A Comparison of Inverse-Wishart Prior Specifications for Covariance Matrices in Multilevel Autoregressive Models

N. K. Schuurman, R. P. P. P. Grasman & E. L. Hamaker

Routledge
Taylor & Francis Group

A Comparison of Inverse-Wishart Prior Specifications for Covariance Matrices in Multilevel Autoregressive Models

N. K. Schuurman[a], R. P. P. P. Grasman[b], and E. L. Hamaker[a]

[a]Utrecht University; [b]University of Amsterdam

ABSTRACT

Multilevel autoregressive models are especially suited for modeling between-person differences in within-person processes. Fitting these models with Bayesian techniques requires the specification of prior distributions for all parameters. Often it is desirable to specify prior distributions that have negligible effects on the resulting parameter estimates. However, the conjugate prior distribution for covariance matrices—the Inverse-Wishart distribution—tends to be informative when variances are close to zero. This is problematic for multilevel autoregressive models, because autoregressive parameters are usually small for each individual, so that the variance of these parameters will be small. We performed a simulation study to compare the performance of three Inverse-Wishart prior specifications suggested in the literature, when one or more variances for the random effects in the multilevel autoregressive model are small. Our results show that the prior specification that uses plug-in ML estimates of the variances performs best. We advise to always include a sensitivity analysis for the prior specification for covariance matrices of random parameters, especially in autoregressive models, and to include a data-based prior specification in this analysis. We illustrate such an analysis by means of an empirical application on repeated measures data on worrying and positive affect.

Introduction

Psychological processes occur within individuals: stress affecting a person's mood, a mother's self-esteem influencing her teenage daughter's self-esteem, an individual's job satisfaction affecting job performance, and so on. It is likely that many of these dynamical processes also differ across individuals (see for instance Adolf, Schuurman, Borkenau, Borsboom, & Dolan, 2014; Hamaker, 2012; Lodewyckx, Tuerlinckx, Kuppens, Allen, & Sheeber, 2011; Molenaar, 2004; Rovine & Walls, 2006; Wang, Hamaker, & Bergeman, 2012). For instance, stressful situations may strongly affect the mood of one individual, while they have little effect on the mood of another individual. Multilevel autoregressive models are ideal for investigating these types of processes, because they allow for modeling how variables affect themselves and each other over time. Moreover, they allow for modeling these effects for each individual separately in the form of random parameters, and for the individuals on average, as a result of the inclusion of fixed effects.

Multilevel autoregressive models are complex models that can prove difficult to fit with software based on traditional maximum likelihood modeling, especially when considering multivariate or latent variable extensions, or

models that include random residual variances. In contrast, with Bayesian modeling software, such as Win-BUGS or OpenBUGS (Lunn, Spiegelhalter, Thomas, & Best, 2009; Lunn, Thomas, Best, & Spiegelhalter, 2000), fitting these complex multilevel models is relatively trivial (see for instance Song & Ferrer, 2012; Wang et al., 2012; or Lodewyckx et al., 2011, for an implementation in R). Other benefits of Bayesian modeling are its flexibility in handling missing data, and that it directly provides the researcher with the estimated random parameters. To benefit from this flexibility of Bayesian modeling, it is necessary to specify prior distributions for the parameters that are to be estimated. The prior distributions may be specified based on a researcher's prior knowledge about the parameters in question, such as results from previous research. However, when there is little or no prior information available, or when the researcher wishes to take a more objective approach, it may be desirable to specify uninformative prior distributions, prior distributions that have a negligible influence on the estimated parameters.

In certain cases it can be difficult to specify uninformative prior distributions. One of these cases is specifying priors for variances or for covariance matrices when the variances are small (close to zero). Typical prior

CONTACT N. K. Schuurman ✉ n.k.schuurman@uu.nl 📧 Methodology and Statistics, P. O. Box 80140, 3508 TC Utrecht, The Netherlands.

distributions chosen for variances and covariance matrices are Inverse-Gamma (IG) distributions and Inverse-Wishart (IW) distributions respectively (e.g., Gelman & Hill, 2007). These prior distributions, which are usually uninformative with certain hyperparameters, are quite informative when variances are small, resulting in a strong effect of the prior distribution on the parameter estimates (Gelman, 2006; see also Song & Ferrer, 2012, for an example). Gelman (2006) and Browne and Draper (2006) show that when a single variance is modeled, choosing a uniform distribution for the standard deviation or variance instead of the IG distribution results in parameter estimates that are negligibly affected by the prior distribution. However, the problem is harder to solve in the case of specifying a prior for a covariance matrix.

The issue is particularly relevant when considering multilevel models, because multilevel models are prone to having small variances in the covariance matrix of the random parameters. Small variances for the random parameters will result when the interindividual differences in the parameters are not very large, that is, the individuals have similar parameter estimates. However, it is important to note that the size of the variances also depends on the scale of the random parameters. For example, small variances for the random parameters may result as an artifact when the random parameters are restricted in range, which also restricts the size of their variance. For example, this may be the case when the random parameters are proportions, or probabilities. In the case of multilevel autoregressive modeling, the regression parameters are restricted in range as a result of the stationarity of the model (Hamilton, 1994). For instance, in a lag 1 autoregressive model (AR(1) model), where a variable is regressed only on itself at the preceding time point, the autoregression coefficient lies in the range from -1 to 1, which necessarily results in a small variance for this coefficient across individuals. As a result, it is difficult to specify uninformative priors for the covariance matrix of the random effects in multilevel autoregressive models (cf. Song & Ferrer, 2012).

Proper estimation of the covariance matrix of random parameters is essential for psychological research, in order to get an accurate impression of the magnitude of interindividual differences in the dynamics of individuals, and proper estimations of the covariances are necessary for getting an accurate impression of the associations among these interindividual differences. Therefore, the aim of this study is to compare the performance of several prior specifications for covariance matrices suggested in the literature, when one or more of the variances in the covariance matrix are close to zero. Specifically, we compare three IW prior specifications: (a) a prior specification that is based on an identity matrix, and is often used as an uninformative prior in practice, (b) a data-based prior that uses input from maximum likelihood estimations, and (c) the default conjugate prior proposed by Kass and Natarajan (2006). Although we are especially interested in the Bayesian estimation of multivariate autoregressive models, in the simulation study we use univariate autoregressive models with one outcome variable and a time varying predictor variable for practical reasons (explained further in the following section). We illustrate a full multivariate model in an empirical example on repeated measures of positive affect and worrying for 129 participants.

The remainder of this article is organized as follows. We start by discussing the multilevel autoregressive model in more detail, followed by a section on the IW distribution and the three prior specifications for the covariance matrix of the random parameters. After that we present our simulation studies and their results, and we present an empirical application in which we compare the effects of the different prior specifications for a multivariate model. We end with our main conclusions and a discussion.

Multilevel autoregressive model

In autoregressive models variables are regressed on themselves and each other on a previous time point. In such a model, the autoregression coefficient reflects the influence the previous state of a variable has on its current state, and crossregression coefficients reflect the influence of the previous value of another variable has on the current state of this variable (Hamilton, 1994; Kim & Nelson, 1999). Multilevel extensions of these models allow for modeling these dynamic processes for multiple persons, and to model the average intraindividual effects over the multiple subjects, which helps generalize the results to a larger population.

Although our main interest is in specifying uninformative priors for full multivariate multilevel autoregressive models with multiple outcome variables, for the simulation study we focus on a univariate multilevel autoregressive model with one outcome variable and a lagged predictor variable instead, for practical reasons: a bivariate multilevel autoregressive model contains six random effects (i.e., two autoregression parameters, two crossregression parameters, and two means), such that the covariance matrix of the random effects contains six variances and 15 covariances. Estimating such a model using WinBUGS is time intensive and computationally demanding, which would make a simulation study based on such a model challenging. Instead, we focus on a univariate multilevel autoregressive model with a lagged predictor, which contains only three random effects, such that the covariance matrix of the random effects contains three variances and three covariances. However, we emphasize that the model
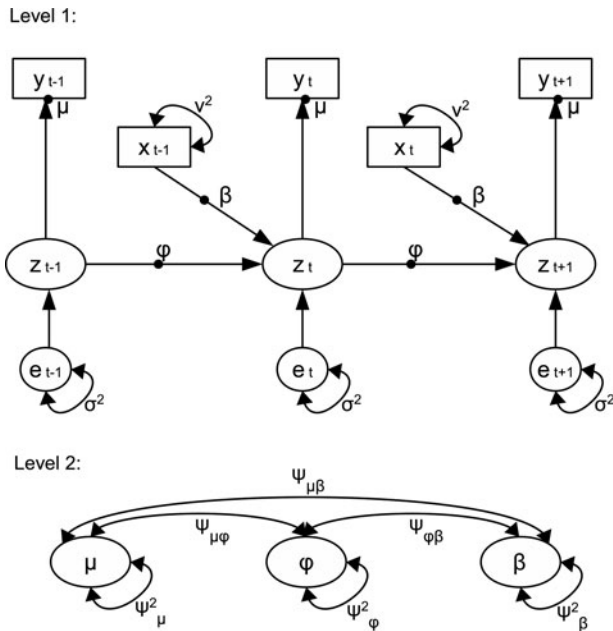
Figure 1. Multilevel AR(1) model with time-varying predictor. At level 1, the outcome variable y for individual j is regressed on itself at the previous time point $t - 1$, and on a time varying predictor x at the previous time point $t - 1$. The mean of y, μ, the autoregression coefficient ϕ, and the crossregression coefficient β, are allowed to vary over individuals j (indicated by the black dots). At level two, the random coefficients are multivariate normally distributed, with the covariance structure as indicated in the figure.

can be generalized to include more than one outcome variable. We illustrate such a multivariate model in the empirical example, for which we fit a bivariate model (this took approximately 24 hours with three chains, without parallel computing for the three chains). For a graphical representation of the univariate model with a lagged predictor, see Figure 1.

In the univariate multilevel AR(1) model with a time-varying predictor, y_{tj} is a score on outcome variable y for person j at time point t. The scores y_{tj} are split in individual means μ_j, and a residual score z_{tj}. The autoregression and crossregression effects are modeled using residual scores z_{tj}: z_{tj} is regressed on z_{t-1j}, the residual score for outcome variable y_{t-1j} for person j at previous time point $t - 1$, and on x_{t-1j}, the score on a time-varying predictor variable x for person j at time point $t - 1$. Modeling the autoregression effects on z_{tj} rather than on y_{tj} directly allows us to estimate the means μ_j directly rather than the intercepts. The means represent the baseline score for an individual, which is more intuitive than the intercept, which represents the score of an individual when the predictor variables are zero. The autoregression coefficients ϕ_j represent the association between the outcome variable y at time t with itself at time $t - 1$. The larger the absolute autoregression coefficient, the better future values of y can be predicted by the previous value of y.

Positive autoregression coefficients are also interpreted as a measure of inertia—the larger the autoregression coefficient, the slower it will take for y to return to its baseline μ after a perturbation of the system (Suls, Green, & Hillis, 1998). The crossregression coefficients β_j indicate how well a past value of a predictor x predicts the future value of y. In multivariate models the crossregression coefficients can be used to investigate the reciprocality of the effects between multiple variables (Moberly & Watkins, 2008; Nezlek & Allen, 2006; Nezlek & Gable, 2001). Innovation e_{tj} represents anything that is not directly measured that may influence the system. These innovations are assumed to be normally distributed with a mean of zero and variance σ^2. In other words, at level 1 the multilevel model can be specified as:

$$y_{tj} = \mu_j + z_{tj}$$
$$z_{tj} = \phi_j z_{t-1j} + \beta_j x_{t-1j} + e_{tj} \qquad (1)$$
$$e_{tj} \sim N(0, \sigma^2). \qquad (2)$$

In this model, three parameters are allowed to vary over individuals: μ_j, the mean for person j; ϕ_j, the autoregression coefficient for person j; and β_j, the crossregression coefficient for person j. We will refer to these individual parameters as random parameters, and assume that they are multivariate normally distributed, with means γ_μ, γ_ϕ, and γ_β, and 3×3 covariance matrix $\mathbf{\Psi}$. The means describe the average effects (i.e., fixed effects) for the group of individuals, and the covariance matrix describes the variations around these means for the group of individuals. Hence, at level 2 we have:

$$\begin{bmatrix} \mu_j \\ \phi_j \\ \beta_j \end{bmatrix} \sim MvN \left\{ \begin{bmatrix} \gamma_\mu \\ \gamma_\phi \\ \gamma_\beta \end{bmatrix}, \begin{bmatrix} \psi_\mu^2 & & \\ \psi_{\mu\phi} & \psi_\phi^2 & \\ \psi_{\mu\beta} & \psi_{\phi\beta} & \psi_\beta^2 \end{bmatrix} \right\}. \qquad (3)$$

