# Bayesian Variable Selection for Median Latent Variable Model*

Zhibin Xu† Yifan Wang‡ Hong Ji†

**Abstract**

In biomedical, psychological, social, and behavioral sciences, it is very common to encounter latent variables along with non-normal data. We propose a median latent variable model to deal with this kind of data in a Bayesian framework. The normal-gamma prior distribution is applied here for simultaneous estimation and model selection. A Markov chain Monte Carlo (MCMC) algorithm for obtaining Bayesian estimates is developed. Simulation studies are carried out to examine the finite sample performance of the proposed estimators. We illustrate the proposed method with a real data set from a longitudinal study of polydrug use.

**Key Words:** median regression, confirmatory factor analysis model, normal-gamma prior, asymmetric Laplace distribution, Markov chain Monte Carlo.

## 1. Introduction

In practical applications, many theoretical concepts, which are called latent variables or factors, such as intelligence, personality, desirability, and welfare, cannot be measured directly or evaluated by a single observed variable, but are inferred from some observable variables instead. The confirmatory factor analysis (CFA) model assesses relationships between latent variables and the corresponding manifest variables, and takes the measurement error into account. It provides a useful statistical tool for explaining and analyzing underlying structure of multivariate data based on the idea that the observable variables are impacted by the underlying unobservable factors. This model has been widely used in behavioral, social, and psychological research (Song and Lee, 2012).

Traditional confirmatory factor analysis models are developed under the assumption that the observed variables are normally distributed, which is not realistic in many practical applications (see, e.g., Cai et al., 2010; Li et al., 2012; Cai et al., 2011; Song et al., 2010). If the non-normality is not tackled properly, the analysis of confirmatory factor analysis model may lead to incorrect inference for model parameters. Hence, it is important to develop more robust methods for analyzing confirmatory factor analysis model under the non-normality assumption. Both parametric and nonparametric methods have been proposed by researchers. Parametric methods using $t$-distribution have received a great deal of attention in the last decades (see, e.g., Bentler, 1983; Shapiro and Browne, 1987; Kano, Berkane, and Bentler, 1993; Lee and Xia, 2006). The drawback of parametric methods is that restricting statistical inference to a specific parametric form may limit the scope and type of the inference. Recently, a semiparametric approach using the truncated Dirichlet process with a stick breaking prior was introduced to relax the normality assumption of residuals in latent variable models (see, e.g., Song et al., 2010; Yang and Dunson, 2010; Yang et al., 2010). Although it has been demonstrated useful for handling non-normal data, a simultaneous estimation and variable selection under this model framework is difficult. In this paper, we develop a median regression method to deal with the problem of non-normality of the

*Supported by a grant GRF 404711 from the Research Grant Council of the Hong Kong Special Administration Region

†School of Statistics, Capital University of Economics and Business, Flower-Town, Fengtai District, Beijing, 100070, P. R. China

‡Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong

observed variables. While the conventional mean regression is suitable for modeling data whose distribution is normal or nearly normal, it may fail to yield efficient estimates when data distributions have heavy-tails or are highly skewed. In contrast, the median regression is robust against outliers, and may be more effective in analyzing non-standard data. Moreover, the median regression does not assume any specific residual distribution form, thus it is able to accommodate more general residual distributions. It expands the usual mean regression model by providing a natural way to deal with data with heteroscedastic, heavy-tailed, or highly skewed error distribution.

The main difficulty of incorporating median regression into the Bayesian framework lies in the fact that the median regression model does not specify a likelihood function, which is indispensable in the Bayesian inference. Some authors employed pseudo-likelihood to the Bayesian quantile regression framework. Kottas and Geland (2001) and Kottas and Krnja-jić (2009) developed a modeling approach for the error distribution in quantile regression based on Dirichlet process mixture models. Reich, Bondell and Wang (2008) assumed the error distribution to be an infinite mixture of normals equipped with stochastic constraints. Dunson and Taylor (2005) introduced an approximate approach which relies on a substitution likelihood for quantiles. Yu and Moyeed (2001) used the asymmetric Laplace distribution as an way for modeling Bayesian quantile regression. Yang and He (2012) proposed the Bayesian empirical likelihood for quantile regression in Bayesain inference. Sriram, Ramamoorthi, and Ghosh (2012) provided theoretical justification for the widely used approach using asymmetric Laplace distribution in Bayesian quantile regression, even if the true underlying distribution may be different. Due to the merits of asymmetric Laplace distribution, in this paper, we proposed a fully Bayesian method for the median latent variable model based on the asymmetric Laplace distribution.

Variable selection plays an important role in model buildings. In practice, it is usual to include a large number of candidate predictor variables at the primary stage of model building for the sake of removing serious modeling bias. However, spurious predictors in the final model make it difficult to interpret the resultant model and degrade prediction ability. Classical variable selection methods such as subset selection are time-consuming and often suffer from numerical instability (Breiman, 1996). On the other hand, the variable selection procedure by penalized likelihood attracted a lot of attention in last decades, whose major advantage is to select variables and estimate coefficients simultaneously, for example, Lasso (Tibshirani, 1996), SCAD (Fan and Li, 2001), adaptive Lasso (Zou, 2006), and minimax concave penalty (MCP, Zhang, 2007). There exist corresponding variable selection approaches in the literature of Bayesian method, including the Bayesian Lasso (Park and Casella, 2008), the horseshoe prior (Carvalho et al., 2010), the orthant normal prior (Hans, 2010), the Bayesian adaptive Lasso (Sun et al., 2010), and the normal-gamma prior (Griffin and Brown, 2010), which extended the Bayesian Lasso (Park and Casella, 2008) by placing an absolutely continuous prior distribution on the regression coefficients of model.

Burgette and Reite (2012) proposed a Bayesian quantile regression model, which also used a confirmatory factor structure and the asymmetric Laplace distribution for Bayesian analysis. However, they mainly focus on studying which of the latent variables has a significant effect on a lower quantile of the response variable, which is quite different from our approach for dealing with non-normality. Moreover, to the best of our knowledge, little work on variable selection has been done for the median latent variable model under the Bayesian framework. In this article, we apply the normal-gamma prior to the median latent variable model to achieve simultaneous coefficients estimation and variable selection, which works effectively and avoids the computational difficulty of other methods.

We illustrate the developed methodology through a data set from a longitudinal study of poly-drug use conducted in five California countries in 2004. Since the both drug use

history and drug-related crime history are respectively related to several variables, they are treated as latent variables. Therefore, a confirmatory factor analysis model is used to measure the latent variables via several indicators. Through some preliminary analysis, we found that most of the variables in polydrug use data were extremely non-normal. Hence, it is necessary to develop some robust statistical methods to handle heterogeneity. In addition, some fixed covariates including treatment motivation, services received, the number of drug tests, and the number of drug tests by criminal justice, etc. are collected to take into account their possible effects on the response variable retention. Therefore, it is useful to consider the choice of which subset of variables should be included into the model. In this paper, we apply the normal-gamma shrinkage prior to deal with the variable selection problem.

The remainder of this article is organized as follows. In Section 2, we introduce median latent variable model with normal-gamma prior distribution. We outline the Bayesian MCMC sampler in the Section 3. To evaluate the performance of the proposed method, we conduct the simulation study in Section 4. In Section 5, We illustrate the method by applying it to a data set related to a longitudinal study of polydrug use. Some concluding remarks are provided in Section 6.

