Hyper-parameter selection in Bayesian structural equation models

Hirose, Kei
九州大学大学院数理学府

Kawano, Shuichi
Graduate School of Mathematics, Kyushu University

Miike, Daisuke
Graduate School of Mathematics, Kyushu University

Konishi, Sadanori
Faculty of Mathematics, Kyushu University

https://hdl.handle.net/2324/16257

出版情報：MI Preprint Series. 2010-3, 2010-01-18. 九州大学大学院数理学研究院
バージョン：

MI Preprint Series

Kyushu University
The Global COE Program
Math-for-Industry Education & Research Hub

Hyper-parameter selection in Bayesian structural equation models

Kei Hirose, Shuichi Kawano, Daisuke Miike & Sadanori Konishi

MI 2010-3

(Received January 18, 2010)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Hyper-parameter selection in Bayesian structural equation models

Kei Hirose[1*], Shuichi Kawano[1*], Daisuke Miike[1†] and Sadanori Konishi[2]

[1] *Graduate School of Mathematics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.*

[2] *Faculty of Mathematics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.*

k-hirose@math.kyushu-u.ac.jp (Hirose, K.),
s-kawano@math.kyushu-u.ac.jp (Kawano, S.),
konishi@math.kyushu-u.ac.jp (Konishi, S.).

Abstract

In the structural equation models, the maximum likelihood estimates of error variances can often turn out to be zero or negative. In order to overcome this problem, we take a Bayesian approach by specifying a prior distribution for variances of error variables. Crucial issues in this modeling procedure include the selection of hyper-parameters in the prior distribution. Choosing these parameters can be viewed as a model selection and evaluation problem. We derive a model selection criterion for evaluating a Bayesian structural equation model. Monte Carlo simulations are conducted to investigate the effectiveness of our proposed modeling procedure. A real data example is also given to illustrate our procedure.

Key Words: Bayesian approach, Improper solutions, Model selection criterion, Prior distribution, Structural equation modeling

1 Introduction

Structural equation models that include the factor analysis model and path analysis play an essential role in various fields of research such as social, educational, behavioral and biological sciences, public health, and medical research (see, e.g., Bentler and Stein, 1992; Jörekog and Sörbom, 1996; Pugesek *et al.*, 2003; Xiong *et al.*, 2004; Liu *et al.*, 2008). The structural equation model is usually estimated by maximum likelihood methods under the assumption that the observations are normally distributed. In practice, however, the maximum likelihood estimates of error variances can often turn out to be

*Research Fellow of the Japan Society for the Promotion of Science.

†Present address: Soft service Co., Ltd. 3-3-22, Hakataeki-Higashi, Hakata-ku, Fukuoka 812-0013, Japan.

zero or negative. Such estimates are known as improper solutions, and many authors have studied these inappropriate estimates both from a theoretical point of view and also by means of numerical examples (see, e.g., van Driel, 1978; Anderson and Gerbing, 1984; Boomsma, 1985; Gerbing and Anderson, 1987; Kano, 1998; Chen *et al.*, 2001; Flora and Curran, 2004). In order to prevent the occurrence of improper solutions in structural equation model, we employ a Bayesian approach by specifying a prior distribution for error variances.

An essential point in the Bayesian approach is the choice of a prior distribution. In the factor analysis model which is the special case of structural equation model, some prior distributions have been proposed by earlier authors (see, e.g., Martin and McDonald, 1975; Akaike, 1987; Hirose *et al.*, 2008). Akaike (1987) introduced a prior distribution using the information extracted from the knowledge of likelihood function. However, it is difficult to apply his prior distribution to the structural equation models. Hirose *et al.* (2008) derived a prior distribution according to the basic idea given by Akaike (1987), and numerical examples showed that their prior distributions can prevent the occurrence of improper solutions. In this paper, we use an inverse exponential distribution for error variances, which is based on the prior distribution given by Hirose *et al.* (2008), and estimate parameters by posterior modes.