We focus here on autoregressive processes that are stationary for each individual, meaning that the mean and variance of the outcome variable are stable over time for each individual. If the AR(1) process is stationary, the autoregression parameters ϕ_j will lie within a range of -1 to 1 (Hamilton, 1994).[1] When this holds for every individual,

---

[1] Note, however, because we assume that ϕ comes from a multivariate normal distribution, that technically autoregressive parameters outside of this range can occur. In our simulation study we chose the mean vectors and covariance matrices for the multivariate normal distribution so that parameters not in line with the stationary assumption are extremely unlikely. We chose these parameters to be in line with what we generally have encountered for autoregressive modeling in psychological practice: stationary processes with autoregressive parameters in a range of about 0 to.5. Note that encountering a nonstationary parameter value during estimation is not problematic for the estimation procedure, so that using a multivariate normal distribution rather than, for instance, a truncated multivariate normal distribution should not result in any technical (estimation) problems. In practice, encountering such a nonstationary parameter value would simply imply that the process is not stationary for that person. It may then be useful to consider different or extended models that models nonstationarity in an informative way (c.f. Hamilton, 1994).

the variance of the autoregression coefficients ψ_ϕ^2 will be small. For instance, the mean of .4 and a variance of .04 for ϕ_j would result in a relatively large range of possible values for ϕ_j, namely a 95% interval of [.008,.792], whereas a variance of .01 would still result in a relatively large 95% interval of [.204,.596]. Similar ranges are found empirically, for instance by Wang et al. (2012): they studied daily measures of negative affect, and found a γ_ϕ of .15 and a variance ψ_ϕ^2 of .04. The means μ_j and crossregression coefficients β_j are not restricted in range. Of course, the variances of μ_j and the random crossregression coefficients β_j may be small as well, due to the scale of the variables resulting in a small coefficient,[2] or simply due to minimal individual differences in these coefficients. Given that the standard priors for covariance matrices are very informative when variances are small, it will be difficult to specify the prior distribution for the covariance matrix of the random parameters Ψ such that it has a negligible influence on the results. In the next section we will go into more detail about the priors for covariance matrices, and why they are informative when variances are small.

Priors for the covariance matrix of the random parameters

For Bayesian estimation of the multilevel autoregressive model, prior distributions have to be specified for the random parameters (i.e., μ_j, ϕ_j, β_j for $j = 1, \ldots, n$), for the fixed effects (i.e., γ_μ, γ_ϕ, γ_β), for the innovation variance (i.e., σ^{22}), and for the covariance matrix of the random parameters (i.e., Ψ). When an influence of the prior distributions on the results is undesirable, for instance when no relevant prior information is available, it is desirable to specify uninformative prior distributions that have a negligible influence on the end results. However, prior specifications that are uninformative in specific circumstances, may become informative under different circumstances. Our main interest here is in how to specify an uninformative prior distribution for Ψ, so that the influence of the prior specification on the estimates of the variances and covariances of the random effects is minimal, under the specific circumstance that the true sizes of some of these variances are small, as would be the case for the autoregression coefficients ϕ_j.

For this purpose we will first discuss the IW distribution, which is the conjugate prior for covariance matrices given normally distributed data, then we will go into more detail about the prior specification problem for covariance matrix Ψ, and we will discuss three prior specifications for Ψ suggested in the literature.

The IW prior distribution

The prior distribution that is typically used for the covariance matrix of multivariate normally distributed variables, such as the covariance matrix Ψ for the random effects, is the IW distribution (Gelman et al., 2013; Gelman & Hill, 2007). The IW distribution is a conjugate prior for the covariance matrix of multivariate normal distributed variables, which implies that when it is combined with the likelihood function, it will result in a posterior distribution that belongs to the same distributional family. Another important advantage of the IW distribution is that it ensures positive definiteness of the covariance matrix.

The IW distribution is specified with an $r \times r$ scale matrix $\mathbf{S}$, where r is equal to the number of random parameters, and with a number of degrees of freedom df, with the restriction that $df > r - 1$. $\mathbf{S}$ is used to position the IW distribution in parameter space, and the df set the certainty about the prior information in the scale matrix. The larger the df, the higher the certainty about the information in $\mathbf{S}$, and the more informative is the distribution (Gelman et al., 2013; Gelman & Hill, 2007). The least informative specification then results when $df = r$, which is the lowest possible number of df.

The means and covariance matrix of the IW distribution are a function of the elements s_{kl} on row k and column l from $\mathbf{S}$, r, and the df. That is, the density of the IW distribution is

$$\frac{\left|\mathbf{S}^{\frac{df}{2}}\right|}{2^{\frac{dfr}{2}}\Gamma_r\left(\frac{df}{2}\right)}\,|\mathbf{X}|^{-\frac{df+r+1}{2}}\,e^{-\frac{1}{2}tr(\mathbf{S}\mathbf{X}^{-1})} \qquad (4)$$

where $tr()$ stands for the trace function, and Γ_r stands for the multivariate Gamma function. The mean of the IW distribution is

$$E[\mathbf{X}] = \frac{\mathbf{S}}{df - r - 1} \qquad (5)$$

and the variance of each element of the IW distribution is

$$Var[x_{kl}] = \frac{(df - r + 1)s_{kl}^2 + (df - r - 1)s_{kk}s_{ll}}{(df - r)(df - r - 1)^2(df - r - 3)}). \qquad (6)$$

---

[2] Note that it is possible to increase the means μ_j and their variance, by transforming the relevant outcome variable (e.g., multiply the variable by 10). When the variance for the mean is increased so it is no longer close to zero, specifying the IW prior distribution for this coefficient will be relatively trivial. However, this is not possible for the autoregressive coefficient, because it is standardized as a result of the stationarity assumption that results in equal variances for y_t and y_{t-1} (Hamilton, 1994). While the transformation is possible for the crossregression coefficients, in a multivariate model increasing one crossregression coefficient results in decreasing the other crossregression coefficient—merely shifting the problem to another coefficient. Further, it may be difficult to determine in advance by how much to increase a coefficient, since its value is unknown a priori.

The variances for the diagonal elements of the IW distribution simplify to

$$Var[x_{kk}] = \frac{2s_{kk}^2}{(df - r - 1)^2(df - r - 3)}. \qquad (7)$$

It can be seen from Equation (6) that when the *df* increase, the denominator will increase more rapidly than the numerator, so that the variance will become smaller, which implies that the IW distribution will become more informative. It can also be seen that the size of the variance is partly determined by **S**: The smaller the elements of **S**, the smaller the variance of the IW distribution, and hence the more informative the prior will be. However, setting the scale to large values also influences the position of the IW distribution in parameter space, as can be seen from Equation (5). In other words, specifying a IW prior distribution requires balancing the size of **S** and the *df*.

A typically used relatively uninformative IW prior is a prior with small *df* and an identity matrix **S**. In many situations this prior specification will be uninformative enough for the the data to dominate the prior, so that the influence of the prior on the results will be minimal. However, when the variances are quite small, IW priors are informative, so that the estimates for the variances will be sensitive to the IW prior specification, resulting in over- or under-estimation of the variances depending on the specification of the prior distribution. The reason for this sensitivity when the variances are close to zero is that the IW distribution is bounded at zero for the variances: in consequence of this boundedness, slightly changing the central tendency of the distribution can have large effects on the weights placed on values close to zero.

We illustrate this in Figure 2, which shows eight plotted marginal densities for one of the diagonal elements of a bivariate IW distribution with varying *df* and **S**. The four panels include two densities with the same diagonal **S**, with respectively.001,.01,.1, or 1 as diagonal elements. For each panel, the density plotted in black has a larger *df* than the corresponding density plotted in gray. These plots further demonstrate that the IW-distribution tends to place either a lot of weight on a specific value close to zero (as in the upper panels), or place almost no weight close to zero (as in the lower panels). This shows that the IW-distribution is easily misspecified when variances are small. When the prior is specified too far from zero (e.g., IW prior with **S** as an identity matrix), this will result in an overestimation of the variances. However, specifying the central tendencies too close to zero will result in an underestimation of the variances, firstly because too much weight is shifted towards zero, and secondly because an element of the scale matrix set close to zero will also have a small variance (the density is more peaked). This is the case for instance for an IG distribution—which is basically a univariate simplification of the IW distribution—with a shape and scale close to zero (e.g., IG[$10^{-5},10^{-5}$]). This IG distribution is often considered as an uninformative prior specification for a single variance, however it has been shown that this indeed results in an underestimation of the variance when this variance is small (Browne & Draper, 2006; Gelman, 2006). Although Gelman (2006) demonstrates that in the univariate case it is possible to use a Uniform or Inverse Half-Cauchy distribution instead of the conjugate IG distribution, giving good results, the solution to this problem is less simple for multivariate (IW prior) cases. In the following we will discuss three IW prior specifications that have been suggested in the literature.

Three ways to choose S for the IW prior distribution

In order to find the best way to specify the prior for $\mathbf{\Psi}$ when some of the variances are close to zero, we will evaluate the performance of three IW priors for $\mathbf{\Psi}$ that have been suggested in the literature, using a simulation study.[3] Note that for most Bayesian software, including the software WinBUGS that we use for the simulation study (Lunn et al., 2000), one actually specifies a Wishart distribution for the precisions, rather than the IW for the variances. The relation of the IW and the Wishart distribution is that if X (here, the precision matrix) is Wishart distributed with scale matrix V and degrees of freedom df, then variable X^{-1} (here, the covariance matrix) is IW distributed with scale matrix V^{-1} and degrees of freedom df. Below, we discuss the prior specifications in terms of the IW distribution. For the corresponding Wishart specification, the scale matrix is simply inverted. For all three specifications the df are set equal to r (here $r = 3$), so that the priors are minimally informative (Gelman et al., 2013; Gelman & Hill, 2007).

The first IW prior specification we will examine is the one that is commonly used as an uninformative prior specification, and which we will refer to as the Identity Matrix (ID) specification. In this specification the diagonal elements of scale matrix **S** are set to 1 and the off-diagonal elements are set to zero. We expect that this prior specification will prove to be quite informative in the current context where the variance of ϕ_j is small.

---

[3] We considered the scaled Wishart described by Gelman and Hill (2007) as well; however, this specification resulted in traps in WinBUGS (e.g., the estimation procedure would crash). Further, we considered specifying the variances and covariances in a regression structure avoiding the use of the IW prior specification, specifying the model with univariate priors, and to transform the random parameters so that they have a larger variance, and specifying an IW prior for the covariance matrix of the transformed parameters. For this work however, we decided to focus on different specifications of IW specifications suggested in the literature. More information on the other specifications is available from the first author.
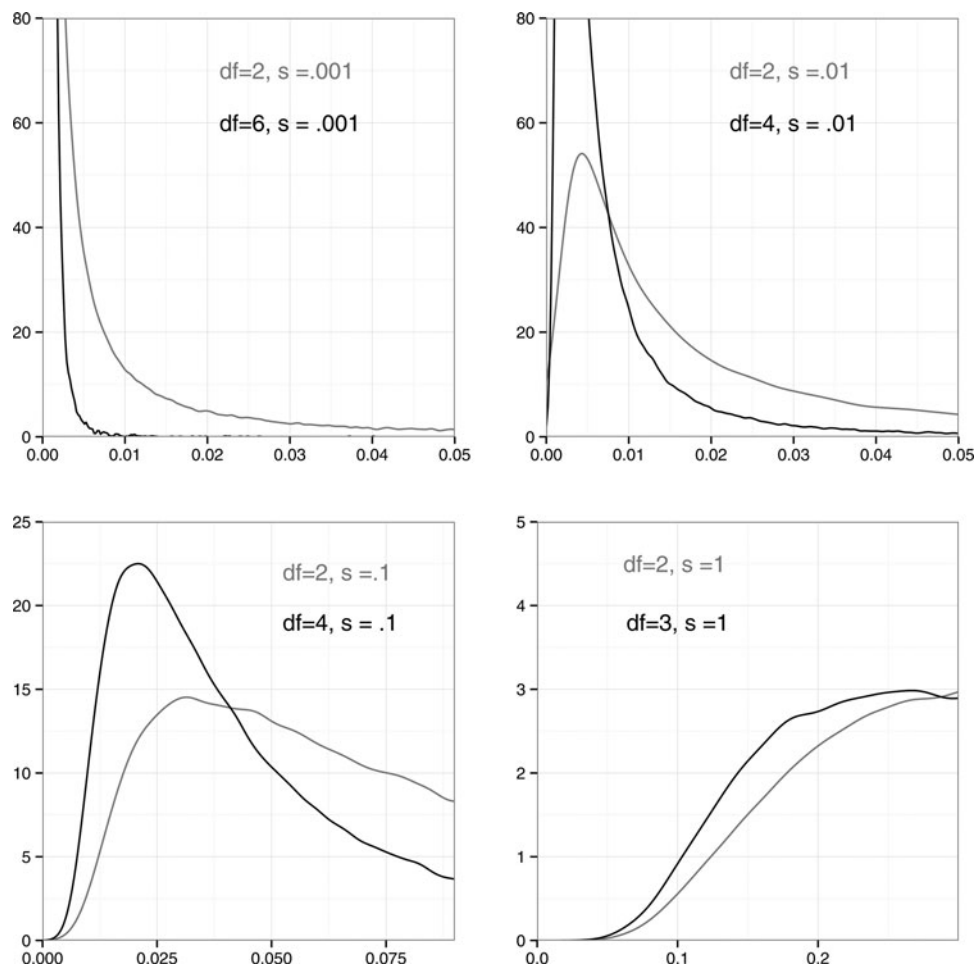
Figure 2. Eight Inverse-Wishart (IW) marginal probability densities, each specified with specific degrees of freedom (*df*) and scale matrix *S*. The densities are based on samples from a bivariate IW distribution with a diagonal *S*. All eight specifications are informative: in the area around zero, each specification gives either a lot of weight to a specific value around zero, or gives almost no weight around zero. As such, the IW-distribution is easily misspecified if a IW-distributed variance is close to zero.

