A BAYESIAN APPROACH TO NONLINEAR LATENT VARIABLE MODELS USING THE GIBBS SAMPLER AND THE METROPOLIS-HASTINGS ALGORITHM

GERHARD ARMINGER

BERGISCHE UNIVERSITÄT WUPPERTAL
DEPARTMENT OF ECONOMICS

BENGT O. MUTHÉN

UNIVERSITY OF CALIFORNIA, LOS ANGELES
GRADUATE SCHOOL OF EDUCATION & INFORMATION STUDIES

Nonlinear latent variable models are specified that include quadratic forms and interactions of latent regressor variables as special cases. To estimate the parameters, the models are put in a Bayesian framework with conjugate priors for the parameters. The posterior distributions of the parameters and the latent variables are estimated using Markov chain Monte Carlo methods such as the Gibbs sampler and the Metropolis-Hastings algorithm. The proposed estimation methods are illustrated by two simulation studies and by the estimation of a non-linear model for the dependence of performance on task complexity and goal specificity using empirical data.

Key words: Gibbs Sampler, LISREL model, Metropolis-Hastings algorithm, Non-linear functions of latent regressors.

1. Introduction

Research scientists in psychometrics, education and marketing often would like to consider models which contain quadratic and/or interaction terms in the latent variables. A typical problem is the estimation of the simple quadratic model

$$y = \gamma_0 + \gamma_1 \xi + \gamma_2 \xi^2 + \zeta, \tag{1}$$

where $\zeta \sim \mathcal{N}(0, \psi)$ is a disturbance term which is uncorrelated with the unobserved random variable $\xi \sim \mathcal{N}(0, 1)$. ξ is a latent variable. γ_0, γ_1 and γ_2 are regression coefficients and y is a random variable that is conditionally normal given ξ with $y \sim \mathcal{N}(\gamma_0 + \gamma_1 \xi + \gamma_2 \xi^2, \psi)$. However, y is unconditionally not normal, but follows a distribution which is a convolution of a standard normal and a χ_1^2 variable.

The measurement model for ξ is given as a factor analytic model with

$$\mathbf{x} = \boldsymbol{\nu} + \boldsymbol{\Lambda}\xi + \boldsymbol{\delta}, \tag{2}$$

where $\mathbf{x} \sim r \times 1$ is a vector of observed variables, $\boldsymbol{\delta} \sim r \times 1$ is a vector of measurement errors that are uncorrelated with ξ. $\boldsymbol{\nu}$ is a $r \times 1$ vector of regression constants and $\boldsymbol{\Lambda}$ is a $r \times 1$ vector of factor loadings. As in classical factor analysis (see for instance Lawley & Maxwell, 1971) it is assumed that the errors are normal and uncorrelated, that is $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta})$ with $\boldsymbol{\Theta} = \text{diag}\{\vartheta_{11}, \ldots, \vartheta_{rr}\}$.

The model (1) and (2) may be extended to a random vector $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_{\tilde{m}})'$ allowing to deal with quadratic or higher order polynomial terms, ξ_1^2, ξ_1^3 and/or interaction of the form $\xi_1\xi_2$, $\xi_1\xi_3$ and so forth. It is assumed that the $\tilde{m} \times 1$ vector $\boldsymbol{\xi}$ follows a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Phi})$.

Until now, estimation of the parameters $\boldsymbol{\gamma}$, $\boldsymbol{\Phi}$, ψ, $\boldsymbol{\nu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Theta}$ has been carried out (Hayduk, 1987; Kenny & Judd, 1984) in the context of LISREL models by writing the observed variables in the model of (1) and (2) as $x_j = \nu_j + \lambda_j\xi + \delta_j, j = 1, \ldots, r$ and then including the variables $x_j^2 = (\nu_j + \lambda_j\xi_j + \delta_j)^2$ as additional indicators for ξ^2 in the measurement model. Then, a LISREL model with ξ and ξ^2 as latent regressor variables is estimated. This solution is often not optimal in the following ways:

- Under the assumption that ζ, ξ and $\delta_j, j = 1, \ldots, r$ are normally distributed, the joint distribution of $(y, x_1, \ldots, x_p, x_1^2 \ldots, x_p^2)$ is not multivariate normal. ML estimation based on the normal distribution of $(y, x_1, \ldots, x_p, x_1^2 \ldots, x_p^2)$ therefore yields consistent estimators for the parameters $(\boldsymbol{\gamma}, \boldsymbol{\Phi}, \psi, \boldsymbol{\nu}, \boldsymbol{\Lambda}, \boldsymbol{\Theta})$ collected in a vector $\boldsymbol{\vartheta}$ but not for the asymptotic covariance matrix of the ML estimator $\boldsymbol{\vartheta}_{\text{ML}}$. To our knowledge, the natural remedy of Pseudo ML estimation (Arminger & Schoenberg, 1989) has not been applied to this model to deal with non-normality. However, weighted least squares (WLS) estimation (Browne, 1984) has recently been applied to this model by Jöreskog and Yang (1996). These authors therefore obtain consistent estimates of the chi-square statistic for model fit and the standard errors of the estimated parameters.

- Computing parameter estimates with the LISREL program (Jöreskog & Sörbom 1993; Jöreskog & Yang 1996) or equivalent programs such as LISCOMP (Muthén, 1988) and MECOSA 3 (Arminger, Wittenberg & Schepers, 1996) is rather cumbersome. Although one of the main obstacles, namely the inclusion of all possible cross products of observed variables as indicators for a latent interaction model, has been removed by Jöreskog and Yang by showing that the inclusion of one product variable is sufficient for identification, it is still rather tedious to write the nonlinear equations for the covariances of cross products as a function of the model parameters. Good examples are found again in Jöreskog and Yang.

- The ML and WLS estimation techniques draw on large sample theory and may not be suitable for small or medium sample sizes. An additional problem is here the number of variables. If the number of observed variables is already large, this number is further increased by the addition of squares and cross products of observed variables.

- The nonlinear models using squares and interactions of latent variables may not be general enough to specify more complicated substantive models. Therefore it should be possible to specify general nonlinear models as in nonlinear regression analysis.

We propose to deal with the problems of nonnormality and a general specification of nonlinear latent variable models as well as with the problem of small sample sizes by using a Bayesian approach to formulate the model and Markov Chain Monte Carlo (MCMC) methods such as the Gibbs sampler (Gelfand and Smith 1990, Tanner 1993) and the Metropolis-Hastings (M-H) algorithm (Chib & Greenberg, 1995; Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953; Tanner, 1993) to estimate the posterior distributions of the parameters given the data.

2. The Nonlinear Latent Variable Model

We introduce a nonlinear version of the LISREL model (Jöreskog & Sörbom, 1993). Let ξ_i be a $\bar{m} \times 1$ vector of random variables that is multivariate normal with $\xi_i \sim \mathcal{N}(0, \Phi)$. Let $\alpha_i = g(\xi_i)$ a $m \times 1$ deterministic function of ξ_i that is known. A $p \times 1$ random variable vector η_i is connected with α_i through a linear regression model

$$\eta_i = \gamma_0 + \Gamma\alpha_i + \zeta_i, \tag{3}$$

where the $p \times 1$ random variable $\zeta_i \sim \mathcal{N}(0, \Psi)$ is a disturbance, $\gamma_0 \sim p \times 1$ is a vector of regression constants and $\Gamma \sim p \times m$ is a matrix of regression coefficients of η_i on α_i. It should be noted that the model is linear in the parameters, but is nonlinear in the components of ξ_i.

The random variables ξ_i and η_i are connected to observed variables $x_i \sim r \times 1$ and $y_i \sim s \times 1$ with the usual factor analytic measurement models

$$x_i = \nu_x + \Lambda_x\xi_i + \delta_i, \quad \delta_i \sim \mathcal{N}(0, \Theta_\delta), \tag{4}$$

$$y_i = \nu_y + \Lambda_y\eta_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \Theta_\varepsilon). \tag{5}$$

For simplicity, it is assumed that Θ_δ and Θ_ε are diagonal matrices according to the classical factor analytic tradition (Lawley & Maxwell, 1971). The nonlinear latent variable model defined in (3) through (5) is abbreviated as the A-M model.

This model includes some important special cases. The first special case is a polynomial regression model in the latent variable ξ_i for a univariate dependent variable y_i

$$\alpha_i = (\xi_i, \xi_i^2, \ldots, \xi_i^p)' \tag{6}$$

$$\eta_i = \gamma_0 + \Gamma\alpha_i + \zeta_i, \quad \zeta_i \sim \mathcal{N}(0, \psi), \tag{7}$$

$$x_i = \nu_x + \Lambda_x\xi_i + \delta_i, \quad \delta_i \sim \mathcal{N}(0, \Theta_\delta), \tag{8}$$

$$y_i = \eta_i. \tag{9}$$

The usual linear model $\eta_i = \gamma_0 + \gamma_1\xi_i + \zeta_i$ is replaced by a polynomial model $\eta_i = \gamma_0 + \gamma_1\xi_i + \ldots \gamma_p\xi_i^p + \zeta_i$. This model allows formulation of a regression model that is nonlinear in the variables and has errors in the variables. It may also be considered as a partially linear and partially nonlinear factor analytic model where y_i is an indicator variable that is nonlinearly dependent on the factor score ξ_i.

The second case is a model with a generalized second order polynomial of the vector ξ_i and a univariate dependent variable y_i

$$\alpha_i = (\xi_{i1}, \ldots, \xi_{i\bar{m}}, \xi_{i1}^2, \ldots, \xi_{i\bar{m}}^2, \xi_{i1}\xi_{i2}, \ldots, \xi_{i,\bar{m}-1}\xi_{i\bar{m}})', \tag{10}$$

$$\eta_i = \gamma_0 + \Gamma\alpha_i + \zeta_i, \quad \zeta_i \sim \mathcal{N}(0, \psi), \tag{11}$$

$$x_i = \nu_x + \Lambda_x\xi_i + \delta_i, \quad \delta_i \sim \mathcal{N}(0, \Theta_\delta), \tag{12}$$

$$y_i = \eta_i. \tag{13}$$

This model allows for first order interactions of latent variables $\xi_{ij}, j = 1, \ldots, \bar{m}$ in the regression model for y_i. Of course, higher order interactions may be formulated in the same way. Again, this model may be also thought of as a nonlinear errors in the variables model and as a partially linear and partially nonlinear factor analytic model.

Finally we consider α_i as a p-th order generalized polynomial and formulate a polynomial factor analytic model:

$$\alpha_i = \mathrm{P}^\mathrm{P}(\xi_{i1}, \ldots, \xi_{i\bar{m}}) = (\xi_{i1}, \ldots, \xi_{i\bar{m}}, \xi_{i1}^2, \ldots, (\xi_{i(\bar{m}-p+1)} \ldots, \xi_{i\bar{m}}))', \tag{14}$$

$$\eta_i = \alpha_i, \tag{15}$$

$$\mathbf{y}_i = \nu_y + \Lambda_y \eta_i + \varepsilon_i. \tag{16}$$

It is important to note that the A-M model may be overparameterized for certain applications. For instance, if the measurement model for $\mathbf{y}_i = \nu_y + \Lambda_y \eta_i + \varepsilon_i$ contains regression constants that are estimated from the data, γ_0 may not be identified as seen from the expected value of the reduced form of $\mathbf{y}_i$ given ξ_i:

$$\mathrm{E}(\mathbf{y}_i|\xi_i) = \nu_y + \Lambda_y \gamma_0 + \Lambda_y \Gamma g(\xi_i). \tag{17}$$

In this case, γ_0 must be set to a fixed value such as $\gamma_0 = 0$. Additional identification restrictions may be necessary for specific models.

As one of the anonymous reviewers has pointed out, the choice of $g(\xi)$ is not completely arbitrary. Restrictions have to be imposed to ensure a proper posterior for η and ξ. Examples of a bad choice of $g(\xi)$ are $g(\xi) = 0$ which eliminates the data x or to choose $\alpha = g(\xi)$ so that the components of α are linearly dependent. In this case, the posterior distribution would collapse in dimensionality. The issue of proper posteriors is discussed in subsection 3.3.

