samsi

NSF • Duke • NCSU • UNC • NISS

Bayesian Structural Equation Modeling

David B. Dunson, Jesus Palomo, and Ken Bollen

Technical Report #2005-5
July 27, 2005

This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Statistical and Applied Mathematical Sciences Institute
PO Box 14006
Research Triangle Park, NC 27709-4006
www.samsi.info

Bayesian Structural Equation Modeling

David B. Dunson[1,*], Jesus Palomo[2], and Ken Bollen[3]

[1]Biostatistics Branch

MD A3-03, National Institute of Environmental Health Sciences,

P.O. Box 12233, Research Triangle Park, NC 27709

[*]*email*: dunson1@niehs.nih.gov

[2]Statistics and Applied Mathematical Sciences Institute (SAMSI)

[3]Department of Sociology, University of North Carolina at Chapel Hill

Abstract

Structural equation models (SEMs) with latent variables are routinely used in social science research, and are of increasing importance in biomedical applications. Standard practice in implementing SEMs relies on frequentist methods. This chapter provides a simple and concise description of an alternative Bayesian approach. We provide a brief overview of the literature, describe a Bayesian specification of SEMs, and outline a Gibbs sampling strategy for model fitting. Bayesian inferences are illustrated through an industrialization and democratization case study from the literature. The Bayesian approach has some distinct advantages, due to the availability of samples from the joint posterior distribution of the model parameters and latent variables, that we highlight. These posterior samples provide important information not contained in the measurement and structural parameters. As is illustrated using the case study, this information can often provide valuable insight into structural relationships.

1 Introduction

Structural equation models (SEMs) with latent variables provide a very general framework for modeling of relationships in multivariate data (Bollen, 1989). Although SEMs are most commonly used in studies involving intrinsically latent variables, such as happiness, quality of life, or stress, they also provide a parsimonious framework for covariance structure modeling. For this reason, they have become increasingly used outside of the traditional social science applications.

Software available for routine fitting of SEMs, including LISREL (Jöreskog and Sörbom, 1996), MPLUS (Muthén and Muthén, 1998, 2003) and BMDP (Bentler, 1992), rely on frequentist methods. Most commonly, SEMs are fitted using either full information maximum likelihood estimation (Jöreskog and Sörbom, 1985) or generalized least squares procedures (Browne, 1974). Such methods can easily allow mixtures of continuous and ordered categorical observed variables by using an underlying variable structure (Muthén, 1984; Arminger and Küsters, 1988). Recent research has developed extensions to allow interactions and nonlinear structures (Jöreskog and Yang, 1996; Bollen and Paxton, 1998; Wall and Amemiya, 2000). Frequentist inferences are typically based on point estimates and hypothesis tests for the measurement and latent variable parameters, marginalizing at the latent variables.

Although the overwhelming majority of the literature on SEMs is frequentist in nature, Bayesian approaches have been proposed by a number of authors. For factor models, which are a special case of SEMs, there is a long history of Bayesian methods (see, for example, Martin and McDonald, 1975; Lee, 1981; Ansari and Jedidi, 2000; Lopes and West, 2004). For more general SEMs, early work was done by Bauwens (1984) and Lee (1992). Recent articles have focused on the use of Markov chain Monte Carlo (MCMC) methods to implement Bayesian analysis in complex cases, involving nonlinear structures (Arminger and Muthén, 1998; Lee and Song, 2004), heterogeneity (Ansari, Jedidi and Jagpal, 2000; Lee and Song, 2003), and multilevel data (Dunson, 2000; Jedidi and Ansari, 2001). In addition, Raftery (1993) considers the important problem of model selection in SEMs from a Bayesian perspective. Additional articles on Bayesian SEMs have been published by Scheines, Holjtink and Boomsma (1999) and Lee and Shi (2000).

The goal of this chapter is not to review all of these approaches, but instead to provide an easily accessible overview of a Bayesian approach to SEMs, illustrating some of the advantages over standard frequentist practice. Due to the flexibility of the Bayesian approach, it is straightforward to apply the method in a very broad class of SEM-type modeling frameworks, allowing nonlinearity, interactions, missing data, mixed categorical, count, and continuous observed variables, etc. The WinBUGS software package[1], which is freely available, can be used to implement Bayesian SEM analysis.

There are several important differences between the Bayesian and frequentist approaches, which will be highlighted. First, the Bayesian approach requires the specification of prior distributions for each of the model unknowns, including the latent variables and the parameters from the measurement and structural models. Frequentists typically assume Gaussian distributions for the latent variables, but do not specify priors for mean or covariance parameters[2]. Because the posterior distributions upon which Bayesian inferences are based depend both on the prior distribution and the likelihood of the data, the prior plays an important role. In particular, specification of the prior allows for the incorporation of substantive information about structural relationships, which may be available from previous studies or social science theory. In the absence of such information, vague priors can be chosen. As the sample size increases, the posterior distribution will be driven less by the prior, and

[1] `www.mrc=bsu.cam.ac.uk/bugs/`.

[2] Researchers can incorporate observed variables that come from distributions with excess kurtosis by using corrected likelihood ratio tests, bootstrapping methods, or sandwich estimators for asymptotic standard errors (Satorra and Bentler, 1988; Bollen and Stine, 1990, 1993).

frequentist and Bayesian estimates will tend to agree closely.

A second difference is computational. Bayesian model fitting typically relies on MCMC, which involves simulating draws from the joint posterior distribution of the model unknowns (parameters and latent variables) through a computationally intensive procedure. The advantage of MCMC is that there is no need to rely on large sample assumptions (e.g., asymptotic normality), because exact posterior distributions can be estimated for any functional of the model unknowns. In small to moderate samples, these exact posteriors can provide a more realistic measure of model uncertainty, reflecting asymmetry and not requiring the use of a delta method or other approximations. The downside is that it may take a long time (e.g., several hours) to obtain enough samples from the posterior so that Monte Carlo (MC) error in posterior summaries is negligible. This is particularly true in SEMs, because there can be problems with slow mixing producing high autocorrelation in the MCMC samples. This autocorrelation, which can be reduced greatly through careful parametrization or computation tricks (e.g., blocking and parameter expansion), makes it necessary to collect more samples to produce an acceptable level of MC error.

An additional benefit that is gained by paying this computational price is that samples are available from the joint posterior distribution of the latent variables. Often, these samples can be used to obtain important insights into structural relationships, which may not be apparent from estimates (Bayesian or frequentist) of the structural parameters. This is certainly the case in the industrialization and democratization application (Bollen, 1989), which we will use to illustrate the concepts starting in Section 3.

Section 2 reviews the basic SEM modeling framework and introduces the notation. Section 3 describes the Bayesian approach, focusing on normal linear SEMs for simplicity in exposition, introduces the conditionally-conjugate priors for the parameters from the measurement and latent variable models, and outlines a simple Gibbs sampling algorithm for posterior computation. Section 4 applies the approach to the industrialization and democratization case study. Section 5 outlines generalizations to a broader class of SEMs. Section 6 contains a discussion, including recommendations for important areas for future research.

2 Structural Equation Models

SEMs provide a broad framework for modeling of means and covariance relationships in multivariate data. Although extensions are straightforward, particularly taking a Bayesian approach, our focus here is on the usual normal linear SEM, which is often referred to as a linear structural relations or LISREL model. LISREL models generalize many commonly-used statistical models, including ANOVA, MANOVA, multiple linear regression, path analysis, and confirmatory factor analysis. Because SEMs are setup to model relationships among endogenous and exogenous latent variables, accounting for measurement error, they are routinely used in social science applications. Social scientists have embraced latent variable models, realizing that it is typically not possible to obtain one perfect measure of a trait of interest. In contrast, biomedical researchers and epidemiologists tend to collapse multiple items related to a latent variable, such as stress, into a single arbitrarily-defined score prior to analysis (Herring and Dunson, 2004).

In factor models, a vector of observed variables $\boldsymbol{Y}_i$ is considered to arise through random sampling from a multivariate normal distribution denoted by $N(\boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{f}_i, \boldsymbol{\Sigma})$, where $\boldsymbol{f}_i$ is the vector of latent variables; $\boldsymbol{\Lambda}$ is the factor loadings matrix describing the effects of the latent variables on the observed variables; $\boldsymbol{\nu}$ is the vector of intercepts and $\boldsymbol{\Sigma}$ is the covariance matrix. However, in SEMs the focus is also on studying relationships among factors. For this purpose, the distinction between the *measurement* model and *structural* (latent) model is common. The former specifies the relationships of the latent to the observed variables, whereas the latter specifies the relationships among the latent variables. Following the standard LISREL notation, as in Bollen (1989) and Jöreskog and Sörbom (1996), the measurement model is, for $i = 1, \dots, N$ observations,

$$\boldsymbol{y}_i = \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta}_i + \boldsymbol{\delta}_i^y, \tag{1a}$$

$$\boldsymbol{x}_i = \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi}_i + \boldsymbol{\delta}_i^x, \tag{1b}$$

where model (1a) relates the vector of indicators $\boldsymbol{y}_i = (y_{i1}, \dots, y_{ip})'$ to an underlying m-vector of latent variables $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{im})'$, $m \leq p$, through the $p \times m$ factor loadings matrix $\boldsymbol{\Lambda}_y$. Similarly, (1b) relates $\boldsymbol{x}_i = (x_{i1}, \dots, x_{iq})'$ to an n-vector of latent variables $\boldsymbol{\xi}_i = (\xi_{i1}, \dots, \xi_{in})'$, $n \leq q$, through the $q \times n$ matrix $\boldsymbol{\Lambda}_x$. The vectors $\boldsymbol{\delta}_i^y$ and $\boldsymbol{\delta}_i^x$ are the measurement error terms, with dimensions $p \times 1$ and $q \times 1$, respectively. The vectors $\boldsymbol{\nu}_y$, $p \times 1$, and $\boldsymbol{\nu}_x$, $q \times 1$, are the intercept terms of the measurement models.

In equations (1a) and (1b), it is assumed that the observed variables are continuous. However, as in Muthén (1984), the model remains valid for categorical or censored observed variables $(\boldsymbol{y}_i, \boldsymbol{x}_i)$ since they can be linked to their underlying continuous counterparts $(\boldsymbol{y}_i^*, \boldsymbol{x}_i^*)$ through a threshold model. Potentially, one can also define separate generalized linear models for each of the observed variables in the measurement model (as in Sammel, Ryan, Legler, 1997; Moustaki and Knott, 2000; Dunson, 2000; 2003) to allow a broader class of measurement models.

On the other hand, the *structural (latent variable) model* is focused on studying the relationships among latent variables, $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. This is performed by regressing the dependent vector, $\boldsymbol{\eta}$, on the explanatory vector $\boldsymbol{\xi}$ as follows, $i = 1, \dots, N$,

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \ \boldsymbol{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\zeta}_i \tag{2}$$

where the $m \times m$ matrix $\boldsymbol{B}$ describes the relationships among latent variables in $\boldsymbol{\eta}_i$. Clearly, the elements of the diagonal of $\boldsymbol{B}$ are all zero. The $m \times n$ matrix $\boldsymbol{\Gamma}$ quantifies the influence of $\boldsymbol{\xi}_i$ on $\boldsymbol{\eta}_i$. The $m \times 1$ vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}_i$ represent the intercept and the unexplained parts of $\boldsymbol{\eta}_i$ respectively.