## 2. Model specification

### 2.1 Bayesian quantile regression

The linear quantile regression is given by:

$$Q_{y_i}(\tau|X_i) = X_i^T\beta(\tau),$$

where $y_i$ and $X_i$ denote the response variable and covariates respectively, $Q_{y_i}(\tau|X_i)$ is the inverse cumulative distribution of $y_i$ conditional on $X_i$ evaluated at $\tau$, and $\beta(\tau)$ is the unknown regression coefficient. Koenker and Bassett (1978) demonstrated that the regression coefficients can be estimated consistently as the solution to the following minimization problem:

$$\min_{\beta} \sum_{i=1}^{n} \rho_\tau(y_i - X_i^T\beta), \tag{1}$$

where $\rho_\tau(x) = x(\tau - I(x < 0))$ is the so-called check function. Koenker and Machado (1999), and Yu and Moyeed (2001) established that there exits a connection between the minimization problem in (1) and the maximum likelihood estimation theory by assuming the residual distribution is the asymmetric Laplace distribution. A random variable $y$ is distributed as asymmetric Laplace distribution with parameters $\mu$, $\sigma$, $\tau$, if the corresponding probability density function takes the following form:

$$f(y|\mu,\sigma,\tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{\rho_\tau\left(\frac{y-\mu}{\sigma}\right)\right\},$$

in which $\mu$ is the location parameter, $\sigma$ is the scale parameter, $\tau$ is the skewness parameter. Specifically, setting $\mathbf{y} = (y_1, \ldots, y_n)$ and $\mu_i = X_i^T\beta$, and assuming that $y_i \sim ALD(\mu_i, \sigma, \tau)$, then the maximum likelihood for $n$ independent subjects is

$$L(\boldsymbol{\beta}, \sigma; \mathbf{y}, \tau) \propto \frac{1}{\sigma^n} \exp\left\{-\sum_{i=1}^{n} \rho_\tau\left(\frac{y_i - X_i^T\boldsymbol{\beta}}{\sigma}\right)\right\}. \tag{2}$$

If we consider $\sigma$ as nuisance parameter, the minimization of (1) is asymptotically equivalent to the maximum likelihood estimation of the asymmetric Laplace distribution. Thus, the

asymmetric Laplace distribution can be exploited to specify a parametric likelihood, which is needed in the Bayesian framework.

An attractive feature of the asymmetric Laplace distribution is that it can be represented as a scale mixture of a standard normal distribution with an exponential distribution (Kozumi and Kobayashi, 2009):

$$y = \kappa_1 e + \sqrt{\kappa_2 \sigma e}\zeta + \mu,$$

where $\kappa_1 = \frac{1-2\tau}{\tau(1-\tau)}$, $\kappa_2 = \frac{2}{\tau(1-\tau)}$, the random variable $\zeta$ and $e$ is independent, and $\zeta$ follows a standard normal distribution, and $e$ follows an exponential distribution with scale parameter $1/\sigma$. The mixture representation can be used to develop a Gibbs sampling algorithm which avoids the inconvenience to choose the proposal distribution in the Metropolis-Hastings algorithm as well as improving the efficiency of the MCMC sampler.

In this paper, we focus on only the median regression, i.e., $\tau = 50\%$ th quantile regression for the consideration that the median regression can provide enough robustness for non-normal data.

### 2.2 Median latent variable model with normal-gamma prior

For the $i$th ($i = 1, \ldots, n$) subject, let $y_i$ represent response variable. The covariates consist of two parts: fixed covariates and latent factors. Let $\boldsymbol{\omega}_i$ be a $q \times 1$ vector of latent factors which are unobservable, $\mathbf{Z}_i$ be an $r \times 1$ vector of fixed covariates, and $\mathbf{X}_i$ be a $(p-1) \times 1$ vector of manifest variables which are correlated with latent factors $\boldsymbol{\omega}_i$. We assume that, for the $i$th subject:

$$\begin{aligned} \text{Median}(\mathbf{X}_i|\boldsymbol{\omega}_i) &= \boldsymbol{\Lambda}\boldsymbol{\omega}_i \\ \text{Median}(y_i|\boldsymbol{\omega}_i, \mathbf{Z}_i) &= \boldsymbol{\beta}_\omega \boldsymbol{\omega}_i + \boldsymbol{\beta}_z \mathbf{Z}_i, i = 1, \ldots, n, \end{aligned}$$

in which $\boldsymbol{\Lambda}$ is a $(p-1) \times q$ unknown parameter matrix, $\boldsymbol{\beta} = (\boldsymbol{\beta}_\omega^T, \boldsymbol{\beta}_z^T)^T$ is the unknown vector of regression coefficient for $\boldsymbol{\omega}_i$ and $\mathbf{Z}_i$, $\boldsymbol{\omega}_i$ is distributed as $\text{N}(0, \boldsymbol{\Phi})$ where $\boldsymbol{\Phi}$ is $q \times q$ unknown covariance matrix. The median latent variable model takes the form as:

$$\mathbf{X}_i = \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i \tag{3}$$

$$y_i = \boldsymbol{\beta}_\omega \boldsymbol{\omega}_i + \boldsymbol{\beta}_z \mathbf{Z}_i + \delta_i, \tag{4}$$

$\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{ip-1})$ is $(p-1) \times 1$ random vector of residuals and $\epsilon_{i1}, \ldots, \epsilon_{ip-1}$ are mutually independent, $\delta_i$ is a scalar residual. The distributions of the residual term $\boldsymbol{\epsilon}_i$ and $\delta_i$ are assumed as unknown, and are restricted to have the median to zero.

Park and Casella (2008) introduced a Bayesian version of the Lasso approach by Tibshirani (1996). They presented a fully Bayesian analysis using a conditional Laplace prior. The Laplace prior distribution can also be considered as a member of the scale mixture of normals family, which we write as

$$p(\beta_j) = \int \text{N}(\beta_j|0, \varphi_j) dG(\varphi_j),$$

where $\text{N}(x|\mu, \varphi)$ denotes normal density function evaluated at $x$ with mean $\mu$ and variance $\varphi$, and $G$ is a mixing distribution. The prior distribution family can be expressed in a hierarchical form as

$$\beta_j|\varphi_j \sim \text{N}(0, \varphi_j), \quad \varphi_j \sim G.$$

The hierarchical form shows that each regression coefficient has a normal prior distribution conditional on coefficient-specific variance, $\varphi_j$, allowing differences in their scales. The Laplace prior can be expressed in this way if $G$ is an exponential distribution. However,

the drawback of Bayesian Lasso is that it inherits the problem of over-shrinkage large coefficients due to the relatively light tails of the Laplace prior distribution. Griffin and Brown (2010) considered so-called normal-gamma prior which generalizes the Laplace prior of the Bayesian Lasso. They assume that the mixing distribution $G$ has the gamma density $Ga(x|\lambda_\beta, 1/2\gamma^2)$, where

$$Ga(x|c,d) = \frac{d^c}{\Gamma(c)} x^{c-1} e^{-dx}.$$

The normal-gamma prior distribution can be expressed as

$$p(\beta_j|\lambda_\beta, 1/2\gamma^2) = \frac{1}{\sqrt{\pi} 2^{\lambda_\beta - \frac{1}{2}} \gamma^{\lambda_\beta + \frac{1}{2}} \Gamma(\lambda_\beta)} |\beta_j|^{\lambda_\beta - \frac{1}{2}} K_{\lambda_\beta - \frac{1}{2}}(|\beta_j|/\lambda_\beta),$$

in which $K(\cdot)$ is the modified Bessel function of the third kind. The marginal distribution above has an spike at zero and very heavy tails, and places increasing mass near zero when the shape parameter $\lambda_\beta$ decreases. Compared with Laplace prior distribution, the normal-gamma prior has substantially improved performance due to the property of strongly shrinking small coefficients to zero while minimally shrinking large coefficients due to the heavy tails, especially when the sample size is small and the number of covariates is large. In this article, we choose normal-gamma distribution as prior distribution due to its merits compared with Laplace prior distribution.