3. The Bayesian Framework

The usual way to estimate the parameters in the vector $\vartheta = \mathrm{vec}\ \{\gamma_0, \Gamma, \Phi, \Psi, \nu_x, \Lambda_x, \Theta_\delta, \nu_y, \Lambda_y, \Theta_\varepsilon\}$ has been a frequentist approach. Given the normal distribution of ξ_i, ζ_i, δ_i and ε_i, the parameter vector ϑ has been estimated with ML method. This approach works nicely if the function g in $\alpha_i = g(\xi_i)$ is the identity function. However, if g is a nonlinear function, such as a polynomial, the unconditional distribution of η_i, and therefore of $\mathbf{y}_i$, is not normal. To avoid these complications we switch from the frequentist to the Bayesian viewpoint and consider only the data $\{\mathbf{y}_i, \mathbf{x}_i\}$, $i = 1, \ldots, n$ as fixed and—in addition to the random variables ξ_i, ζ_i, δ_i and ε_i—the parameter vector ϑ as a random variable. A Bayesian framework has been used before in the context of linear factor analysis by a number of authors, for instance by Lee (1991) who considered the posterior distribution of the matrix of factor loadings, the covariance matrix of factors and the covariance matrix of errors and Press and Shigemasu (1989) who considered additionally the posterior distribution of factor scores.

Our aim is to estimate the joint and marginal distribution of ξ, η, and ϑ given x and y. We are interested in the posterior density of ξ, η, and ϑ given the data x, y. For conditional density functions we use the formulation of Gelfand and Smith (1990), where $[\mathbf{y}|\mathbf{x}]$ denotes the conditional density of a random vector Y at the value of y conditional on a fixed value x. The joint posterior is therefore denoted by $[\xi, \eta, \vartheta|\mathbf{x}, \mathbf{y}]$.

3.1 Prior Distributions

As noted before, the random variables ξ_i, ζ_i, δ_i and ε_i are multivariate normal with $\xi_i \sim \mathcal{N}(0, \Phi)$, $\zeta_i \sim \mathcal{N}(0, \Psi)$, $\delta_i \sim \mathcal{N}(0, \Theta_\delta)$ and $\varepsilon_i \sim \mathcal{N}(0, \Theta_\varepsilon)$. ξ_i, ζ_i, δ_i, and ε_i are assumed to be independent of each other. Θ_δ and Θ_ε are assumed to be diagonal. The components $\{\gamma_0, \Gamma\}$, Φ, Ψ, $\{\nu_x, \Lambda_x\}$, Θ_δ, $\{\nu_y, \Lambda_y\}$, Θ_ε of ϑ are assumed to be stochastically independent. The priors for the components of ϑ are chosen as conjugate priors and given in detail below.

We define the $p(m + 1)$ dimensional column vector γ as the vectorized form of γ_0 and Γ, that is

$$\gamma = (\gamma_{01}, \gamma_{11}, \ldots, \gamma_{1m}, \gamma_{02}, \gamma_{21}, \ldots, \gamma_{2m}, \ldots, \gamma_{pm})'. \tag{18}$$

In a similar way, we define

$$\tilde{\lambda}_{x,k} = (\nu_{x,k}, \lambda_{x,k,1}, \ldots, \lambda_{x,k,\bar{m}})'. \tag{19}$$

and

$$\tilde{\lambda}_{y,j} = (\nu_{y,j}, \lambda_{y,j,1}, \ldots, \lambda_{y,j,p})'. \tag{20}$$

as the vectorized forms of the constant and factor loading coefficients for the k-th component of x and the j-th component of y. These vectors are assumed to be independent and to follow the conjugate priors

$$\gamma \sim \mathcal{N}(\gamma^*, \Omega_\gamma), \tag{21}$$

$$\tilde{\lambda}_{x,k} \sim \mathcal{N}(\lambda_{x,k}^*, \Omega_{x,k}), \tag{22}$$

$$\tilde{\lambda}_{y,j} \sim \mathcal{N}(\lambda_{y,j}^*, \Omega_{y,j}), \tag{23}$$

where the expected values γ^*, $\lambda_{x,k}^*$, $\lambda_{y,j}^*$ and the covariance matrices Ω_γ, $\Omega_{x,k}$, $\Omega_{y,j}$ or respectively the precision matrices Ω_γ^{-1}, $\Omega_{x,k}^{-1}$, $\Omega_{y,j}^{-1}$ are chosen by the researcher.

Before we turn to the conjugate priors of the covariance matrices Ψ and Φ, we review briefly the Wishart and the inverse Wishart distribution. The following definition of the Wishart distribution is used. The $p \times p$-dimensional positive definite matrix A follows a Wishart distribution with $p \times p$-dimensional positive definite parameter matrix B and d degrees of freedom if the density is given by

$$A \sim \mathcal{W}(B, d) \propto |B|^{-d/2} |A|^{(d-p-1)/2} \exp\left(-\frac{1}{2} \operatorname{tr} (B^{-1}A) \right) \tag{24}$$

where $d \geq p$ (see Anderson, 1984, p. 249; and Carlin & Louis, 1996, p. 168). We note that $E(A) = dB$. The inverse matrix $A^* = A^{-1}$ with parameter matrix $B^* = B^{-1}$ has the density

$$A^* \sim \mathcal{W}^{-1}(B^*, d) \propto |B^*|^{d/2} |A^*|^{-(d+p+1)/2} \exp\left(-\frac{1}{2} \operatorname{tr} (B^*(A^*)^{-1}) \right) \tag{25}$$

as shown by Anderson (1984, p 268). B^* is called the precision matrix. Following Anderson (1984, p. 269) further, his corollary 7.7.1 says:

With an inverse Wishart prior for the covariance matrix Σ,

$$\Sigma \sim \mathcal{W}^{-1}(d\Omega, d) \tag{26}$$

where $(d\Omega)$ is a precision matrix and

$$A = nS \sim \mathcal{W}(\Sigma, n) \tag{27}$$

we find the posterior

$$\Sigma | S \sim \mathcal{W}^{-1}(nS + d\Omega, n + d). \tag{28}$$

The notation of Anderson (1984) is changed slightly in the sense that $(d\Omega)$ is the precision matrix of the inverse Wishart distribution. Hence, a weak prior for Σ has a low d and a small matrix Ω so that nS dominates in the posterior form $nS + d\Omega$. S is the sample covariance matrix computed from n data points centered about the expected value.

The conjugate prior for the covariance matrix Ψ of the random vector $\zeta \sim \mathcal{N}(0, \Psi)$ is given in terms of the inverse Wishart distribution

$$\boldsymbol{\Psi} \sim \mathcal{W}^{-1}(d_\Psi \boldsymbol{\Omega}_\Psi, d_\Psi) \tag{29}$$

The parameter matrix $\boldsymbol{\Omega}_\Psi$ and the degrees of freedom d_Ψ in (29) are chosen by the researcher.

Similarly, the conjugate prior for the covariance matrix $\boldsymbol{\Phi}$ of $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi})$ is given by

$$\boldsymbol{\Phi} \sim \mathcal{W}^{-1}(d_\Phi \boldsymbol{\Omega}_\Phi, d_\Phi). \tag{30}$$

Since the covariance matrices $\boldsymbol{\Theta}_\delta$ and $\boldsymbol{\Theta}_\varepsilon$ are diagonal, we can use the conjugate priors for the variances of the error terms in the measurement equations. The prior is given by the inverse Gamma distribution for a random variable s with $s > 0$

$$s \sim \mathcal{IG}(a, b) = \Gamma(a)^{-1} b^a s^{-(a+1)} \exp\left(-\frac{b}{s}\right) \tag{31}$$

where a ($a > 0$) is a shape and b ($b > 0$) is a scale parameter (see Gelman, Carlin, Stern, & Rubin, 1995, p. 474). The expected value of s is $\mathrm{E}(s) = b/(a - 1)$ for $a > 1$. The inverse chi-square distribution is a special case of the inverse gamma distribution with $a = \mathrm{df}/2$, $b = 1/2$ where df is the number of degrees of freedom. The random variables $\Theta_{\delta,jj}$ and $\Theta_{\varepsilon,kk}$ then follow inverse Gamma distributions

$$\Theta_{\delta,jj} \sim \mathcal{IG}(a_{\delta,jj}, b_{\delta,jj}) \tag{32}$$

Taken together, this implies that the posterior distribution $[\boldsymbol{\vartheta}, \boldsymbol{\xi}, \boldsymbol{\eta}|\mathbf{y}, \mathbf{x}]$ is a proper posterior distribution.

$$\Theta_{\varepsilon,jj} \sim \mathcal{IG}(a_{\varepsilon,jj}, b_{\varepsilon,jj}) \tag{33}$$

where $a_{\delta,jj}$, $b_{\delta,jj}$ and $a_{\varepsilon,jj}$, $b_{\varepsilon,jj}$ are chosen by the researcher.

Several of the conditional distributions used in this paper for Gibbs sampling can be derived using known Bayesian results for the two linear models discussed in subsection 3.3.

3.2 Posterior and Likelihood

It is clarifying to express the joint posterior distribution in terms of the densities of all the distributions and write out the likelihood as the product of marginal and conditional distributions. The issue of whether proper priors lead to a proper posterior distribution can then be also considered.

Let the parameters of the model be denoted $\boldsymbol{\vartheta}$ and let $[\mathbf{z}]$ denote a density for the random variable $\mathbf{z}$. Using the assumptions of the A-M model, we note that $[\boldsymbol{\eta}|\mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\vartheta}] = [\boldsymbol{\eta}|\boldsymbol{\xi}, \boldsymbol{\vartheta}]$ and $[\mathbf{y}|\boldsymbol{\eta}, \mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\vartheta}] = [\mathbf{y}|\boldsymbol{\eta}, \boldsymbol{\vartheta}]$. Denoting the observed data distribution $[\mathbf{y}, \mathbf{x}]$ as c, the posterior distribution $[\boldsymbol{\vartheta}, \boldsymbol{\xi}, \boldsymbol{\eta}|\mathbf{y}, \mathbf{x}]$ for the unknown parameters and the latent variables conditional on the observed data may be expressed as

$$c \times [\boldsymbol{\vartheta}, \boldsymbol{\xi}, \boldsymbol{\eta}|\mathbf{y}, \mathbf{x}] = [\boldsymbol{\vartheta}][\boldsymbol{\xi}|\boldsymbol{\vartheta}][\mathbf{x}|\boldsymbol{\xi}, \boldsymbol{\vartheta}][\boldsymbol{\eta}|\boldsymbol{\xi}, \boldsymbol{\vartheta}][\mathbf{y}|\boldsymbol{\eta}, \boldsymbol{\vartheta}] \tag{34}$$

where $[\boldsymbol{\vartheta}]$ is the prior for the parameters. Here, the conditional distributions $[\boldsymbol{\xi}|\boldsymbol{\vartheta}]$, $[\mathbf{x}|\boldsymbol{\xi}, \boldsymbol{\vartheta}]$, $[\boldsymbol{\eta}|\boldsymbol{\xi}, \boldsymbol{\vartheta}]$, and $[\mathbf{y}|\boldsymbol{\eta}, \boldsymbol{\vartheta}]$ are all multivariate normal according to the A-M model.

The likelihood L is

$$L = [\mathbf{y}, \mathbf{x}|\boldsymbol{\vartheta}] = \int \int [\boldsymbol{\xi}|\boldsymbol{\vartheta}][\mathbf{x}|\boldsymbol{\xi}, \boldsymbol{\vartheta}][\boldsymbol{\eta}|\boldsymbol{\xi}, \boldsymbol{\vartheta}][\mathbf{y}|\boldsymbol{\eta}, \boldsymbol{\vartheta}] \, d\boldsymbol{\eta} \, d\boldsymbol{\xi}. \tag{35}$$

In (34), a proper posterior distribution is obtained if the following integral converges

$$c = [\mathbf{y}, \mathbf{x}] = \int L \times [\boldsymbol{\vartheta}] \, d\boldsymbol{\vartheta}. \tag{36}$$

With $g(\boldsymbol{\xi}) = \boldsymbol{\xi}$, $[\boldsymbol{\xi}, \boldsymbol{\eta}]$ and $[\mathbf{y}, \mathbf{x}]$ are multivariate normal. In this case, it is well-known from Bayesian analysis using conjugate priors that this gives a proper posterior for $\boldsymbol{\vartheta}$, see for example, Gelman, Carlin, Stern and Rubin (1995, pp. 80–81). The propriety of the posterior for other choices $g(\boldsymbol{\xi}) \neq \boldsymbol{\xi}$ needs to be investigated separately for each class of $g(\boldsymbol{\xi})$.