Under this parametrization, common assumptions in SEMs are: i) the elements of $\boldsymbol{\xi}_i$ and $\boldsymbol{\zeta}_i$ are independent and normally distributed, $\boldsymbol{\xi}_i \sim N_n(\boldsymbol{\mu}_\xi, \boldsymbol{\Omega}_\xi)$, $\boldsymbol{\Omega}_\xi = diag(\omega^2_{\xi 1}, \ldots, \omega^2_{\xi n})$, and $\boldsymbol{\zeta}_i \sim N_m(\mathbf{0}, \Omega_\zeta)$, $\boldsymbol{\Omega}_\zeta = diag(\omega^2_{\zeta 1}, \ldots, \omega^2_{\zeta n})$; ii) the measurement error vectors $\boldsymbol{\delta}^y_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_y)$, $\boldsymbol{\Sigma}_y = diag(\sigma^2_{1y}, \ldots, \sigma^2_{py})$, and $\boldsymbol{\delta}^x_i \sim N_q(\mathbf{0}, \boldsymbol{\Sigma}_x)$, $\boldsymbol{\Sigma}_x = diag(\sigma^2_{1x}, \ldots, \sigma^2_{qx})$ are assumed independent; and iii) $\boldsymbol{\delta}' = (\boldsymbol{\delta}^{y\prime}, \boldsymbol{\delta}^{x\prime})$, $Cov(\boldsymbol{\zeta}, \boldsymbol{\delta}') = \mathbf{0}$, $Cov(\boldsymbol{\xi}, \boldsymbol{\delta}') = \mathbf{0}$, $Cov(\boldsymbol{\xi}, \boldsymbol{\zeta}') = \mathbf{0}$, and $(\boldsymbol{I} - \boldsymbol{B})$ is nonsingular. In addition, some constraints need to be placed on $\boldsymbol{\Lambda}_x$ and $\boldsymbol{\Lambda}_y$ for identifiability.

3 Bayesian Approach

Instead of relying on point estimates (MLEs, least squares, etc) and asymptotically-justified confidence bounds and test statistics, the Bayesian approach we describe bases inferences on exact posterior distributions for the parameters and latent variables estimated by Markov chain Monte Carlo. As sample sizes increase, Bayesian and frequentist estimators of the parameters should converge. However, an appealing feature of the Bayesian approach is that posterior distributions are obtained not only for the parameters, but also for the latent variables. Although the posterior distribution for the latent variables is shrunk back towards the normal prior, lack of fit can be captured, including non-normality, non-linearity, and relationships that are not immediately apparent from the parameter estimates. Although frequentist two-stage approaches that fit the measurement model first and then compute factor scores can similarly be used to capture lack of fit, estimates are biased and measures of uncertainty in the factors scores are difficult to obtain (Croon and Bolck, 1997).

In contrast, the Bayesian approach yields estimates of the exact joint posterior distribution of the latent variables. This posterior distribution can be used flexibly to, for example,

1. Obtain point and interval estimates for the factor scores of each individual.

2. Formally compare the factor scores for different subjects (e.g., through a posterior

probability that the score is higher for a particular subject).

3. Assess whether a particular subject's factor score has changed over time.

4. Identify outlying subjects in the tails of the latent variable distribution.

5. Assess relationships that may not be fully captured by the basic modeling structure (e.g., is the association between latent traits linear and apparent across the range of factor scores or predominantly due to the more extreme individuals?)

Potentially, one could use a richer model that allows non-linear and more complex relationships among the latent variables. However, it is often not apparent *a priori* how such relationships should be specified, and important insights can be obtained through careful examination of posterior distributions of the latent variables obtained under a simple LISREL model.

3.1 Specification

The Bayesian model requires the specification of a full likelihood and prior distributions for the parameters. The complete data likelihood, including the latent variables, has the following form:

$$\mathcal{L}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{\Theta}) = \prod_{i=1}^{N} \Big\{ N_p\left(\boldsymbol{y}_i; \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta}_i, \boldsymbol{\Sigma}_y\right) N_q\left(\boldsymbol{x}_i; \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi}_i, \boldsymbol{\Sigma}_x\right) \times \\ \times N_m\left(\boldsymbol{\eta}_i; \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i, \boldsymbol{\Omega}_\zeta\right) N_n(\boldsymbol{\xi}_i; \boldsymbol{\mu}_\xi, \boldsymbol{\Omega}_\xi) \Big\}$$

where $\boldsymbol{\Theta} = \left(\boldsymbol{\alpha}, \boldsymbol{b}, \boldsymbol{\gamma}, \boldsymbol{\nu}_y, \boldsymbol{\nu}_x, \boldsymbol{\lambda}_y, \boldsymbol{\lambda}_x, \boldsymbol{\sigma}_y^2, \boldsymbol{\sigma}_x^2, \boldsymbol{\omega}_\zeta^2, \boldsymbol{\mu}_\xi, \boldsymbol{\omega}_\xi^2\right)$ is the vector of model parameters. Here, the lower case bold letters denote that only the free elements are included in the parameter vector $\boldsymbol{\Theta}$, with the remaining elements being fixed in advance in the model specification process.

To complete a Bayesian specification of the model, we choose priors for each of the parameters in $\boldsymbol{\Theta}$. For convenience in elicitation and computation, we choose normal or truncated normal priors for the free elements of the intercept vectors, $\boldsymbol{\nu}_y$, $\boldsymbol{\nu}_x$ and $\boldsymbol{\alpha}$, the factor loadings, $\boldsymbol{\lambda}_y$ and $\boldsymbol{\lambda}_x$, and the structural parameters $\boldsymbol{b}$ and $\boldsymbol{\gamma}$. For the variance component parameters, including the diagonal elements of $\boldsymbol{\Sigma}_y$, $\boldsymbol{\Sigma}_x$, $\boldsymbol{\Omega}_\zeta$ and $\boldsymbol{\Omega}_\xi$, we choose independent inverse-gamma priors (avoiding high variance priors for the latent variable variances, which have well known problems). The bounds on the truncated normal are chosen to restrict parameters that are known in advance to fall within a certain range. For example, positivity constraints are often appropriate and may be necessary for identifiability based on the data. It is important to distinguish between frequentist identifiability, which implies that all the model parameters can be estimated based on the data given sufficient sample size, and Bayesian identifiability, which implies Bayesian learning. In particular, Bayesian learning occurs when the posterior distributions can differ from the prior distributions, reflecting that we have updated our beliefs based on the current data. Potentially, one can choose informative prior distributions for the parameters in a model that is underidentified from a frequentist perspective, and still obtain Bayesian identifiability for unknowns of interest. However, we prefer to focus on models which are identified in a frequentist sense to avoid relying so strongly on the prior specification.

The joint posterior distribution for the parameters and latent variables is computed, following Bayes' rule, as

$$\pi(\boldsymbol{\Theta}, \boldsymbol{\xi}, \boldsymbol{\eta} | \boldsymbol{y}, \boldsymbol{x}) = \frac{\mathcal{L}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{\Theta})\pi(\boldsymbol{\Theta})}{\int \mathcal{L}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{\Theta})\pi(\boldsymbol{\Theta}) d\boldsymbol{\eta} d\boldsymbol{\xi} d\boldsymbol{\Theta}}, \tag{3}$$

which is simply the complete data likelihood multiplied by the prior and divided by a normalizing constant referred to as the marginal likelihood. Clearly, calculation of the marginal likelihood (the term in the denominator) is very challenging, because it typically involves a high dimensional integration of the likelihood over the prior distribution. Fortunately, MCMC techniques can be used to generate draws from the joint posterior distribution without need to calculate the marginal likelihood. For an overview of MCMC algorithms, refer to the recent books by Robert and Casella (2004), Gilks, Richardson, and Spiegelhalter (1996), Gamerman (1997) and Chen, Shao and Ibrahim (2000). Due to the conditionally normal linear structure of the SEM and to the choice of conditionally conjugate truncated normal and inverse-gamma priors for the parameters, MCMC computation can proceed through a straightforward Gibbs sampling algorithm, see Geman and Geman (1984) or Gelfand and Smith (1990) for more details.

3.2 Gibbs sampler

The Gibbs sampler is an MCMC technique that alternately samples from the full conditional posterior distributions of each unknown, or blocks of unknowns, including the parameters and latent variables. Before proceeding to the next step, the sampled parameter or group of parameters value is updated. Under mild regularity conditions, these samples converge to a stationary distribution, which is the joint posterior distribution. Hence, we can run the Gibbs sampler, discard a burn-in to allow convergence (diagnosed by trace plots and standard tests), and then calculate posterior summaries based on the collected samples. For illustration, we focus here on full conditional posterior distributions for the latent variables and structural parameters. Derivation of the conditional distribution for the remaining parameters follows simpler algebraic results and, in general, is not necessary since black-box sampling algorithms exist. For example, packages such as WinBUGS, Spiegelhalter et al. (2003), can automatically run Gibbs sampler algorithms based only on model specifications and priors beliefs.

We focus now on deriving the conditional posterior distributions for the latent variables, $\boldsymbol{\eta}_i$, $\boldsymbol{\xi}_i$, and the structural parameters $\boldsymbol{\alpha}$, $\boldsymbol{b}$, $\boldsymbol{\gamma}$. As introduced previously, MCMC methods use the joint posterior distribution (3) in terms of $\pi(\boldsymbol{\Theta}, \boldsymbol{\xi}, \boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{x}) \propto \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\eta}, \boldsymbol{\xi}; \boldsymbol{\Theta})\pi(\boldsymbol{\Theta})$. Based on this property and factoring the joint posterior, we compute the conditional posterior for the endogenous latent variable as follows

$$\pi(\boldsymbol{\eta}_i|\boldsymbol{\nu}_y, \boldsymbol{\Lambda}_y, \boldsymbol{\Sigma}_y, \widetilde{\boldsymbol{\mu}}_{\eta i}, \widetilde{\boldsymbol{\Omega}}_\eta,\ \boldsymbol{y}_i) \propto \pi(\boldsymbol{y}_i; \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta}_i, \boldsymbol{\Sigma}_y) \cdot \pi(\boldsymbol{\eta}_i; \widetilde{\boldsymbol{\mu}}_{\eta i}, \widetilde{\boldsymbol{\Omega}}_\eta)$$

with $\widetilde{\boldsymbol{\mu}}_{\eta i} = \boldsymbol{A}[\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\xi}_i]$, $\widetilde{\boldsymbol{\Omega}}_\eta = \boldsymbol{A}\ \Omega_\zeta \boldsymbol{A}'$ and $\boldsymbol{A} = [\boldsymbol{I}_{m\times m} - \ \boldsymbol{B}]^{-1}$. After straightforward computations it is distributed as $N_m\left(\widehat{\boldsymbol{\eta}}_i, \widehat{\boldsymbol{\Omega}}_\eta\right)$ with

$$\begin{aligned}
\widehat{\boldsymbol{\eta}}_i &= \widehat{\boldsymbol{\Omega}}_\eta \left[\boldsymbol{\Lambda}_y'\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{y}_i - \boldsymbol{\nu}_y) + \widetilde{\boldsymbol{\Omega}}_\eta^{-1}\widetilde{\boldsymbol{\mu}}_{\eta i}\right], \\
\widehat{\boldsymbol{\Omega}}_\eta^{-1} &= \boldsymbol{\Lambda}_y'\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Lambda}_y + \widetilde{\boldsymbol{\Omega}}_\eta^{-1}.
\end{aligned}$$