In the Bayesian structural equation models, the hyper-parameters in the prior distribution are often subjectively given. However, the modeling procedure based on such subjective hyper-parameters does not always provide appropriate estimates of parameters. Therefore, crucial points are the selection processes of the hyper-parameters. The selection methods with MCMC outputs (see, e.g., Chib, 1995; Diciccio *et al.*, 1997; Chib and Jeliazkov, 2001; Spiegelhalter *et al.*, 2002) could be used in the structural equation models. Although the MCMC-based selection process is attractive, we take a different approach since it sometimes requires much computational load. We view the selection of the hyper-parameters as a model selection and evaluation problem, and derive a model selection criterion from a Bayesian point of view (Konishi *et al.*, 2004) for evaluating Bayesian structural equation models. The proposed modeling procedure is investigated by analyzing Monte Carlo simulations and a real data example. Numerical results show that our modeling strategy prevents the occurrence of improper solutions and often yields stable estimates.

The remainder of this paper is organized as follows: Section 2 describes maximum likelihood methods for structural equation model. In Section 3, we introduce a Bayesian structural equation modeling. Section 4 describes Monte Carlo simulations to investigate the performance of our modeling procedure. Section 5 illustrates the proposed procedure with a real data example. Some concluding remarks are given in Section 6.

2 Maximum likelihood procedure for structural equation model

A number of models for the analysis of covariance structure, such as LISREL (Bock and Bargmann, 1966; Jöreskog, 1970), EQS (Bentler and Weeks, 1980) and RAM (McArdle, 1980; McArdle and McDonald, 1984), have been proposed. In this paper, we use the RAM model since the description of this model is quite simple, and it generalizes the LISREL and EQS.

First, we define p-dimensional observable random vector, m-dimensional latent variables and p-dimensional error variables given in the following:

$\boldsymbol{f} = (f_1, \cdots, f_m)'$: m-dimensional latent random vector,
$\boldsymbol{x} = (x_1, \cdots, x_p)'$: p-dimensional observable random vector,
$\boldsymbol{t} = (\boldsymbol{f}', \boldsymbol{x}')'$: $q\ (= m + p)$-dimensional structural variables that include latent variables and observable variables, and they satisfy $\mathrm{E}[\boldsymbol{t}] = \boldsymbol{0}$,
$\boldsymbol{d} = (d_1, \cdots, d_m)'$: m-dimensional error variables for latent variables $\boldsymbol{f}$,
$\boldsymbol{e} = (e_1, \cdots, e_p)'$: p-dimensional error variables for observable variables $\boldsymbol{x}$,
$\boldsymbol{u} = (\boldsymbol{d}', \boldsymbol{e}')'$: q-dimensional error variables with $\mathrm{E}[\boldsymbol{u}] = \boldsymbol{0}$ and $\mathrm{cov}[\boldsymbol{u}] = \Sigma_{\boldsymbol{u}} = (\sigma^u_{ij})$. Assume that unknown error variances in $\Sigma_{\boldsymbol{u}}$ are $\sigma^u_{v_1}, \cdots, \sigma^u_{v_{qv}}$.

The structure between latent variables and observable variables in the RAM model is given by

$$\boldsymbol{t} = A\boldsymbol{t} + \boldsymbol{u}, \tag{1}$$

where $A = (a_{ij})$ is a $q \times q$-coefficient matrix for structural variables.

Next, we calculate the variance-covariance matrix of $\boldsymbol{x}$. Suppose that there exists an inverse matrix $T = (I_q - A)^{-1}$, where I_q is a $q \times q$ identity matrix. From Equation (1), we have

$$\boldsymbol{t} = T\boldsymbol{u}. \tag{2}$$

The observable random vector $\boldsymbol{x}$ is then given by

$$\boldsymbol{x} = GT\boldsymbol{u}, \tag{3}$$

where G is a $p \times q$-matrix which extracts the observable variables from the structural variables: $G = [\boldsymbol{O}_{p\times m} \quad I_p]$, with $\boldsymbol{O}_{p\times m}$ being $p \times m$ 0-matrix. Hence, the variance-covariance matrix of $\boldsymbol{x}$ is given by

$$\Sigma(\boldsymbol{\theta}) = GT\Sigma_{\boldsymbol{u}}T'G', \tag{4}$$

where $\boldsymbol{\theta}$ is a k-dimensional unknown parameter vector. The unknown parameters in the structural equation models are the coefficient matrix A and a lower triangular part of variance-covariance matrix $\Sigma_{\boldsymbol{u}}$. Note that most of the elements of A and $\Sigma_{\boldsymbol{u}}$ are fixed by 0 or 1 since researchers make such a hypothesis. Hence, the parameter vector $\boldsymbol{\theta}$ are constructed by eliminating these fixed parts.