3.3 Posteriors for Covariance Matrices and Regression Coefficients

3.3.1 Random Case

We consider a linear model for the p-dimensional vector $\mathbf{v}_i$ for observation $i(i = 1, 2, \ldots, n)$ with known $p \times r$ matrix $\mathbf{A}$, r-dimensional random coefficient vector $\boldsymbol{\kappa}_i$, and p-dimensional residual vector $\mathbf{e}_i$,

$$\mathbf{v}_i = \mathbf{A}\boldsymbol{\kappa}_i + \mathbf{e}_i, \tag{37}$$

for which it is assumed that

$$\boldsymbol{\kappa}_i \sim \mathcal{N}(\boldsymbol{\kappa}_0, \boldsymbol{\Sigma}), \tag{38}$$

where $\boldsymbol{\Sigma}$ is an $r \times r$ covariance matrix. We further assume that the residuals e_{ij} are uncorrelated and have variances

$$V(e_{ij}) = \sigma_j^2; j = 1, \ldots, p. \tag{39}$$

The following proper prior distributions are assumed

$$\boldsymbol{\kappa}_0 \sim \mathcal{N}(\mathbf{c}, \mathbf{D}), \tag{40}$$

$$\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(d\boldsymbol{\Omega}, d), \tag{41}$$

$$\sigma_j^2 \sim \mathcal{IG}(e, f). \tag{42}$$

Here, a vague prior for $\boldsymbol{\kappa}_0$ is obtained by choosing $\mathbf{D}^{-1} = \mathbf{0}$. As seen below, this results in the prior mean $\mathbf{c}$ vanishing from the posterior. As in section 3.1, $(d\boldsymbol{\Omega})$ is the precision matrix for the inverse Wishart prior of $\boldsymbol{\Sigma}$. A low value of d relative to n gives a vague covariance matrix prior; the smallest admissible value is r, the number of rows in $\boldsymbol{\Sigma}$. A vague prior for σ_j^2 is obtained by choosing e close to zero and f large.

Well-known results (see, e.g., Anderson, 1984, p. 269; Carlin & Louis, 1996, pp. 168–169; and Lindley & Smith, 1972) give the conditional posterior distribution for $\boldsymbol{\kappa}_0$ conditional on $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, \boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2, \ldots, \boldsymbol{\kappa}_n, \sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2, \boldsymbol{\Sigma}$

$$\boldsymbol{\kappa}_0 \sim \mathcal{N}(\mathbf{V}(n\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{\kappa}} + \mathbf{D}^{-1}\mathbf{c}), \mathbf{V}), \tag{43}$$

for $\boldsymbol{\Sigma}$ given $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, \boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2, \ldots, \boldsymbol{\kappa}_n, \boldsymbol{\kappa}_0, \sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2$,

$$\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(n\mathbf{S} + d\boldsymbol{\Omega}, n + d), \tag{44}$$

and for σ_j^2 given $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, \boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2, \ldots, \boldsymbol{\kappa}_n, \boldsymbol{\kappa}_0, \sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2, \boldsymbol{\Sigma}^{-1}$

$$\sigma_j^2 \sim \mathcal{IG}\left(\frac{n}{2} + e, \left[\frac{1}{2}\sum_{i=1}^n (\mathbf{v}_i - \mathbf{A}\boldsymbol{\kappa}_i)'(\mathbf{v}_i - \mathbf{A}\boldsymbol{\kappa}_i) + f^{-1}\right]^{-1}\right). \tag{45}$$

Here,

$$\mathbf{V} = (n\mathbf{\Sigma}^{-1} + \mathbf{D}^{-1})^{-1}, \tag{46}$$

$$\bar{\kappa} = \frac{1}{n} \sum_{i=1}^{n} \kappa_i, \tag{47}$$

$$n\mathbf{S} = \sum_{i=1}^{n} (\kappa_i - \kappa_0)(\kappa_i - \kappa_0)'. \tag{48}$$

While the above posterior distributions are conditional on the other unknowns, it is well-known that the joint posterior distribution is also proper (see Gelman et al. 1995, pp. 80–81). Hobert and Casella (in press) further discusses properties of posteriors with various variance priors for the mixed linear model and show conditions under which proper posteriors are obtained even with "power" improper priors on the variance parameters.

3.3.2 Fixed Case

We consider the q-dimensional vector $\mathbf{v}_i$ for observations $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, the known $q \times r$ matrix $\mathbf{A}_i$, the r-dimensional parameter vector κ, and the q-dimensional residual vector $\mathbf{e}_i$,

$$\mathbf{v}_i = \mathbf{A}_i\kappa + \mathbf{e}_i, \tag{49}$$

where $\mathbf{V}(\mathbf{e}_i) = \mathbf{\Sigma}$, a $q \times q$ covariance matrix which is assumed known.

We assume the proper prior distribution

$$\kappa \sim \mathcal{N}(\kappa_0, \mathbf{\Omega}), \tag{50}$$

where κ_0 and $\mathbf{\Omega}$ are known. Using well-known results (see, e.g., Carlin & Louis, 1996, p. 41; Lindley & Smith, 1972), we find that the posterior distribution for κ conditional on $\mathbf{v}_1$, $\mathbf{v}_2, \ldots, \mathbf{v}_n$ is

$$\kappa \sim \mathcal{N}(\mathbf{Dd}, \mathbf{D}) \tag{51}$$

where

$$\mathbf{D}^{-1} = \left(\sum_{i=1}^{n} \mathbf{A}_i'\mathbf{\Sigma}^{-1}\mathbf{A}_i \right) + \mathbf{\Omega}^{-1}, \text{ and} \tag{52}$$

$$\mathbf{d} = \left(\sum_{i=1}^{n} \mathbf{A}_i'\mathbf{\Sigma}^{-1}\mathbf{v}_i \right) + \mathbf{\Omega}^{-1}\kappa_0 \tag{53}$$

A weak, proper prior is obtained by choosing the precision matrix $\mathbf{\Omega}^{-1}$ close to a zero matrix.

4. The Components of Markov Chain Monte Carlo Sampling

To describe the joint and marginal posterior distributions of ξ, η and ϑ given x and y we use the Gibbs sampler (Gelfand & Smith 1990; Geman & Geman, 1984), the M-H algorithm (Chib & Greenberg, 1995; Hastings 1970; Metropolis et al., 1993; Tanner, 1993) and a combination of both (Müller 1994). Easily accessible descriptions of the Gibbs sampler are found in Arnold (1993) and Casella and George (1992). Statistical theory underlying Markov Chain Monte Carlo (MCMC) methods such as the Gibbs sampler and

the M-H algorithm is found in Chib and Greenberg and Tierney (1994). General references on Bayesian computations are Besag, Green, Higdon and Mengersen (1995) and Gilks, Richardson and Spiegelhalter (1996). The relationship between iterative simulation techniques and the more familiar EM algorithm is discussed in Rubin (1991).

In the Gibbs sampler one samples iteratively from the conditional distributions

$$[\xi^{(k+1)}, \eta^{(k+1)}|x, y, \vartheta^{(k)}],$$ (54)

where $k + 1$ denotes the present iteration and the distributions

$$[\vartheta_q^{(k+1)}|x, y, \xi^{(k)}, \eta^{(k)}, \vartheta_{l<q}^{(k+1)}, \vartheta_{l>q}^{(k)}],$$ (55)

where $\vartheta_q^{(k+1)}$ is the q-th subvector of the vector ϑ at the present iteration. After a burn in phase, the conditional distributions converge to the posterior distributions of the parameters of interest.

We start by assuming that ξ_i and η_i are known for the k-th iteration taking on the values $\xi_i^{(k)}$, $\eta_i^{(k)}$. Then we can compute

$$\alpha_i^{(k)} = g(\xi_i^{(k)}).$$ (56)

4.1 Parameters of the Regression Model

The conditional distribution of the regression parameters in the model (3)–(5) is

$$[\gamma_0, \Gamma|x, y, \xi, \eta, \vartheta - \{\gamma_0, \Gamma\}].$$ (57)

Here, $\vartheta - \{\gamma_0, \Gamma\}$ is the set of all parameters in ϑ without the set of parameter that are elements of γ_0 or Γ.

Since the prior distribution of γ from (18) is multivariate normal with $\gamma \sim \mathcal{N}(\gamma^*, \Omega_\gamma)$ where γ^* is the known a priori expected value and Ω_γ is the known a priori covariance matrix of γ, the posterior distribution of γ given η_i and α_i may be derived using the results of subsection 3.2.2. The values α_i are written in a regressor matrix $A_i \sim p \times (p(m + 1))$ defined by

$$A_i = I_{p \times p} \otimes (1, \alpha_i),$$ (58)

where $\otimes$ denotes the Kronecker product. With this notation

$$\eta_i = A_i \gamma + \zeta_i$$ (59)

where η_i, A_i and $V(\zeta_i) = \Psi$ are known in the k-th step of the Gibbs sampler. Application of (51) and substituting η_i for v_i, γ for κ, and Ψ for Σ yields that the posterior distribution of γ is multivariate normal with expected value.

$$E(\gamma) = \left(\left[\sum_{i=1}^{n} A_i' \Psi^{-1} A_i \right] + \Omega_\gamma^{-1} \right)^{-1} \left(\left[\sum_{i=1}^{n} A_i' \Psi^{-1} \eta^{-1} \right] + \Omega^{-1} \gamma^* \right),$$ (60)

and covariance matrix

$$V(\gamma) = \left(\left[\sum_{i=1}^{n} A_i' \Psi^{-1} A_i \right] + \Omega_\gamma^{-1} \right)^{-1}.$$ (61)

Hence $\gamma^{(k+1)}$ is drawn from the posterior distribution $\mathcal{N}(E(\gamma), V(\gamma))$ where α_i, η_i and Ψ are replaced by $\alpha_i^{(k)}$, $\eta_i^{(k)}$ and $\Psi^{(k)}$.

A weak proper prior is obtained by choosing the precision matrix $\Omega_\gamma^{-1} = 0$. Then, the equations (60) and (61) simplify to

$$E(\gamma) = \left(\sum_{i=1}^{n} A_i' \Psi^{-1} A_i \right)^{-1} \left(\sum_{i=1}^{n} A_i' \Psi^{-1} \eta_i \right), \tag{62}$$

and

$$V(\gamma) = \left(\sum_{i=1}^{n} A_i' \Psi^{-1} A_i \right)^{-1}. \tag{63}$$

Alternatively, one might set γ^* to correspond to the expectations of the researcher, for instance as a vector of zeros and set Ω_γ^{-1} to a diagonal matrix with small values, for instance 0.01.

The conditional distribution of

$$[\Phi | x, y, \xi, \eta, \vartheta - \{\Phi\}] \tag{64}$$

is found using the inverse Wishart distribution. Since $\xi \sim \bar{m} \times 1$ is multivariate normal with $\mathcal{N}(0, \Phi)$ and Φ follows the inverse Wishart prior $\mathcal{W}^{-1}(d_\Phi \Omega_\Phi, d_\Phi)$, the posterior covariance matrix $\Phi^{(k+1)}$ follows an inverse Wishart distribution with parameter matrix

$$B_\Phi = \left(\sum_{i=1}^{n} \xi^{(k)} \xi^{(k)'} + d_\Phi \Omega_\Phi \right) \tag{65}$$

and $n + d_\Phi$ degrees of freedom using equations (44) and (48) and substituting ξ_i for κ_i, 0 for κ_0, d_Φ for d and Ω_Φ for Ω. A weak, proper prior for Φ is obtained by choosing d_Φ small relative to n, say $d_\Phi = \bar{m}$ and setting Ω_Φ to a diagonal matrix with small positive values.

The conditional distribution of

$$[\Psi | x, y, \xi, \eta, \vartheta - \{\Psi\}] \tag{66}$$

is also found using the inverse Wishart distribution. Since $\zeta \sim p \times 1$ is multivariate normal with $\mathcal{N}(0, \Psi)$ and Ψ follows the inverse Wishart prior $\mathcal{W}^{-1}(d_\Psi \Omega_\Psi, d_\Psi)$, the posterior covariance matrix $\Psi^{(k+1)}$ follows an inverse Wishart distribution with parameter matrix

$$B_\Psi = \left(\sum_{i=1}^{n} (\eta_i^{(k)} - A_i^{(k)} \gamma^{(k)})(\eta_i^{(k)} - A_i^{(k)} \gamma^{(k)})' + d_\Psi \Omega_\Psi \right), \tag{67}$$

and $n + d_\Psi$ degrees of freedom using (44) and (48) and substituting $\eta_i - A_i \gamma$ for κ_i, d_Ψ for d and Ω_Ψ for Ω. A weak, proper prior for Ψ is obtained by choosing d_Ψ small relative to n, say $d_\Phi = p$ and setting Ω_Ψ to a diagonal matrix with small positive values.