The conditional posterior for the exogenous latent variable is obtained as follows

$$\begin{aligned}
\pi(\boldsymbol{\xi}_i|\boldsymbol{\eta}_i, \boldsymbol{\Omega}_\zeta, \boldsymbol{\nu}_x, \boldsymbol{\lambda}_x, \boldsymbol{\Sigma}_x, \boldsymbol{\alpha}, \boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\mu}_\xi, \boldsymbol{\Omega}_\xi, \boldsymbol{x}_i) &\propto \pi(\boldsymbol{x}_i; \boldsymbol{\nu}_x + \boldsymbol{\lambda}_x\boldsymbol{\xi}_i, \boldsymbol{\Sigma}_x)\times \\
&\times \pi(\boldsymbol{\eta}_i; \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta}_i - \boldsymbol{\Gamma}\boldsymbol{\xi}_i, \boldsymbol{\Omega}_\zeta)\pi(\boldsymbol{\xi}_i; \boldsymbol{\mu}_\xi, \boldsymbol{\Omega}_\xi)
\end{aligned}$$

which, after computations, is distributed as $N_n(\widehat{\boldsymbol{\xi}}_i, \widehat{\boldsymbol{\Omega}}_\xi)$ with

$$\begin{aligned}
\widehat{\boldsymbol{\xi}}_i &= \widehat{\boldsymbol{\Omega}}_\xi \left[\boldsymbol{\Lambda}_x'\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}_i - \boldsymbol{\nu}_x) + \boldsymbol{\Gamma}'\boldsymbol{\Omega}_\zeta^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\alpha} - \boldsymbol{B}\boldsymbol{\eta}_i) + \boldsymbol{\Omega}_\xi^{-1}\boldsymbol{\mu}_\xi\right], \\
\widehat{\boldsymbol{\Omega}}_\xi^{-1} &= \boldsymbol{\Lambda}_x'\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Lambda}_x + \boldsymbol{\Gamma}'\boldsymbol{\Omega}_\zeta^{-1}\boldsymbol{\Gamma} + \boldsymbol{\Omega}_\xi^{-1}.
\end{aligned}$$

The structural parameters have the following conditional posteriors:

- For the vector of intercepts,

$$\pi(\boldsymbol{\alpha}|\boldsymbol{B}, \boldsymbol{\eta}_i, \boldsymbol{\Gamma}, \boldsymbol{\xi}_i, \boldsymbol{\Omega}_\zeta, \boldsymbol{\mu}_\alpha, \boldsymbol{\Omega}_\alpha) \propto \prod_{i=1}^{N} \pi(\boldsymbol{\eta}_i; \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i, \boldsymbol{\Omega}_\zeta)\pi(\boldsymbol{\alpha}; \boldsymbol{\mu}_\alpha, \boldsymbol{\Omega}_\alpha)$$

  which is distributed as $N_m(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\Omega}}_\alpha)$ with

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}} &= \widehat{\boldsymbol{\Omega}}_\alpha \left[\boldsymbol{\mu}_\alpha'\boldsymbol{\Omega}_\alpha^{-1} + \sum_{i=1}^{N}(\boldsymbol{\eta}_i - \boldsymbol{B}\boldsymbol{\eta}_i - \boldsymbol{\Gamma}\boldsymbol{\xi}_i)'\boldsymbol{\Omega}_\zeta^{-1}\right], \\
\widehat{\boldsymbol{\Omega}}_\alpha^{-1} &= N\boldsymbol{\Omega}_\zeta^{-1} + \boldsymbol{\Omega}_\alpha^{-1}.
\end{aligned}$$

- For the coefficient b_{rj}, which measures the impact of $\boldsymbol{\eta}^j$ on $\boldsymbol{\eta}^r$, with $r, j = 1, \ldots, m$ and $b_{rr} = 0$,

$$\pi(b_{rj}|\boldsymbol{b}_{-rj}, \boldsymbol{\alpha}, \boldsymbol{\eta}_i, \boldsymbol{\Gamma}, \boldsymbol{\xi}_i, \boldsymbol{\Omega}_\zeta, \mu_b, \omega_b^2) \propto \prod_{i=1}^{N} \pi(\boldsymbol{\eta}_i; \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i, \boldsymbol{\Omega}_\zeta)\pi(b_{rj}; \mu_b, \omega_b^2)$$

which is distributed as $N(\widehat{b}_{rj}, \widehat{\omega}_b)$ with

$$\widehat{b}_{rj} = \widehat{\omega}_b \left[\frac{\mu_b}{\omega_b^2} + \sum_{i=1}^{N} \frac{\eta_i^j}{\omega_{\zeta^r}^2} \left(\eta_i^r - \alpha^r - \sum_{s=1}^{n} (\gamma_{rs} \cdot \xi_i^s) - \sum_{\substack{t=1 \\ t \neq j}}^{m} b_{rt} \cdot \eta_i^t \right) \right],$$
$$\widehat{\omega}_b^{-1} = \frac{\sum_{i=1}^{N} (\eta_i^j)^2}{\omega_{\zeta^r}^2} + \frac{1}{\omega_b^2}.$$

- For the coefficient γ_{rj}, which measures the effect of $\boldsymbol{\xi}^j$ on $\boldsymbol{\eta}^r$, with $r = 1, \ldots, m$, $j = 1, \ldots, n$,

$$\pi(\gamma_{rj}|\boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{\gamma}_{-rj}, \boldsymbol{\eta}_i, \boldsymbol{\xi}_i, \boldsymbol{\Omega}_\zeta, \mu_\gamma, \omega_\gamma^2) \propto \prod_{i=1}^{N} \pi(\boldsymbol{\eta}_i; \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i, \boldsymbol{\Omega}_\zeta)\pi(\gamma_{rj}; \mu_\gamma, \omega_\gamma^2)$$

which is distributed as $N(\widehat{\gamma}_{rj}, \widehat{\omega}_\gamma)$ with

$$\widehat{\gamma}_{rj} = \widehat{\omega}_\gamma \left[\frac{\mu_\gamma}{\omega_\gamma^2} + \sum_{i=1}^{N} \frac{\xi_i^j}{\omega_{\zeta^r}^2} \left(\eta_i^r - \alpha^r - \sum_{s=1}^{m} (b_{rs} \cdot \eta_i^s) - \sum_{\substack{t=1 \\ t \neq j}}^{n} \gamma_{rt} \cdot \xi_i^t \right) \right],$$
$$\widehat{\omega}_\gamma^{-1} = \frac{\sum_{i=1}^{N} (\xi_i^j)^2}{\omega_{\zeta^r}^2} + \frac{1}{\omega_\gamma^2}.$$

Once all the full conditional posteriors are computed, the following Gibbs sampling algorithm can be implemented:

Given $\boldsymbol{\Theta}^0, \boldsymbol{\xi}^0, \boldsymbol{\eta}^0$

for (k = 1,...,# iterations)

for (i = 1,...,N)

Sample $\boldsymbol{\eta}_i^k \sim \pi(\boldsymbol{\eta}_i|\widetilde{\boldsymbol{\mu}}_{\eta i}^{k-1}, \widetilde{\boldsymbol{\Omega}}_{\eta}^{k-1}, \boldsymbol{\xi}_i^{k-1}, \boldsymbol{\Theta}^{k-1}, \boldsymbol{y}_i)$

Sample $\boldsymbol{\xi}_i^k \sim \pi(\boldsymbol{\xi}_i|\boldsymbol{\eta}_i^k, \boldsymbol{\Theta}^{k-1}, \boldsymbol{x}_i)$

Sample $\left(\boldsymbol{\nu}_y^k, \boldsymbol{\lambda}_y^k\right) \sim \pi\left(\boldsymbol{\nu}_y, \boldsymbol{\lambda}_y|\boldsymbol{\Theta}^{k-1}, \boldsymbol{\eta}^k, \boldsymbol{y}\right)$

Sample $\left(\boldsymbol{\nu}_x^k, \boldsymbol{\lambda}_x^k\right) \sim \pi\left(\boldsymbol{\nu}_x, \boldsymbol{\lambda}_x|\boldsymbol{\Theta}^{k-1}, \boldsymbol{\xi}^k, \boldsymbol{x}\right)$

Sample $\left(\boldsymbol{\alpha}^k,\ \boldsymbol{b}^k, \boldsymbol{\gamma}^k\right) \sim \pi\Big(\boldsymbol{\alpha}, \boldsymbol{b}, \boldsymbol{\gamma}|\boldsymbol{\eta}^k, \boldsymbol{\xi}^k, \boldsymbol{\Theta}^{k-1}, \boldsymbol{x}, \boldsymbol{y}\Big)$

Sample $(\boldsymbol{\sigma}_y^2)^k \sim \pi\left(\boldsymbol{\sigma}_y^2|\boldsymbol{\eta}^k, \boldsymbol{\nu}_y^k, \boldsymbol{\lambda}_y^k, \boldsymbol{\Theta}^{k-1}, \boldsymbol{y}\right)$

Sample $(\boldsymbol{\sigma}_x^2)^k \sim \pi\left(\boldsymbol{\sigma}_x^2|\boldsymbol{\xi}^k, \boldsymbol{\nu}_x^k, \boldsymbol{\lambda}_x^y, \boldsymbol{\Theta}^{k-1}, \boldsymbol{x}\right)$

Sample $\boldsymbol{\mu}_\xi^k \sim \pi\left(\boldsymbol{\mu}_\xi|\boldsymbol{\xi}^k,\ \boldsymbol{m}, \boldsymbol{M}\right)$

Sample $\boldsymbol{\Omega}_\xi^k \sim \pi\left(\boldsymbol{\Omega}_\xi|\boldsymbol{\xi}^k, \boldsymbol{a}_\xi, \boldsymbol{\beta}_\xi\right)$

Sample $\boldsymbol{\Omega}_\zeta^k \sim \pi\left(\boldsymbol{\Omega}_\zeta|\boldsymbol{\eta}^k, \boldsymbol{a}_\eta, \boldsymbol{\beta}_\eta\right)$

Output=$\left\{\boldsymbol{\eta}^k, \boldsymbol{\xi}^k, \boldsymbol{\Theta}^k\right\}$

Along with the benefits of Bayesian SEMs come the need to carefully consider certain computational issues. A particular concern is *slow mixing* of the MCMC algorithm, which can lead to very high autocorrelation in the samples and slow convergence rates. Parametrization has a large impact on computation in hierarchical models, including SEMs. For a given implied multivariate normal model, there is an equivalence class of SEMs having identical MLEs, but with different constraints made to ensure identifiability. The level of slow mixing can vary dramatically across SEMs in such an equivalence class, ranging from autocorrelation values near 1 to values near 0. Fortunately, it is easy to preselect an SEM in each equivalence class to limit slow mixing by choosing a *centered parametrization.* This simply involves incorporating free mean and variance parameters for each of the latent variables, with constraints instead incorporated in the intercepts and factor loadings in the measurement model. Following such a rule of thumb has a dramatic impact on computational efficiency without limiting inferences - one can always obtain posterior samples under a different parametrization by appropriately transforming draws obtained under the centered parametrization. In addition to centering, techniques that can be used to improve mixing include data augmentation or parameter expansion (Hills and Smith, 1992), updating parameters in blocks instead of one by one, and randomizing the order of updating (Liu, Wong and Kong, 1994; Roberts and Sahu, 1997). Techniques to determine the effective number of Gibbs samples necessary to produce a given level of precision in a posterior quantile of interest are available (Raftery and Lewis (1992). In addition, there are many tests to diagnose convergence of the Markov chain (cf., Brooks and Gelman, 1998; Brooks and Giudici, 2000).

4 Democratization and Industrialization Application

We will illustrate the Bayesian approach and highlight differences with frequentist methods using a democratization and industrialization example from the literature (Bollen, 1980, 1989). There has long been interest in studying relationships between industrialization in developing countries and democratization. To obtain insight into this relationship, our focus is on assessing whether industrialization level (IL) in Third World countries is positively associated with current and future political democracy level (PDL). The common political instabilities make these associations unclear. In the proposed model, it is assumed that some of the consequences of industrialization, for example societal wealth, an educated population, advances in living standards, etc, enhance the chances of democracy. To test this theory, measures of PDL (in 1960 and 1965) and IL indicators (in 1960) were collected in 75 developing countries. These include all developing countries, excluding micro-states, for which complete data were available.

Since political democracy refers to the extent of political rights and political liberties, we define a vector $\boldsymbol{y}$ of measures based on expert ratings: freedom of the press ($\boldsymbol{y}_1^{1960}$,$\boldsymbol{y}_5^{1965}$), freedom of group opposition ($\boldsymbol{y}_2^{1960}$,$\boldsymbol{y}_6^{1965}$), fairness of elections ($\boldsymbol{y}_3^{1960}$,$\boldsymbol{y}_7^{1965}$), and elective nature of the legislative body ($\boldsymbol{y}_4^{1960}$,$\boldsymbol{y}_8^{1965}$). Each of the rates were arbitrarily linearly transformed to the scale $[0, 10]$.