The second IW prior specification that will be examined is an IW prior specification in which the scale matrix is based on prior estimates of the variances of the random parameters. Using estimates of the variances as input for the IW prior specification ensures that the prior specification will be close to the data, and therefore should limit bias. However, this requires us to use the data twice; once for estimating the input for the prior, and again for the likelihood. When the data are used twice, the certainty about the estimated parameters is exaggerated (Kass & Steffey, 1989; see also discussions on the use of Empirical Bayes by Gill, 2014, p. 276–270, and Lindley, 1969, p. 420–421). This can have statistical repercussions: because certainty about the point estimates is exaggerated, the standard deviations of the point estimates and their credible intervals will become too small (Kass & Steffey, 1989). How much the estimates will be influenced by using the data twice, will depend on how, and how much of the data is used. When the information in the used data is little, the effect will be negligible asymptotically (see for instance

Berger & Pericchi, 1996; O'Hagan, 1995, on training data). Setting the *df* of the IW specification as small as possible minimizes the information value of the data-based prior, and therefore limits the effects of exaggerating the certainty about the point estimates. We will examine the effect of using the data twice as part of the simulation study, for instance by examining the coverage rates for the credible intervals of the estimated variances.

For the simulation study we will use a maximum likelihood procedure to obtain prior estimates of the variances, and we will refer to the prior specification based on these estimates as the maximum likelihood input (ML) specification. In this ML prior specification, we specify the Wishart prior distribution in WinBUGS so that the ML estimates of the variances of the random parameters are plugged into the Wishart distribution scale matrix S^{-1} so that the mean of the Wishart distribution equals the estimated precisions (inverted variances). Note however that it is also possible to obtain estimates of the variances by other means—for instance by fitting a Bayesian

model with uniform priors on the variances (ignoring any covariances), and base the IW or Wishart scale matrix on those estimates of the variances. Another option would be to fit an autoregressive model for each individual separately, provided that there are enough repeated measures per person to do this. Afterwards, the variances can be estimated by computing the variances of the estimated individual coefficients, which can then be used for the IW or Wishart prior specification. For our simulation study we opt for ML mainly because of its speed. We obtain the ML estimates by fitting the model in R (R Development Core Team, 2012) with the R-package lme4 (Bates, Maechler, & Bolker, 2012). In order to estimate the variance of the mean and not the variance of the intercept, the model in lme4 was fit on person-centered data. We used only the ML estimated variances as input for the prior specification, while setting the prior covariances to zero, because preliminary results showed that using estimates of the full covariance matrix decreased performance, probably because the ML-estimates for the covariances were not close to the true values.

The third IW specification we consider is the default conjugate (DC) prior proposed by Kass and Natarajan (2006). In the DC prior specification, the mean of the IW distribution is set to

$$\left(\frac{1}{n} \sum_{j=1}^{k} \boldsymbol{Z}_j' \boldsymbol{W}_j \boldsymbol{Z}_j \right)^{-1} \tag{8}$$

where n is the number of participants, $\boldsymbol{Z}_j$ is the design matrix for person j, and $\boldsymbol{W}_j$ is the generalized linear model weight matrix for person j. The latter is based on (estimates of) model parameters. In the case of a normal model, $\boldsymbol{W}_j$ is a diagonal matrix with $1/\sigma^2$ on the diagonal (see Fahrmeir, Tutz, & Hennevogl, 2010). Given that we need an estimate of the residual variance σ^2 for the specification of the DC prior specification, we fit the multilevel model with maximum likelihood techniques in R-package lme4, and use ML-estimates of σ^2 as input for the generalized linear model weight matrix. Therefore, this specification is also data-based. However, the information in the data used will be little, so that the effect of using the data twice should be negligible, as is shown by Kass and Natarajan (2006).

The effect of the DC prior specification is that half of the prior weight on the random parameters is given to the common effects ($\gamma_\mu, \gamma_\phi, \gamma_\beta$), and half of the weight is given to estimates for each individual separately (μ, ϕ, β for each individual) as if a model was fit for each individual separately. This approach is directly related to shrinkage estimates (see Bryk & Raudenbush, 1992; Kass & Natarajan, 2006, but note that the weight on the random

parameters is not necessarily one half for shrinkage estimates). In other words, the prior information is specified so that the prior weight is in between a parameter variance of zero (i.e., no individual differences) and the maximum parameter variance (i.e., maximum individual differences).

Kass and Natarajan (2006) compare the performance of the DC prior specification and ML specification for a Poisson model in a simulation study. In their study the DC prior outperformed the ML prior in terms of coverage rates, and squared and entropy loss. However, the model used was univariate with respect to the random parameters: only one parameter was random, so only one variance had to be modeled. Hence it remains unclear how the DC prior performs with regard to the estimation of the covariances between the random parameters. It also remains unclear how the DC prior performs when variances are close to zero. We will investigate these issues in the simulation study.

Simulation study

Our simulation study consists of two parts: in the first part we examine the performance of the Wishart priors for different sizes of (small) variances in $\boldsymbol{\Psi}$, and in the second part we examine the performance of the IW prior specifications for different sample sizes and covariance structures when one or more variances in the covariance matrix are small. We compare three prior specifications for $\boldsymbol{\Psi}$ as discussed previously: the ID specification, the ML specification, and the DC specification (Kass & Natarajan, 2006). We will evaluate the performance of these three specifications against a specification that has the df set to 3, and the means of the IW distribution set to the true values. In practice this benchmark (BM) specification of course cannot be used, but we use it in the simulation study to determine optimal performance. For both parts the data are simulated according to the previously described model in open source software R (R Development Core Team, 2012). For both parts of the study the models are simulated 1,000 times (1,000 replications). In both parts of the simulation study γ_μ is set to 3, ψ_μ^2 is set to 0.25, γ_ϕ is set to 0.3, γ_β is set to 0.35, and σ^2 is set to 1. The variance of the predictor variable x, ν^2, is set to 1.2.

For both parts we implemented and estimated all models in free Bayesian modeling software WinBUGS (Lunn et al., 2000), in combination with the R-package R2WinBUGS (Sturtz, Ligges, & Gelman, 2005). We chose $Normal(0, 10^{-9})$ priors (specified in terms of precision instead of variance, which is required in WinBUGS) for the fixed effects $\gamma_\phi, \gamma_\beta$, and γ_μ, and a $Uniform(0, 10)$ prior for σ^2, the residual variance at level 1. We evaluated the

convergence of each model based on the visual inspection of the mixing of the three chains, and the Gelman-Rubin convergence diagnostic (Gelman & Rubin, 1992). We also evaluated the autocorrelations for the samples. Practically it was not possible to evaluate the convergence for each replication of each model in the simulation study (e.g., run each replication with three chains and visually inspect the convergence). Instead, we fitted and evaluated the convergence for one replication for each different condition in Part I and II of the study.

For all models convergence results were fairly similar. The three chains mixed well for all models and parameters. The Gelman-Rubin diagnostic was 1, or very close to 1, for the parameters in all models. Generally, the autocorrelations decreased exponentially to zero for the parameters σ^2, γ_μ, and ψ_μ, and the individual μ_js. For the remaining parameters, the autocorrelations were generally a bit slower to decrease, depending on the sample size and size of the variances of the random parameters. For most model specifications autocorrelations for these parameters diminished to zero after about 20 lags. Autocorrelations diminished to zero after about 40 to 60 lags when variance size is the smallest (0.0025), and sample size is the smallest (25 persons and repeated measures). Based on the convergence results and the autocorrelations, we judged 40,000 iterations with 30,000 burn-in iterations as sufficient for convergence.

As point estimates for the parameters we used the means of the posterior distributions. Performance is evaluated using: a) coverage rates of the 95% (equal tailed) credibility intervals (CIs), which we would expect to be about .95 when the priors are uninformative; (b) bias, which is computed by taking the average of the difference between the true value and the point estimate across all 1,000 replications; and (c) the ratio of the average posterior standard deviation and the standard deviation of the posterior averages, which should be about 1 if the posterior standard deviations reflect the actual sampling variation.

Part I: The effect of small variances in Ψ

In order to study the effect of the (small) size of the variances of the random parameters per prior for $\boldsymbol{\Psi}$, the variances ψ_ϕ^2 and ψ_β^2 were either set to .0025, .01, or .0225. These variances result in 95% intervals for the autoregression coefficients of respectively [.202, .398], [.104, .496], and [.006, .594]. These ranges are in line with autoregressive coefficients reported in the literature, which are usually small and positive (e.g, Moberly and Watkins (2008), Nezlek and Gable (2001), Suls et al. (1998), and Wang et al. (2012) report fixed autoregression effects between .08

and .3 approximately). The sample size is set to 50 individuals and 50 time points. All the correlations between the random parameters are set to .3. This results in a 4 × 3 (i.e., prior specification × size variance) simulation design. Below, we summarize the results for Part I of the simulation study. More detailed results for the simulation study are available in the supplementary materials, and the simulated data are available upon request from the first author.

The results show that overall, the ML prior specification performs best. The bias of the ML specification is quite close to that of the BM specification. It can be seen from Figure 3 that even though coverage rates are lower than 95% for ψ_ϕ^2 and ψ_β^2 for this prior specification, it outperforms both the ID specification and DC specification. The coverage rates for ψ_ϕ^2 and ψ_β^2 are lower than .95 likely as a result of the double use of data: the data is used once in the prior and again in the likelihood, and as a result the information about the estimation is exaggerated, which in turn results in smaller credible intervals. The ID specification severely overestimates ψ_ϕ^2 and ψ_β^2. The DC specification performs well only if the prior specification is close to the true values of ψ_ϕ^2 and ψ_β^2. In this simulation study this was the case when the true variances were .01 or .0225, but not when they were .0025. Since in practice it is unknown if the DC prior is close to the true values of ψ_ϕ^2 and ψ_β^2, it is an undependable prior to use when the aim is to use an uninformative prior for the covariance matrix, while some variances are close to zero. The ML specification on the other hand is by definition close to the information in the data and therefore performs relatively well.

Note that when a variance is further away from zero, all prior specifications perform reasonably well: the true value for the variance ψ_μ^2 for random effect μ_j was .25, and it can be seen from Table 1 to 3 that all prior specifications perform well for this parameter. We discuss the results per prior specification in more detail below.