If η is univariate, then $\zeta \sim \mathcal{N}(0, \Psi)$ and Ψ is assumed to follow an inverse gamma distribution $\mathcal{IG}(e, f)$ as prior. From equations (42) and (45) we find that Ψ has the posterior distribution $\mathcal{IG}(a, b)$ with parameters

$$a = \frac{n}{2} + e \tag{68}$$

and

$$b = \frac{1}{2} \sum_{i=1}^{n} (\eta_i^{(k)} - A_i^{(k)} \gamma^{(k)})^2 + f^{-1}. \tag{69}$$

If e is chosen as an integer $\times 1/2$, for instance $e = 1/2$, then the variance $\Psi^{(k+1)}$ can be sampled from an inverse χ^2 distribution with $n + 2e$ degrees of freedom. f^{-1} is set to a small value, for instance 0.01.

4.2 Parameters of the Factor Analytic Models

Given η_i and the fact that Θ_ε is diagonal, one finds the posterior distribution of the vector $\tilde{\lambda}_{y,j} = (\nu_{y,j}, \lambda_{y,j,1}, \ldots, \lambda_{y,j,p})'$ for the jth component of y_i from the factor analytic regression equation

$$y_{ij} = \nu_{y,j} + \lambda_{y,j,1}\eta_{i1} + \cdots + \lambda_{y,j,p}\eta_{ip} + \varepsilon_{ij} \qquad (70)$$

and the prior for $\tilde{\lambda}_{y,j} \sim \mathcal{N}(\tilde{\lambda}^*_{y,j}, \Omega_{y,j})$. Application of (51) and substituting $\tilde{\lambda}_{y,j}$ for κ, $\Theta^{-1}_{\varepsilon,jj}$ for Σ^{-1} and η_j for A_i and $y_{1j}, \ldots, y_{nj}$ for $v_1, \ldots, v_n$ shows that the conditional distribution

$$[\nu_y, \Lambda_y | x, y, \xi, \eta, \vartheta - \{\nu_y, \Lambda_y\}] \qquad (71)$$

is found by drawing each vector $\tilde{\lambda}_{y,j}$ from a $p + 1$ dimensional normal distribution with expected value

$$E(\tilde{\lambda}_{y,j}) = (\Theta^{-1}_{\varepsilon,jj} Z'Z + \Omega^{-1}_{y,j})^{-1} (\Theta^{-1}_{\varepsilon,jj} Z'y_j + \Omega^{-1}_{y,j} \tilde{\lambda}^*_{y,j}) \qquad (72)$$

and covariance matrix

$$(\Theta^{-1}_{\varepsilon,jj} Z'Z + \Omega^{-1}_{y,j})^{-1}. \qquad (73)$$

The $n \times (p + 1)$ regressor matrix Z is defined as

$$Z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \text{ with } z_i = (1, \eta_{i1}, \ldots, \eta_{ip}). \qquad (74)$$

The regressand vector $y_j = (y_{1j}, y_{2j}, \ldots, y_{nj})'$.

A weak proper prior is obtained by setting the precision matrix $\Omega^{-1}_{y,j} = 0$. Alternatively, one might choose $\tilde{\lambda}_{y,j}$ to reflect the expectation of the researcher, for instance as a vector of ones and set $\Omega^{-1}_{y,j}$ to a diagonal matrix with small values, for instance 0.01.

The conditional distribution

$$[\Theta_\varepsilon | x, y, \xi, \eta, \vartheta - \{\Theta_\varepsilon\}] \qquad (75)$$

is found by considering that Θ_ε is a diagonal matrix and each element $\Theta_{\varepsilon,jj}$ follows an inverse Gamma distribution (cf. equation (33)) $\Theta_{\varepsilon,jj} \sim \mathcal{IG}(a_{\varepsilon,jj}, b_{\varepsilon,jj})$. From (42) and (45) we find that $\Theta_{\varepsilon,jj}$ has the posterior distribution $\mathcal{IG}(a, b)$ with parameters

$$a = \frac{n}{2} + a_{\varepsilon,jj} \qquad (76)$$

and

$$b = \frac{1}{2} \sum_{i=1}^{n} (y_{ij} - \nu_{y,j} - \Lambda_{y,j}\eta_i)^2 + b^{-1}_{\varepsilon,jj} \qquad (77)$$

where $\Lambda_{y,j}$ is the j-th row of Λ_y. To compute $\Theta^{(k+1)}_{\varepsilon,jj}$ in the Gibbs sampler, $\nu_{y,j}$, $\Lambda_{y,j}$ and η_i are replaced by $\nu^{(k)}_{y,j}$, $\Lambda^{(k)}_{y,j}$ and $\eta^{(k)}$. If $a_{\varepsilon,jj}$ is chosen as an integer $\times 1/2$, for instance $a_{\varepsilon,jj} = 1/2$, then the variance $\Theta^{(k+1)}_{\varepsilon,jj}$ can be sampled from an inverse χ^2 distribution with $n + 2a_{\varepsilon,jj}$ degrees of freedom. $b^{-1}_{\varepsilon,jj}$ is set to a small value, for instance 0.01.

Given ξ_i and the fact that Θ_δ is diagonal, similar calculations can be made to find the marginal posterior distributions of (ν_x, Λ_x) and Θ_δ. Since $\tilde{\lambda}_{x,k} = (\nu_{x,k}, \lambda_{x,k,1}, \ldots, \lambda_{x,k,\tilde{m}})'$ follows the prior $\mathcal{N}(\tilde{\lambda}^*_{x,k}, \Omega_{x,k})$, (51) may be applied again by substituting $\tilde{\lambda}_{x,k}$ for κ, $\Theta^{-1}_{\delta,kk}$ for Σ^{-1}, ξ_k for A_i and $x_{1k}, \ldots, x_{nk}$ for $v_1, \ldots, v_n$.

Then the conditional distribution

$$[\nu_x, \Lambda_x | x, y, \xi, \eta, \vartheta - \{\nu_x, \Lambda_x\}] \tag{78}$$

is found by drawing each vector $\tilde{\lambda}_{x,k}$ from a $(\tilde{m} + 1)$ dimensional normal distribution with expected value

$$E(\tilde{\lambda}_{x,k}) = (\Theta^{-1}_{\delta,kk} U'U + \Omega^{-1}_{x,k})^{-1}(\Theta^{-1}_{\delta,kk} U'x_k + \Omega^{-1}_{x,k}\lambda^*_{x,k}) \tag{79}$$

and covariance matrix

$$V(\tilde{\lambda}_{x,k}) = (\Theta^{-1}_{\delta,kk} U'U + \Omega^{-1}_{x,k})^{-1}. \tag{80}$$

The $n \times (\tilde{m} + 1)$ regressor matrix U is defined as

$$U = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \text{ with } u_i = (1, \xi_{i1}, \ldots, \xi_{i\tilde{m}}). \tag{81}$$

The regressand vector $x_k = (x_{1k}, x_{2k}, \ldots, x_{nk})'$.

Sometimes, special identification restrictions have to be taken into consideration. For instance, if only a single factor $\xi \sim \mathcal{N}(0, \Phi)$ is assumed with measurement model $x_i = \nu_x + \Lambda_x \xi_i + \delta_i$ with $\delta_i \sim \mathcal{N}(0, \Theta_\delta)$ and error variances greater than 0, then Φ is only identified, if Λ_x is restricted, for instance by setting $\lambda_{x,11} = 1$, so that the equation

$$x_{i1} = \nu_{x,1} + \xi_i + \delta_{i1} \tag{82}$$

holds. Since $\lambda_{x,11}$ need not be estimated, the parameter vector $\tilde{\lambda}_{x,1}$ consists only of the element $\nu_{x,1}$. The $n \times 1$ regressor matrix U in (81) consists of a vector of ones. The regressand vector is given by $x_k = (x_{1k} - \xi_1, \ldots, x_{nk} - \xi_n)'$.

The conditional distribution

$$[\Theta_\delta | x, y, \xi, \eta, \vartheta - \{\Theta_\delta\}] \tag{83}$$

is found by considering that Θ_δ is a diagonal matrix where each element $\Theta_{\delta,kk}$ follows an inverse Gamma distribution. (see (32)) $\Theta_{\delta,kk} \sim \mathcal{IG}(a_{\delta,kk}, b_{\delta,kk})$. From (42) and (45) we find that $\Theta_{\delta,kk}$ has the posterior distribution $\mathcal{IG}(a, b)$ with parameters

$$a = \frac{n}{2} + a_{\delta,kk} \tag{84}$$

and

$$b = \frac{1}{2} \sum_{i=1}^{n} (x_{ik} - \nu_{x,k} - \Lambda_{x,k}\xi_i)^2 + b^{-1}_{\delta,kk}, \tag{85}$$

where $\Lambda_{x,k}$ is the k-th row Λ_x. To compute $\Theta^{(k+1)}_{\delta,kk}$ in the Gibbs sampler, $\nu_{x,k}$, $\Lambda_{x,k}$ and ξ are replaced by $\nu^{(k)}_{x,k}$, $\Lambda^{(k)}_{x,k}$ and $\xi^{(k)}_i$. If $a_{\delta,kk}$ is chosen as an integer $\times 1/2$, for instance $a_{\delta,kk} = 1/2$, then the variance $\Theta^{(k+1)}_{\delta,kk}$ can be sampled from an inverse χ^2 distribution with $n + 2a_{\delta,kk}$ degrees of freedom. $b^{-1}_{\delta,kk}$ is set to a small value, for instance 0.01.

4.3   Drawing from the Distribution of Latent Variables

The conditional distribution of ξ and η given x, y and ϑ is now considered. Since independence of (x_i, y_i) and ξ_i, η_i, $i = 1, \ldots, n$ across units has been assumed, it suffices to look at the conditional distribution

$$[\xi_i, \eta_i | x_i, y_i, \vartheta]. \tag{86}$$

Unlike for the conditional distributions considered before, we have not been able to derive a form of the conditional distribution from which values of (ξ_i, η_i) given x_i, y_i, ϑ can be sampled easily. Therefore, we use the Metropolis-Hastings algorithm as described in Tanner (1993), and Chib and Greenberg (1995) to generate values of (ξ_i, η_i) that follow the conditional distribution of (86). Let $(\xi^{(0)}, \eta^{(0)})$ be the current value and $(\xi^{(1)}, \eta^{(1)})$ denote the value generated by a distribution from which one can easily sample. This easily sampled distribution is called the driver (driving or proposal distribution). Let $\pi^{(0)}$ and $\pi^{(1)}$ denote the density of (86) evaluated at $(\xi^{(0)}, \eta^{(0)})$ and $(\xi^{(1)}, \eta^{(1)})$. Then the value

$$c = \min \left\{ \frac{\pi^{(1)}}{\pi^{(0)}}, 1 \right\} \tag{87}$$

is calculated. If $c = 1$, then the value $(\xi^{(1)}, \eta^{(1)})$ is accepted and becomes the current value. If $c < 1$, then the value $(\xi^{(1)}, \eta^{(1)})$ is accepted only with probability c, otherwise $(\xi^{(0)}, \eta^{(0)})$ stays the current value. This algorithm ensures that the accepted values form a sample from the distribution of (86). To use the M-H algorithm we have to compute $\pi^{(0)}$ and $\pi^{(1)}$ and to sample from a simple driving distribution. First, we deal with the computation of $\pi^{(0)}$ and $\pi^{(1)}$. We note that $[\xi_i, \eta_i | x_i, y_i, \vartheta]$ may be written as

$$[\xi_i, \eta_i | x_i, y_i, \vartheta] \propto [\xi_i, \eta_i | \vartheta][x_i | \xi_i, \eta_i, \vartheta][y_i | x_i, \xi_i, \eta_i, \vartheta]$$

$$= [\xi_i, \eta_i | \vartheta][x_i | \xi_i, \vartheta][y_i | \eta_i, \vartheta]. \tag{88}$$

The last equality holds because x_i depends only on ξ_i and not on η_i and because y_i given η_i does not depend on ξ_i or x_i. Substituting

$$[\xi_i, \eta_i | \vartheta] = [\xi_i | \vartheta][\eta_i | \xi_i, \vartheta] \tag{89}$$

yields the expression:

$$[\xi_i, \eta_i | x_i, y_i, \vartheta] \propto [\xi_i | \vartheta][\eta_i | \xi_i, \vartheta][x_i | \xi_i, \vartheta][y_i | \eta_i, \vartheta]. \tag{90}$$

Since the ratio $\pi^{(1)}/\pi^{(0)}$ does not depend on the proportionality constant, it suffices to evaluate $[\xi_i | \vartheta][\eta_i | \xi_i, \vartheta][x_i | \xi_i, \vartheta][y_i | \eta_i, \vartheta]$ at the points $(\xi^{(0)}, \eta^{(0)})$ and $(\xi^{(1)}, \eta^{(1)})$. The components of this joint density of ξ_i, η_i, x_i and y_i given ϑ are implied by the model in (3) through (5).