Industrialization is defined as the degree to which a society's economy is characterized by mechanized manufacturing processes, and the following vector of indicators $\boldsymbol{x}$ is compiled for consideration: gross national product per capita ($\boldsymbol{x}_1^{1960}$), inanimate energy consumption per capita ($\boldsymbol{x}_2^{1960}$) and the percentage of the labor force in industry ($\boldsymbol{x}_3^{1960}$). For simplicity in the notation, we will hereafter remove the superscripts indicating the year.

The data collected are summarized in Table 1 and plotted in Figure 1.

Indicator	Min	1st Qu.	Median	Mean	3rd. Qu.	Max	Sd.
$\boldsymbol{y}_1$	1.250	2.900	5.400	5.465	7.500	10.000	2.623
$\boldsymbol{y}_2$	0	0	3.333	4.256	8.283	10.000	3.947
$\boldsymbol{y}_3$	0	3.767	6.667	6.563	10.000	10.000	3.281
$\boldsymbol{y}_4$	0	1.581	3.333	4.453	6.667	10.000	3.349
$\boldsymbol{y}_5$	0	3.692	5.000	5.136	7.500	10.000	2.613
$\boldsymbol{y}_6$	0	0	2.233	2.978	4.207	10.000	3.373
$\boldsymbol{y}_7$	0	3.478	6.667	6.196	10.000	10.000	3.286
$\boldsymbol{y}_8$	0	1.301	3.333	4.043	6.667	10.000	3.246
$\boldsymbol{x}_1$	3.784	4.477	5.075	5.054	5.515	6.737	0.733
$\boldsymbol{x}_2$	1.386	3.663	4.963	4.792	5.830	7.872	1.511
$\boldsymbol{x}_3$	1.002	2.300	3.568	3.558	4.523	6.425	1.406

Table 1: Summary of the Industrialization and Democratization data.

4.1 Model structure

We show in the path diagram of Figure 2 the assumed model, where, for the countries under study, the PDL in 1960 and 1965 is represented by $\boldsymbol{\eta}^{60}$ and $\boldsymbol{\eta}^{65}$ respectively, and the IL in 1960 is symbolized by $\boldsymbol{\xi}$. Following the convention, circles represent latent variables, the squares the observed variables and the arrows linear relations. The relations assumed imply that the IL in 1960, $\boldsymbol{\xi}$, affects the PDL both in 1960, $\boldsymbol{\eta}^{60}$, and 1965, $\boldsymbol{\eta}^{65}$, through the regression coefficients γ^{60} and γ^{65} respectively. The impact of the PDL in 1960 on the level in 1965 is rep-

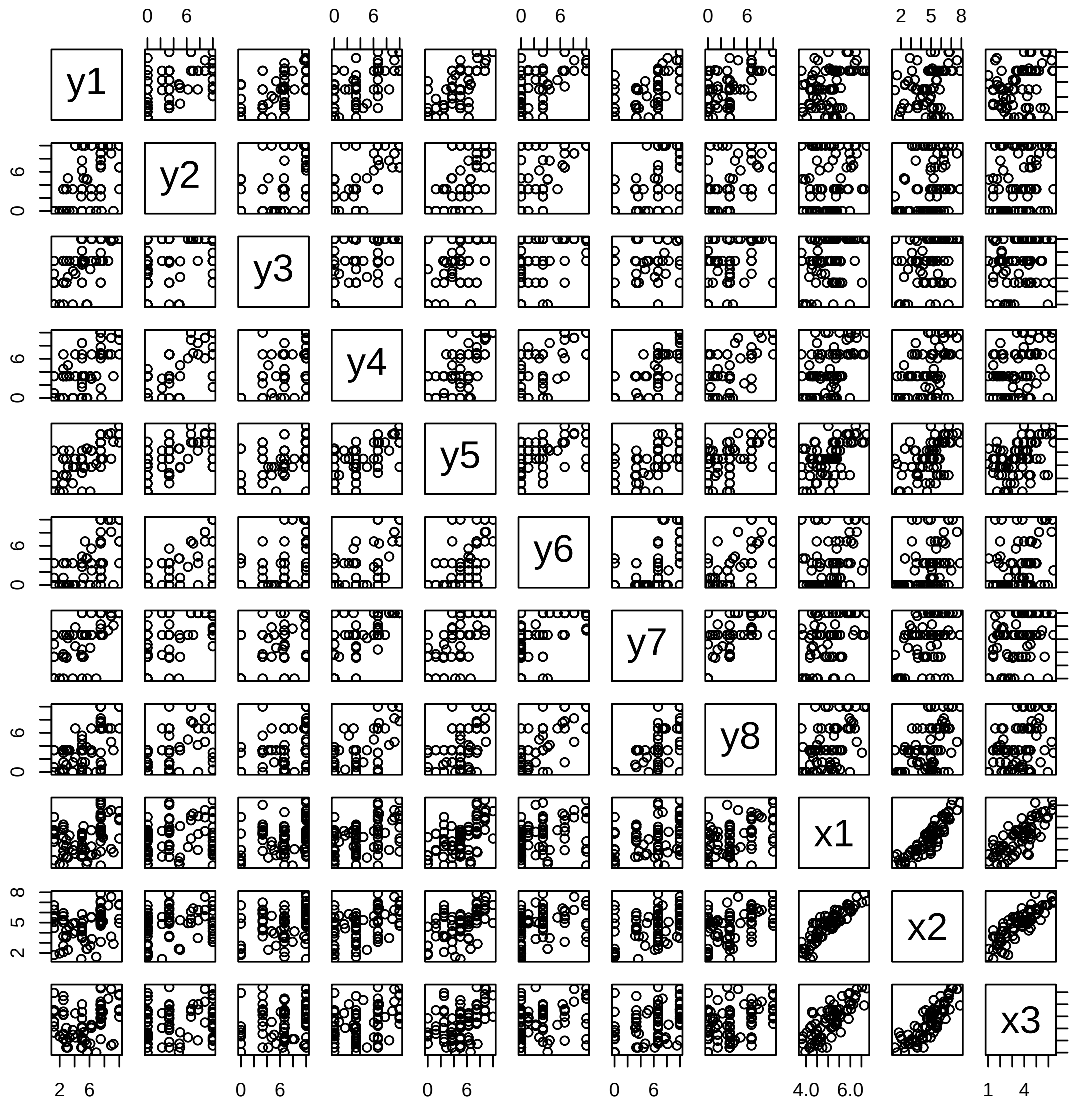

Figure 1: Industrialization and Democratization data.

resented by the arrow b_{21}, and the pseudo-latent variables. $(\boldsymbol{D}^{15}, \boldsymbol{D}^{24}, \boldsymbol{D}^{26}, \boldsymbol{D}^{37}, \boldsymbol{D}^{48}, \boldsymbol{D}^{68})$ are used to represent the correlation among the errors in the ratings that were elicited by the same expert in two points of the time.

For the i-th country, the latent variable model, as introduced in (2), is now formulated

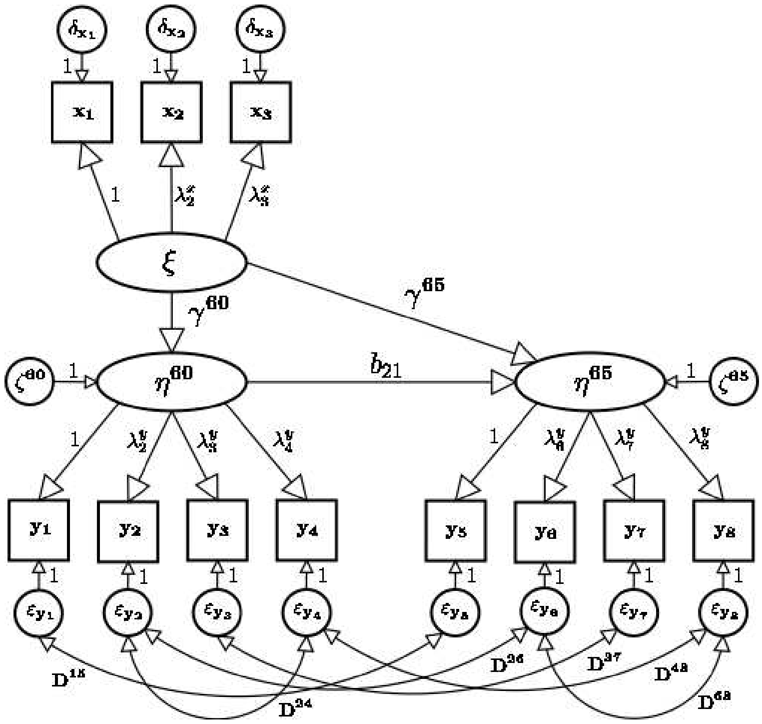

Figure 2: Path diagram for the democratization study

in matrix form as follows,

$$\begin{pmatrix}\eta_i^{60}\\ \eta_i^{65}\end{pmatrix} = \begin{pmatrix}\alpha^{60}\\ \alpha^{65}\end{pmatrix} + \begin{pmatrix}0 & 0\\ b_{21} & 0\end{pmatrix}\begin{pmatrix}\eta_i^{60}\\ \eta_i^{65}\end{pmatrix} + \begin{pmatrix}\gamma^{60}\\ \gamma^{65}\end{pmatrix}\xi_i + \begin{pmatrix}\zeta_i^{60}\\ \zeta_i^{65}\end{pmatrix} \tag{4}$$

where the disturbances $\boldsymbol{\zeta}_i = (\zeta_i^{60}, \zeta_i^{65})$ are assumed to be independent normally distributed with mean zero and precision parameters $\omega_{\zeta^{60}}^{-1}$ and $\omega_{\zeta^{65}}^{-1}$ respectively. The measurement model, as introduced in (1), is now formulated as follows

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

$$\begin{pmatrix}x_{1i}\\ x_{2i}\\ x_{3i}\end{pmatrix} = \begin{pmatrix}0\\ \nu_2^x\\ \nu_3^x\end{pmatrix} + \begin{pmatrix}1\\ \lambda_2^x\\ \lambda_3^x\end{pmatrix}\xi_i + \begin{pmatrix}\delta_{1i}^x\\ \delta_{2i}^x\\ \delta_{3i}^x\end{pmatrix} \tag{5b}$$

where λ_j^y is the influence of PDL on the indicator $\boldsymbol{y}_j$, $j = 1, \ldots, 8$, $\boldsymbol{D}^{rs}$ is a pseudo-latent variable to model the correlations among the measurement errors $\boldsymbol{\delta}^y_r$ and $\boldsymbol{\delta}^y_s$. We fix the intercepts, $\nu^y_1 = \nu^y_5 = \nu^x_1 = 0$, and factor loadings, $\lambda^y_1 = \lambda^y_5 = \lambda^x_1 = 1$, for identifiability of the model and to scale the latent variables. Therefore, PDL will be scaled in terms of freedom in press and IL in terms of gross national product per capita. Furthermore, this approach results in a centered parametrization, which has appealing computational properties as discussed in Section 3.2.