How to choose the value of $\lambda_\beta$ and $\gamma$ plays an important role in the sparsity estimation. Park and Casella (2008) proposed an empirical Bayes procedure for the hyperparameters of Laplace prior distribution. However, the empirical Bayes approach is very difficult to implement due to the complexity of posterior distribution induced by the normal-gamma prior. Therefore, we takes a fully Bayesian method and suggests data-driven priors.

## 3. Bayesian analysis

In this section, we develop a Bayesian approach for obtaining the estimation of unknown parameters. We consider the application of Markov chain Monte Carlo (MCMC) methods to obtain the Bayesian estimate by drawing samples from the joint posterior distribution. The full conditional distributions in the implementation of the MCMC algorithm involve the prior distributions of unknown parameters.

For the reason we have mentioned in section 2.1, we assume the residuals $\delta_i$ and $\epsilon_{i1}, \ldots, \epsilon_{ip-1}$ are from the asymmetric Laplace distribution in Bayesian analysis. Note that this assumption is by no means based on the belief that the true data follow this specific distribution, but rather the equivalence of maximizing the likelihood function (2) and minimizing (1). From the mixture representation of asymmetric Laplace distribution, the model can be equivalently rewritten as

$$x_{ij} = \mathbf{\Lambda_j} \boldsymbol{\omega}_i + \sqrt{8\sigma_j e_{ij}}\zeta_{ij}, \tag{5}$$

$$y_j = \boldsymbol{\beta}_\omega \boldsymbol{\omega}_i + \boldsymbol{\beta}_z \mathbf{Z}_i + \sqrt{8\sigma_y e_{yi}}\zeta_{yi}, \quad j = 1, \ldots, p-1, \tag{6}$$

where the random variables $\zeta_{ij}$ and $e_{ij}$ are independent, $\zeta_{yi}$ and $e_{yi}$ are independent, and $\zeta_{ij}, \zeta_{yi}$ follow standard normal distribution, and $e_{ij}, e_{yi}$ follow exponential distribution with scale parameter $1/\sigma_j, 1/\sigma_y$ respectively.

### 3.1 Prior specifications

Inspired by the work of the statistician who taken the Bayesian method in researching CFA model or other latent variable model (Lee, 2007), the following conjugate prior distributions

will be used. Let $\mathbf{\Lambda}_j$ be $j$th row of $\mathbf{\Lambda}$. The prior distributions of the unknown parameters in model (3) are given by

$$p(\mathbf{\Lambda}_j) \overset{D}{=} N(\mathbf{\Lambda}_{0j}, \mathbf{H}_{0j}), \quad j = 1, \ldots, p-1,$$

$$p(\mathbf{\Phi}) \overset{D}{=} \text{IW}_q(\rho_0, \mathbf{R}_0),$$

$$p(\sigma_j) \overset{D}{=} \text{Inverse Gamma}(\alpha_{0\sigma j}, \beta_{0\sigma j}), \quad j = 1, \ldots, p-1,$$

where '$p(\cdot) \overset{D}{=}$' is defined as the distribution of $p(\cdot)$ is equal to, $\text{IW}_q(\rho_0, \mathbf{R}_0)$ denotes a $q$-dimensional Inverse-Wishart distribution with degrees of freedom $\rho_0$ and scale matrix $\mathbf{R}_0$, and $\mathbf{\Lambda}_{0j}$, $\mathbf{H}_{0j}$, $\alpha_{0j}$, $\beta_{0j}$ $(j = 1, \ldots, p-1)$, $\rho_0$, $\mathbf{R}_0$ are hyperparameters whose values is prespecified.

For $\beta_j$ $(j = 1, \ldots, q+r)$, we assign a normal-gamma prior as follows:

$$p(\beta_j | \varphi_j) \overset{D}{=} N(0, \varphi_j),$$

$$p(\varphi_j | \lambda_\beta, \gamma^2) \overset{D}{=} \text{Gamma}(\lambda_\beta, 1/2\gamma^2),$$

$$p(\gamma^2 | c, d) \overset{D}{=} \text{Gamma}\left(c, \frac{2}{d}\right).$$

The sparsity parameters $\lambda_\beta$ and $c$ are given exponential prior distribution with mean 1. We impose a vague prior of the form $p(d) \propto (1+d)^{-2}$ on $d$. Furthermore, the prior distribution of $\sigma$ is given as follows

$$p(\sigma_y) \overset{D}{=} \text{Inverse Gamma}(\alpha_{0y\sigma}, \beta_{0y\sigma}).$$

### 3.2 Posterior computation

The posterior distribution of the parameters can be simulated using a Gibbs sampler which is implemented by iteratively sampling observations from the full conditional distributions of the parameters with additional Metropolis-Hastings algorithms for the non-standard conditional distributions. The convergence of MCMC algorithm is monitored by the estimated potential scale reduction (EPSR) values (Gelman, 1996). As suggested by Gelman (1996), the convergence of the MCMC procedure is achieved if all the EPSR values of the unknown parameters are less than 1.2.

## 4. Simulation study

In this section, we examine the empirical performance of the proposed Bayesian method through simulation. We compare the proposed method (NG-Median) with the usual mean regression method which assumes that the non-normal residual terms follow normal distribution (NG-Normal), i.e., $\boldsymbol{\epsilon}_i \sim N(0, \mathbf{\Psi}_{\boldsymbol{\epsilon}})$, $\delta_i \sim N(0, \psi_\delta)$, where $\mathbf{\Psi}_{\boldsymbol{\epsilon}} = \text{diag}(\psi_{\epsilon 1}, \ldots, \psi_{\epsilon p-1})$. The data sets are generated from the latent variable model defined by Model (3) and (4) with $p = 7$, $q = 2$, $r = 6$. Specifically, the latent variables $\boldsymbol{\omega} = (\omega_1, \omega_2)$ are drawn from $N(0, \mathbf{\Phi})$, where $\phi_{11} = \phi_{22} = 1.0$, and $\phi_{12} = \phi_{21} = 0.3$. The fixed covariates $\mathbf{Z}$ follows a multivariate normal distribution $N(\mathbf{0}, \mathbf{\Sigma})$ with $(\mathbf{\Sigma})_{ij} = 0.5^{|i-j|}$. The structure of the loading matrix is defined as follows:

$$\Lambda^T = \begin{pmatrix} 1.0 & \lambda_{21} & \lambda_{31} & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & \lambda_{52} & \lambda_{62} \end{pmatrix},$$

where the zeros and ones are fixed to achieve an identified model, and $\lambda_{21}$, $\lambda_{31}$, $\lambda_{52}$, $\lambda_{62}$ are unknown parameters whose true values are taken as: $\lambda_{21} = \lambda_{31} = \lambda_{52} = \lambda_{62} = 0.8$. We simulate the residual terms $\delta_i$ and $\boldsymbol{\epsilon}_i$ from four possible residual distributions:

**Design 1:** $\epsilon_{i1}, \epsilon_{i3}, \epsilon_{i5}, \delta_i \sim t(3)$; $\epsilon_{i2}, \epsilon_{i4}, \epsilon_{i6} \sim \text{Laplace}(0,1)$.