An important special case occurs if $y = \eta$, that is, the measurement model for y is given by $y = \eta$. In this case, we consider the conditional distribution:

$$[\xi_i | x_i, y_i, \vartheta] \propto [\xi_i | \vartheta][x_i | \xi_i, \vartheta][y_i | \xi_i, \vartheta]. \tag{91}$$

For the proposal distribution, we take the joint distribution $[\xi_i, \eta_i | \vartheta] = [\xi_i | \vartheta][\eta_i | \xi_i, \vartheta]$ by first sampling ξ from the a priori distribution $\mathcal{N}(0, \Phi)$ and then sampling η from $\mathcal{N}(\gamma_0 + \Gamma \alpha(\xi), \Psi)$ where Φ, γ_0, Γ, and Ψ are evaluated at the kth iteration step. Alternatively, we sample ξ from the conditional distribution $[\xi | x, \vartheta]$ which may be derived from the joint normal distribution of x and ξ.

$$\begin{pmatrix} x \\ \xi \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \nu_x \\ 0 \end{bmatrix}, \begin{bmatrix} \Lambda_x \Phi \Lambda_x' + \Theta_\delta, & \Lambda_x \Phi \\ \Phi \Lambda_x', & \Phi \end{bmatrix} \right) \tag{92}$$

Hence, the conditional distribution of ξ given $\mathbf{x}_i$ is multivariate normal with

$$E(\xi|\mathbf{x}_i, \boldsymbol{\nu}_x, \boldsymbol{\Phi}, \boldsymbol{\Lambda}_x, \boldsymbol{\Theta}_\delta) = \boldsymbol{\Phi}\boldsymbol{\Lambda}_x'(\boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta)^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_x) \tag{93}$$

and covariance matrix

$$V(\xi|\mathbf{x}_i, \boldsymbol{\nu}_x, \boldsymbol{\Phi}, \boldsymbol{\Lambda}_x, \boldsymbol{\Theta}_\delta) = \boldsymbol{\Phi} - \boldsymbol{\Phi}\boldsymbol{\Lambda}_x'(\boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta)^{-1}\boldsymbol{\Lambda}_x\boldsymbol{\Phi}$$

$$= (\boldsymbol{\Phi}^{-1} + \boldsymbol{\Lambda}_x'\boldsymbol{\Theta}_\delta^{-1}\boldsymbol{\Lambda}_x)^{-1}. \tag{94}$$

Since the proposal distribution does not depend on $(\xi^{(0)}, \boldsymbol{\eta}^{(0)})$, the M-H chain generated in this way is called an independence chain (Tierney, 1994).

5. Analysis of Simulated Data

The Bayesian analysis described above is investigated using two simulated data sets, the first coming from a model with one latent regressor variable and its quadratic term, the second coming from a model with two latent regressor variables and an interaction term between the two regressor variables. The sample sizes used are in both cases $n = 100$, $n = 250$ and $n = 500$. A total of 100 Monte Carlo replications were used to study the variation across different generated samples.

Important questions for the practical implementation of Gibbs sampling are whether to use multiple runs with different start values or to use a single run, whether to use—after a burn in phase—each value of the Gibbs sampler or to use a subsample and how to judge convergence of the Gibbs sampler. Each of these questions has been debated in the literature. Following Chib and Greenberg (1996), we have implemented the single run method because it is less wasteful in the number of iterations needed. Following the arguments of McEachern and Berliner (1994) we use the full Gibbs sample rather than subsampling because subsampling is inefficient in comparison to full sampling. Methods of judging convergence from multiple runs are found in Gelman and Rubin (1992) and from a single run in Ritter and Tanner (1992). Given the massive amount of computing for analyzing the Monte Carlo replications and that the number of iterations for each analysis should be comparable we have settled for a burn in phase of 500 Gibbs cycles and additional 2000 Gibbs cycles from which parameters of the posterior distribution are estimated. The number of M-H cycles is critical. We started with 5 cycles yielding unsatisfactory results while 20 cycles gave satisfactory results in all simulations performed. The check, how many cycles are necessary, can only be done by varying the numbers of cycles systematically and checking for convergence of results. In the simulation, the convergence to the true values suffices.

All computations are performed using the program BALAM (Bayesian Analysis of Latent Variable Models) which is written in GAUSS 3.2.13 for DOS and 3.2.29 for Unix and is run on a Sun Sparc 10 station. The time for one run is approximately 10.00, 18.20, and 36.00 minutes (sample size $n = 100, 250, 500$) in the first model, 10.20, 23.10 and 44.40 minutes (sample size $n = 100, 250, 500$) in the second model and 45.20 minutes in the third model discussed below.

The regression model for the first data set is given by

$$y_i = \gamma_{10} + \gamma_{11}\xi_i + \gamma_{12}\xi_i^2 + \zeta_i, \tag{95}$$

with $\xi_i \sim \mathcal{N}(0, \Phi_{11})$ and $\zeta_i \sim \mathcal{N}(0, \Psi_{11})$. The factor analytic model connecting the latent regressors ξ_i and the observed indicators $\mathbf{x}_i$ is

$$\mathbf{x}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\xi_i + \boldsymbol{\delta}_i, \tag{96}$$

with $\boldsymbol{\delta}_i \sim \mathcal{N}(0, \boldsymbol{\Theta}_\delta)$. For the simulation we use four indicator variables yielding the structure

$$\boldsymbol{\nu} = \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{pmatrix}, \quad \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \end{pmatrix}, \quad \boldsymbol{\Theta}_\delta = \operatorname{diag} \{\Theta_{\delta,11}, \ \Theta_{\delta,22}, \ \Theta_{\delta,33}, \ \Theta_{\delta,44}\}. \tag{97}$$

The value of λ_{11} is set to 1 to fix the scale of ξ. The following proper priors are used for the analysis: $\boldsymbol{\gamma} = (\gamma_{10}, \gamma_{11}, \gamma_{12})' \sim \mathcal{N}(\boldsymbol{\gamma}^*, \boldsymbol{\Omega}_\gamma)$ with $\boldsymbol{\gamma}^* = (1, 1, 1)$ and $\boldsymbol{\Omega}_\gamma^{-1} = \operatorname{diag} \{0.01,$ $0.01, 0.01\}$; $\Phi_{11} \sim \mathcal{I}\mathcal{G}(a_\Phi, b_\Phi)$ with $a_\Phi = 1/2$, $b_\Phi^{-1} = 0.01$, $\Psi_{11} \sim \mathcal{I}\mathcal{G}(a_\Psi, b_\Psi)$ with $a_\Psi = 1/2, b_\Psi^{-1} = 0.01$, $\nu_{x,j} \sim \mathcal{N}(\nu_{x,j}^*, \Omega_{\nu,x,j})$ with $\nu_{x,j}^* = 0$, and $\Omega_{\nu,x,j}^{-1} = 0.01$ for $j = 1, \dots,$ 4; $\lambda_{x,j} \sim \mathcal{N}(\lambda_{x,j}^*, \Omega_{\lambda,x,j})$ with $\lambda_{x,j}^* = 1$, and $\Omega_{\lambda,x,j}^{-1} = 0.01$ for $j = 2, \dots, 4$; $\Theta_{\delta,jj} \sim \mathcal{I}\mathcal{G}(a_{\delta,jj}, b_{\delta,jj})$ with $a_{\delta,jj} = 1/2, b_{\delta,jj}^{-1} = 0.01$ for $j = 1, \dots, 4$.

The results of the simulation are given in Tables 1–3. In these tables, the true parameters and the averages of the median, the mean and the standard deviation of the posterior distributions of the individual parameters over the 100 Monte Carlo replications are shown. These tables also give the 90 percent coverage for each parameter. In Bayesian analysis the coverage is computed by considering for each Monte Carlo replication whether the interval from the 5th to the 95th percentile in the distribution of the 2000 values of the Gibbs sampler covers the true parameter and then computing the proportion out of the 100 replications for which this event occurred.

Tables 1 through 3 show close agreement between the true parameter and the average median and mean. Judging from the values of the average median and mean, the posterior distributions are already symmetric for $n = 100$. The coverage properties are acceptable for all parameters already for $n = 100$.

Plotting the frequency estimate of the posterior distribution for individual parameters shows that these distributions are approximately normal even for the small sample size $n = 100$. As an example we consider the posterior distribution for the parameter of greatest interest in this specific nonlinear model, that is γ_{12}. The estimated posterior density is plotted using a histogram in Figure 1 and the k-nearest neighbor (kNN) smooth estimate (compare Loftsgaarden & Quesenberry 1965; and Härdle, 1990) in Figure 2. Figure 3 shows how the algorithm samples across the support of the posterior density from iteration to iteration indicating fast transition from one area of support to other areas.

The regression model for the second data set includes two regressor variables and their interaction. The regression constant has been set to 0.

$$y_i = \gamma_{11}\xi_{i1} + \gamma_{12}\xi_{i2} + \gamma_{13}\xi_{i1}\xi_{i2} + \zeta_i, \tag{98}$$

with $\xi_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi})$ and $\zeta_i \sim \mathcal{N}(0, \Psi_{11})$. $\boldsymbol{\Phi}$ is now a 2×2 covariance matrix with elements $\{\Phi_{11}, \Phi_{21}, \Phi_{22}\}$. The factor analytic model connecting ξ_i with five indicators $\mathbf{x}_i$ is now

$$\mathbf{x}_i = \boldsymbol{\Lambda}\xi_i + \boldsymbol{\delta}_i, \tag{99}$$

where $\boldsymbol{\delta}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta}_\delta)$ and the regression constants $\boldsymbol{\nu}$ are set to $\mathbf{0}$. The matrix of factor loadings is given by a 5×2 matrix patterned as

$$\boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \end{pmatrix}. \tag{100}$$

$\boldsymbol{\Theta}_\delta$ is a diagonal matrix with $\boldsymbol{\Theta}_\delta = \{\Theta_{\delta,11}, \dots, \Theta_{\delta,55}\}$.

The following proper priors are used for the analysis: $\boldsymbol{\gamma} = (\gamma_{11}, \gamma_{12}, \gamma_{13})' \sim \mathcal{N}(\boldsymbol{\gamma}^*, \boldsymbol{\Omega}_\gamma)$

TABLE 1

Averages of 100 Monte Carlo Replications

Quadratic Model, Sample Size = 100

Gibbs: Cycles = 2000, Burn-In = 500

Metropolis: Cycles = 20

Parameter	True	Median	Mean	Std.Dev.	Cover
γ_{10}	0.5	0.4802	0.4747	0.1598	0.9500
γ_{11}	1.0	1.0337	1.0378	0.1713	0.9300
γ_{12}	-0.6	-0.6146	-0.6196	0.0747	0.8500
Φ_{11}	1.4	1.3842	1.4050	0.2325	0.8900
Ψ_{11}	0.5	0.5117	0.5217	0.1001	0.9000
ν_1	-0.4	-0.4106	-0.4113	0.1285	0.9700
ν_2	-0.2	-0.2144	-0.2151	0.1174	0.9600
ν_3	0.2	0.1856	0.1852	0.1193	0.9500
ν_4	0.4	0.3752	0.3749	0.1097	0.9200
λ_{11}	1.0	1.0000	1.0000	—	1.0000
λ_{21}	0.9	0.9132	0.9132	0.0567	0.8700
λ_{31}	0.8	0.8034	0.8034	0.0719	0.9000
λ_{41}	0.7	0.6995	0.6995	0.0694	0.9200
$\Theta_{\delta 11}$	0.2	0.1945	0.1984	0.0448	0.9200
$\Theta_{\delta 22}$	0.2	0.2006	0.2042	0.0412	0.8900
$\Theta_{\delta 33}$	0.5	0.4989	0.5065	0.0802	0.8700
$\Theta_{\delta 44}$	0.5	0.4910	0.4984	0.0773	0.9200

with $\gamma^* = (1, 1, 1)$ and $\Omega_\gamma^{-1} = $ diag $\{0.01, 0.01, 0.01\}$; $\Phi \sim \mathcal{W}^{-1}(d_\Phi \Omega_\Phi, d_\Phi)$ where $\Omega_\Phi = $ diag $\{0.01, 0.01\}$ and $d_\Phi = 2$; $\Psi_{11} \sim \mathcal{IG}(a_\Psi, b_\Psi)$ with $a_\Psi = 1/2$, $b_\Psi^{-1} = 0.01$; $\lambda_{x,j} \sim \mathcal{N}(\lambda_{x,j}^*, \Omega_{\lambda,x,j})$ with $\lambda_{x,j}^* = 1$, and $\Omega_{\lambda,x,j}^{-1} = 0.01$ for $j = 2, 3, 5$; $\Theta_{\delta,jj} \sim \mathcal{IG}(a_{\delta,jj}, b_{\delta,jj})$ with $a_{\delta,jj} = 1/2$, $b_{\delta,jj}^{-1} = 0.01$ for $j = 1, \ldots, 5$.