Under expressions (4) and (5), and for $i = 1, \ldots, 75$ developing countries, the complete data likelihood including the latent variables $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ is as follows

$$\mathcal{L}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\eta}, \xi, \boldsymbol{D}; \boldsymbol{\Theta}) = \prod_{i=1}^{75} \Big\{ N_8 \left(\boldsymbol{y}_i; \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta}_i, \boldsymbol{\Sigma}_y\right) N_3 \left(\boldsymbol{x}_i; \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x \xi_i, \boldsymbol{\Sigma}_x\right) N(\xi_i; \mu_\xi, \omega^2_\xi) \times \\ \times N_2 \left(\boldsymbol{\eta}_i; \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\xi_i, \boldsymbol{\Omega}_\zeta\right) N_6 \left(\boldsymbol{D}_i; \boldsymbol{0}, \boldsymbol{\Omega}_D\right) \Big\}$$

with $\boldsymbol{\Sigma}_y = diag(\sigma^2_{y1}, \ldots, \sigma^2_{y8})$, $\boldsymbol{\Sigma}_x = diag(\sigma^2_{x1}, \sigma^2_{x2}, \sigma^2_{x3})$, $\boldsymbol{\Omega}_\zeta = \sigma^2_{y1} \cdot diag\left(\omega^{-1}_{\zeta^{60}}, \omega^{-1}_{\zeta^{65}}\right)$, $\boldsymbol{\Omega}_D = diag\left(\omega^2_{D^{15}}, \omega^2_{D^{24}}, \omega^2_{D^{26}}, \omega^2_{D^{37}}, \omega^2_{D^{48}}, \omega^2_{D^{68}}\right)$, and $\boldsymbol{\Theta}$ includes the free elements of $(\boldsymbol{\nu}_y, \boldsymbol{\Lambda}_y, \boldsymbol{\nu}_x, \boldsymbol{\Lambda}_x, \boldsymbol{B})$ and the parameters $(\boldsymbol{\sigma}^2_y, \boldsymbol{\sigma}^2_x, \mu_\xi, \omega^2_\xi, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\Omega}_\zeta, \boldsymbol{\omega}^2_D)$.

In the Bayesian analysis, the prior specification involves quantifying expert's uncertainty in the model parameters $\boldsymbol{\Theta}$. In the cases where not much information is available beyond the observed data, non-informative or objective priors are the usual selection (Berger, 1985; Bernardo and Smith, 1994). Here, we consider a variety of alternative priors, with the primary choice based on expert elicitation, choosing a specification that assigns high probability to a plausible range for the parameter values based on Ken Bollen's experience in this area. We also consider priors centered on the MLEs, but with inflated variance, for sake of comparison. Refer to Appendix A for more details on the hyperparameters used.

In this case, the joint posterior is computed, following Bayes' rule, as

$$\pi(\boldsymbol{\Theta}, \xi_i, \boldsymbol{\eta}_i, \boldsymbol{D}_i | \boldsymbol{x}, \boldsymbol{y}) \propto \mathcal{L}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{D}; \boldsymbol{\Theta}) \cdot \pi(\boldsymbol{\Theta}).$$

Although this joint posterior distribution is complex, all the corresponding full conditional posterior distributions have simple conjugate forms due to the model assumed. A Gibbs sampling algorithm based on the general scheme introduced before is used to obtain samples from the posterior distributions of the parameters of interest, for example, PDL and IL for every single country in the study in both periods (1960 and 1965), or the impact of the PDL in 1960 on the PDL in 1965. The implementation of the algorithm was written in R, and run 50,000 iterations, discarding the first $10,000$ for burn-in, and keeping one every 400 iterations to reduce the correlation among the posterior samples. WinBUGS could also be used, but our R implementation gave us greater flexibility regarding the computational algorithms and priors we could consider.

4.2 Results

We start by comparing the frequentist (Maximum Likelihood) and Bayesian estimates for the aforementioned parameters of interest, see Appendix B for a full list of parameters estimates. In Figure 3 and 4 we show graphically the histograms of the posterior samples for the parameters b_{21}, γ^{60}, γ^{65}, μ_ξ and σ^2_ξ along with a confidence interval for the MLEs. The learning process experimented in updating the prior to the posterior beliefs based on the observed data are presented in Table 2. For example, a prior 95% probability interval for the

	MLE		Centered MLE					Subjective				
			Prior beliefs		Posteriors			Prior beliefs		Posteriors		
	Mean	Sd	Mean	Sd	Median	Mean	Sd	Mean	Sd	Median	Mean	Sd
b_{21}	0.837	0.098	0.837	2.449	0.744	0.741	0.109	1	1.414	0.814	0.811	0.144
γ^{60}	1.483	0.399	1.483	2.449	1.455	1.454	0.313	1.5	1.414	1.083	1.077	0.209
γ^{65}	0.572	0.221	0.572	2.449	0.774	0.774	0.278	0.5	1.414	0.297	0.322	0.205
μ_ξ	5.054	0.084	5.054	2.5	5.054	5.053	0.087	5	1	5.040	5.035	0.098
ω^2_ξ	0.448	0.087	0.448	0.224	0.433	0.442	0.077	1	0.5	0.655	0.667	0.119

Table 2: Parameters of interest estimated under the frequentist (MLE) and Bayesian approach (summary of the posterior distributions).

influence of PDL in 1960 on the level in 1965 is: $[-4.062, 5.736]$, under the centered MLE priors scheme, and $[-1.828, 3.828]$ under the subjective priors scheme. These probability intervals, a posteriori, are narrowed to $[0.523, 0.959]$ and $[0.522, 1.1]$ respectively. This shows a convergence, after observing the data, regardless of the starting prior knowledge.

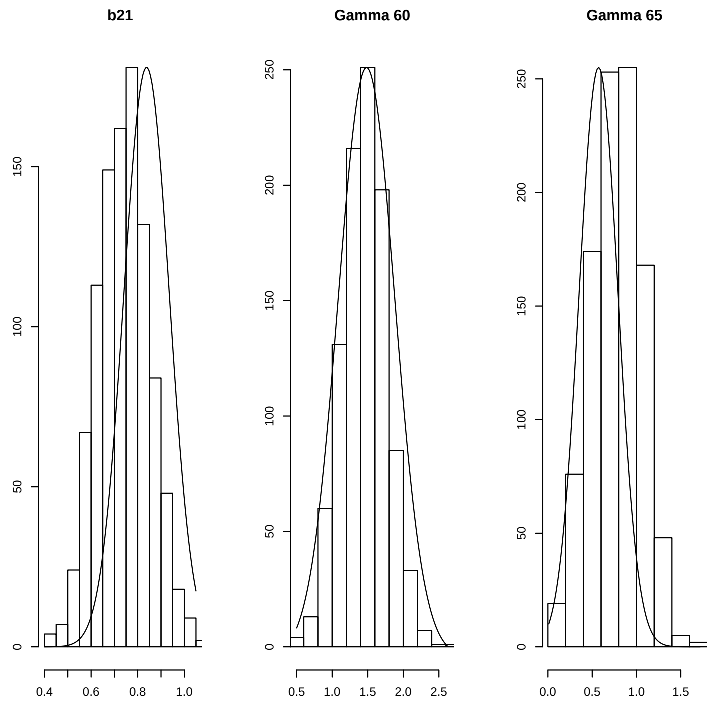

Figure 3: Histograms for the posterior samples under the subjective priors scheme. From left to right: b_{21}, γ^{60} and γ^{65}. The confidence intervals for the MLEs are represented with straight lines.

As measures of the goodness of fit of the frequentist model, we report the R-square indicators in Table 3. In Bayesian SEMs, we shall use some loss function $L(y_i, \widehat{y_i})$ to measure

R^2 Estimates												
y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	x_1	x_2	x_3	η^{60}	η^{65}
0.723	0.514	0.522	0.715	0.653	0.557	0.678	0.685	0.846	0.947	0.761	0.200	0.961

Table 3: R-square indicators for the frequentist model.

the goodness of the predictive distribution. For example, the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) are common measures computed as follows,

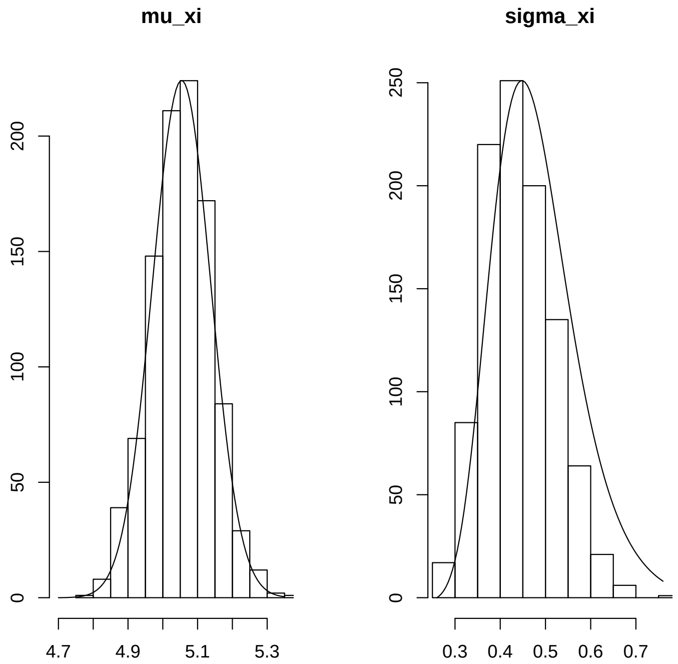

Figure 4: Histograms of the posterior samples for μ_ξ (left) and ω_ξ^2 (right) under the subjective priors scheme. The confidence intervals for the MLEs are represented with straight lines.

$i = 1, \ldots, 75$ and $j = 1, \ldots, 8$,

$$\begin{array}{ll} RMSE_i = \frac{\sqrt{\sum_{j=1}^{8}(y_{ij}-E(\widehat{y}_{ij}))^2}}{8} & RMSE_j = \frac{\sqrt{\sum_{i=1}^{75}(y_{ij}-E(\widehat{y}_{ij}))^2}}{75} \\ MAE_i = \frac{\sum_{j=1}^{8}|y_{ij}-E(\widehat{y}_{ij})|}{8} & MAE_j = \frac{\sum_{i=1}^{75}|y_{ij}-E(\widehat{y}_{ij})|}{75} \end{array}$$

where $E(\widehat{y}_{ij}) = \frac{1}{MN}\sum_{l=1}^{M}\sum_{k=1}^{N}\widehat{y}_{ij}^{lk}$ is the average of the posterior predictions for the j-th PDL indicator and i-th country. Those for the indicators of IL follow symmetrically. We report these estimates in Table 4, where countries $\{46, 61, 22\}$ are samples from each of the three industrialization clusters identified.

So far, we have presented summaries of the results obtained following both the frequentist's and Bayesian's approaches. However, there are more posterior information that can be obtained from the Bayesian SEM methods introduced in Section 3. In particular, the benefit of having posterior samples from the joint posterior distribution of the latent variables is

	$Country_{46}$	$Country_{61}$	$Country_{22}$
$RMSE_{Xi}$	0.94	0.24	0.323
MAE_{Xi}	0.13	0.33	0.384
$RMSE_{Yi}$	0.511	0.277	0.355
MAE_{Yi}	1	0.583	0.783

	IL_1	IL_2	IL_3
$RMSE_{Xj}$	0.029	0.032	0.075
MAE_{Xj}	0.21	0.214	0.54

	PDL_1	PDL_2	PDL_3	PDL_4	PDL_5	PDL_6	PDL_7	PDL_8
$RMSE_{Yj}$	0.096	0.193	0.172	0.116	0.131	0.134	0.131	0.114
MAE_{Yj}	0.689	1	1	0.818	0.892	0.912	0.931	0.765

Table 4: Measures of the goodness of the predictive distribution.

large. They provide important information not contained in the measurement and structural models. We highlight these issues on our case study in the next section.

4.3 Democratization results

The average across countries of the posterior samples of IL in 1960 are summarized as follows

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Sd.
3.498	4.374	5.117	5.035	5.560	6.593	0.798

Recall that the main goal is to determine if the IL of a country has an impact on the change of its PDL. In Figure 5 and 6 we show the PDL in 1960 (gray boxes) and in 1965 (black boxes) for each country in the study, along with their IL (posterior mean) in 1960 (black circles). To facilitate the interpretation, we have sorted the countries by increasing IL.