**Design 2:** $\delta_i$ and $\epsilon_i \sim t(3)$.

**Design 3:** $\epsilon_i \sim \text{Laplace}(0,1)$, $\delta_i \sim \text{lognormal}(0,1)$.

**Design 4:** $\epsilon_i \sim \text{Laplace}(0,1)$, $\delta_i\ (i = 1,\ldots,n-10) \sim \text{lognormal}(0,1)$, $\delta_i\ (i = n-9,\ldots,n) \sim N(30,1)$.

We use $t$ distribution to reflect the heavy-tailed characteristics of data, and lognormal distribution to reflect the highly skewed characteristics of data. Design 4 is used to investigate the sensitivity of the median latent variable model to outliers.

The true values of the $\beta$'s are set as follows:

**Dense case:** $\beta = (0.8, 0.0, 0.5, 0.5, 0.5, 0.5, 0.5, 0.0)$.

**Sparse case:** $\beta = (0.8, 0.0, 0.5, 0.5, 0.0, 0.0, 0.0, 0.0)$.

Based on the above settings, the sample size of $n = 800$ is considered, the simulation study is repeated 100 times. To decide the number of burn-in iterations required for achieving convergence, we run a few chains and found that all the EPSR values of the unknown parameters are less than 1.2 after 5000 iterations. We generate additional 10,000 observations to obtain Bayesian estimates after discarding a burn-in of 6,000 iterations.

In each simulation study, the prior inputs with following hyperparameters are employed for the proposed Bayesian approach. To investigate the sensitivity of the Bayesian results to the choice of prior distributions, we consider the following two different prior choices:

**Prior I:** All the elements in $\mathbf{\Lambda}_{0j}$ $(j = 1,\ldots,p-1)$ are taken to be 0.0, $\mathbf{H}_{0j}$ $(j = 1,\ldots,p-1)$ are diagonal matrices with diagonal elements 1.0; $\rho_0 = 10$, $\mathbf{R}_0 = 6\mathbf{\Phi}^{-1}$; The parameters $\alpha_{0y\sigma}$, $\alpha_{0\sigma j}$ $(j = 1,\ldots,p-1)$ and $\beta_{0y\sigma}$, $\beta_{0\sigma j}$ $(j = 1,\ldots,p-1)$ of the Inverse Gamma distribution are set to be 0.01.

**Prior II:** All the elements in $\mathbf{\Lambda}_{0j}$ $(j = 1,\ldots,p-1)$ are taken to be 1.0, $\mathbf{H}_{0j}$ $(j = 1,\ldots,p-1)$ are diagonal matrices with diagonal elements 10.0; $\rho_0 = 15$, $\mathbf{R}_0 = 11\mathbf{\Phi}^{-1}$; The parameters $\alpha_{0y\sigma}$, $\alpha_{0\sigma j}$ $(j = 1,\ldots,p-1)$ and $\beta_{0y\sigma}$, $\beta_{0\sigma j}$ $(j = 1,\ldots,p-1)$ of the Inverse Gamma distribution are set to be 0.001.

For the mean regression model, we specify the conjugate priors as follows:

$$p(\psi_{\epsilon j}^{-1}) \overset{D}{=} \text{Gamma}(\alpha_{0\epsilon j}, \beta_{0\epsilon j}), \quad j = 1,\ldots,p-1,$$

$$p(\psi_{\delta}^{-1}) \overset{D}{=} \text{Gamma}(\alpha_{0\delta}, \beta_{0\delta}),$$

$$p(\mathbf{\Lambda}_j | \psi_{\epsilon j}) \overset{D}{=} N(\mathbf{\Lambda}_{0j}, \psi_{\epsilon j}\mathbf{H}_{0j}), \quad j = 1,\ldots,p-1,$$

$$p(\mathbf{\Phi}) \overset{D}{=} \text{IW}_q(\rho_0, \mathbf{R}_0).$$

The following two different prior inputs are used:

**Prior I:** All the elements in $\mathbf{\Lambda}_{0j}$ $(j = 1,\ldots,p-1)$ are taken to be 0.0, $\mathbf{H}_{0j}$ $(j = 1,\ldots,p-1)$ are diagonal matrices with diagonal elements 1.0; $\rho_0 = 10$, $\mathbf{R}_0 = 6\mathbf{\Phi}^{-1}$; The parameters $\alpha_{0\epsilon j}$ $(j = 1,\ldots,p-1)$ are set to be 7, $\beta_{0\epsilon j}$ $(j = 1,\ldots,p-1)$ are set to be 6, $\alpha_{0\delta} = \beta_{0\delta}$=0.001.

**Prior II:** All the elements in $\mathbf{\Lambda}_{0j}$ $(j = 1,\ldots,p-1)$ are taken to be 1.0, $\mathbf{H}_{0j}$ $(j = 1,\ldots,p-1)$ are diagonal matrices with diagonal elements 10.0; $\rho_0 = 15$, $\mathbf{R}_0 = 11\mathbf{\Phi}^{-1}$; The parameters $\alpha_{0\epsilon j}$ $(j = 1,\ldots,p-1)$ are set to be 13, $\beta_{0\epsilon j}$ $(j = 1,\ldots,p-1)$ are set to be 12, $\alpha_{0\delta} = \beta_{0\delta}$=0.01.

Let $\beta$ be one of elements of the unknown parameters, $\beta_0$ and $\widehat{\beta}$ are true value and Bayesian estimate of $\beta$, respectively. The performance of the Bayesian method is investigated through the bias (Bias) and root mean squares (RMS) of the estimates and their true values on the basis of 100 replications as follows:

$$\text{Bias of } \widehat{\beta} = \frac{1}{100}\sum_{k=1}^{100}|\widehat{\beta}^{(k)} - \beta_0|, \quad \text{RMS of } \widehat{\beta} = \left\{\sum_{k=1}^{100}(\widehat{\beta}^{(k)} - \beta_0)^2\right\}^{\frac{1}{2}}.$$

In order to evaluate the shrinkage estimation of two methods, we calculate the average numbers of those true coefficients correctly estimated to be 0 (Corr.), and the average numbers of those nonzero coefficients erroneously estimated to be 0 (Inco.).

The simulation results are listed in Tables 1-5. From Tables 1-4, we observe that all of the Bayesian results obtained from our procedure are close to true values of parameters and root mean square (RMS) values of the estimates are reasonably small across the four error distributions for dense and sparse cases while some Bayesian estimates obtained from the mean regression method are biased. As expected, the Bayesian estimation is not sensitive to the prior choice with a fairly large sample size, so only the results of prior I are listed for each group. The results of Bayesian variable selection for each method are presented in Table 5. It can be seen from Table 5 that, in terms of the average correct zero coefficients, our method behaves better than the mean regression approach.