The results of the Monte Carlo study are collected in Tables 4 through 6. The average means and medians of these tables show close agreement with the true parameters. The plots of the posterior density estimates of the interaction term γ_{13} and the iteration behavior are similar to the plots shown for the previous example.

6. Performance Regressed on Task Complexity and Goal Specificity

For empirical illustration we use data collected in 1990 from 158 police officers during a training course in Münster, Germany. Details of data collection are found in Holling

TABLE 2

Averages of 100 Monte Carlo Replications
Quadratic Model, Sample Size = 250
Gibbs: Cycles = 2000, Burn-In = 500
Metropolis: Cycles = 20

Parameter	True	Median	Mean	Std.Dev.	Cover
γ_{10}	0.5	0.4983	0.4962	0.0975	0.8600
γ_{11}	1.0	1.0075	1.0089	0.1057	0.8600
γ_{12}	-0.6	-0.6126	-0.6146	0.0446	0.8700
Φ_{11}	1.4	1.3906	1.3989	0.1443	0.8800
Ψ_{11}	0.5	0.5083	0.5121	0.0610	0.8500
ν_1	-0.4	-0.4018	-0.4024	0.0796	0.8800
ν_2	-0.2	-0.1981	-0.1982	0.0731	0.9000
ν_3	0.2	0.1942	0.1938	0.0745	0.8900
ν_4	0.4	0.3969	0.3968	0.0686	0.8900
λ_{11}	1.0	1.0000	1.0000	—	1.0000
λ_{21}	0.9	0.9096	0.9096	0.0347	0.9200
λ_{31}	0.8	0.8051	0.8051	0.0444	0.8800
λ_{41}	0.7	0.7018	0.7018	0.0433	0.9000
$\Theta_{\delta 11}$	0.2	0.1986	0.2001	0.0276	0.9600
$\Theta_{\delta 22}$	0.2	0.2000	0.2014	0.0255	0.9000
$\Theta_{\delta 33}$	0.5	0.4977	0.5006	0.0494	0.8900
$\Theta_{\delta 44}$	0.5	0.4996	0.5025	0.0484	0.9100

(1995). Background for the psychological theories to be tested are found in Early, Lee and Hanson (1990) and Holling. The dependent variable η is performance, regressor variables are task complexity ξ_1, goal specificity ξ_2, and their interaction ($\xi_1\xi_2$). The corresponding model is a nonlinear regression model in latent variables

$$\eta_i = \gamma_{11}\xi_{i1} + \gamma_{12}\xi_{i2} + \gamma_{13}(\xi_{i1}\xi_{i2}) + \zeta_i \qquad (101)$$

with $\xi \sim \mathcal{N}(0, \Phi)$ where Φ is a 2×2 covariance matrix with elements $\{\Phi_{11}, \Phi_{21}, \Phi_{22}\}$ and $\zeta_i \sim \mathcal{N}(0, \Psi_{11})$. The proper priors for this part of the model are chosen as: $\gamma = (\gamma_{11}, \gamma_{12}, \gamma_{13})' \sim \mathcal{N}(\gamma^*, \Omega_\gamma)$ with $\gamma^* = (1, 1, -1)$. The value of -1 is chosen since a negative interaction coefficient is expected theoretically. $\Omega_\gamma^{-1} = \text{diag} \{0.01, 0.01, 0.01\}$. $\Phi \sim \mathcal{W}^{-1}(d_\Phi\Omega_\Phi, d_\Phi)$ where Ω_Φ is a diagonal matrix with $\Omega_{\Phi,11} = 1.9646$ and $\Omega_{\Phi,22} = 0.6890$.

TABLE 3

Averages of 100 Monte Carlo Replications

Quadratic Model, Sample Size = 500

Gibbs: Cycles = 2000, Burn-In = 500

Metropolis: Cycles = 20

Parameter	True	Median	Mean	Std.Dev.	Cover
γ_{10}	0.5	0.4848	0.4836	0.0686	0.8900
γ_{11}	1.0	1.0147	1.0153	0.0739	0.8800
γ_{12}	-0.6	-0.6047	-0.6056	0.0309	0.8800
Φ_{11}	1.4	1.3918	1.3957	0.1018	0.8200
Ψ_{11}	0.5	0.4983	0.5002	0.0417	0.9100
ν_1	-0.4	-0.4100	-0.4101	0.0563	0.9100
ν_2	-0.2	-0.2099	-0.2097	0.0518	0.9000
ν_3	0.2	0.1887	0.1886	0.0528	0.8900
ν_4	0.4	0.3911	0.3910	0.0487	0.9200
λ_{11}	1.0	1.0000	1.0000	—	1.0000
λ_{21}	0.9	0.9094	0.9094	0.0243	0.8600
λ_{31}	0.8	0.8026	0.8026	0.0313	0.9100
λ_{41}	0.7	0.7019	0.7019	0.0305	0.8900
$\Theta_{\delta 11}$	0.2	0.1994	0.2002	0.0192	0.8200
$\Theta_{\delta 22}$	0.2	0.1955	0.1962	0.0174	0.8900
$\Theta_{\delta 33}$	0.5	0.5007	0.5022	0.0349	0.9500
$\Theta_{\delta 44}$	0.5	0.5026	0.5040	0.0342	0.9200

$d_\Phi = 2$; $\Psi_{11} \sim \mathcal{IG}(a_\Psi, b_\Psi)$ with $a_\Psi = 1/2$, $b_\Psi^{-1} = 0.6782$. The values of Ω_Φ and b_Ψ^{-1} have been chosen from the variances of the reference indicators x_1, x_4 and y_1 of the latent variables ξ_1, ξ_2 and η.

The indicator variables y_1, y_2, y_3 for the performance η are five point Likert scales with the questions A57A ("Is the amount of your work achieved during the last year (1) far below average to (5) far above average"), A57B ("Is the quality or precision of your work achieved during the last year (1) far below average to (5) far above average"), and A57E ("Is the fulfillment of your goals in your work achieved during the last year (1) far below average to (5) far above average"). The factor analytic model connecting the indicator variables y_1, y_2, y_3 with η is given by:

$$\mathbf{y}_i = \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y \eta_i + \boldsymbol{\varepsilon}_i \tag{102}$$
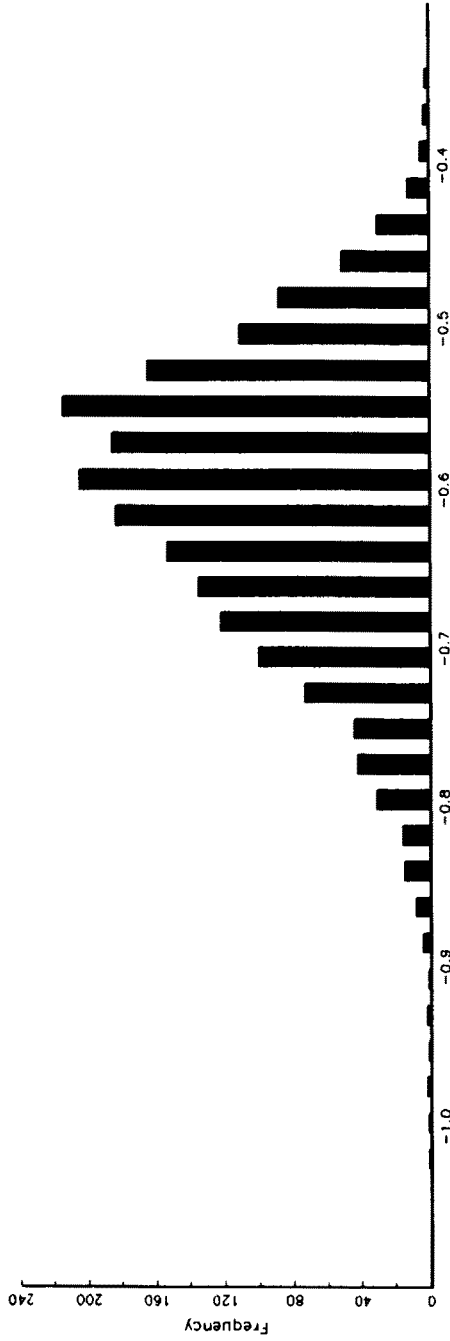
FIGURE 1.

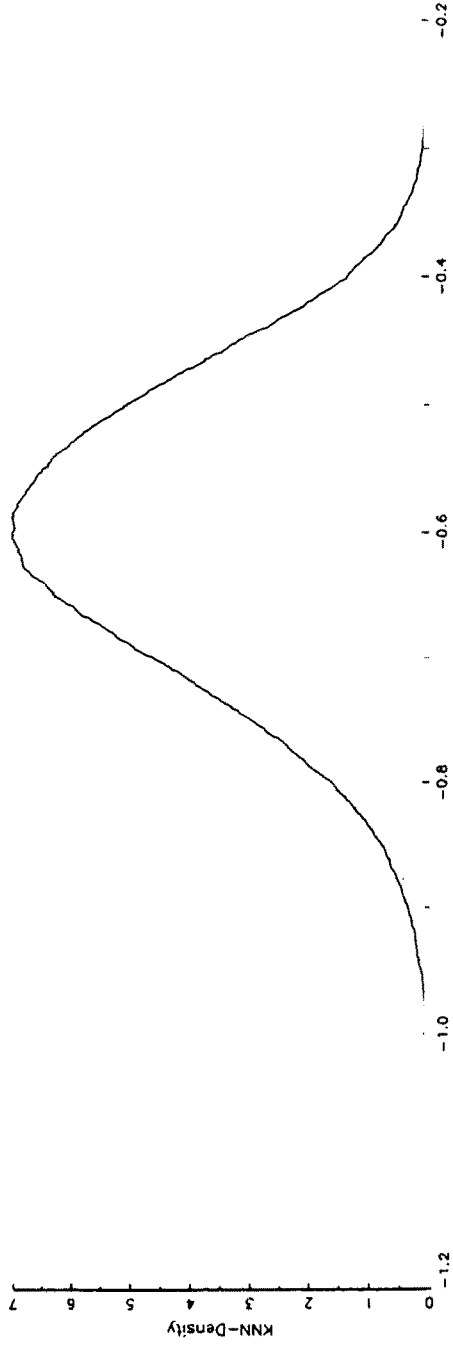Histogram of the estimated posterior density of γ_{12} in the quadratic model.

FIGURE 2.
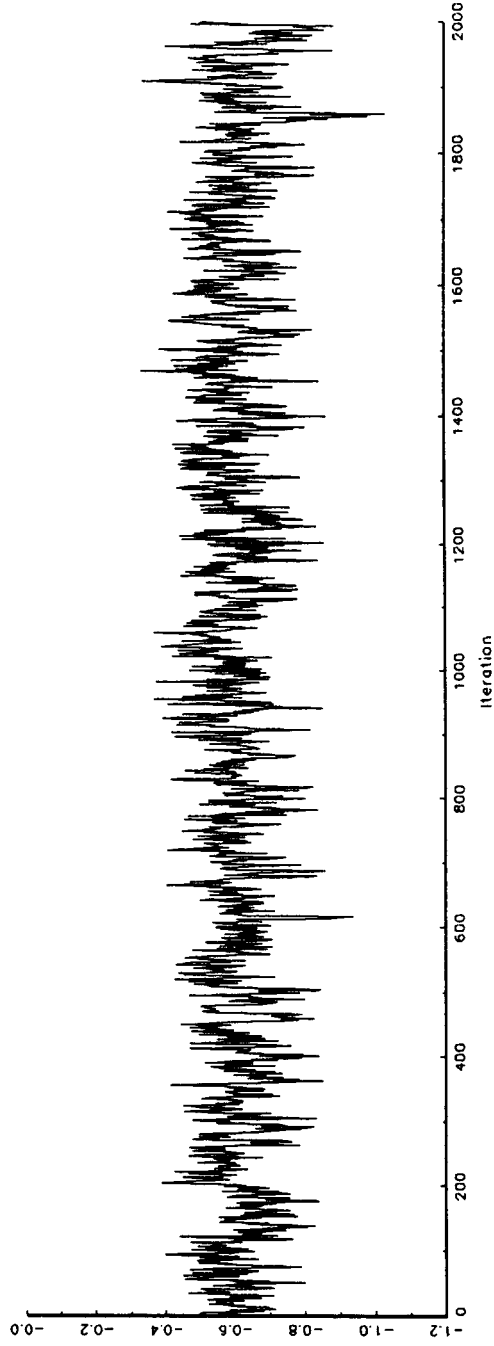Smooth estimate of the posterior density of γ_{12} in the quadratic model.