We notice a generalized reduction in PDL from 1960 to 1965 — in Figure 6 the gray circles are mostly below the black diamonds. In Figure 7 we show this behavior for the countries in the study, where the straight black line represents the average across-countries of the PDL change; the PDL average reduced amount is 0.314.

As a first approach, we have linearly regressed each posterior sample of the IL against the square of the PDL change for each country. We have estimated by least squares the slope of the regression line, finding that the posterior probability of having a negative slope is 0.94. This indicates that an increase in the IL will almost surely cause a positive or negative change in the PDL. However, we notice that this behavior is not homogeneous among countries, and consequently further analysis is required. We define three clusters of countries, corresponding to: *poorly industrialized*, those countries in the first quartile; *mildly industrialized*, those in the second and third quartile; and *highly industrialized*, those in the forth quartile. These clusters are represented with vertical dotted lines in Figures 5, 6 and 7, yielding 19, 37 and 19 countries respectively on each group. By regressing within each group, as previously introduced, we obtain the following posterior probabilities

	Poorly Indust.	*Mildly* Indust.	*Highly* Indust.
$Prob\left(\widehat{\beta} < 0 \mid data\right)$	0.55	0.443	0.644

Recall that a negative slope indicates that a variation in the IL produces a variation in the opposite direction of the square of the PDL change. Therefore, we find that the impact of the IL is different among groups. Figure 6 shows this behavior, where the horizontal dashed lines indicate the first and forth quartile of the average, across countries, of the PDL in 1960 (grey) and 1965 (black). For the poorly industrialized countries, we find that the IL has caused, in most of the cases, the PDL to remain within the bands in 1965 when previously it was within the bands in 1960. The same pattern is found in the most industrialized group and is, in fact, stronger in this case since there are no countries outside the PDL bands that were not present in the previous period. In the case of the mildly industrialized countries, the variability is considerably higher and no pattern is detected.

Another issue of note is the difference in variability among the groups with regard to changes in the PDL. Only in the highly industrialized countries, has the generalized democracy reduction trend been lessened. This is particularly clear in Figure 7, where the horizontal dashed red lines indicate the first and forth quartile of the average across-countries of the square of the PDL change. For the poorly industrialized countries, the level change is mainly negative out of the lower band, indicating that the low levels of industrialization cannot com-

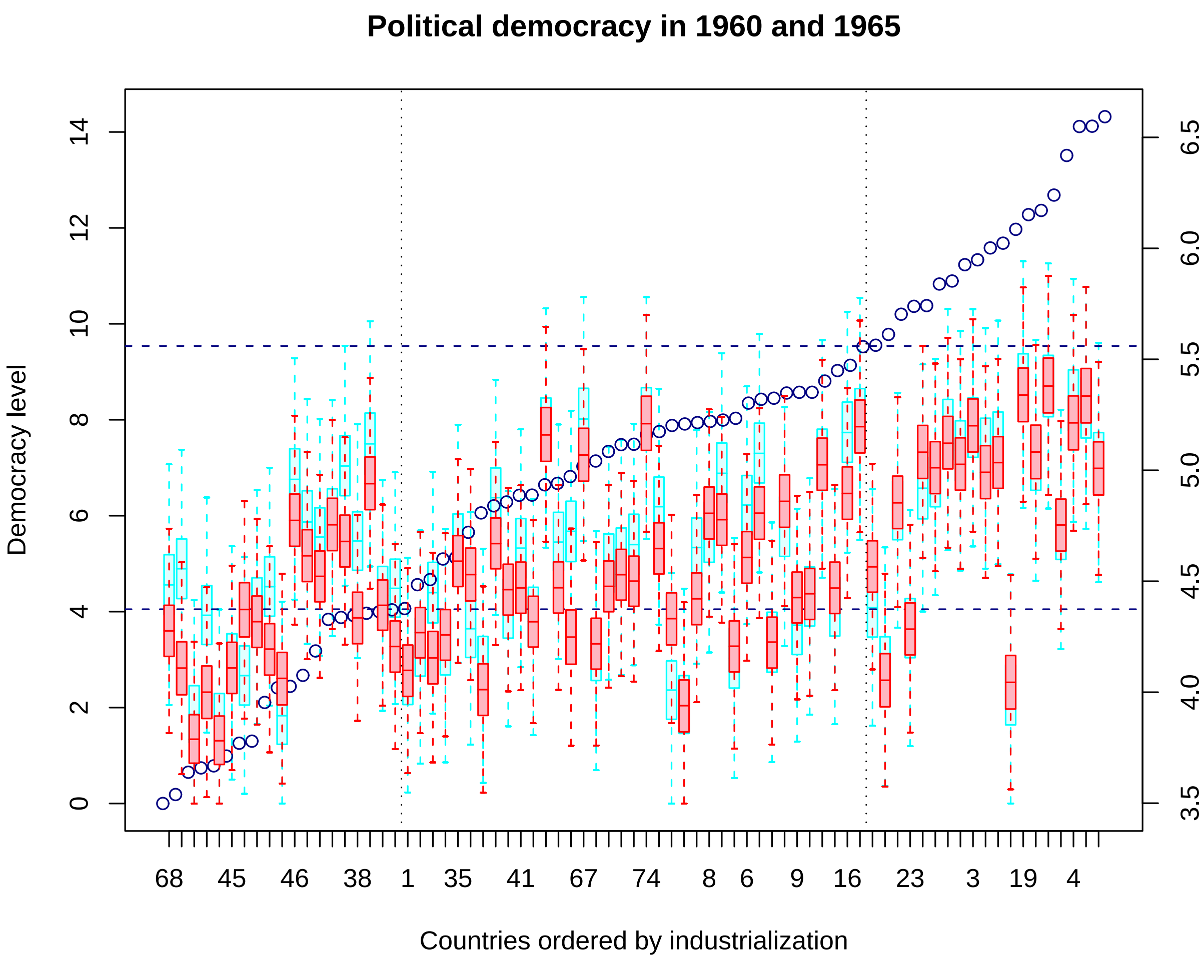

Figure 5: Boxplots of the PDL in 1960, gray, and in 1965, black. The countries are sorted, black circles, by increasing IL (posterior mean) in 1960. Vertical dashed lines separate the three clusters using IL as criteria. The horizontal dashed lines show the 95% bands for the posterior means of PDL in 1965.

pensate the general trend. This is not the case in the highly industrialized countries, where there are just a few countries where the PDL change is below the average. In the mildly industrialized countries, the aforementioned large variability results in alternating changes around the average and outside of the bands.

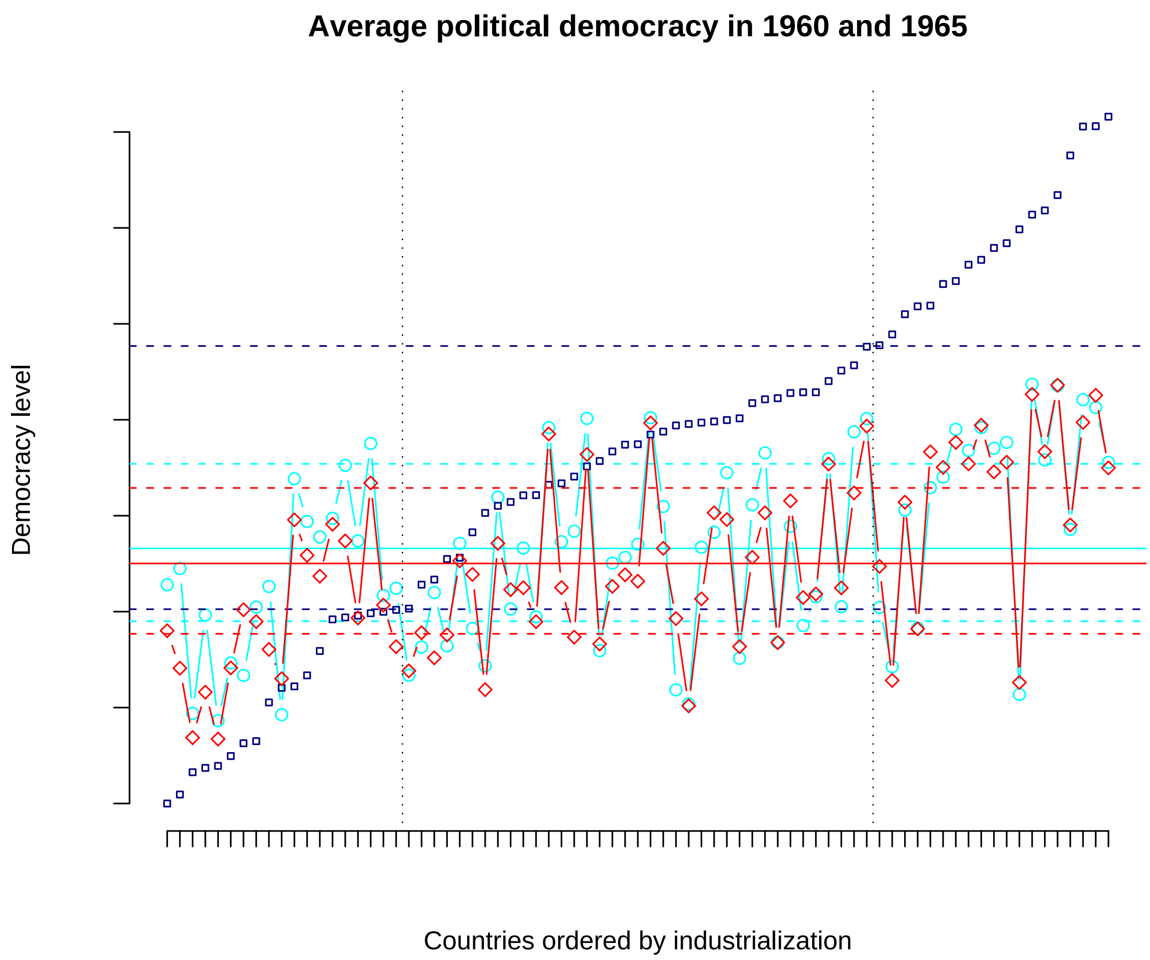

Figure 6: Chart that shows the posterior mean of PDL in 1960, gray dots, and in 1965, black diamonds. The countries are sorted, black circles, by increasing IL (posterior mean) in 1960. Vertical dashed lines separate the three clusters using IL as criteria. The horizontal gray dashed lines show the 95% bands for the posterior means of PDL in 1960, and black dashed lines show these bands in 1965.