Performance ID specification

Overall, the ID specification performs poorly. From Figure 3 it can be seen that the coverage rates for ψ_ϕ^2 and ψ_β^2 are equal to zero regardless of the true sizes for ψ_ϕ^2 and ψ_β^2, indicating that the true values were never within the credible interval. This is due to a large bias in these parameter estimates: the parameters ψ_ϕ^2 and ψ_β^2 are severely overestimated. The coverage rates for the remaining parameters on the other hand are equal or close to 1 (see Table 1), as a result of too conservative standard deviations for all parameters for the ID specification. This is illustrated by the the ratios of the average posterior standard deviations and calculated standard deviations of the posterior means, over 1,000 replications, reported in
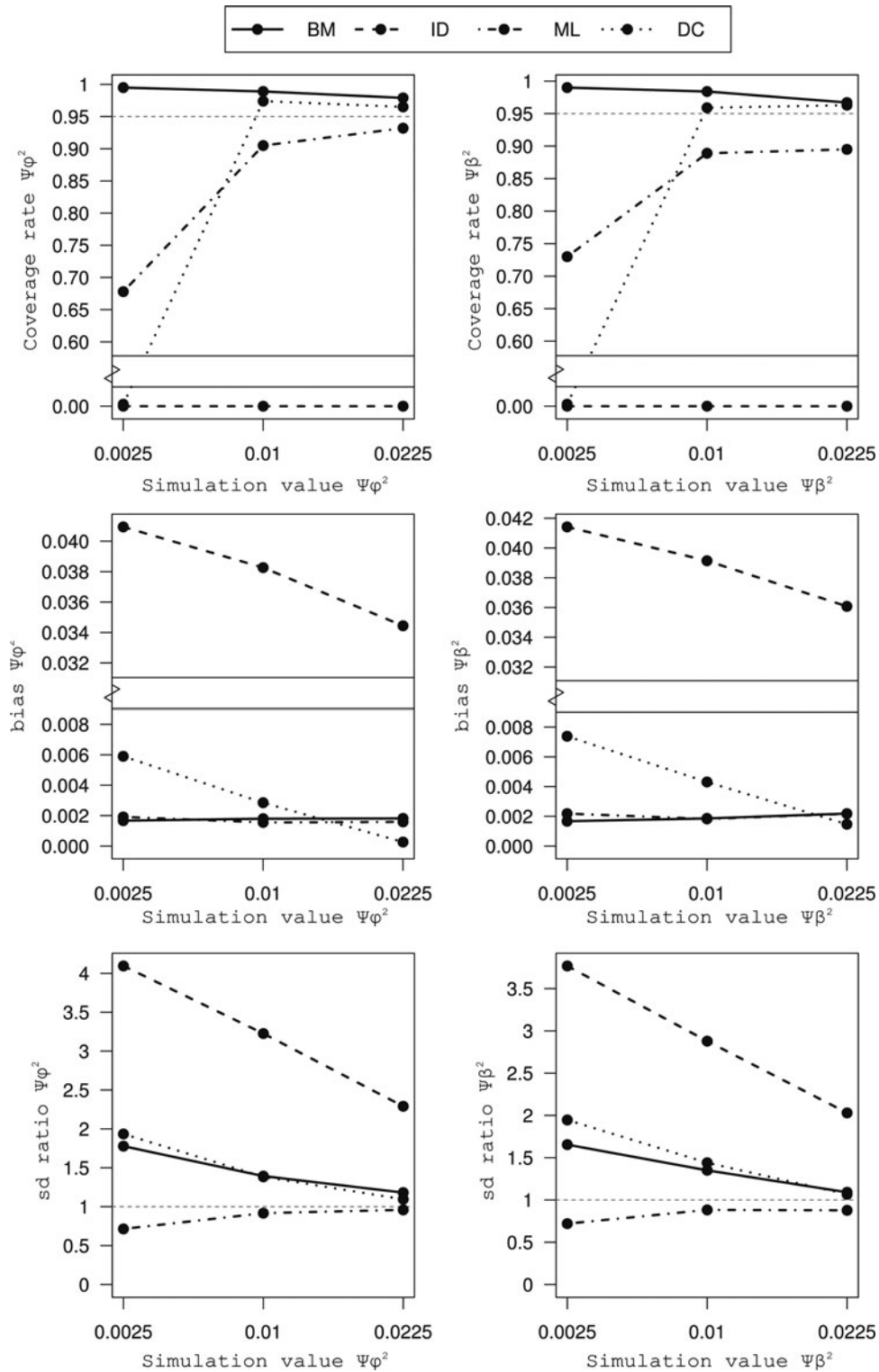
Figure 3. Part I coverage rates, estimated bias, and ratios of the average estimated posterior standard deviations and the standard deviations of the estimated posterior means, for the variance ψ_ϕ^2 of the autoregression coefficient and the variance ψ_β^2 of the cross-lagged coefficient calculated over 1,000 replications. The coverage rates, bias, and ratios of standard deviations are shown for the benchmark (BM), identity matrix (ID), maximum likelihood (ML) input, and default conjugate (DC) prior specification, for true values for ψ_ϕ^2 and ψ_β^2 of .0025, .01, and .0225. Overall, the ID specification clearly performs the worst. The performance of the ML specification performs the best, as it is closest overall to the benchmark specification. The performance of the DC specification depends on if the specification is close or not to the true value of ψ_ϕ^2 and ψ_β^2, as can be seen most clearly from the graphs of the bias and coverage rates. The DC specification of performs relatively well when ψ_ϕ^2 is .01 or .0225, but its performance dramatically decreases when ψ_ϕ^2 is .0025.

Table 1. Part I: Coverage rates for the 95% credible intervals, calculated over 1,000 replications.

	$\psi_\beta^2 = \psi_\phi^2 = .0025$				$\psi_\beta^2 = \psi_\phi^2 = .01$				$\psi_\beta^2 = \psi_\phi^2 = .0225$			
	BM	ID	ML	DC	BM	ID	ML	DC	BM	ID	ML	DC
μ_j	.949	.951	.950	.948	.949	.952	.950	.949	.947	.949	.947	.946
β_j	.970	.977	.920	.990	.947	.972	.929	.959	.941	.964	.937	.939
ϕ_j	.971	.979	.893	.984	.948	.974	.928	.953	.941	.965	.937	.936
γ_μ	.937	.939	.935	.932	.967	.966	.965	.962	.948	.948	.946	.940
γ_β	.966	.999	.946	.981	.968	.999	.963	.972	.948	.991	.946	.948
γ_ϕ	.964	.999	.937	.979	.957	.999	.956	.962	.950	.992	.946	.947
ψ_μ^2	.948	.952	.930	.941	.937	.940	.917	.946	.965	.968	.945	.967
ψ_β^2	.990	.000	.730	.000	.984	.000	.889	.959	.967	.000	.895	.963
ψ_ϕ^2	.995	.000	.678	.003	.989	.000	.905	.974	.979	.000	.932	.965
$\psi_{\mu\beta}$	.991	1.000	.950	.988	.980	1.000	.968	.975	.978	.998	.973	.968
$\psi_{\mu\phi}$	.987	.999	.876	.832	.977	1.000	.953	.872	.975	.998	.964	.912
$\psi_{\phi\beta}$	1.000	1.000	.907	1.000	.985	1.000	.962	.986	.971	1.000	.958	.957
$\rho_{\mu\beta}$	.996	.848	.984	.951	.986	.940	.981	.975	.975	.959	.976	.966
$\rho_{\mu\phi}$	.991	.851	.978	.681	.980	.929	.984	.857	.976	.952	.970	.928
$\rho_{\phi\beta}$	.999	.676	.998	.987	.987	.849	.985	.980	.978	.921	.979	.971

Note. A coverage rate of .95 is considered optimal. The coverage rates are calculated for three different true values of variances ψ_ϕ^2 and ψ_β^2. Coverage rates are shown for the benchmark (BM) prior specification, the identity matrix (IM) prior specification, the maximum likelihood (ML) input specification, and the default conjugate (DC) prior specification, and for the following parameters: the random effects μ_j, ϕ_j, and β_j, the fixed effects γ_{μ_j}, γ_{ϕ_j}, and γ_{β_j}, all elements from the covariance matrix Ψ for the random effects, and the correlations between the random effects $\rho_{\mu\phi}$, $\rho_{\mu\beta}$, and $\rho_{\phi\beta}$.

Table 3, which are much larger than 1—indicating too large posterior standard deviations.

Performance ML specification

The coverage rates for the ML specification are similar to those for the BM specification, except for the estimates of the random effects ϕ_j, β_j, and their variances ψ_ϕ^2 and ψ_β^2. The coverage rates for ψ_ϕ^2 and ψ_β^2 for the ML specification are low compared to the BM specification, ranging from approximately .70 when ψ_ϕ^2 and ψ_β^2 are .0025 to approximately .90 when ψ_ϕ^2 and ψ_β^2 are .01 or .0225. These coverage rates are however considerably better than those for the ID specification. The coverage rates for the random parameters ϕ_j and β_j range from .893 to .936 (lowest rates for a true variance of .0025), which is relatively low compared to the other prior specifications, including the BM specification. These results are consistent with the relatively small posterior standard deviations for ψ_ϕ^2, ψ_β^2, ϕ_j, and β_j for the ML specification compared to the other prior specifications, as can be seen from the ratios of standard deviations in Table 3. The relatively small posterior standard deviations for these parameters are likely the consequence of the double use of data. Further, it can be seen from Table 2 that the ML specification results in very little bias compared to the ID and DC specification, and that the amount of bias is actually similar to that of the BM specification.

Performance DC specification

The performance of the DC specification varies depending on the true values for the variances ψ_ϕ^2 and ψ_β^2, as can

be most clearly seen from a plot of the bias and coverage rates in Figure 3. The DC specification performs well when ψ_ϕ^2 and ψ_β^2 are .01 and .0225: the coverage rates, bias, and ratios of the posterior standard deviations and standard errors for the DC specification are then close to those for the BM specification. However, when the variances are equal to .0025, performance strongly declines, with coverage rates for ψ_ϕ^2 and ψ_β^2 that are equal to zero. Closer inspection of the results indicates that this low coverage rate is due to an upward bias for these parameters. The ratio of the posterior standard deviations and standard errors also strongly increases when ψ_ϕ^2 and ψ_β^2 are .0025, indicating that the posterior standard deviations are overestimated. The upward bias for the parameters is so large however that it is not compensated by the relatively wide credible intervals. The DC prior has low coverage rates for the covariance and correlation between μ_j and ϕ_j. This is the result of a downward bias for this parameter, which seems due to the DC prior specification that sets $\rho_{\phi\mu}$ to approximately $-.90$ based on Equation (8). The coverage rates for the remaining parameters are high compared to the BM specification, due to relatively large posterior standard deviations for these parameters, as shown in Table 3.

The discrepancy in performance of the DC prior specification across the three values of ψ_ϕ^2 and ψ_β^2 probably results because the DC specification does not depend directly on ψ_ϕ^2 and ψ_β^2, so that it does not change much in accordance with ψ_ϕ^2 and ψ_β^2. Therefore, the input for the DC specification may be similar regardless of the true value for ψ_ϕ^2 and ψ_β^2. When the information in the DC specification for ψ_ϕ^2 and ψ_β^2 is not close to the information

Table 2. Part I: Estimated bias for the estimated means of the posterior distributions, calculated over 1,000 replications.

True	$\psi_\beta^2 = \psi_\phi^2 = .0025$				$\psi_\beta^2 = \psi_\phi^2 = .01$				$\psi_\beta^2 = \psi_\phi^2 = .0225$			
	BM	ID	ML	DC	BM	ID	ML	DC	BM	ID	ML	DC
$\mu_j =^a$	.001	−.001	.000	−.001	−.001	−.003	.000	−.001	−.001	−.004	−.001	−.002
$\beta_j =^a$	−.000	.000	−.001	−.000	−.000	−.000	−.000	−.000	−.000	−.000	−.000	−.000
$\phi_j =^a$	−.002	−.007	−.002	−.002	−.001	−.006	−.001	−.001	.000	−.004	.000	.001
$\gamma_\mu = 3$	−.000(0%)	−.002(0%)	−.001(0%)	−.002(0%)	−.003(0%)	−.007(0%)	−.004(0%)	−.005(0%)	.000(0%)	−.003(0%)	.000(0%)	−.001(0%)
$\gamma_\beta = 0.35$	−.001(0%)	−.000(0%)	−.001(0%)	−.000(0%)	−.000(0%)	−.000(0%)	−.000(0%)	−.000(0%)	−.000(0%)	−.000(0%)	−.000(0%)	−.000(0%)
$\gamma_\phi = 0.3$	−.002(−1%)	−.007(−2%)	−.002(−1%)	−.003(−1%)	−.000(0%)	−.005(−2%)	−.000(0%)	−.000(0%)	.000(0%)	−.004(−1%)	.000(0%)	.001(0%)
$\psi_\mu^2 = 0.25$	.021(9%)	.023(9%)	.024(9%)	.009(3%)	.023(9%)	.026(10%)	.026(10%)	.012(5%)	.024(9%)	.027(11%)	.028(11%)	.014(5%)
$\psi_\beta^2 =^b$	.002(67%)	.041(1657%)	.002(87%)	.007(295%)	.002(19%)	.039(391%)	.002(18%)	.004(43%)	.002(10%)	.036(160%)	.002(10%)	.001(7%)
$\psi_\phi^2 =^b$	.002(67%)	.041(1638%)	.002(77%)	.006(236%)	.002(18%)	.038(383%)	.002(16%)	.003(29%)	.002(8%)	.034(153%)	.002(7%)	.000(1%)
$\psi_{\mu\beta}^2 =^b$	−.000(−3%)	−.001(−18%)	−.003(−37%)	−.001(−13%)	.002(14%)	−.001(−5%)	−.001(−9%)	.001(4%)	.002(9%)	−.002(−9%)	−.001(−6%)	.000(1%)
$\psi_{\mu\phi}^2 =^b$	−.001(−9%)	−.001(−13%)	−.002(−31%)	−.012(−156%)	.001(9%)	−.002(−13%)	−.002(−14%)	−.010(−64%)	.001(6%)	−.003(−15%)	−.002(−9%)	−.008(−35%)
$\psi_{\phi\beta}^2 =^b$	−.000(−13%)	−.000(−21%)	−.000(−58%)	−.000(−46%)	−.001(−24%)	−.001(−22%)	−.001(−30%)	−.001(−23%)	.000(5%)	−.001(−19%)	−.001(−14%)	−.000(−7%)
$\rho_{\mu\beta} = 0.3$	−.060(−20%)	−.245(−82%)	−.172(−57%)	−.175(−58%)	.002(1%)	−.181(−60%)	−.061(−20%)	−.052(−17%)	.009(3%)	−.139(−46%)	−.034(−11%)	−.005(−2%)
$\rho_{\mu\phi} = 0.3$	−.053(−18%)	−.243(−81%)	−.165(−55%)	−.422(−141%)	−.011(−4%)	−.190(−63%)	−.073(−24%)	−.235(−78%)	.001(0%)	−.148(−49%)	−.042(−14%)	−.118(−39%)
$\rho_{\phi\beta} = 0.3$	−.091(−30%)	−.287(−96%)	−.249(−83%)	−.263(−88%)	−.111(−37%)	−.254(−85%)	−.125(−42%)	−.138(−46%)	.003(1%)	−.206(−69%)	−.051(−17%)	−.021(−7%)