FIGURE 3.
Sample path of MCMC interactions for γ_{12} of the quadratic model.

TABLE 4

Averages of 100 Monte Carlo Replications

Interaction Model, Sample Size = 100

Gibbs: Cycles = 2000, Burn-In = 500

Metropolis: Cycles = 20

Parameter	True	Median	Mean	Std.Dev.	Cover
γ_{11}	0.8	0.8138	0.8176	0.1245	0.9200
γ_{12}	1.7	1.8175	1.8476	0.2743	0.8200
γ_{13}	0.5	0.5400	0.5514	0.1747	0.8700
Φ_{11}	1.2	1.1522	1.1679	0.2179	0.8500
Φ_{21}	0.1	0.0829	0.0857	0.0999	0.9400
Φ_{22}	0.7	0.6153	0.6298	0.1638	0.8800
Ψ_{11}	0.6	0.6687	0.6858	0.1966	0.9000
λ_{11}	1.0	1.0000	1.0000	—	1.0000
λ_{21}	0.7	0.7361	0.7361	0.0857	0.9100
λ_{31}	-0.5	-0.5181	-0.5181	0.0773	0.9200
λ_{42}	1.0	1.0000	1.0000	—	1.0000
λ_{52}	1.6	1.8160	1.8160	0.2862	0.8400
$\Theta_{\delta 11}$	0.2	0.2130	0.2195	0.0887	0.9000
$\Theta_{\delta 22}$	0.3	0.2885	0.2921	0.0626	0.8900
$\Theta_{\delta 33}$	0.4	0.4084	0.4142	0.0656	0.9600
$\Theta_{\delta 44}$	0.5	0.5250	0.5331	0.1000	0.8400
$\Theta_{\delta 55}$	0.6	0.5321	0.5396	0.1844	0.8800

with $\varepsilon_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Theta}_\varepsilon)$ and the parameter specification

$$\boldsymbol{\nu}_y = \begin{pmatrix} \nu_{y,1} \\ \nu_{y,2} \\ \nu_{y,3} \end{pmatrix}, \ \boldsymbol{\Lambda}_y = \begin{pmatrix} 1 \\ \lambda_{y,21} \\ \lambda_{y,31} \end{pmatrix}, \ \boldsymbol{\Theta}_\varepsilon = \operatorname{diag}\{\Theta_{\varepsilon,11}, \Theta_{\varepsilon,22}, \Theta_{\varepsilon,33}\}. \tag{103}$$

The value of $\lambda_{y,11}$ is set to 1 to fix the scale of η. The proper priors for this part of the model are specified as: $\nu_{y,j} \sim \mathcal{N}(\nu^*_{y,j}, \Omega_{\nu y,j})$ where $\nu^*_{y,j} = 0$, and $\Omega^{-1}_{\nu y,j} = 0.01$ for $j = 1$, 2, 3; $\lambda_{y,j1} \sim \mathcal{N}(\lambda^*_{y,j1}, \Omega_{\lambda y,j1})$ where $\lambda^*_{y,j1} = 1$, and $\Omega^{-1}_{\lambda y,j1} = 0.01$ for $j = 2, 3$; $\Theta_{\varepsilon,jj} \sim \mathcal{IG}(a_{\varepsilon,jj}, b_{\varepsilon,jj})$ with $a_{\varepsilon,jj} = 1/2$, $b^{-1}_{\varepsilon,jj} = 0.01$ for $j = 1, \ldots, 3$.

The indicator variables x_1, x_2, x_3 for task complexity ξ_1 are A18 ("How many of your

TABLE 5

Averages of 100 Monte Carlo Replications

Interaction Model, Sample Size = 250

Gibbs: Cycles = 2000, Burn-In = 500

Metropolis: Cycles = 20

Parameter	True	Median	Mean	Std.Dev.	Cover
γ_{11}	0.8	0.8004	0.8017	0.0738	0.9100
γ_{12}	1.7	1.6925	1.6992	0.1375	0.9500
γ_{13}	0.5	0.4971	0.5001	0.0913	0.9600
Φ_{11}	1.2	1.1687	1.1751	0.1327	0.9200
Φ_{21}	0.1	0.1041	0.1053	0.0654	0.8900
Φ_{22}	0.7	0.6698	0.6754	0.1036	0.8800
Ψ_{11}	0.6	0.6728	0.6792	0.1138	0.8300
λ_{11}	1.0	1.0000	1.0000	—	1.0000
λ_{21}	0.7	0.7087	0.7087	0.0483	0.9100
λ_{31}	-0.5	-0.5055	-0.5055	0.0452	0.9000
λ_{42}	1.0	1.0000	1.0000	—	1.0000
λ_{52}	1.6	1.6651	1.6651	0.1384	0.8700
$\Theta_{\delta 11}$	0.2	0.1930	0.1947	0.0532	0.8300
$\Theta_{\delta 22}$	0.3	0.3003	0.3019	0.0379	0.8900
$\Theta_{\delta 33}$	0.4	0.3898	0.3919	0.0388	0.9000
$\Theta_{\delta 44}$	0.5	0.5058	0.5090	0.0590	0.8800
$\Theta_{\delta 55}$	0.6	0.5538	0.5567	0.1061	0.8300

main tasks are the same from day to day" ranging from (1) almost all to (5) almost none), A19 ("Are the daily situations in which you have to fulfill your tasks (1) very much the same to (5) usually very different"), and A28 ("Are the daily problems which you have to solve to fulfill your tasks (1) very much the same to (5) usually very different").

The indicator variables x_4, x_5, x_6 for goal specificity ξ_2 are A45 ("Are the performance requirements set by your superior officer (1) very unclear to (5) very clear"), A55A ("Are the goals set for your work always clear and specific" ranging from (1) not true at all to (5) almost always true), and B09 ("Are the specific goals for your work (1) very unclear to (5) very clear").

The factor analytic model connecting the indicator variables x_1, x_2, x_3 with ξ_1 and x_4, x_5, x_6 with ξ_2 is:

TABLE 6

Averages of 100 Monte Carlo Replications

Interaction Model, Sample Size = 500

Gibbs: Cycles = 2000, Burn-In = 500

Metropolis: Cycles = 20

Parameter	True	Median	Mean	Std.Dev.	Cover
γ_{11}	0.8	0.7969	0.7975	0.0505	0.8500
γ_{12}	1.7	1.7153	1.7188	0.0941	0.9100
γ_{13}	0.5	0.4983	0.4997	0.0616	0.9000
Φ_{11}	1.2	1.1997	1.2030	0.0952	0.9000
Φ_{21}	0.1	0.1073	0.1079	0.0473	0.9300
Φ_{22}	0.7	0.6858	0.6889	0.0729	0.8500
Ψ_{11}	0.6	0.6429	0.6459	0.0779	0.9100
λ_{11}	1.0	1.0000	1.0000	—	1.0000
λ_{21}	0.7	0.7010	0.7010	0.0331	0.8900
λ_{31}	-0.5	-0.4966	-0.4966	0.0310	0.8900
λ_{42}	1.0	1.0000	1.0000	—	1.0000
λ_{52}	1.6	1.6363	1.6363	0.0903	0.8600
$\Theta_{\delta 11}$	0.2	0.1995	0.2006	0.0375	0.9100
$\Theta_{\delta 22}$	0.3	0.3026	0.3034	0.0267	0.9000
$\Theta_{\delta 33}$	0.4	0.3927	0.3938	0.0275	0.8500
$\Theta_{\delta 44}$	0.5	0.5013	0.5027	0.0405	0.8900
$\Theta_{\delta 55}$	0.6	0.5759	0.5773	0.0714	0.8700

$$\mathbf{x}_i = \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi}_i + \boldsymbol{\delta}_i \tag{104}$$

with $\boldsymbol{\delta}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta}_\delta)$ and the parameter specification

$$\boldsymbol{\nu}_x = \begin{pmatrix} \nu_{x,1} \\ \nu_{x,2} \\ \nu_{x,3} \\ \nu_{x,4} \\ \nu_{x,5} \\ \nu_{x,6} \end{pmatrix}, \boldsymbol{\Lambda}_x = \begin{pmatrix} 1 & 0 \\ \lambda_{x,21} & 0 \\ \lambda_{x,31} & 0 \\ 0 & 1 \\ 0 & \lambda_{x,52} \\ 0 & \lambda_{x,62} \end{pmatrix}, \boldsymbol{\Theta}_\delta = \operatorname{diag}\{\Theta_{\delta,11}, \ldots, \Theta_{\delta,66}\}. \tag{105}$$

Again, the values of $\lambda_{x,11}$ and $\lambda_{x,42}$ are set to 1 to fix the scales of ξ_1 and ξ_2. The proper priors for this part of the model are specified as: $\nu_{x,k} \sim \mathcal{N}(\nu_{x,k}^*, \Omega_{\nu,x,k})$ where $\nu_{x,k}^* = 0$, and $\Omega_{\nu,x,k}^{-1} = 0.01$ for $k = 1, \ldots, 6$; $\lambda_{x,k1} \sim \mathcal{N}(\lambda_{x,k1}^*, \Omega_{\lambda,x,k1})$ where $\lambda_{x,k1}^* = 1$, and $\Omega_{\lambda,x,k1}^{-1} = 0.01$ for $k = 2, 3$; $\lambda_{x,k2} \sim \mathcal{N}(\lambda_{x,k2}^*, \Omega_{\lambda,x,k2})$ where $\lambda_{x,k2}^* = 1$, and $\Omega_{\lambda,x,k2}^{-1} = 0.01$ for $k = 5, 6$; $\Theta_{\delta,kk} \sim \mathcal{IG}(a_{\delta,kk}, b_{\delta,kk})$ with $a_{\delta,kk} = 1/2$, $b_{\delta,kk}^{-1} = 0.01$ for $k = 1, \ldots, 6$.

The analysis has been performed with a burn in phase of 500 Gibbs iterations and 2000 additional Gibbs iterations and 50 Metropolis-Hastings iterations within each Gibbs cycle to estimate the posterior distribution. To check convergence we have also used 20000 Gibbs cycles and have found identical results.

We first report in Table 7 median, mean and standard deviation of the posterior distributions of the individual parameters of a linear regression model (without the interaction term $\xi_1\xi_2$) as well as of the non-linear regression model including the interaction term with the additional parameter γ_{13}.

While most of the medians and means of most parameters show close agreement, the values of γ_{11} and γ_{12} are somewhat higher in the non-linear regression model. The sign of the interaction parameter γ_{13} is negative, the standard deviation indicates that the interaction term may be considered as significantly different from 0 using a conventional 95 percent coverage. The median coefficient of determination R^2 increases from 0.1016 in the linear model to 0.2750 in the non-linear model indicating a higher predictive power of the nonlinear model. Substantively, this result implies that increases in task complexity and in goal specificity predict an increase in performance, but the increase is dampened by the negative sign of the interaction term. For the interaction parameter γ_{13} we plot the smoothed (kNN) density estimate in Figure 4 and the sample path of the MCMC algorithm in Figure 5.

In Table 8 the results are reported for the factor analytic model for (y_1, y_2, y_3) as indicators for the performance η.

The results agree closely for the linear and the non-linear regression model implying that the relations in the factor analytic model for the performance are not changed by the underlying regression model for η. Similar results hold for the factor analytic model connecting (x_1, x_2, x_3) to task complexity ξ_1 and (x_4, x_5, x_6) to goal specificity ξ_2.

7. Conclusion

We have formulated a nonlinear latent variable model including quadratic functions and interactions of latent regressors. Estimation of model parameters using conventional covariance structure analysis is almost impossible yielding unwieldy mean and covariance structures and cumbersome numerical algorithms. Putting the nonlinear latent variable model in a Bayesian framework allows the use of the Gibbs sampler to estimate the posterior distributions of the parameters and the latent variables given the data. The Bayesian framework also allows the incorporation of prior information by specifying prior distributions for the model parameters.