5 Discussion and Future Research

This chapter has provided an overview of a Bayesian approach to structural equation modeling, highlighting some aspects of the Bayesian approach that have not been fully explored

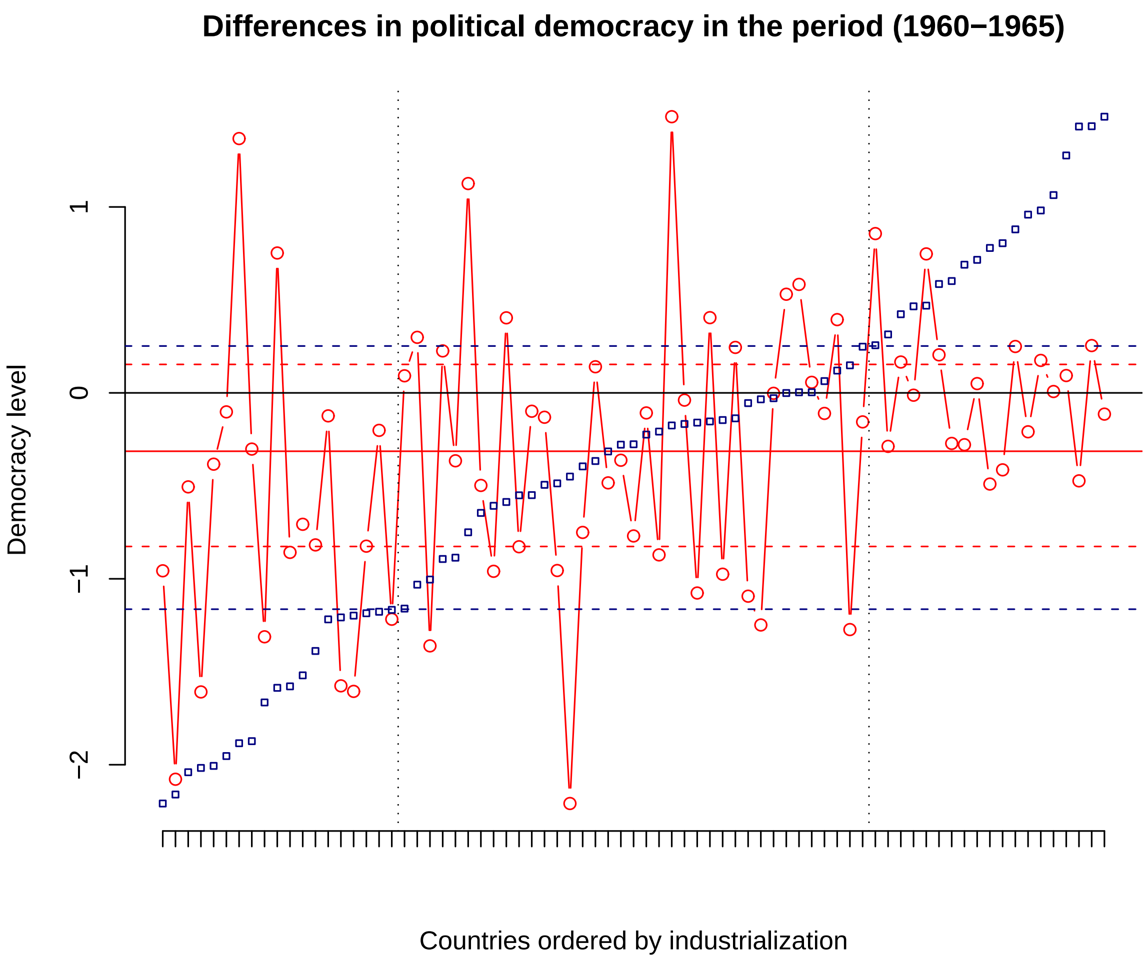

Figure 7: Chart that shows the differences in PDL between 1960 and 1965. The countries are sorted, black circles, by increasing IL (posterior mean) in 1960. Vertical dashed lines separate the three clusters using IL as criteria.

in previous articles on Bayesian SEMs. In particular, previous authors have not carefully considered the issue of parametrization, which really does have an enormous practical impact on Bayesian computation. The most important points to remember in conducting Bayesian SEM analysis are (1) use a centered parametrization allowing the latent variables to have free intercepts and variances; (2) do not use diffuse (high variance) or improper uniform priors for the latent variable variances; and (3) examine the posterior distributions of the latent variables, because they often provide additional information and insight not apparent from the estimated population parameters.

There is a need for additional research into computationally efficient algorithms for SEMs. There has been a lot of interest in computation for simpler variance component models, and a variety of approaches have been proposed, including not only centering but also clever parameter expansion techniques (refer to Gelman et al, 2004, for a recent reference). The parameter expansion approach can potentially be applied directly to SEMs, but practical details remain to be worked out.

Another very important issue is the prior specification, particularly in cases in which limited information is available *a priori* or one wishes to choose a noninformative prior in conducting a reference analysis. Previous authors suggested using a uniform improper prior for the vector of parameters, including the variance components, to choose a non-informative specification (Scheines, Hoijtink and Boomsma, 1999). Unfortunately, uniform improper priors on the latent variable variances will result in an improper posterior, so that the results under such analysis are meaningless. This problem is not solved by using highly diffuse inverse-gamma priors, because the posterior is then close to improper (i.e., it might as well be improper as far as MCMC behavior and interpretation). In addition, as illustrated by Gelman (2004) for simple variance component models, there can be enormous sensitivity to the prior variance chosen in the diffuse gamma prior. Better reference priors for variance component models were suggested by Gelman (2004) and similar specifications can be used in SEMs.

Additional areas in need of further research, include model selection/averaging and semi-parametric methods. The Bayesian approach has the major advantage that it can allow for uncertainty in different aspects of the model specification in performing inferences about structural relations of interest. Raftery has developed Bayesian methods for model selection and averaging in SEMs in a series of papers, primarily based on the BIC and Laplace approximations to the marginal likelihood. Such approaches are not expected to have good performance when comparing models with different variance component structures due to the constraints needed. Expanding the class of models considered to allow unknown latent variable and measurement error distributions can potentially be accomplished within a Bayesian approach using Dirichlet process priors, but again details remain to be worked out.

References

Agresti, A. (1999). Modelling ordered categorical data: Recent advances and future challenges. *Statistics in Medicine* **18**, 2191-2207

Ansari, A. and Jedidi, K. (2000). Bayesian factor analysis for multilevel binary observations. *Psychometrika* **64**, 475-496.

Ansari, A., Jedidi, K. and Jagpal, S. (2000). A hierarchical Bayesian methodology for treating heterogeneity in structural equation models. *Marketing Science*, 19, 328-347.

Arminger, G. and Küsters, U. (1988), "Latent trait models with indicators of mixed measurement level," In *Latent Trait and Latent Class Models* (eds R. Langeheine and J. Rost), pp. 51-73. New York: Plenum.

Arminger, G. and Muthén, B.O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika*, 63, 271-300.

Bauwens, L. (1984). Bayesian full information analysis of simultaneous equation models using integration by Monte Carlo. New York: Springer-Verlag.

Bentler, P.M. (1992). *EQS Structural Equation Program Manual.* Los Angeles: BMDP Statistical Software.

Berger, J.O. (1985) Statistical Decision Theory and Bayesian Analysis, 2nd edn. New York: Springer- Verlag.

Bernardo, J. M. and Smith, A. F. M. (1994). Bayesian Theory. Chichester: Wiley.

Bock, R.D., Gibbons, R. and Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, 12, 261-280.

Bollen, K.A. (1980). Issues in the Comparative Measurement of Political Democracy. American Sociological Review, 80, 370-390.

Bollen, K.A. (1989). Structural Equations with Latent Variables. New York: Wiley.

Bollen, K.A. and Paxton, P. (1998). Two-stage least squares estimation of interaction effects. In R.E. Schumacker and G.A. Marcoulides (eds), *Interaction and Nonlinear Effects in Structural Equation Models*, 125-151. Mahwah, NJ: Lawrence Erlbaum Associates.

Brooks, S.P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*,**7**,434-455.

Brooks, S.P. and Giudici, P. (2000). MCMC convergence assessment via two-way ANOVA. *Journal of Computational and Graphical Statistics*, **9**,266-285.

Browne, M.W. (1974). Generalized least squares estimations in the Analysis of covariance structures. South African Statistical Journal, 8, 1-24.

Chen, M.H, Shao, Q.M. and Ibrahim, J.G (2000). Monte Carlo methods in Bayesian computation. New York : Springer.

Cohen, P., Cohen, J., Teresi, J., Marchi, M. and Velez, C.N. (1990). Problems in the Measurement of Latent Variables in Structural Equations Causal Models. *Applied Psychological Measurement*, 14, 183-196.

Croon, M. and Bolck, A. (1997). On the use of factor scores in structural equations models. *Tech. Rep. 97.10.102/7*, Work and Organization Research Centre, Tilburg University.

Dunson, D.B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society* **B 62**, 355-366.

Dunson, D.B., Chen, Z. and Harry, J. (2002). A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. *Biometrics*, under revision.

Dunson, D. B. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association*, **98**, 555-563.

Edwards, J.R. and Bagozzi, R.P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5, 155-174.

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

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. I.E.E.E. Trans. Pattern Anal. Machine Intell., 6, 721-741.

Gamerman, D. (1997). Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, London: Chapman and Hall.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996). Markov Chain Monte Carlo in Practice. Chapman and Hall.

Hayashi, K. and Yuan, K.-H. (2003). Robust Bayesian factor analysis. *Structural Equation Modeling*, 10, 525-533.

Herring, A.H. and Dunson, D.B. (2004). Modeling the effects of a bi-directional latent predictor from multivariate questionnaire data. *Biometrics*, 60, 926-935.

Hills, S.E. and Smith, A.F.M. (1992). Parametrization issues in Bayesian inference (with discussion). In, *Bayesian Statistics 4*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds.), pp. 227 246. Oxford University Press.

Jedidi, K. and Ansari, A. (2001). Bayesian structural equation models for multilevel data. In *New Developments and Techniques in Structural Equation Modeling* (edited by G.A. Marcoulides and R.E. Schumacker)

Jöreskog, K.G. and Sörbom, D. (1985), LISREL VI; Analysis of Linear structural relationships by Maximum Likelihood, Instrumental variables, and Least Squares, Uppsala: University of Uppsala.

Jöreskog, K.G. and Sörbom, D. (1996). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Hove and London: Scientific Software International.

Jöreskog, K.G. and Yang, F. (1996). Nonlinear structural equation models: the Kenny-Judd model with interaction effects. In G.A. Marcoulides and R.E. Schumacker (eds), *Advanced Structural Equation Modeling Techniques*, 57-88, Hillsdale, NJ: Lawrence Erlbaum Associates.

Lee, S.-Y. (1981). A Bayesian approach to confirmatory factor analysis. *Psychometrika*, 46, 153-160.

Lee, S.-Y. (1992). Bayesian analysis of stochastic constraints in structural equation models. *British Journal of Mathematical and Statistical Psychology*, 45, 93-107.

Lee, S.-Y. and Shi, J.-Q. (2000). Bayesian analysis of structural equation model with fixed covariates. *Structural Equation Modeling: A Multidisciplinary Journal*, 7, 411-430.

Lee, S.-Y. and Song, X.-Y. (2003). Bayesian model selection for mixtures of structural equation models with an unknown number of components. *British Journal of Mathematical and Statistical Psychology*, 56, 145-165.

Lee, S.-Y. and Song, X.-Y. (2004). Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal data. *British Journal of Mathematical and Statistical Psychology*, 57, 131-150.

Liu, J.S., Wong, W.H. and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* **81**, 27-40.

Martin, J.K. and McDonald, R.P. (1975). Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika*, 40, 505-517.

Moustaki, I. and Knott, M. (2000). Generalized latent trait models. *Psychometrika*, 65, 391-411.

Muthén, B.O. (1984), "A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators," *Psychometrika*, 49, 115-132.

Muthén, L.K. and Muthén, B.O. (1998), Mplus User's Guide. Los Angeles, CA: Muthén & Muthén.

Muthén, L.K. and Muthén, B.O. (2003), Mplus Version 2.13. Addendum to the Mplus User's Guide. Los Angeles, CA: Muthén and Muthén. Downloadable from `http://www.statmodel`

Press, S.J. and Shigemasu, K. (1989). Bayesian inference in factor analysis. In L.J. Gleser, M.D. Perlman, S.J. Press and A.R. Sampson (eds.) *Contributions to Probability and Statistics.* New York: Springer-Verlag, 271-287.

Raftery, A.E. (1993). Bayesian model selection in structural equation models. In *Testing Structural Equation Models* (edited by K.A. Bollen and J.S. Long)

Raftery, A.E. and Lewis, S. (1992). How many iterations in the Gibbs sampler?. In, *Bayesian Statistics 4*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds.), pp. 763 773. Oxford University Press.

Robert, C.P. and Casella, G. (2004). Monte Carlo Statistical Methods. 2nd ed. Springer Texts in Statistics.

Roberts, G.O. and Sahu, S.K. (1997). Updating schemes; correlation structure; blocking and parametrization for the Gibbs sampler. *Journal of the Royal Statistical Society*, **B**, 59, 291-317.