Note. A bias of zero is considered optimal. The estimated bias is calculated for three different true values of variances ψ_ϕ^2 and ψ_β^2. The estimated bias is shown for the benchmark (BM) prior specification, the identity matrix (IM) prior specification, the maximum likelihood (ML) input specification, and the default conjugate (DC) prior specification per parameter. Relative bias to the true value of the parameter is presented in percentages within brackets for the fixed effects $\gamma_{\mu_j}, \gamma_{\phi_j}$, and γ_{β_j}, all elements from the covariance matrix Ψ for the random effects, and the correlations between the random effects $\rho_{\mu\phi}, \rho_{\mu\beta}$, and $\rho_{\phi\beta}$. The simulation values for these parameters vary over individuals and replications. ^{a}The simulation values for these parameters vary over individuals and replications. This is why the relative bias is not presented for some parameters. ^{b}The simulation values for these parameters vary for study I (see top row of table).

Table 3. Part I: Ratio of estimated average posterior standard deviations and calculated standard deviations of the estimated posterior means over 1,000 replications.

	$\psi_\beta^2 = \psi_\phi^2 = .0025$				$\psi_\beta^2 = \psi_\phi^2 = .01$				$\psi_\beta^2 = \psi_\phi^2 = .0225$			
	BM	ID	ML	DC	BM	ID	ML	DC	BM	ID	ML	DC
γ_μ	.983	.992	.986	.964	1.088	1.100	1.091	1.069	1.028	1.042	1.035	1.014
γ_β	1.048	1.770	.820	1.189	1.083	1.625	1.074	1.129	1.031	1.385	1.030	1.022
γ_ϕ	1.087	1.847	1.037	1.196	1.062	1.627	1.057	1.080	1.038	1.404	1.028	1.012
ψ_μ^2	1.100	1.130	1.020	1.046	1.080	1.108	1.000	1.037	1.170	1.197	1.084	1.128
ψ_β^2	1.654	3.768	.720	1.946	1.350	2.879	.882	1.440	1.091	2.030	.877	1.067
ψ_ϕ^2	1.778	4.094	.714	1.934	1.394	3.226	.917	1.384	1.181	2.290	.959	1.095
$\psi_{\mu\beta}$	1.392	2.058	1.261	1.304	1.201	1.758	1.183	1.159	1.146	1.551	1.151	1.095
$\psi_{\mu\phi}$	1.284	1.976	1.137	1.125	1.181	1.788	1.162	1.066	1.152	1.612	1.158	1.050
$\psi_{\phi\beta}$	1.769	3.762	1.556	1.950	1.332	2.912	1.301	1.365	1.133	2.124	1.130	1.075
$\rho_{\mu\beta}$	1.456	2.029	1.553	1.296	1.200	1.764	1.256	1.153	1.124	1.551	1.153	1.070
$\rho_{\mu\phi}$	1.382	1.977	1.510	1.065	1.189	1.801	1.234	1.027	1.105	1.601	1.120	1.000
$\rho_{\phi\beta}$	2.071	3.716	2.411	2.035	1.388	2.889	1.453	1.407	1.143	2.136	1.157	1.080

Note. A ratio of 1 is considered optimal, with a ratio > 1 indicating an overestimation of the posterior standard deviations, and a ratio < 1 indicating an underestimation of the posterior standard deviations. The ratios are calculated for three different true values of variances ψ_ϕ^2 and ψ_β^2. The ratios are shown for the benchmark (BM) prior specification, the identity matrix (IM) prior specification, the maximum likelihood (ML) input specification, and the default conjugate (DC) prior specification, and for the following parameters: the random effects μ_j, ϕ_j, and β_j, the fixed effects γ_{μ_j}, γ_{ϕ_j}, and γ_{β_j}, all elements from the covariance matrix Ψ for the random effects, and the correlations between the random effects $\rho_{\mu\phi}$, $\rho_{\mu\beta}$, and $\rho_{\phi\beta}$.

from the data, it biases the estimates for ψ_ϕ^2 and ψ_β^2. This can be seen most clearly from Figure 3, which shows that the bias increases when the true variance diverges from the DC prior specification: when ψ_ϕ^2 and ψ_β^2 are .0025 or .01 the bias is positive, and when they are .0025 the bias turns negative. Apparently the DC specification was close enough to ψ_ϕ^2 and ψ_β^2 when their true values were .01 and .0225, but not close enough when their true values were .0025.

Part II: The effects of sample size and covariance structure

For Part II of our simulation study, we aim to study the effect of sample size and the sizes of the covariances or correlations on the parameter estimates for each prior specification for Ψ. For this purpose we vary sample sizes between 25, 50, and 75 for both number of individuals and time points, and the correlations between the random parameters are either all set to 0 or all set to .3. The variances for both the crossregression and autoregression coefficients are set to .01 for this study, which is the medium value for the variances in Part I. This results in a $4 \times 3 \times 3 \times 2$ (i.e., prior specification $\times$ time points $\times$ individuals $\times$ correlation matrix specification) simulation design. The results for Part II of the simulation study are presented below.

Effects of sample size

In general, when sample size increased, parameter estimates improved as would be expected: the bias became smaller, the coverage rates became closer to .95, and the

ratios of the posterior standard deviations and standard errors became closer to 1. Figure 4 contains graphs for the coverage rates, bias, and ratios of standard deviations for ψ_ϕ^2, and ψ_β^2 for sample sizes of 25 time point and individuals, 50 time points and individuals, and 75 time points and individuals. The results for different combinations of time points and individuals, such as 25 time points and 50 individuals, were not included in Figure 4 to save space; these results, as well as the results for the other parameters are available in the supplementary materials, and the simulated data are available upon request from the first author.

The estimates for μ_j, ϕ_j, and β_j improved when the number of time points increased, as would be expected for within-subject parameters. For the remaining parameters, estimates improved both when time points and number of individuals increased, as would be expected for between-subject parameters. Increasing the number of individuals seems most advantageous for these parameters. Noteworthy is that for all sample sizes and all prior specifications, including the BM specification, the credible intervals and posterior standard deviations for the correlations were quite large: for the BM specification the posterior standard deviations ranged from approximately .30 for the smallest sample size to .16 for the largest sample size. Although the accuracy of the estimates of the correlations increases as sample size increases, efficient estimates of the correlations clearly will require even larger sample sizes. We discuss the results per prior specification in more detail below.

The ID specification did not perform well regardless of sample size, as can be seen from Figure 4. The coverage rates for ψ_ϕ^2 and ψ_β^j were equal to zero, regardless of
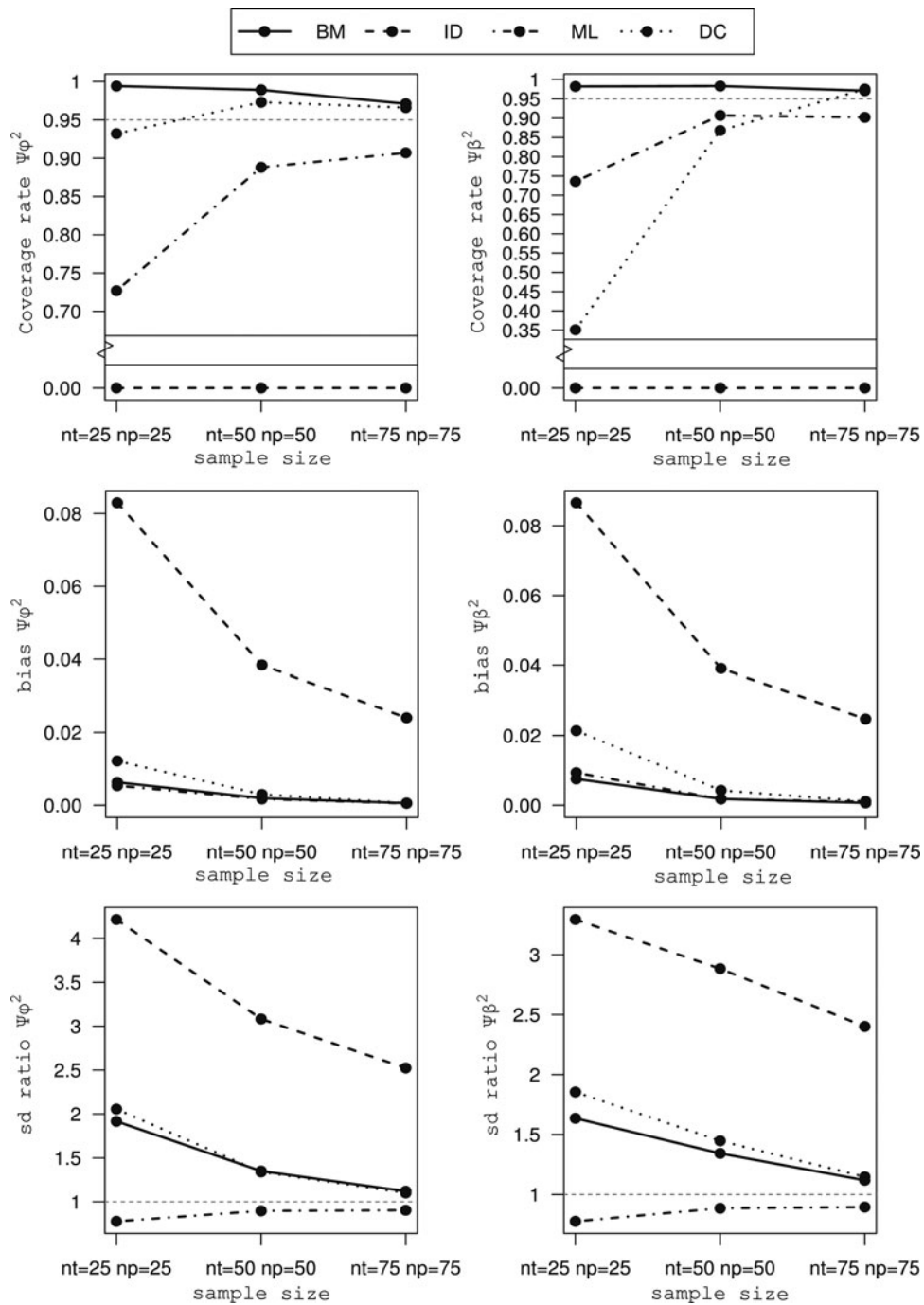
Figure 4. Part II coverage rates, estimated bias, and ratios of the average estimated posterior standard deviations and the standard deviations of the estimated posterior means, for ψ_ϕ^2 and ψ_β^2 calculated over 1,000 replications for the models with correlations set to .3. The coverage rates, bias, and ratios of standard deviations are shown for the benchmark (BM), identity matrix (ID), maximum likelihood (ML) input, and default conjugate (DC) prior specification, for sample sizes of 25 time points (nt) and 25 individuals (np), 50 time points and 50 individuals, and 75 time points and 75 individuals. We choose to present only these combinations of number of time points and persons in order to save space. The results for the the remaining combinations are available as supplementary materials. For each prior specification, performance improves as sample size increases and the data dominates the prior specification more.

number of time points or individuals. Although the bias in the parameter estimates decreased when sample sizes increased, it remained large, which was reflected in the coverage rates. The ratios of standard deviations are larger than 1, and large compared to the other prior specifications, which indicates that the posterior standard deviations are relatively large across sample sizes, resulting in relatively large credible intervals.