The structural equation model is usually estimated by the maximum likelihood procedure. Suppose that we have a random sample of N observations $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ from the p-dimensional normal population $N_p(\boldsymbol{0}, GT\Sigma_{\boldsymbol{u}}T'G')$. The log-likelihood function is then given by

$$\log f(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N|\boldsymbol{\theta}) = -\frac{N}{2}\left\{p\log(2\pi) + \log|\Sigma(\boldsymbol{\theta})| + \mathrm{tr}(\Sigma(\boldsymbol{\theta})^{-1}S)\right\}, \tag{5}$$

where S is a sample variance-covariance matrix

$$S = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n\boldsymbol{x}_n'. \tag{6}$$

The maximum likelihood estimates of $\boldsymbol{\theta}$ are given as the solutions of

$$\frac{\partial \log f(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{0}. \tag{7}$$

Since the solutions cannot be expressed in a closed form, the quasi-Newton's method is usually used to obtain the maximum likelihood estimates.

In practice, however, the maximum likelihood estimates of error variances can often turn out to be zero or negative, which have been called improper solutions. In order to overcome this difficulty, we take a Bayesian approach by specifying a prior distribution for error variances.

3 Bayesian structural equation modeling

In this section, we investigate the prior distribution for the variances of error variables, and then illustrate a selection procedure of the hyper-parameters in the prior distribution.

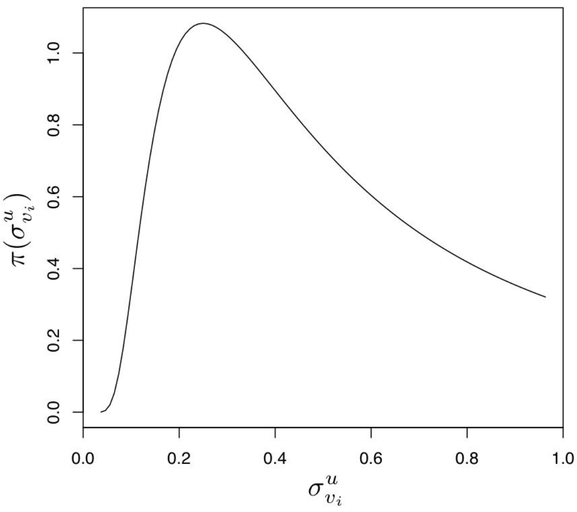

Figure 1: The inverse exponential distribution when $N = 100$ and $\lambda_i = 0.005$.

3.1 Prior distributions

An important point in the Bayesian structural equation models is the selection of a prior distribution. In the factor analysis model which is the special case of the structural equation model, Hirose *et al.* (2008) derived exponential distributions for the inverses of unique variances, and numerical examples showed that their prior distributions can prevent the occurrence of improper solutions. On the basis of their prior distributions, we use an inverse exponential distribution for error variances given by

$$\pi(\boldsymbol{\theta}|\boldsymbol{\lambda}) = \prod_{i=1}^{q_v} \frac{N\lambda_i}{(\sigma^u_{v_i})^2} \exp\left(-\frac{N\lambda_i}{\sigma^u_{v_i}}\right) \qquad (\sigma^u_{v_i} > 0 \ \text{ for } \ i = 1, \cdots, q_v), \tag{8}$$

where $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_{q_v})'$ is a q_v-dimensional hyper-parameter vector with $\lambda_i > 0$ ($i = 1, \cdots, q_v$). Note that this prior distribution is an inverse gamma prior distribution

$$\pi(\sigma^u_{v_i}|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^u_{v_i})^{-(\alpha+1)} \exp\left(-\frac{\beta}{\sigma^u_{v_i}}\right) \tag{9}$$

with $\alpha = 1$ and $\beta = N\lambda_i$, where $\Gamma(\cdot)$ is a gamma function. Figure 1 shows the inverse exponential distribution when $N = 100$ and $\lambda_i = 0.005$. It can be seen from Figure 1 that a probability that $\sigma^u_{v_i}$ is close to 0 is extremely low. Thus, this prior distribution may prevent the occurrence of improper solutions.