The application of the Gibbs sampler is complicated by the fact that no closed mathematical form of the distribution of the latent variables could be found. Therefore, the M-H algorithm is used within iterations of the Gibbs sampler. The proposed model and the numerical algorithms are illustrated by two simulation studies and an empirical example analyzing the possibly non-linear dependence of performance on task complexity and goal specificity. In the future, further checks of the behavior of the proposed algorithm and an analysis of the sensitivity of the results to the specification of strong priors will have to be performed.

TABLE 7

Linear and Non-linear Regression Model

Gibbs: Cycles = 2000, Burn-In = 500 Metropolis: Cycles = 50

Parameter	Linear regression model			Non-linear regression model		
	Median	Mean	Std.Dev.	Median	Mean	Std.Dev.
γ_{11}	0.0921	0.0931	0.0540	0.1323	0.1336	0.0619
γ_{12}	0.3649	0.3745	0.2078	0.6099	0.6304	0.2382
γ_{13}	—	—	—	-0.4131	-0.4363	0.1979
Φ_{11}	1.2575	1.2716	0.2506	1.1491	1.1685	0.2321
Φ_{21}	-0.0164	-0.0153	0.0525	-0.0233	-0.0239	0.0538
Φ_{22}	0.1827	0.1950	0.0687	0.1983	0.2072	0.0699
Ψ_{11}	0.3403	0.3438	0.0760	0.2902	0.2961	0.0695
R^2_η	0.1016	0.1129	0.0706	0.2750	0.2797	0.1157

TABLE 8

Factor Analytic Model for Performance

Gibbs: Cycles = 2000, Burn-In = 500 Metropolis: Cycles = 50

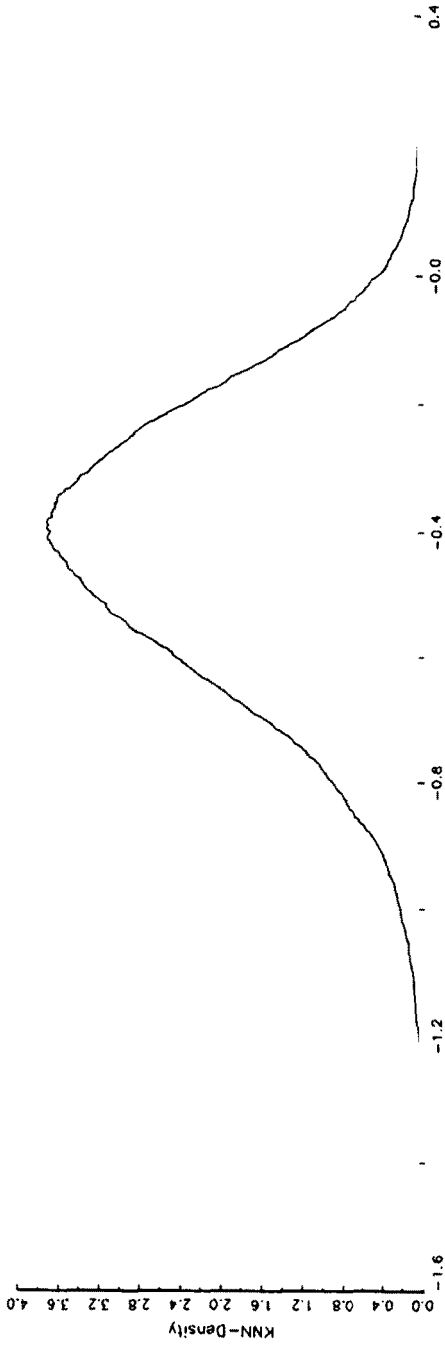| Parameter	Linear regression model			Non-linear regression model		
	Median	Mean	Std.Dev.	Median	Mean	Std.Dev.
$\nu_{y,1}$	4.3197	4.3215	0.0652	4.3191	4.3174	0.0718
$\nu_{y,2}$	4.3534	4.3527	0.0574	4.3521	4.3505	0.0623
$\nu_{y,3}$	4.2219	4.2211	0.0583	4.2194	4.2158	0.0620
$\lambda_{y,11}$	1.0000	1.0000	—	1.0000	1.0000	—
$\lambda_{y,21}$	0.9040	0.9174	0.1211	0.8839	0.8915	0.1019
$\lambda_{y,31}$	0.9403	0.9522	0.1181	0.9049	0.9102	0.1004
$\Theta_{\epsilon,11}$	0.3135	0.3158	0.0525	0.3015	0.3028	0.0482
$\Theta_{\epsilon,22}$	0.2063	0.2070	0.0409	0.2037	0.2050	0.0356
$\Theta_{\epsilon,33}$	0.1821	0.1819	0.0411	0.1883	0.1902	0.0359

FIGURE 4.

Smooth estimate of the posterior density of γ_{13} of the empirical example.
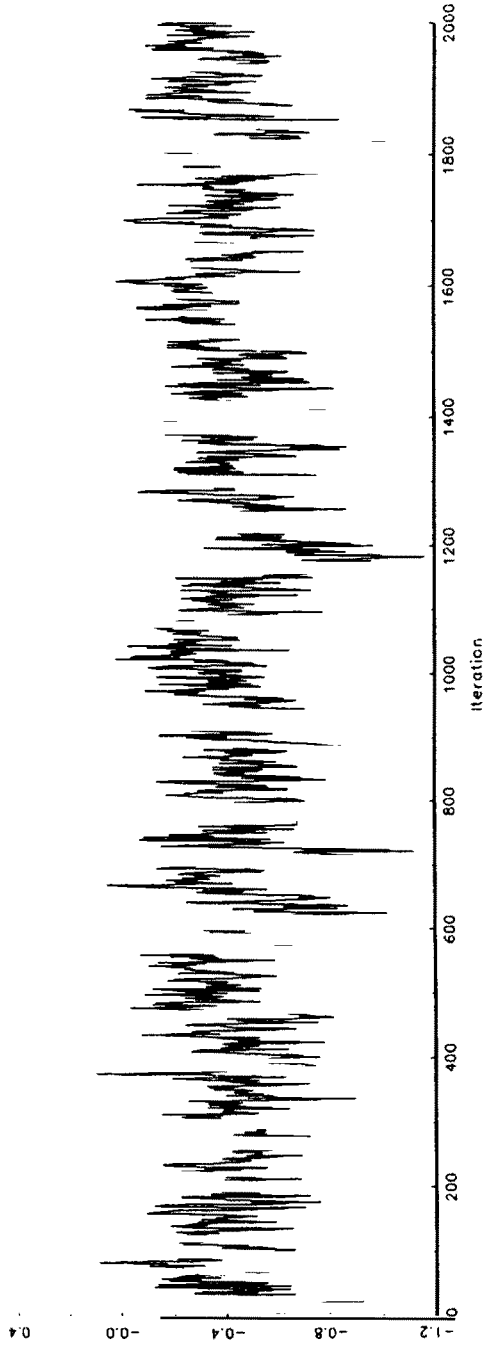
FIGURE 5.
Sample path of MCMC iterations for γ_{13} of the empirical example.

TABLE 9

Factor Analytic Model for Task Complexity and Goal Specificity

Gibbs: Cycles = 2000, Burn-In = 500 Metropolis: Cycles = 50

Parameter	Linear regression model			Non-linear regression model		
	Median	Mean	Std.Dev.	Median	Mean	Std.Dev.
$\nu_{x,1}$	2.6146	2.6195	0.1188	2.6137	2.6119	0.1161
$\nu_{x,2}$	2.8971	2.9050	0.1103	2.8950	2.8920	0.1106
$\nu_{x,3}$	2.7570	2.7588	0.0924	2.7491	2.7496	0.0929
$\nu_{x,4}$	4.0758	4.0735	0.0694	4.0737	4.0735	0.0690
$\nu_{x,5}$	4.3325	4.3305	0.0566	4.3302	4.3310	0.0582
$\nu_{x,6}$	3.8182	3.8163	0.0845	3.8135	3.8147	0.0893
$\lambda_{x,11}$	1.0000	1.0000	—	1.0000	1.0000	—
$\lambda_{x,21}$	1.0117	1.0272	0.1348	1.0985	1.1060	0.1325
$\lambda_{x,31}$	0.6590	0.6622	0.0886	0.6735	0.6795	0.0914
$\lambda_{x,42}$	1.0000	1.0000	—	1.0000	1.0000	—
$\lambda_{x,52}$	0.8117	0.8413	0.2813	0.9112	0.9312	0.2716
$\lambda_{x,62}$	1.5850	1.7142	0.6228	1.2884	1.3140	0.3287
$\Theta_{\delta,11}$	0.7218	0.7259	0.1553	0.8099	0.8127	0.1412
$\Theta_{\delta,22}$	0.4570	0.4457	0.1512	0.3443	0.3524	0.1374
$\Theta_{\delta,33}$	0.7287	0.7361	0.1059	0.7516	0.7585	0.1054
$\Theta_{\delta,44}$	0.5357	0.5396	0.0915	0.5092	0.5110	0.0796
$\Theta_{\delta,55}$	0.3742	0.3694	0.0750	0.3377	0.3386	0.0615
$\Theta_{\delta,66}$	0.6961	0.6330	0.2812	0.8334	0.8357	0.1369

References

Anderson, T. W. (1984). *An Introduction to multivariate statistical analysis* (2nd ed.). New York: Wiley.

Arminger, G., & Schoenberg, R. J. (1989). Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models. *Psychometrika, 54*, 409–425.

Arminger, G., Wittenberg, J., & Schepers, A. (1996). *MECOSA3 User Guide.* Friedrichsdorf, Germany: ADDITIVE GmbH.

Arnold, S. F. (1993). Gibbs Sampling. In C. R. Rao (Ed.), *Handbook of statistics* (Vol. 9, pp. 599–625). Amsterdam: North Holland.

Besag, J., Green, P., Higdon, D., & Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science, 19*(1), 3–66.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis.* Reading: Addison-Wesley.

Browne, M. W. (1984). Asymptotic distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62–83.

Carlin, B. P., & Louis, T. A. (1996). *Bayes and empirical bayes methods for data analysis*. London: Chapman & Hall.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistican, 46*, 167–174.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistican, 49*(4), 327–335.

Chib, S., & Greenberg, E. S. (1996). Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory, 12*(3), 409–431.

Early, P. C., Lee, C., & Hanson, L. A. (1990). Joint moderating effects of job experience and task component complexity: relations among goal setting, task strategies, and performance. *Journal of Organizational Behavior, 11*, 3–15.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85*, 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman & Hall.

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences (with discussion). *Statistical Science, 7*, 457–511.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721–741.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*, London: Chapman and Hall.

Härdle, W. (1990). *Applied nonparametric regression*. Cambridge, MA: Cambridge University Press.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*, 97–109.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances*. Baltimore, MD: Johns Hopkins Press.

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

Holling, H. (1995). *Goal setting and performance*. Unpublished manuscript, Universität Münster, Department of Psychology, Germany.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ, Lawrence Earlbaum Associates.

Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *Advanced structural equation modeling techniques* (pp. 57–88). Hillsdale, NJ, Lawrence Erlbaum Associates.

Kenny, D., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96*, 201–210.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworth.

Lee, S. Y. (1991). A Bayesian Approach to Confirmatory Factor Analysis. *Psychometrika, 46*, 153–160.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society, Series B, 34*, 1–41.

Loftsgaarden, D. O., & Quesenberry, G. P. (1965). A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics, 36*, 1049–1051.

MacEachern, S. N., & Berliner, M. L. (1994). Subsampling the Gibbs sampler. *The American Statistican, 48*, 188–190.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics, 21*, 1087–1091.

Müller, P. (1994). Metroplis Based Posterior Integration Schemes. Unpublished manuscript, Duke University, Durham.

Muthén, B. O. (1988). *LISCOMP—Analysis of linear structural equations with a comprehensive measurement model*. Mooresville: Scientific Software.

Press, S. J., & Shigemasu, K. (1989). Bayesian inference in factor analysis. In L. Gleser, M. D. Perlman, S. J. Press, & A. R. Sampson (Eds.), *Contributions to probability and statistics* (pp. 271–287). New York: Springer Verlag.

Ritter, C., & Tanner, M. A. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sampler. *Journal of the American Statistical Association, 87*, 861–868.

Rubin, D. B. (1991). EM and Beyond. *Psychometrika, 56*, 241–254.

Tanner, M. A. (1993). *Tools for statistical inference* (2nd ed.). Heidelberg and New York: Springer Verlag.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with Discussion). *Annals of Statistics, 22*, 1701–1762.