For the ML specification, the coverage rates for ϕ_j and β_j were low when sample sizes were small, but they improved as sample size increased, from approximately .88 for 25 time points and participants to .94 for 75 time points and individuals. The coverage rates for ψ_ϕ^2 and ψ_β^2 also improved when sample size increased, from approximately .73 for 25 time points and individuals to .90 for 75 time points and individuals (see Figure 4).

The performance of the DC prior specification also increased when sample size increased. However, Figure 4 shows that for small sample sizes the DC specification shows an especially sharp drop in coverage rates ψ_β^2, indicating that for this parameter the small sample was not enough to dominate the prior. In general, for all prior specifications and across both parts of the simulation study, the estimates for ψ_ϕ^2 seem to be slightly less biased than those for ψ_β^2. In this case the true value for ψ_β^2 seems to lie just outside the credible interval, whereas the true value for ψ_ϕ^2 lies just within the credible interval, resulting in this sharp drop in coverage rates for ψ_β^2, but less so for ψ_ϕ^2. The estimates for the covariance $\psi_{\mu\phi}$ and correlation $\rho_{\mu\phi}$ improve strongly as sample size increases, with coverage rates ranging from approximately .55 to .92 for $\psi_{\mu\phi}$ and from .43 to .93 for $\rho_{\mu\phi}$, and bias ranging from approximately from −.04 to −.003 for $\psi_{\mu\phi}$, and from −.7 to −.08 for $\rho_{\mu\phi}$, for the smallest to the largest sample sizes respectively.

Effects of covariance structure

In general, performance did not differ much when the covariance structure was altered, except for the estimates of the covariances and correlations of the random parameters for the ID and ML prior specification. Note that the performance for the correlations and covariances will not necessarily be the same because the correlations are also affected by the estimates of the variances. However, for both the correlations and covariances estimates were better when the true values of the covariances were set to zero, which is not surprising since these prior specifications had covariances set to zero. When correlations of .3 were used to generate the data, the covariance and correlation estimates were downward biased for these specifications compared to the benchmark specification. This relatively large bias compared to the benchmark specification was absent for the ID and ML specification when the true values of the correlations were equal to zero, and decreased when sample size increased so the data dominated the prior more. For the smallest to the largest sample sizes, for the BM specification this bias ranged from approximately −.06 to .004 for $\rho_{\mu\phi}$ and $\rho_{\mu\phi}$, and from −.18 to −.04 for $\rho_{\mu\phi}$. For the ID specification the bias ranged from approximately −.24 to −.16 for $\rho_{\mu\phi}$ and $\rho_{\mu\phi}$,

and from −.27 to −.23 for $\rho_{\beta\phi}$. The bias for the ML specification was considerably less with the bias for $\rho_{\mu\phi}$ and $\rho_{\mu\phi}$ ranging from −.19 to −.03, and from −.26 to −.04 for $\rho_{\beta\phi}$. For the DC specification, the bias for the correlations was generally in between the bias for the ML and BM specification, except for $\rho_{\mu\phi}$ as described in section 4.2.3 and 4.3.1. For all prior specifications, the coverage rates and ratios of standard deviations were not clearly affected by the different true correlation values. We briefly evaluated the performance of the BM and ML model for correlations equal to .7 rather than .3, with sample sizes of 25 occasions and persons, and 75 occasions and persons. As would be expected, the bias in the correlations for the ML specification was more severe when the correlations were equal to .7. For the rest, results were comparable to the condition for which the correlations were equal to .3.

Conclusion

Overall, the ML prior specification outperformed the other prior specifications. The ID specification, which is probably one of the most common choices in specifying uninformative priors for covariance matrices in practice, is not a good choice when variances may be small, because it results in severely overestimated variances even for relatively large sample sizes. The DC specification performs better than the ID specification, but gives inconsistent results. That is, it strongly influences the results when the DC prior information is not close enough to the information in the data. Given that there is no guarantee that the prior information from the DC will be close to the information in the data, the performance of the DC prior is unreliable when variances are small. The ML specification on the other hand, is directly based on maximum likelihood estimates of the variances from the data, which provide a good guess of the true value of the variances. As a result, the ML specification performs relatively well. The double usage of the data in the ML specification however does have consequences for the standard deviations and credible intervals of the variances: these are too small. However, this effect diminishes when sample size increases.

A disadvantage of the ML specification is that when the models of interest become more complex, it may be difficult to fit these models with traditional ML procedures and software—in fact, this may be one of the reasons to opt for Bayesian estimation in the first place. For instance, multivariate multilevel modeling is often not available in frequentist software whereas it is relatively easy to fit with Bayesian software. Two other examples are multilevel multivariate autoregressive models that include latent variables, and models that include residual variances—both may not be possible with frequentist

software, whereas it is relatively trivial within Bayesian software. There are several ways to get estimates for the covariance matrix of the random parameters, or the variances in this matrix, when the models are too complex to fit with traditional techniques: firstly, it may be possible to use simpler models that allow ML estimation to get preliminary estimates of the variances in question. For instance, if one aims to fit a multivariate multilevel autoregressive model, one option is to fit multiple univariate models with ML techniques in order to get preliminary estimates for the variances, and use these estimates in the prior specification. Secondly, an option is to fit state-space models per individual instead of one multilevel model, and calculate the variances based on the estimates of the individual parameters. Thirdly, one can fit the model with Bayesian estimation using uniform priors for the variances, while disregarding potential covariances, and use these estimates for the variances in the specification of the IW prior distribution. In the following section we will illustrate this last option, together with the ID, DC, and ML prior specification, using an empirical example.

Empirical application on positive affect and worrying

The data for this empirical illustration consist of ESM measurements (see Geschwind et al., 2011). Each participant was alerted randomly throughout each day to fill out the provided questionnaires, for six days, resulting in approximately 45 repeated measures per participant. Here we focus on baseline measures for 129 participants of positive affect (PA), measured with principal component scores for seven items (I feel 'happy', 'satisfied', 'strong', 'enthusiastic', 'curious', 'animated', and 'inspired') (for details, see Geschwind et al., 2011), and on baseline measures of worrying, measured with the item 'I am worrying'. All items were answered on a scale from 1 to 7 (with 1 being 'not at all' and 7 being 'very much so'). Because an assumption for AR(1) models is that time intervals between measurement are about equal, we added observations and coded these observations as missing between measurements, when time intervals between random measurements were especially large (e.g., between the last observation of a day and the first observation of the following day), resulting in an average time lag of about 1.7 hours.

In the psychological literature worrying is considered to be both potentially productive and potentially destructive. That is, worrying is productive when it results in solving a (potential) problem, reducing negative affect that accompanied the problem. In that case, worrying is considered an adaptive emotion-regulating strategy

(Ehring & Watkins, 2008; Nolen-Hoeksema et al., 2008; Pyszcynski & Greenberg, 1987; Watkins, 2008). On the other hand, it may become destructive when the problem cannot be solved, and worrying becomes repetitive or compulsive in continuously trying to solve the problem, exacerbating negative emotions related to the problem. This repetitive worrying has been considered as a maladaptive strategy to regulate emotions, and has been related to affect, especially negative affects such as sadness and anxiety, to rumination, and to various depressive and anxiety disorders (Aldao et al., 2010; Ehring & Watkins, 2008; Nolen-Hoeksema et al., 2008; Querstret & Cropley, 2013; Pyszcynski & Greenberg, 1987; Watkins, 2008). Within the current modeling framework, a positive autoregressive effect for worrying may serve as an indication of such repetitive or compulsive thinking — reflecting that a person tends to 'get stuck' in their worrying across multiple observations. Here, we will explore this autoregressive effect of worrying, and that of PA, and we will explore the reciprocal effects of worrying and PA on each other, by means of fitting a multilevel multivariate autoregressive model. Furthermore, we will investigate whether there are any associations between the individual autoregressive effects, cross-lagged effects, and individual means. Note that this is possible because we are using a model with a multilevel and multivariate structure.

Modeling approach

Applications of multivariate multilevel autoregressive models are sparse (see Lodewyckx et al., 2011, for an exception). Univariate applications are more commonly seen in the psychological literature (e.g., Cohn & Tronick, 1989; Moberly & Watkins, 2008; Nezlek & Allen, 2006; Nezlek & Gable, 2001; De Haan-Rietdijk et al., 2014; Rovine & Walls, 2006; Suls et al., 1998). When researchers are interested in reciprocal lagged effects between two or more variables, they typically estimate several univariate models instead (e.g., Moberly & Watkins, 2008; Nezlek & Allen, 2006; Nezlek & Gable, 2001). The reason for this may be that it is difficult to estimate multivariate multilevel models using traditional software. Here, a Bayesian approach is extremely valuable, because it can be easily extended to multivariate processes. An additional advantage of the Bayesian approach that is especially important for longitudinal designs is that it handles missing data well. We have some missing data for the measures on worrying and PA as a result of nonresponse, as well as the observations we added and coded as missing as noted previously. As such, the Bayesian approach will be quite helpful here.

In order to illustrate the effect of different prior specifications for the covariance matrix Ψ of the random parameters (two means, two cross-lagged parameters, and two autoregressive parameters per person), we fit the model with the ID prior, the DC prior, and the ML prior. Because fitting a multivariate multilevel autoregressive model currently is not an option in ML software, we fit two univariate models with package lme4 in R (Bates et al., 2012; R Development Core Team, 2012) in order to obtain estimates for the variances in Ψ to plug into the ML and DC prior. This may not be ideal, because 1) fitting two univariate models ignores any residual correlation between PA and worrying; and 2) In lme4 any missing observations are discarded from the analysis by means of listwise deletion, so that many observations are disregarded in this analysis: one missing value in the dependent variable, also means a missing value in the predictor at the preceding occasion, resulting in the list-wise deletion of two observations. Therefore, we also specify a second data-based prior, based on first fitting a bivariate Bayesian model with Uniform priors on the variances of the random parameters. Although this model ignores any correlation between the random parameters, it allows for a residual correlation between PA and worrying, and more importantly, it efficiently deals with the missing observations. We plug the estimates for the variances of the random parameters of this model into an Inverse-Wishart prior for a full multivariate model, and we will refer to this Inverse-Wishart prior specification as the Bayesian data based (BDB) prior specification.

In the first six panels of Figure 5 we provide plots of the resulting (marginal) IW prior distributions for the variances of the random parameters. For the variances of the means (two top most panels) the ML, BDB, and DC prior specifications are quite similar, with the exception of the DC prior specification for the mean of worrying which is more similar to the ID prior specification. For the variances of the autoregressive and cross-lagged coefficients there are more dissimilarities between the prior specifications, as we would expect. The ML, DC, and BDB prior distributions all peak in the area close to zero for the variances of these regression coefficients. For these parameters, the prior distributions for the ML and BDB specification are most similar (except for $\beta_{Wo->PAj}$), but they do not overlap completely, especially in the area close to zero. As expected, the ID prior peaks quite far away from zero, and is most dissimilar to the other prior specifications. The final two panels of Figure 5 show plots for two of the fifteen correlations between the random parameters, specifically between $\beta_{Wo->PAj}$ and ϕ_{Woj}, and between ϕ_{Woj} and μ_{Woj}. For the correlation between $\beta_{Wo->PAj}$ and ϕ_{Woj} the prior covariance was set to zero for all prior specifications, resulting in a symmetric,

saddle-shaped distribution. For the correlation between ϕ_{Woj} and μ_{Woj}, the prior covariance was set to zero for all specifications except the DC prior, for which the prior is shifted in favor of a negative correlation.

We fitted each model with three chains, with each 40,000 samples of which 20,000 were burn-in. We evaluated the convergence of each model through the visual inspection of the mixing of the three chains, the Gelman-Rubin convergence diagnostic (Gelman & Rubin, 1992), and autocorrelations. Based on these results we judged 40,000 iterations with 20,000 burn-in iterations as sufficient for convergence. Code for R and WinBUGS for simulating data and fitting the bivariate model, based on the ML prior specification and the BDB prior specification, is provided as supplementary materials.

Results

From Table 4 it can be seen that for most parameters, the estimates are quite similar across the different prior specifications. As would be expected, the largest differences are between the estimated variances of the random autoregressive and cross-lagged parameters (see the random effect for ϕ_{PA} and ϕ_{Wo}, and $\beta_{PA \to Wo}$ and $\beta_{Wo \to PA}$ in Table 4), and therefore, between the estimates of the individual random parameters. For the models with the ML based prior and the BDB prior we find very similar estimates for the variances. For the DC prior, we also find similar results, albeit a somewhat smaller point estimate for the variance of the cross-lagged effect of PA on worrying, compared to the models with ML and BDB priors. In the model with the ID prior specification, the variances are consistently estimated to be about twice as large compared to the estimates for the other prior specifications.