## 5. A real data example

We applied the proposed methodology to a data set from a longitudinal study of polydrug use conducted in five California countries in 2004. The study is designed to access how treatment retention is affected by various variables such as severity of drug use, criminal history, and so on. In this study, the data from self-reported and administrative questionnaires on the retention of drug treatment, drug use history, drug-related crime history, motivation of drug treatment and received service and test were recorded for 1170 subjects at intake, 3-month, and 12-month follow-up interviews. Due to the possible heterogeneity among the participants, the data set is of specific interest to us. There exist six manifest variables which are grouped into two latent variables: 'drug severity' and 'crime'. Therefore, the median confirmatory factor analysis model is utilized to analyze the effects of various factors on response variables 'treatment retention'. The response variable is 'retention (Retent), $y$', which was measured at 12-month follow-up interview and which indicated the days of stay in the treatment. The manifest variables which are correlated with two latent factors include: 'Drug use in past 30 days at intake (drugday30), $x_1$', 'Drug problems in past 30 days at intake (Drgplm30), $x_2$', 'The number of drugs used in past 30 days at intake (DrgN30), $x_3$', 'The age of first arrest (Agefirstarrest), $x_4$', 'The number of incarceration in lifetime at intake (Incar), $x_5$', and 'The number of arrests in lifetime at intake (ArrN), $x_6$'. These variables are considered as continuous. Since $x_1$, $x_2$, and $x_3$ are related with severity of drug use, they are grouped into the latent factor, 'drug severity, $\omega_1$', and since $x_4$, $x_5$, and $x_6$ are related with drug-related crime history, they are grouped into the latent factor, 'crime, $\omega_2$'. The fixed covariates include 'Mtsum01, $z_1$', 'Mtsum02, $z_2$', and 'Mtsum03, $z_3$' which are about treatment motivation, 'services received in past 3 months at TSI 3 month interview

(Servicem), $z_4$', 'The number of drug tests by Tx in past 3 months at TSI 3 month interview (DrugtestTX), $z_5$', 'The number of drug tests by criminal in past 3 months at TSI 3 month interview (DrugtestCJ), $z_6$', 'Treatment mode (Modality), $z_7$', and 'The number of prior treatments the subject has taken at intake (TXcode), $z_8$'. We found through some data analyses that most of the observed variables in longitudinal study of polydrug use were extremely non-normal such as highly skewed or U-shaped, etc. We applied the logarithm transformations to those extremely non-normal data to alleviate the non-normality problem. In addition, the continuous variables are standardized in order to unify the scale. The path diagram to depict the inter-relationships between retention and its important predictors is presented in Figure 1.

A median confirmatory model with two latent variables and six manifest variables is proposed with the following specifications

$$\Lambda^T = \begin{pmatrix} 1.0 & \lambda_{21} & \lambda_{31} & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & \lambda_{52} & \lambda_{62} \end{pmatrix},$$

where the ones and zeros were treated as fixed parameters for model identification.

To examine the sensitivity of the Bayesian results to the inputs of prior distributions, we considered the following two different prior inputs:

**For the proposed method**

**Prior I:** All the elements in $\mathbf{\Lambda}_{0j}$ $(j = 1,\ldots,6)$ are taken to be 0.0, $\mathbf{H}_{0j} = 6\mathbf{I}$ $(j = 1,\ldots,6)$; $\rho_0 = 10$, $\mathbf{R}_0 = 6\mathbf{I}$; the parameters $\alpha_{0y\sigma}$, $\alpha_{0\sigma j}$ $(j = 1,\ldots,6)$ and $\beta_{0y\sigma}$, $\beta_{0\sigma j}$ $(j = 1,\ldots,6)$ are set to be 0.01.

**Prior II:** All the elements in $\mathbf{\Lambda}_{0j}$ $(j = 1,\ldots,6)$ are taken to be 1.0, $\mathbf{H}_{0j} = 10\mathbf{I}$ $(j = 1,\ldots,6)$; $\rho_0 = 15$, $\mathbf{R}_0 = 11\mathbf{I}$; the parameters $\alpha_{0y\sigma}$, $\alpha_{0\sigma j}$ $(j = 1,\ldots,6)$ and $\beta_{0y\sigma}$, $\beta_{0\sigma j}$ $(j = 1,\ldots,6)$ are set to be 0.001.

**For the mean regression method**

**Prior I:** All the elements in $\mathbf{\Lambda}_{0j}$ $(j = 1,\ldots,p-1)$ are taken to be 0.0, $\mathbf{H}_{0j}$ $(j = 1,\ldots,p-1)$ are diagonal matrices with diagonal elements 1.0; $\rho_0 = 10$, $\mathbf{R}_0 = 6\mathbf{\Phi}^{-1}$; The parameters $\alpha_{0\epsilon j}$ $(j = 1,\ldots,p-1)$ and $\alpha_{0\delta}$ are set to be 3, $\beta_{0\epsilon j}$ $(j = 1,\ldots,p-1)$ and $\beta_{0\delta}$ are set to be 2.

**Prior II:** All the elements in $\mathbf{\Lambda}_{0j}$ $(j = 1,\ldots,p-1)$ are taken to be 1.0, $\mathbf{H}_{0j}$ $(j = 1,\ldots,p-1)$ are diagonal matrices with diagonal elements 10.0; $\rho_0 = 15$, $\mathbf{R}_0 = 11\mathbf{\Phi}^{-1}$; The parameters $\alpha_{0\epsilon j}$ $(j = 1,\ldots,p-1)$ and $\alpha_{0\delta}$ are set to be 7, $\beta_{0\epsilon j}$ $(j = 1,\ldots,p-1)$ and $\beta_{0\delta}$ are set to be 6.

After checking the convergence, we found that MCMC chains converged within 6,000 iterations. To be conservative, 10,000 observations generated by the MCMC algorithm were used to obtain Bayesian estimates after discarding 7,000 burn-in iterations.

Table 10 summarizes the Bayesian estimates of all the unknown parameters and their corresponding standard error estimates (SE) for the two methods. We can see from Table 10 that (i) The Bayesian estimation is not sensitive to the prior inputs. (ii) The proposed procedure indicates two latent factor 'drug severity' and 'crime' have strong effect on the response variable 'treatment retention', while the fixed covariates including 'Mtsum01', 'Mtsum02', 'Mtsum03', and 'TXcode' have insignificant effect on the response variable 'treatment retention'. (iii) The mean regression method fails to select any significant variables, which is unrealistic in the real application.

To investigate the influence of the latent variables on response variable, we treat the variables related to the two latent variables as the fixed covariates. Table 11 gives Bayesian estimates of all parameters and their corresponding standard error estimates (SE). Table 11 shows that although the latent variables are significant, the effect of the individual variables which are associated with the latent variables are insignificant. This indicates that the latent variable can pool the information from these multiple individual variables to more accurately reflect the influence of explained variables on the response variable.

## 6. Concluding Remarks

In this article, we develop a Bayesian method for estimation and variable selection under median latent variable model. We consider a normal-gamma prior distribution which extend the double exponential prior of the Bayesian Lasso. On the basis of a theoretic derivation of the asymmetric Laplace distribution and normal-gamma prior distribution as a scale mixture of normal distributions, a Gibbs sampler augmented by a Metropolis-Hastings step is introduced to fit the model. We validated the proposed method through extensive simulation study, and a real example demonstrates that the proposed method yielded satisfactory results in both parameter estimation and variable selection. There exist several directions for future research. First, we assume linear relations among the response variable and the explained variables in this paper which is very limited in practical application. The nonlinear median latent variable model which allows nonlinear relations among the response variable and the explained variables such as the interactions and quadratic terms can be considered. Second, the proposed method assumes normality for the distribution of latent variables. It is very useful to develop more robust methodology to relax the assumption in the future study.

## References


[1] Bentler, P.M. (1983). Some contributions to efficient statistics in structural models: specification and estimation of moment structures. *Psychometrika*. **48**, 493-517.

[2] Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *Annals of Statistics*. **24**, 2350-238.

[3] Burgette, L.F. and Reiter, J.P. (2012). Modeling Adverse Birth Outcomes via Confirmatory Factor Quantile Regression. *Biometrics*. **68**, 92-100.

[4] Cai, J., Song, X., and Hser, Y. (2010). A Bayesian analysis of mixture structural equation models with non-ignorable missing responses and covariates. *Statistics in Medicine*. **29**, 1861-1874.