The posterior distribution based on the prior distribution given by (8) is

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N;\boldsymbol{\lambda}) &= \frac{f(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})}{\displaystyle\int f(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})d\boldsymbol{\theta}} \\ &\propto f(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda}). \end{aligned} \tag{10}$$

In this paper, the parameters $\boldsymbol{\theta}$ are estimated through modes of the posterior distribution. The procedure is equivalent to obtain estimates by maximizing the penalized log-likelihood function

$$\ell_{\boldsymbol{\lambda}}(\boldsymbol{\theta}) = \log f(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N|\boldsymbol{\theta}) - H_{N,\boldsymbol{\lambda}}(\boldsymbol{\theta}), \tag{11}$$

where $H_{N,\boldsymbol{\lambda}}(\boldsymbol{\theta})$ is a penalty term given by the following:

$$H_{N,\boldsymbol{\lambda}}(\boldsymbol{\theta}) = \sum_{i=1}^{q_v}\left(\frac{N\lambda_i}{\sigma_{v_i}^u} + 2\log\sigma_{v_i}^u\right), \tag{12}$$

and the hyper-parameters $\boldsymbol{\lambda}$ can be considered as regularization parameters. Since it is difficult to obtain the parameters that maximize the function in (11) analytically, we use a quasi-Newton's method to obtain the maximum penalized likelihood estimates.

3.2 Model selection criterion

This subsection describes a selection process of hyper-parameters in the prior distribution. For example, the maximum likelihood estimate of $\sigma_{v_1}^u$ becomes negative. When the value of λ_1 is very small, the penalized maximum likelihood estimates of $\sigma_{v_1}^u$ may be close to 0. On the other hand, when the value of λ_1 is large, the penalized maximum likelihood estimates of $\sigma_{v_1}^u$ also becomes large. Therefore, a crucial aspect of model construction is the choice of the regularization parameter $\lambda_1,\cdots,\lambda_{q_v}$. In this paper, we derive a model selection criterion GBIC (Konishi *et al.*, 2004) for evaluating Bayesian structural equation models. The proposed procedure selects the values of hyper-parameters objectively.

The generalized Bayesian information criterion (GBIC), proposed by Konishi *et al.* (2004), enables us to choose adjusted hyper-parameters $\lambda_1,\cdots,\lambda_{q_v}$ by extending the Bayesian information criterion (BIC) proposed by Schwarz (1978). The basic idea of BIC is to select a model from a set of candidate models by maximizing the posterior probability. The BIC only deals with models estimated by the maximum likelihood method, whereas the model selection criterion GBIC can be applied to models estimated by the maximum penalized likelihood method. For model selection criteria we refer to Konishi and Kitagawa (2008) and references given therein.

The model selection criterion GBIC for evaluating the Bayesian structural equation model is given by

$$\mathrm{GBIC} = -k\log(2\pi) + k\log N + \log|J_{\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}})| + N\left\{p\log(2\pi) + \log|\Sigma(\hat{\boldsymbol{\theta}})| + \mathrm{tr}(\Sigma(\hat{\boldsymbol{\theta}})^{-1}S)\right\}$$
$$-2\sum_{i=1}^{q_v}\log\left\{\frac{N\lambda_i}{(\hat{\sigma}_{v_i}^u)^2}\right\} + 2\sum_{i=1}^{q_v}\frac{N\lambda_i}{\hat{\sigma}_{v_i}^u}, \qquad (13)$$

where $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}_{v_i}^u$ $(i = 1, \cdots, q_v)$ are the posterior modes, and $J_{\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}})$ is $k \times k$ matrix given by

$$J_{\boldsymbol{\lambda}}(\hat{\boldsymbol{\theta}}) = -\frac{1}{N}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\left\{\log f(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N|\boldsymbol{\theta}) + \log\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})\right\}\Bigg|_{\hat{\boldsymbol{\theta}}}\right]. \qquad (14)$$

The derivation of $J_{\boldsymbol{\lambda}}(\boldsymbol{\theta})$ is given in Appendix A.