The fixed autoregressive effects are positive, which implies that on average, a participants' current PA is likely to be similar to their PA of the previous occasion, and a participants' current worrying is likely to be similar to their worrying of the previous occasion. Based on these point estimates for fixed effects, and the corresponding estimates of the variances based on BDB prior, we find an approximate 95% interval of .065 to .647 for the random autoregressive parameters of PA, and of -.041 to .591 for the autoregressive parameters of worrying. This indicates that the autoregressive coefficients are expected to be positive for most individuals. The average cross-lagged effect for the effect of worrying on PA is near zero, which implies that on average worrying on the preceding occasion does not affect PA at the current occasion. However, the variation around this average effect implies that for some persons the effect is actually positive, whereas for others it is negative: The point estimate of the fixed effect and of the
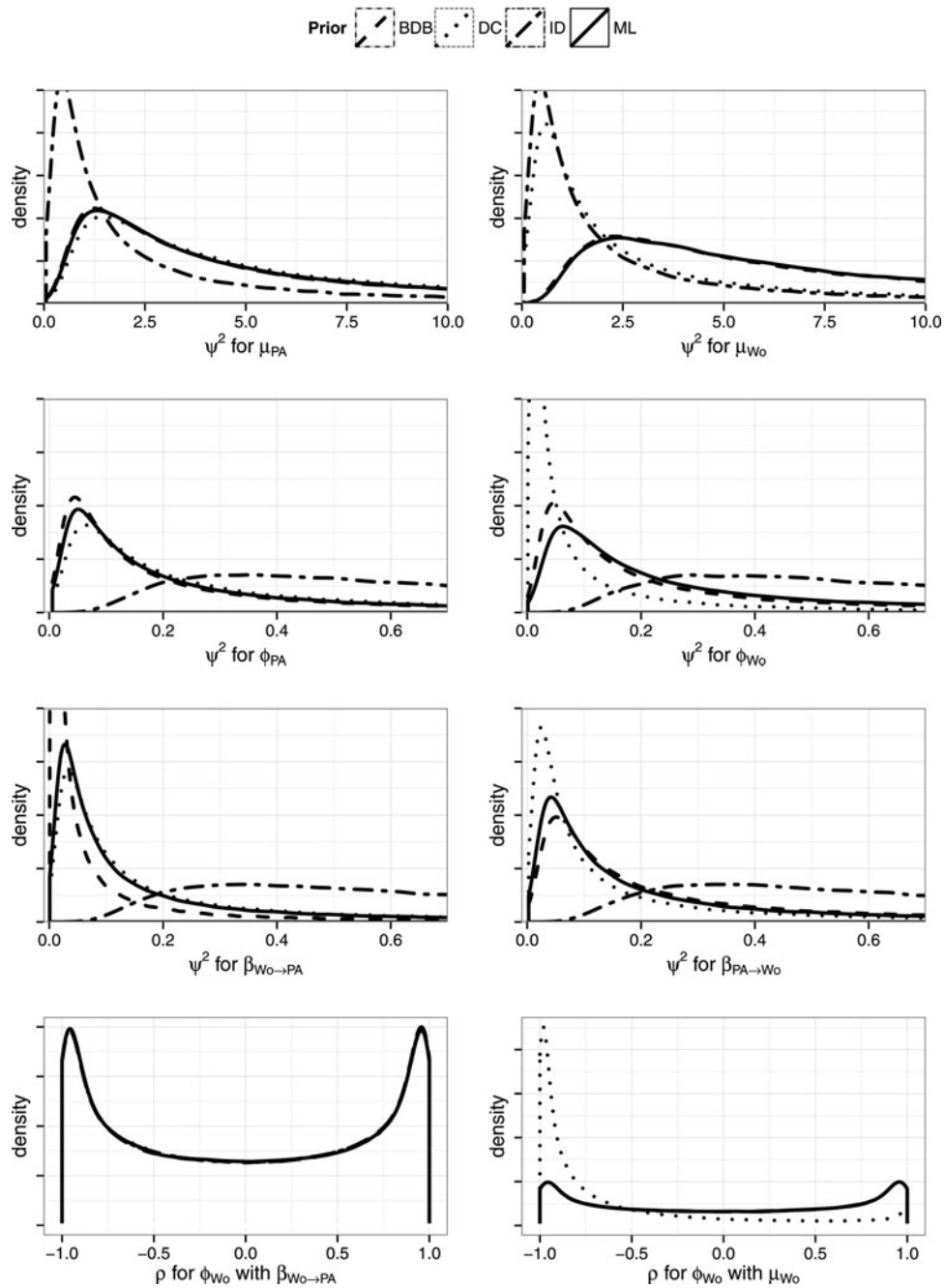
Figure 5. Plots of the (marginal) Inverse-Wishart prior distributions based on the maximum likelihood (ML), Bayesian data-based (BDB), default conjugate (DC), and identity matrix (ID) specification for the variances of the random parameters, and for the correlation between $\beta_{Wo->PAj}$ and ϕ_{Woj}, and $\beta_{Wo->PAj}$ and ϕ_{Woj}, and between ϕ_{Woj} and μ_{Woj}.

corresponding variance imply a 95% interval of $-.172$ to $.132$ for the cross-lagged effects of worrying on PA. This may indicate that for some persons worrying is mostly a productive problem solving behavior, with successful problem solving leading to more positive affect, whereas for others worrying is ineffective, leading to less positive affect. The average cross-lagged effect of PA on worrying is negative, which implies that on average (across persons), higher PA on the preceding occasion is likely to lead

to less worrying at the current occasion, whereas diminished PA is likely to lead to more worrying. Based on the estimated fixed effect and corresponding variance, we find a 95% interval of $-.485$ to $.171$ for the random cross-lagged effects of PA on worrying, indicating that this effect is expected to be negative for most persons. This seems a logical result if worrying is problem-oriented: When there are problems to be solved, this may lead to lower PA, and to worrying in order to solve the problem, and vice versa.

Table 4. Parameter estimates for the multilevel bivariate autoregressive model on positive affect and worrying (posterior means and 95% CI), for four different prior specifications for the covariance matrix of the random parameters.

Parameter	ID	DC	ML	BDB
Fixed Effects for:				
μ_{PA}	3.691 (3.551, 3.831)	3.687 (3.547, 3.828)	3.687 (3.545, 3.83)	3.689 (3.548, 3.83)
μ_{Wo}	2.937 (2.744, 3.13)	2.936 (2.741, 3.131)	2.940 (2.742, 3.137)	2.939 (2.743, 3.135)
ϕ_{PA}	.343 (.294, .390)	.354 (.312, .394)	.354 (.313, .395)	.356 (.316, .396)
ϕ_{Wo}	.267 (.217, .316)	.280 (.237, .323)	.277 (.235, .318)	.275 (.230, .318)
$\beta_{Wo \to PA}$	-.023 (-.058, .011)	-.020 (-.048, .007)	-.021 (-.047, .007)	-.020 (-.045, .006)
$\beta_{PA \to Wo}$	-.160 (-.223, -.096)	-.150 (-.205, -.094)	-.156 (-.213, -.100)	-.157 (-.213, -.102)
Random Effects for:				
μ_{PA}	.590 (.456, .774)	.618 (.481, .810)	.612 (.474, .801)	.611 (.475, .800)
μ_{Wo}	1.126 (.869, 1.486)	1.132 (.874, 1.491)	1.187 (.918, 1.559)	1.186 (.915, 1.555)
ϕ_{PA}	.042 (.030, .060)	.024 (.015, .038)	023 (.015, .037)	.022 (.013, .035)
ϕ_{Wo}	.045 (.032, .064)	.024 (.013, .039)	.028 (.017, .044)	.026 (.015, .041)
$\beta_{Wo \to PA}$	.025 (.018, .034)	.009 (.006, .015)	.009 (.005, .014)	.006 (.003, .011)
$\beta_{PA \to Wo}$	.059 (.039, .090)	.017 (.007, .038)	.025 (.011, .049)	.028 (.013, .054)
Residuals:				
σ^2_{PA}	.815 (.785, .845)	.822 (.793, .853)	.822 (.792, .852)	.823 (.794, .854)
σ^2_{Wo}	1.918 (1.849, 1.990)	1.943 (1.873, 2.017)	1.935 (1.865, 2.008)	1.934 (1.865, 2.008)
ρ_{PAWo}	-.479 (-.498, -.458)	-.478 (-.498, -.458)	-.479 (-.499, -.459)	-.479 (-.499, -.459)

Note. Included are the average effects (fixed effects) and the variances (random effects) for the means of PA and worrying (μ_{PA}, μ_{Wo}), autoregressive effects of PA and worrying (ϕ_{PA}, ϕ_{Wo}), the cross–lagged effect of worrying on PA ($\beta_{Wo \to PA}$) and of PA on worrying ($\beta_{PA \to Wo}$). Further included are the residual variances (σ^2_{PA}, σ^2_{Wo}), and the residual correlation (ρ_{PAWo}).

For the correlations between the random parameters (not reported in Table 4 for reasons of space) we find that most correlations have quite wide credible intervals, with values ranging from strongly negative to strongly positive, so that we have too little information to draw conclusions about these correlations (similar to our findings in the simulation study). However, for three correlations we find credible intervals that include only negative values across the DC, ML, and DBD prior specifications, namely between the means for PA μ_{PAj} and worrying μ_{Woj} ($-.293$, 95% CI:$[-.453, -.115]$), between the mean of worrying μ_{Woj} and the cross-lagged effect of worrying on PA $\beta_{Wo \to PAj}$ ($-.360$, 95% CI:$[-.615, -.047]$), and between the autoregressive parameter for worrying ϕ_{Woj} and the cross-lagged effect of worrying on PA $\beta_{Wo \to PAj}$ ($-.551$, 95% CI:$[-.771, -.197]$; here we report the results based on the BDB prior, though results are similar across the other specifications). To gain more insight in the meaning of these correlations we made scatter plots of the individual parameters (see Figure 6), and we discuss each correlation in more detail below. First, the negative correlation between the mean of PA and of worrying (left panel of Figure 6), indicates that persons with higher average PA are likely to worry less on average compared to persons who generally have lower average PA. Second, the correlation between the cross-lagged effect of worrying on PA with the mean of worrying (middle panel of Figure 6) implies that individuals who worry a lot on average tend to have a negative cross-lagged effect of worrying on PA at the next occasion, whereas individuals who do not worry a lot on average tend to have a positive cross-lagged effect of worrying on PA. This may reflect

the dual nature of worrying. For persons for whom worrying is effective in solving problems, worrying results in a higher positive affect because problems are being dealt with (i.e., a positive cross-lagged effect), and therefore may not need to worry as much (i.e., a low mean of worrying). In contrast, for persons for whom worrying is not effective, worrying may result in a lower PA (i.e., a negative cross-lagged effect) without the relief and accomplishment of solving the problem, and they may worry relatively a lot on average (i.e., a high mean for worrying), because their problems are not going away. Third, the correlation between the cross-lagged effect of worrying on PA with the autoregressive effect of worrying (right panel of Figure 6) implies that persons who have high inertia in worrying (i.e., get stuck in worrying), tend to have a negative cross-lagged effect of worrying on PA, whereas persons that have little or no inertia in their worrying tend to have a positive cross-lagged effect of worrying on PA. This correlation also seems to illustrate the potential problem solving nature of worrying: When worrying results in solving the problem, worrying may result in a higher PA (i.e., a positive cross-lagged effect), and for persons for whom this is the case there may be little need to keep worrying (i.e., a relatively low inertia in worrying). In contrast, when worrying is ineffective, the futile worrying may result in a lower PA (i.e., a negative cross-lagged effect), and the persons for whom this is the case may continuously worry in order to keep trying to solve the problem, resulting in a relatively high positive inertia in worrying.