Sammuel, M.D., Ryan, L.M. and Legler, J.M. (1997). Latent Variable Models for Mixed Discrete and Continuous outcomes. In *Journal of the Royal Statistical Society*. **B**, 59, 667-675.

Scheines, R., Hoijtink, H. and Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, 64, 37-52.

Wall, M.M. and Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the American Statistical Association*, 95, 929-940.

Spiegelhalter, D.J., Thomas, A., Best, N. and Gilks, W. (2003). WinBUGS. Version 1.4 User Manual. MRC Biostatistics Unit. URL: `http://weinberger.mrc-bsu.cam.ac.uk/bugs/Welcome.html`.

Appendix

A Prior specifications

We consider the following gamma and inverse-gamma formulations for the prior distribution of the precision and variance parameters respectively, where

$$
\begin{aligned}
\sigma^2 \sim InvGamma(\sigma^2; \alpha, \beta) \\
f(\sigma^2) &= \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-(\alpha+1)} \exp\left(-\frac{\beta}{\sigma^2}\right) \\
\mu &= \frac{\beta}{\alpha - 1} \qquad \sigma^2 = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} \\
\sigma^{-2} \sim Gamma(\sigma^{-2}; \alpha, \beta) \\
f(\sigma^{-2}) &= \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^{-2})^{\alpha-1} \exp\left(-\beta\sigma^{-2}\right) \\
\mu &= \frac{\alpha}{\beta} \qquad \sigma^2 = \frac{\alpha}{\beta^2}
\end{aligned}
$$

See Table 5 for the prior parameters used.

B Results: posterior parameters estimates

	MLE		Sub. Priors			Centered MLE		
	Mean	Sd	Median	Mean	Sd	Median	Mean	Sd
α^{60}	-2.031	2.037	-0.021	-0.005	1.051	-1.865	-1.876	1.587
α^{65}	-2.332	1.119	-0.722	-0.763	0.944	-2.788	-2.806	1.287
b_{21}	0.837	0.098	0.814	0.811	0.144	0.744	0.741	0.109
γ^{60}	1.483	0.399	1.083	1.077	0.209	1.455	1.454	0.313
γ^{65}	0.572	0.221	0.297	0.322	0.205	0.774	0.774	0.278
ν_2^y	-2.611	1.064	-1.235	-1.205	0.708	-2.251	-2.289	0.945
ν_3^y	0.783	0.883	0.208	0.195	0.637	0.630	0.593	0.796
ν_4^y	-2.459	0.843	-1.415	-1.4	0.646	-2.249	-2.281	0.799
ν_6^y	-3.112	0.928	-1.292	-1.296	0.563	-2.555	-2.595	0.82
ν_7^y	-0.376	0.878	0.539	0.529	0.54	-0.014	-0.031	0.757
ν_8^y	-2.459	0.868	-0.841	-0.85	0.536	-2.024	-2.051	0.714
λ_2^y	1.257	0.182	1.039	1.04	0.131	1.186	1.193	0.165
λ_3^y	1.058	0.151	1.173	1.176	0.119	1.086	1.091	0.138
λ_4^y	1.265	0.145	1.111	1.106	0.118	1.218	1.228	0.14
λ_6^y	1.186	0.169	0.850	0.852	0.107	1.076	1.077	0.150
λ_7^y	1.280	0.160	1.095	1.094	0.1	1.207	1.208	0.135
λ_8^y	1.266	0.158	0.962	0.963	0.1	1.175	1.180	0.128
$\sigma^2_{y_1}$	1.891	0.444	1.486	1.509	0.215	0.777	0.799	0.178
$\sigma^2_{y_2}$	7.373	1.374	4.330	4.409	0.958	4.763	4.896	1.025
$\sigma^2_{y_3}$	5.067	0.952	3.534	3.590	0.7305	3.929	4.029	0.843
$\sigma^2_{y_4}$	3.148	0.739	2.581	2.620	0.55	2.061	2.118	0.537
$\sigma^2_{y_5}$	2.351	0.480	2.567	2.635	0.532	1.868	1.899	0.421
$\sigma^2_{y_6}$	4.954	0.914	2.801	2.869	0.619	2.662	2.739	0.647
$\sigma^2_{y_7}$	3.431	0.713	2.767	2.811	0.611	2.495	2.578	0.642
$\sigma^2_{y_8}$	3.254	0.695	2.422	2.467	0.479	1.937	2.012	0.497
ν_2^x	-6.228	0.705	-4.059	-4.059	0.48	-6.364	-6.405	0.624
ν_3^x	-5.634	0.774	-3.361	-3.380	0.536	-5.706	-5.734	0.732
λ_2^x	2.180	0.139	1.759	1.760	0.095	2.209	2.215	0.123
λ_3^x	1.819	0.152	1.382	1.381	0.105	1.836	1.838	0.144
$\sigma^2_{x_1}$	0.082	0.019	0.106	0.109	0.022	0.083	0.085	0.017
$\sigma^2_{x_2}$	0.120	0.070	0.172	0.179	0.056	0.113	0.118	0.04
$\sigma^2_{x_3}$	0.467	0.090	0.445	0.454	0.083	0.466	0.478	0.085
$\omega^{-1}_{\zeta^{60}}$	3.956	0.921	2.047	2.091	0.465	4.735	5.046	1.726
$\omega^{-1}_{\zeta^{65}}$	0.172	0.215	3.756	3.826	0.687	2.750	2.840	0.621
μ_ξ	5.054	0.084	5.040	5.035	0.098	5.054	5.053	0.087
ω^2_ξ	0.448	0.087	0.655	0.667	0.119	0.433	0.442	0.077
$\omega^2_{D^{15}}$	0.624	0.358	0.648	0.677	0.204	0.625	0.662	0.22
$\omega^2_{D^{24}}$	1.313	0.702	1.215	1.319	0.565	1.409	1.508	0.566
$\omega^2_{D^{26}}$	2.153	0.734	1.406	1.504	0.571	1.756	1.816	0.499
$\omega^2_{D^{37}}$	0.795	0.608	0.885	0.939	0.336	0.79	0.857	0.329
$\omega^2_{D^{48}}$	0.348	0.442	0.719	0.756	0.251	0.317	0.351	0.15
$\omega^2_{D^{68}}$	1.356	0.568	0.89	0.968	0.376	1.125	1.189	0.384

	MLE Centered Parameters		Subjective Parameters		MLE Estimates	
$\sigma^2_{y_1} \sim IGamma(\cdot,\cdot)$	6.0	9.455	10	36	1.891	0.444
$\sigma^2_{y_2} \sim IGamma(\cdot,\cdot)$	6.0	36.865	10	36	7.373	1.374
$\sigma^2_{y_3} \sim IGamma(\cdot,\cdot)$	6.0	25.335	10	36	5.067	0.952
$\sigma^2_{y_4} \sim IGamma(\cdot,\cdot)$	6.0	15.74	10	36	3.148	0.739
$\sigma^2_{y_5} \sim IGamma(\cdot,\cdot)$	6.0	11.755	10	36	2.351	0.480
$\sigma^2_{y_6} \sim IGamma(\cdot,\cdot)$	6.0	24.77	10	36	4.954	0.914
$\sigma^2_{y_7} \sim IGamma(\cdot,\cdot)$	6.0	17.155	10	36	3.431	0.713
$\sigma^2_{y_8} \sim IGamma(\cdot,\cdot)$	6.0	16.27	10	36	3.254	0.695
$\sigma^2_{x_1} \sim IGamma(\cdot,\cdot)$	6.0	0.41	6	1	0.082	0.019
$\sigma^2_{x_2} \sim IGamma(\cdot,\cdot)$	6.0	0.6	6	1	0.120	0.070
$\sigma^2_{x_3} \sim IGamma(\cdot,\cdot)$	6.0	2.335	6	1	0.467	0.090
$\mu_\xi \sim N(\cdot,\sigma^2)$	5.054	6.25	5	1	5.054	0.084
$\omega^2_\xi \sim IGamma(\cdot,\cdot)$	6.0	2.24	6	5	0.448	0.087
$\omega^{-1}_{\zeta^{60}} \sim Gamma(\cdot,\cdot)$	4.0	1.912	16	16	2.092	
$\omega^{-1}_{\zeta^{65}} \sim Gamma(\cdot,\cdot)$	4.0	43.977	4.0	93.023	0.091	
$\omega^2_{D^{15}} \sim IGamma(\cdot,\cdot)$	6.0	3.12	6.0	5.0	0.624	0.358
$\omega^2_{D^{24}} \sim IGamma(\cdot,\cdot)$	6.0	6.565	6.0	5.0	1.313	0.702
$\omega^2_{D^{26}} \sim IGamma(\cdot,\cdot)$	6.0	10.765	6.0	5.0	2.153	0.734
$\omega^2_{D^{37}} \sim IGamma(\cdot,\cdot)$	6.0	3.975	6.0	5.0	0.795	0.608
$\omega^2_{D^{48}} \sim IGamma(\cdot,\cdot)$	6.0	1.74	6.0	5.0	0.348	0.442
$\omega^2_{D^{68}} \sim IGamma(\cdot,\cdot)$	6.0	6.78	6.0	5.0	1.356	0.568
$\nu^y_2 \sim N(\cdot,\sigma^2)$	-2.611	6.0	0	1	-2.611	1.064
$\nu^y_3 \sim N(\cdot,\sigma^2)$	0.783	6.0	0	1	0.783	0.883
$\nu^y_4 \sim N(\cdot,\sigma^2)$	-2.459	6.0	0	1	-2.459	0.843
$\nu^y_6 \sim N(\cdot,\sigma^2)$	-3.112	6.0	0	1	-3.112	0.928
$\nu^y_7 \sim N(\cdot,\sigma^2)$	-0.376	6.0	0	1	-0.376	0.878
$\nu^y_8 \sim N(\cdot,\sigma^2)$	-2.459	6.0	0	1	-2.459	0.868
$\lambda^y_2 \sim N(\cdot,\sigma^2)$	1.257	6.0	1	2	1.287	0.182
$\lambda^y_3 \sim N(\cdot,\sigma^2)$	1.058	6.0	1	2	1.058	0.151
$\lambda^y_4 \sim N(\cdot,\sigma^2)$	1.265	6.0	1	2	1.265	0.145
$\lambda^y_6 \sim N(\cdot,\sigma^2)$	1.186	6.0	1	2	1.186	0.169
$\lambda^y_7 \sim N(\cdot,\sigma^2)$	1.280	6.0	1	2	1.280	0.160
$\lambda^y_8 \sim N(\cdot,\sigma^2)$	1.266	6.0	1	2	1.266	0.158
$\nu^x_2 \sim N(\cdot,\sigma^2)$	-6.228	6.0	0	1	-6.228	0.705
$\nu^x_3 \sim N(\cdot,\sigma^2)$	-5.634	6.0	0	1	-5.634	0.774
$\lambda^x_2 \sim N(\cdot,\sigma^2)$	2.180	6.0	1	2	2.180	0.139
$\lambda^x_3 \sim N(\cdot,\sigma^2)$	1.819	6.0	1	2	1.819	0.152
$\alpha^{60} \sim N(\cdot,\sigma^2)$	-2.031	6.0	1	2	-2.031	2.037
$\alpha^{65} \sim N(\cdot,\sigma^2)$	-2.332	6.0	1	2	-2.332	1.119
$b_{21} \sim N(\cdot,\sigma^2)$	0.837	6.0	1	2	0.837	0.098
$\gamma^{60} \sim N(\cdot,\sigma^2)$	1.483	6.0	1.5	2	1.483	0.399
$\gamma^{65} \sim N(\cdot,\sigma^2)$	0.572	6.0	0.5	2	0.572	0.221

Table 5: Prior distributions for the model parameters and their corresponding ML estimates.