[5] Cai, J., Song, X., Lam, L., and Ip, E. (2011) A mixture of generalized latent variable models for mixed mode and heterogeneous data. *Computational Statistics & Data Analysis*. **55**, 2889-2907.

[6] Carvalho, C., Polson, N. and Scott, J. (2010). The horseshoe estimator for sparse signals. *Biometrika*. **97**, 465-480.

[7] Chhikara, R.S. and Folks, L.(1989) The inverse gaussian distribution: theory, methodology, and applications. New York: Marcel Dekker.

[8] Devroye, L. (1986). Non-Uniform Random Variate Generation. New York: Springer.

[9] Dunson, D.B. and Taylor, J. A. (2005). Approximate Bayesian inference for quantiles. *Journal of Nonparametric Statistics*. **17**, 385-400.

[10] Fan, J. and Li, R.Z. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*. **96**, 1348-60.

[11] Gelman, A. (1996). Inference and monitoring convergence. In: Gilks, W.R., Richardson, S., Speigelhalter, D.J. (Eds.), Markov chain Monte Carlo in Practice. Chapman & Hall, London, pp. 131-144.

[12] Gelman, A., and Meng, X.L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science*. **13**, 163-185.

[13] Gelman, A., Roberts, G.O., and Gilks, W.R. (1995). Efficient Metropolis Jumping Rules. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 5. Oxford University Press, pp. 599-607.

[14] Griffin, J.E. and Brown, P.J. (2010). Inference with Normal-Gamma prior distributions in regression problems. *Bayesian Analysis*. **5**, 171-188.

[15] Hans, C. (2011). Elastic net regression modeling with the orthant normal prior. *Journal of the American Statistical Association*. **106**, 1383-1393.

[16] Lee, S.Y. (2007). *Structural equation modeling: A Bayesian approach*. New York, NY: Wiley.

[17] Lee, S.Y. and Xia, Y.M. (2006). Maximum likelihood methods in treating outliers and symmetrically heavy-tailed distributions for nonlinear structural equation models with missing data. *Psychometrika*. **71**, 565-585.

[18] Li, Y., Kano, Y., Pan, J., and Song, S. (2012). A criterion-based model comparison statistic for structural equation models with heterogeneous data. *Journal of Multivariate Analysis*. **112**, 92-107.

[19] Kano, Y., Berkane, M. and Bentler, P. M. (1993). Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations. *Journal of the American Statistical Association*. **88**, 135-143.

[20] Koenker, R. and Bassett, G. Jr. (1978). Regression quantiles. *Econometrica*. **46**, 33-50.

[21] Koenker, R. and Machado, J. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*. **94**, 1296-1309.

[22] Kottas, A. and Gelfand, A.E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*. **96**, 1458-1468.

[23] Kottas, A. and Krnja jić, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics*. **36**, 297-319.

[24] Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*. **81**, 1565-1578.

[25] Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*. **103**, 681-6.

[26] Reich, B. J., Bondell, H. D. and Wang, H. J. (2008). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics*. **11**, 337-352.

[27] Shapiro, A. and Browne, M.W. (1987). Analysis of covariance structures under elliptical distributions. *Journal of the American Statistical Association*. **82**, 1092-1097.

[28] Sriram, K., Ramamoorthi, R. V., and Ghosh, P. (2011). Posterior Consistency of Bayesian Quantile Regression Under a Mis-Specified Likelihood Based on Asymmetric Laplace Density.

[29] Song, X.Y., and Lee, S.Y. (2012). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*. Wiley, London.

[30] Song, X.Y., Pan, J.H., Kwok, T., Vandenput, L., Ohlsson, C., and Leung, P.C. (2010). A semiparametric Bayesian approach for structural equation models. *Biometrical Journal*. **52**, 314-332.

[31] Sun, W., Ibrahim, J.G., and Zou, F. (2010). Genome-wide multiple loci mapping in experimental crosses by the iterative adaptive penalized regression. *Genetics*. **185**, 349-59.

[32] Tibshirani, R.(1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society (Series B)*. **58**, 267-288.

[33] Yang, M. and Dunson, D.B. (2010). Bayesian semiparametric structural equation models with latent variables. *Psychometrika*, **75**, 675-693.

[34] Yang, M., Dunson, D.B., Baird, D. (2010). Semiparametric Bayes hierarchical models with mean and variance constraints. *Computational Statistics and Data Analysis*. **54**, 2172-2186.

[35] Yang, Y. and He, X. (2012). Bayesian empirical likelihood for quantile regression. *The Annals of Statistics*. **40**, 1102-1131.

[36] Yu, K. and Moyeed, R.A. (2001). Bayesian quantile regression. *Statistics and Probability Letters*. **54**, 437-447.

[37] Zhang, C. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*. **38**, 894-942.

[38] Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*. **101**, 1418-29.

**Table 1**: Bayesian estimates of two methods for median latent variable model, Design 1.

| Par | NG-Median | | | | NG-Normal | | | |
|---|---|---|---|---|---|---|---|---|
| | Dense | | Sparse | | Dense | | Sparse | |
| | Bias | RMS | Bias | RMS | Bias | RMS | Bias | RMS |
| $\lambda_{21}$ | 0.027 | 0.075 | 0.023 | 0.074 | 0.051 | 0.136 | 0.050 | 0.136 |
| $\lambda_{31}$ | 0.013 | 0.069 | 0.009 | 0.067 | 0.039 | 0.119 | 0.038 | 0.118 |
| $\lambda_{52}$ | 0.011 | 0.089 | 0.012 | 0.088 | 0.078 | 0.173 | 0.077 | 0.172 |
| $\lambda_{62}$ | 0.018 | 0.085 | 0.019 | 0.085 | 0.072 | 0.157 | 0.071 | 0.156 |
| $\beta_1$ | 0.018 | 0.081 | 0.009 | 0.077 | 0.038 | 0.144 | 0.032 | 0.141 |
| $\beta_2$ | 0.005 | 0.066 | 0.002 | 0.058 | 0.009 | 0.103 | 0.006 | 0.097 |
| $\beta_3$ | 0.002 | 0.060 | 0.000 | 0.060 | 0.007 | 0.079 | 0.006 | 0.079 |
| $\beta_4$ | 0.004 | 0.066 | 0.010 | 0.064 | 0.010 | 0.083 | 0.013 | 0.082 |
| $\beta_5$ | 0.005 | 0.068 | 0.010 | 0.051 | 0.007 | 0.088 | 0.010 | 0.076 |
| $\beta_6$ | 0.004 | 0.070 | 0.001 | 0.048 | 0.005 | 0.092 | 0.003 | 0.077 |
| $\beta_7$ | 0.016 | 0.071 | 0.007 | 0.050 | 0.007 | 0.082 | 0.003 | 0.068 |
| $\beta_8$ | 0.003 | 0.047 | 0.001 | 0.040 | 0.005 | 0.069 | 0.007 | 0.065 |
| $\phi_{11}$ | 0.055 | 0.118 | 0.047 | 0.115 | 0.058 | 0.192 | 0.054 | 0.193 |
| $\phi_{12}$ | 0.006 | 0.059 | 0.006 | 0.059 | 0.019 | 0.077 | 0.019 | 0.076 |
| $\phi_{22}$ | 0.041 | 0.131 | 0.042 | 0.132 | 0.083 | 0.197 | 0.084 | 0.197 |