Finally, we note that there remains a strong negative association ($-.479$, 95% CI:$[-0.499, -0.459]$) between
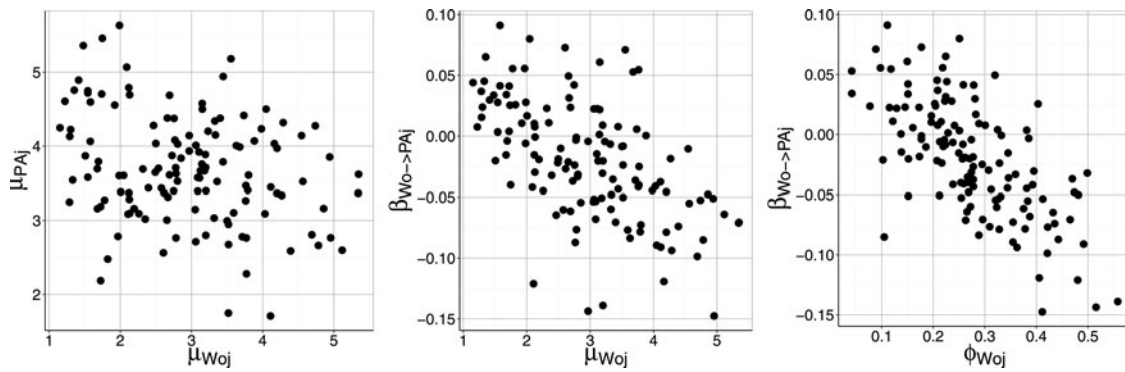
Figure 6. Scatter-plots of the point estimates of the random parameters based on the Bayesian data-based (BDB) prior specification, indicating negative correlations between the means for PA and worrying (μ_{PA_j} and μ_{Wo_j}), the cross-lagged effects of worrying on PA and the means of worrying ($\beta_{Wo->PA_j}$ and μ_{Wo_j}), and cross-lagged effects of worrying on PA and the autoregressive effects for worrying ($\beta_{Wo->PA_j}$ and ϕ_{Wo_j}).

the residuals of PA and worrying. This residual correlation between PA and worrying after the lagged effects are taken into account, indicates that there is more to the relationship between PA and worrying. As such, it may be worthwhile to look at the relationship between PA and worrying at other time intervals than the interval of about 1.7 hours that was considered here, or to look for additional explanatory variables, for instance specific negative events, social interactions, stress, or psycho-physiological factors.

In sum, these results provide interesting considerations for future (confirmatory) research on the relationship between worrying and PA, and individual differences in this relationship. Based on the correlations between the random effects, we found that individuals who worry a lot on average, as well as individuals who get stuck in worrying, tend to have a negative cross-lagged effect of worrying on PA, indicating that for them worrying is a maladaptive coping strategy. In contrast, individuals who do not worry a lot on average, or who bounce back from worrying quickly, tend to have a positive cross-lagged effect from worrying to PA, indicating that for them worrying is a adequate tactic to solve current problems. Note that we were only able to find these results because we made use of a multivariate multilevel model, allowing for all random effects to be correlated. That is, had we used two separate multilevel models (i.e., for PA and worrying as dependent variables separately), we would not have obtained estimates of the correlations between these random effects. This illustrates the unique opportunities offered by the multivariate approach. Furthermore, fitting several (data-based) priors helps evaluate the influence of specifying certain priors: The results for the ID specification considerably diverged from the results from the other prior specifications. The remaining three prior specifications however converged to approximately the same results, so that we feel that we can be reasonably

confident about our results and conclusions based on these specifications.

Discussion

The multivariate multilevel autoregressive model is a valuable model for studying between-person differences in within-person processes. The Bayesian modeling framework provides a flexible environment for fitting this complex multilevel model. However, when some variances of the random parameters in the model are close to zero, the conjugate IW prior distribution for the covariance matrix of the random parameters becomes quite informative, unintentionally influencing the parameter estimates. In this study we evaluated the performance of three different IW prior specifications for the covariance matrix of the random parameters by means of a simulation study. In addition, by means of an empirical data set we demonstrated a sensitivity analysis for the IW prior specification, and illustrated the added value of the multivariate, multilevel modeling approach provided by the flexible Bayesian modeling environment.

The results from the simulation study indicate that the data-based ML prior specification for the covariance matrix of the random parameters performs the best compared to the ID specification and the DC specification. The ML specification performs well because it is based on estimates of the variances from the data. There are multiple ways to obtain estimates of the variances based on the data besides the ML procedure, that we discussed in the conclusion section of the simulation study. The consequence of using the data twice is that the certainty about the parameter estimates is overestimated, resulting in too small posterior standard deviations. A solution to this problem may be to use a small part of the data for calibrating the prior specification (also referred to as training data, see Berger & Pericchi, 1996; O'Hagan, 1995), and

using the remainder of the data for the model fitting procedure. Of course, this raises questions for future research on exactly how to do this. To cite two examples: should you use part of the persons in your sample for calibrating the prior, or part of the repeated measures of each person? What sample size would provide good enough estimates for calibrating the prior?

An alternative specification that we considered, but was inoperative for the multilevel autoregressive model in WinBUGS (see footnote 3), is the scaled Wishart discussed in Gelman and Hill (2007). However, this may be a viable specification for other multilevel models. Other alternative IW prior specifications for Ψ that we did not investigate consist of specifying improper IW priors with df smaller than r, or to use a specification suggested in a recent study by Huang and Wand (2013), in which Half-Cauchy distributions for the standard deviations, and Uniform distributions for the correlations, are specified via an IW distribution. However, both WinBUGS and OpenBUGS require proper IW priors, and do not allow for setting priors within IW priors, so that these specifications are not available within this software. Still another option may be to transform the random parameters so that they have a larger variance, and specifying an IW prior for the covariance matrix of the transformed parameters. Finally, two potential specifications that circumvent the use of an IW distribution, are to specify the variances and covariances in a regression structure, or to specify uniform distributions for the variances of the random parameters, if disregarding the covariances does not affect the parameter estimates of interest. If the covariances between the random covariances are of primary interest, a possibility for the latter specification may be to correlate the random parameters a posteriori. Possible directions for future research are to compare the performance of the alternative specifications with the ML specification in a simulation study (in other software).

In conclusion, this study demonstrates that the IW prior specification for covariance matrices should not be taken lightly. When variances are small, the prior specification can have considerable consequences for the parameter estimates. In the multilevel autoregressive model, it is known in advance that some variances will be close to zero. We expect that our results will generalize to any multilevel model that has small variances in the covariance matrix of the random parameters, either as a result of the scale of the variables or parameters, or simply because there are only small individual differences in the parameters. Therefore, it seems imperative to include a prior specification sensitivity analysis for the covariance matrix of the random parameters in multilevel studies in psychology. Our empirical application provides an example of such an analysis, in which we compared the

results for four different prior specifications: three different (data-based) priors converged to approximately the same results, whereas the ID specification showed divergent results. Finally, we advise to include a data-based prior in such a prior sensitivity analysis. Although it may not be ideal to use the data twice in order to calibrate the prior, our simulation study results indicate that a prior distribution based on estimates of the variances of the random parameters performs the best in this specific situation that some variances in the covariance matrix may be close to zero.

Article information

References

Adolf, J., Schuurman, N.K., Borkenau, P., Borsboom, D., & Dolan, C.V. (2014). Measurement invariance within and

between subjects: A distinct problem in testing the equivalence of intra- and inter-individual model structures. *Frontiers in Psychology*, 5. doi: doi:10.3389/fpsyg.2014.00883

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, 30(2), 217–237. doi: doi:10.1016/j.cpr.2009.11.004

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=lme4 (R package version 0.999999-0)

Berger, J.O., & Pericchi, L.R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the Royal Statistical Society*, 91, 109–122. doi: doi:10.1080/01621459.1996.10476668

Browne, W.J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1, 473–514. doi: doi:10.1214/06-ba117

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Cohn, J.F., & Tronick, E. (1989). Specificity of Infants' Response to Mothers' Affective Behavior. *Adolescent Psychiatry*, 28, 242–248. doi: doi:10.1097/00004583-198903000-00016

De Haan-Rietdijk, S., Gottman, J.M., Bergeman, C.S., & Hamaker, E.L. (2014). Get over it! A multilevel threshold autoregressive model for state-dependent affect regulation. *Psychometrika*, 1–25. doi: doi:10.1007/s11336-014-9417-x

Ehring, T., & Watkins, E.R. (2008). Repetitive negative thinking as a transdiagnostic process. *International Journal of Cognitive Therapy*, 1(3), 192–205. doi: doi:10.1680/ijct.2008.1.3.192

Fahrmeir, L., Tutz, G., & Hennevogl, W. (2010). *Multivariate statistical modelling based on generalized linear models*. New York, NY: Springer-Verlag.

Gelman, A. (2006). Comment on Browne and Draper, Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 514–534. doi: doi:10.1214/06-ba117a

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

Gelman, A., & Hill, L.S. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–511. doi:10.1214/ss/1177011136

Geschwind, N., Peeters, F., Drukker, M., van Os, J., & Wichers, M. (2011). Mindfulness training increases momentary positive emotions and reward experience in adults vulnerable to depression: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 79, 618–628. doi: doi:10.1037/a0024595

Gill, J. (2014). *Bayesian methods: A social and behavioral sciences approach (3d ed.)*. Boca Raton, FL: Chapman & Hall/CRC.

Hamaker, E.L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. Mehl & T. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). New York, NY: Guilford. doi: doi:10.5860/choice.50-1159

Hamilton, J.D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Huang, A., & Wand, M.P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8, 439–452. doi: doi:10.1214/13-ba815

Kass, R.E., & Natarajan, R. (2006). Comment on Browne and Draper, A default conjugate prior for variance components in generalized linear mixed models. *Bayesian Analysis*, 1, 535–542. doi: doi:10.1214/06-ba117b

Kass, R.E., & Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association*, 84, 717–726. doi: doi:10.2307/2289653

Kim, C.-J., & Nelson, C.R. (1999). *State–space models with regime switching*. Cambridge, MA: MIT Press.

Lindley, D.V. (1969). Commentary on J. B. Copas "Compound Decisions and Empirical Bayes". *Journal of the Royal Statistical Society, Series B*, 31, 397–425.

Lodewyckx, T., Tuerlinckx, F., Kuppens, P., Allen, N.B., & Sheeber, L. (2011). A hierarchical state space approach to affective dynamics. *Journal of Mathematical Psychology*, 55, 68–83. doi: doi:10.1016/j.jmp.2010.08.004

Lunn, D.J., Spiegelhalter, D.J., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique, and future directions. *Statistics in Medicine*, 28, 3049–3067. doi: doi:10.1002/sim.3680

Lunn, D.J., Thomas, A., Best, N., & Spiegelhalter, D.J. (2000). WinBUGS – a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10, 325–337. doi: doi:10.1023/A:1008929526011

Moberly, N.J., & Watkins, E.R. (2008). Ruminative self–focus and negative affect: An experience sampling study. *Journal of Abnormal Psychology*, 117, 314–323. doi: doi:10.1037/0021-843x.117.2.314

Molenaar, P. C.M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives*, 2, 201–218. doi: doi:10.1207/s15366359mea0204_1

Nezlek, J.B., & Allen, M.R. (2006). Social support as a moderator of day-to-day relationships between daily negative events and daily psychological well-being. *European Journal of Personality*, 20, 53–68. doi: doi:10.1002/per.566

Nezlek, J.B., & Gable, S.L. (2001). Depression as a moderator of relationships between positive daily events and day-to-day psychological adjustment. *Personality and Social Psychology Bulletin*, 27, 1692–1704. doi: doi:10.1177/01461672012712012

Nolen-Hoeksema, S., Wisco, B., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, 3(5), 400–424. doi: doi:10.1111/j.1745-6924.2008.00088.x

O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society, Series B*, 57, 99–138.

Pyszcynski, T., & Greenberg, J. (1987). Self-regulatory perseveration and the depressive self-focusing style: A self-awareness theory of reactive depression. *Psychological Bulletin*, 102(1), 122–138. doi: doi:10.1007/978-1-4613-9191-3_6

Querstret, D., & Cropley, M. (2013). Assessing treatments used to reduce rumination and/or worry: A systematic review. *Clinical Psychology Review*, 33(8), 996–1009. doi: doi:10.1016/j.cpr.2013.08.004

(2012). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/ (ISBN 3-900051-07-0)

Rovine, M.J., & Walls, T.A. (2006). A multilevel autoregressive model to describe interindividual differences in the stability of a process. In J. Schafer & T. Walls (Eds.), *Models for intensive longitudinal data* (pp. 124–147). New York, NY: Oxford.

Song, H., & Ferrer, E. (2012). Bayesian estimation of random coefficient dynamic factor models. *Multivariate Behavioral Research*, *47*, 26–60. doi: doi:10.1080/00273171.2012.640593

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, *12*, 1–16.

Suls, J., Green, P., & Hillis, S. (1998). Emotional reactivity to everyday problems, affective inertia, and neuroticism. *Personality and Social Psychology Bulletin*, *24*, 127–136. doi: doi:10.1177/0146167298242002

Wang, L., Hamaker, E.L., & Bergeman, C.S. (2012). Investigating inter-individual differences in short-term intra-individual variability. *Psychological Methods*, *17*, 567–581. doi: doi:10.1037/a0029317

Watkins, E.R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin*, *134*(2), 163–206. doi: doi:10.1037/0033-2909.134.2.163

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. doi: doi:10.1037/0022-3514.54.6.1063