Tversky, A. (1974). Assessing uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)*, 148–159.
- Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological methods*, 14(4), 301-322.