**Table 2**: Bayesian estimates of two methods for median latent variable model, Design 2.

| Par | NG-Median | | | | NG-Normal | | | |
|---|---|---|---|---|---|---|---|---|
| | Dense | | Sparse | | Dense | | Sparse | |
| | Bias | RMS | Bias | RMS | Bias | RMS | Bias | RMS |
| $\lambda_{21}$ | 0.025 | 0.093 | 0.015 | 0.076 | 0.049 | 0.152 | 0.044 | 0.149 |
| $\lambda_{31}$ | 0.030 | 0.094 | 0.010 | 0.085 | 0.053 | 0.139 | 0.049 | 0.137 |
| $\lambda_{52}$ | 0.034 | 0.085 | 0.008 | 0.079 | 0.104 | 0.217 | 0.104 | 0.215 |
| $\lambda_{62}$ | 0.037 | 0.103 | 0.018 | 0.080 | 0.095 | 0.223 | 0.097 | 0.225 |
| $\beta_1$ | 0.020 | 0.089 | 0.034 | 0.103 | 0.054 | 0.150 | 0.043 | 0.144 |
| $\beta_2$ | 0.001 | 0.066 | 0.012 | 0.063 | 0.016 | 0.104 | 0.012 | 0.098 |
| $\beta_3$ | 0.001 | 0.052 | 0.016 | 0.083 | 0.002 | 0.066 | 0.001 | 0.067 |
| $\beta_4$ | 0.008 | 0.069 | 0.012 | 0.075 | 0.003 | 0.084 | 0.007 | 0.082 |
| $\beta_5$ | 0.000 | 0.067 | 0.009 | 0.052 | 0.001 | 0.082 | 0.005 | 0.069 |
| $\beta_6$ | 0.007 | 0.071 | 0.001 | 0.052 | 0.013 | 0.079 | 0.010 | 0.066 |
| $\beta_7$ | 0.005 | 0.061 | 0.004 | 0.063 | 0.001 | 0.073 | 0.002 | 0.060 |
| $\beta_8$ | 0.004 | 0.049 | 0.002 | 0.050 | 0.011 | 0.069 | 0.008 | 0.064 |
| $\phi_{11}$ | 0.054 | 0.144 | 0.028 | 0.123 | 0.057 | 0.192 | 0.048 | 0.189 |
| $\phi_{12}$ | 0.012 | 0.063 | 0.004 | 0.058 | 0.024 | 0.089 | 0.023 | 0.089 |
| $\phi_{22}$ | 0.075 | 0.163 | 0.021 | 0.119 | 0.087 | 0.227 | 0.088 | 0.225 |

**Table 3**: Bayesian estimates of two methods for median latent variable model, Design 3.

| Par | NG-Median | | | | NG-Normal | | | |
|---|---|---|---|---|---|---|---|---|
| | Dense | | Sparse | | Dense | | Sparse | |
| | Bias | RMS | Bias | RMS | Bias | RMS | Bias | RMS |
| $\lambda_{21}$ | 0.018 | 0.077 | 0.015 | 0.076 | 0.062 | 0.140 | 0.062 | 0.139 |
| $\lambda_{31}$ | 0.012 | 0.086 | 0.010 | 0.085 | 0.062 | 0.132 | 0.062 | 0.132 |
| $\lambda_{52}$ | 0.009 | 0.078 | 0.008 | 0.079 | 0.040 | 0.123 | 0.039 | 0.124 |
| $\lambda_{62}$ | 0.020 | 0.079 | 0.018 | 0.080 | 0.057 | 0.144 | 0.056 | 0.144 |
| $\beta_1$ | 0.024 | 0.101 | 0.034 | 0.103 | 0.078 | 0.171 | 0.073 | 0.169 |
| $\beta_2$ | 0.011 | 0.070 | 0.012 | 0.063 | 0.026 | 0.143 | 0.024 | 0.140 |
| $\beta_3$ | 0.012 | 0.082 | 0.016 | 0.083 | 0.008 | 0.124 | 0.009 | 0.124 |
| $\beta_4$ | 0.001 | 0.074 | 0.012 | 0.075 | 0.007 | 0.120 | 0.004 | 0.119 |
| $\beta_5$ | 0.002 | 0.080 | 0.009 | 0.052 | 0.006 | 0.123 | 0.002 | 0.110 |
| $\beta_6$ | 0.006 | 0.085 | 0.001 | 0.052 | 0.002 | 0.126 | 0.001 | 0.110 |
| $\beta_7$ | 0.016 | 0.092 | 0.004 | 0.063 | 0.000 | 0.121 | 0.004 | 0.107 |
| $\beta_8$ | 0.004 | 0.059 | 0.002 | 0.050 | 0.005 | 0.108 | 0.002 | 0.105 |
| $\phi_{11}$ | 0.034 | 0.124 | 0.028 | 0.123 | 0.100 | 0.176 | 0.100 | 0.175 |
| $\phi_{12}$ | 0.005 | 0.057 | 0.004 | 0.058 | 0.023 | 0.067 | 0.023 | 0.067 |
| $\phi_{22}$ | 0.024 | 0.117 | 0.021 | 0.119 | 0.052 | 0.161 | 0.049 | 0.163 |

**Table 4**: Bayesian estimates of two methods for median latent variable model, Design 4.

| Par | NG-Median | | | | NG-Normal | | | |
|---|---|---|---|---|---|---|---|---|
| | Dense | | Sparse | | Dense | | Sparse | |
| | Bias | RMS | Bias | RMS | Bias | RMS | Bias | RMS |
| $\lambda_{21}$ | 0.020 | 0.072 | 0.019 | 0.072 | 0.068 | 0.142 | 0.070 | 0.144 |
| $\lambda_{31}$ | 0.012 | 0.082 | 0.011 | 0.083 | 0.069 | 0.147 | 0.071 | 0.149 |
| $\lambda_{52}$ | 0.009 | 0.079 | 0.008 | 0.078 | 0.039 | 0.123 | 0.040 | 0.124 |
| $\lambda_{62}$ | 0.019 | 0.081 | 0.017 | 0.081 | 0.057 | 0.141 | 0.058 | 0.142 |
| $\beta_1$ | 0.039 | 0.108 | 0.031 | 0.103 | 0.041 | 0.234 | 0.040 | 0.233 |
| $\beta_2$ | 0.027 | 0.105 | 0.022 | 0.093 | 0.017 | 0.232 | 0.016 | 0.230 |
| $\beta_3$ | 0.012 | 0.095 | 0.016 | 0.098 | 0.012 | 0.169 | 0.012 | 0.170 |
| $\beta_4$ | 0.016 | 0.106 | 0.028 | 0.106 | 0.002 | 0.185 | 0.002 | 0.183 |
| $\beta_5$ | 0.007 | 0.109 | 0.018 | 0.078 | 0.015 | 0.159 | 0.019 | 0.147 |
| $\beta_6$ | 0.006 | 0.111 | 0.001 | 0.078 | 0.005 | 0.182 | 0.007 | 0.168 |
| $\beta_7$ | 0.023 | 0.107 | 0.003 | 0.076 | 0.011 | 0.154 | 0.005 | 0.141 |
| $\beta_8$ | 0.021 | 0.085 | 0.010 | 0.071 | 0.023 | 0.167 | 0.019 | 0.163 |
| $\phi_{11}$ | 0.047 | 0.123 | 0.044 | 0.122 | 0.105 | 0.191 | 0.108 | 0.193 |
| $\phi_{12}$ | 0.006 | 0.057 | 0.006 | 0.057 | 0.024 | 0.068 | 0.025 | 0.068 |
| $\phi_{22}$ | 0.023 | 0.120 | 0.021 | 0.119 | 0.052 | 0.161 | 0.053 | 0.162 |

**Table 5**: Bayesian variable selection results of two methods for median latent variable model

| Par | Error | NG-Median | | | | NG-Normal | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | prior I | | prior II | | prior I | | prior II | |
| | | Corr. | Inco. | Corr. | Inco. | Corr. | Inco. | Corr. | Inco. |
| dense | design 1 | 1.860 | 0.000 | 1.850 | 0.000 | 1.600 | 0.000 | 1.610 | 0.000 |
| dense | design 2 | 1.810 | 0.000 | 1.800 | 0.000 | 1.540 | 0.000 | 1.590 | 0.000 |
| dense | design 3 | 1.760 | 0.000 | 1.790 | 0.000 | 1.180 | 0.000 | 1.200 | 0.000 |
| dense | design 4 | 1.510 | 0.000 | 1.500 | 0.000 | 0.870 | 0.080 | 0.860 | 0.080 |
| sparse | design 1 | 4.810 | 0.000 | 4.820 | 0.000 | 4.160 | 0.000 | 4.170 | 0.000 |
| sparse | design 2 | 4.740 | 0.000 | 4.750 | 0.000 | 4.270 | 0.000 | 4.270 | 0.000 |
| sparse | design 3 | 4.620 | 0.000 | 4.620 | 0.000 | 3.260 | 0.000 | 3.240 | 0.000 |
| sparse | design 4 | 4.160 | 0.000 | 4.110 | 0.000 | 2.440 | 0.030 | 2.430 | 0.040 |

**Table 6**: Bayesian estimates and standard error of parameters for a longitudinal study of polydrug.

| | NG-QR | | | | NG-Normal | | | |
|---|---|---|---|---|---|---|---|---|
| | prior I | | prior II | | prior I | | prior II | |
| Par | EST | SE | EST | SE | EST | SE | EST | SE |
| $\lambda_{21}$ | 0.936 | 0.004 | 0.937 | 0.004 | 0.879 | 0.022 | 0.884 | 0.022 |
| $\lambda_{31}$ | 0.954 | 0.011 | 0.954 | 0.011 | 1.062 | 0.015 | 1.065 | 0.016 |
| $\lambda_{52}$ | 0.118 | 0.012 | 0.118 | 0.013 | 0.969 | 0.042 | 0.978 | 0.044 |
| $\lambda_{62}$ | 0.197 | 0.010 | 0.198 | 0.011 | 2.428 | 0.121 | 2.510 | 0.129 |
| $\beta_1$ | -0.145* | 0.040 | -0.149* | 0.041 | -0.018 | 0.022 | -0.020 | 0.023 |
| $\beta_2$ | 0.193* | 0.021 | 0.195* | 0.022 | -0.011 | 0.028 | -0.009 | 0.029 |
| $\beta_3$ | -0.017 | 0.027 | -0.016 | 0.028 | 0.004 | 0.006 | 0.004 | 0.006 |
| $\beta_4$ | -0.028 | 0.034 | -0.027 | 0.033 | -0.007 | 0.006 | -0.008 | 0.006 |
| $\beta_5$ | 0.010 | 0.032 | 0.009 | 0.030 | 0.008 | 0.006 | 0.008 | 0.006 |
| $\beta_6$ | 0.109* | 0.033 | 0.109* | 0.033 | 0.054 | 0.009 | 0.054 | 0.009 |
| $\beta_7$ | 0.140* | 0.032 | 0.139* | 0.033 | 0.037 | 0.006 | 0.038 | 0.006 |
| $\beta_8$ | 0.183* | 0.030 | 0.183* | 0.031 | 0.051 | 0.007 | 0.051 | 0.007 |
| $\beta_9$ | -0.211* | 0.037 | -0.213* | 0.036 | -0.089 | 0.062 | -0.088 | 0.062 |
| $\beta_{10}$ | -0.049 | 0.041 | -0.049 | 0.041 | -0.025 | 0.010 | -0.025 | 0.010 |
| $\phi_{11}$ | 2.014 | 0.084 | 2.006 | 0.083 | 1.823 | 0.083 | 1.805 | 0.085 |
| $\phi_{12}$ | 2.630 | 0.138 | 2.614 | 0.135 | 0.993 | 0.056 | 0.978 | 0.057 |
| $\phi_{22}$ | 5.998 | 0.350 | 5.939 | 0.352 | 0.822 | 0.064 | 0.786 | 0.066 |

* marks relatively large values indicating the significance of the corresponding term.

**Table 7**: Bayesian estimates and standard error of parameters for a longitudinal study of polydrug without latent variables.

| | NG-QR | |
|---|---|---|
| Par | EST | SE |
| $\beta_1$ | -0.061 | 0.058 |
| $\beta_2$ | 0.023 | 0.032 |
| $\beta_3$ | -0.089 | 0.062 |
| $\beta_4$ | 0.069 | 0.035 |
| $\beta_5$ | -0.042 | 0.037 |
| $\beta_6$ | -0.060 | 0.056 |
| $\beta_7$ | -0.004 | 0.024 |
| $\beta_8$ | 0.022 | 0.037 |
| $\beta_9$ | 0.037 | 0.042 |
| $\beta_{10}$ | 0.147* | 0.032 |
| $\beta_{11}$ | 0.132* | 0.033 |
| $\beta_{12}$ | 0.209* | 0.033 |
| $\beta_{13}$ | -0.162* | 0.039 |
| $\beta_{14}$ | -0.029 | 0.040 |

* marks relatively large values indicating the significance of the corresponding term.

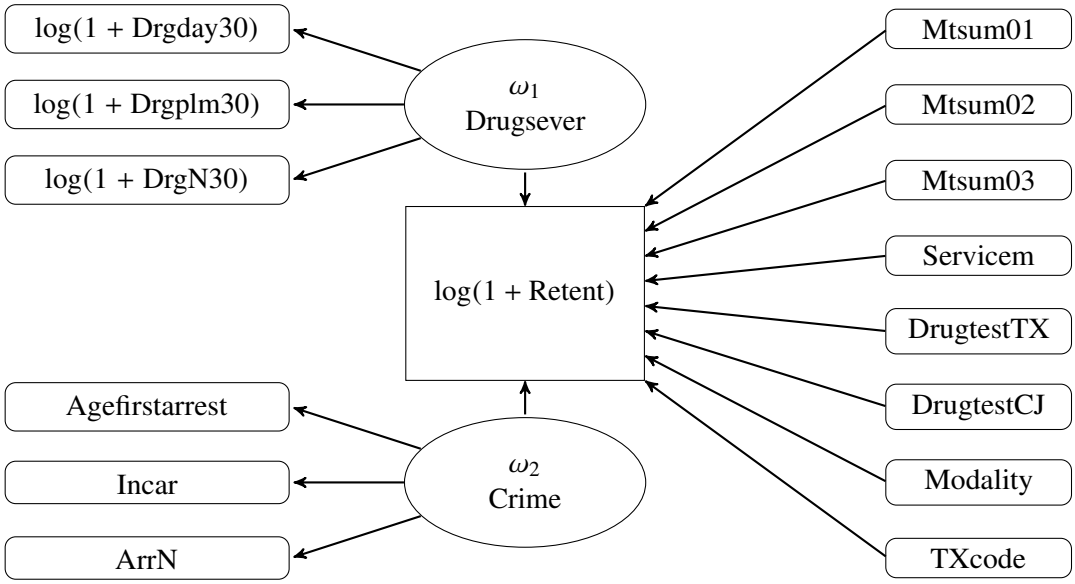


**Figure 1**: The path diagram of the median confirmatory factor analysis model in the illustrative example.