When we have several candidates for hyper-parameter vectors $\boldsymbol{\lambda} = (\lambda_{v_1}, \cdots, \lambda_{v_{q_v}})'$, the GBIC is calculated for each candidate, and then the model which minimizes the value of GBIC is selected. However, if the dimension of hyper-parameters q_v is large, it is difficult to calculate the GBIC for all possible candidates because it requires extremely computational load. Thus, we restrict the number of hyper-parameters in the following way:

PMLE$_1$: All error variances have the same hyper-parameter λ_1.

PMLE$_2$: The error variances for observable variables have a hyper-parameter λ_1 and those for latent variables have a hyper-parameter λ_2.

It can be seen that PMLE$_1$ is useful when all of the variances have similar values while PMLE$_2$ can be used when the error variances for observable variables and those for latent variables are completely different.

4 Monte Carlo simulations

Monte Carlo simulations are conducted to investigate the performance of our proposed procedure. In this simulation study, latent variables are ξ, η, and observable variables are given by y_1, y_2, x_1, x_2. The true model is given by

$$\begin{bmatrix} \xi \\ \eta \\ y_1 \\ y_2 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \eta \\ y_1 \\ y_2 \\ x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_\xi \\ \varepsilon_\eta \\ \varepsilon_{y_1} \\ \varepsilon_{y_2} \\ \varepsilon_{x_1} \\ \varepsilon_{x_2} \end{bmatrix}, \tag{15}$$

and the true variance-covariance matrix $\Sigma_{\boldsymbol{u}}$ of $\boldsymbol{u}$ is

$$\Sigma_{\boldsymbol{u}} = \mathrm{diag}(1.0, 0.6, 0.1, 0.7, 0.5, 0.5). \tag{16}$$

To estimate the true model, we assume the hypothetic model

$$\begin{bmatrix} \xi \\ \eta \\ y_1 \\ y_2 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & a_{y_2} & 0 & 0 & 0 & 0 \\ a_{x_1} & 0 & 0 & 0 & 0 & 0 \\ a_{x_2} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \eta \\ y_1 \\ y_2 \\ x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_\xi \\ \varepsilon_\eta \\ \varepsilon_{y_1} \\ \varepsilon_{y_2} \\ \varepsilon_{x_1} \\ \varepsilon_{x_2} \end{bmatrix}, \tag{17}$$

with variance-covariance matrix

$$\Sigma_{\boldsymbol{u}} = \mathrm{diag}(1, \sigma_\eta, \sigma_{y_1}, \sigma_{y_2}, \sigma_{x_1}, \sigma_{x_2}). \tag{18}$$

In this case, the parameter vector is given by $\boldsymbol{\theta} = (\gamma, a_{y_2}, a_{x_1}, a_{x_2}, \sigma_\eta, \sigma_{y_1}, \sigma_{y_2}, \sigma_{x_1}, \sigma_{x_2})'$. The true variance-covariance matrix $\Sigma = GT\Sigma_{\boldsymbol{u}}T'G'$ is calculated by using Equations (15) and (16), and then the data were generated 100 times with sample size N ($N = 100, 150, 200$).

We compare the performance of our proposed procedure with that of maximum likelihood method. Table 1 shows the frequency of improper solutions, mean squared error (MSE) for parameters, mean value of hyper-parameters ($\overline{\lambda_1}$ and $\overline{\lambda_2}$), and the mean value of the GBIC ($\overline{\mathrm{GBIC}}$) for MLE, PMLE_1 and PMLE_2. The mean squared error (MSE) is given by

$$\mathrm{MSE} = \frac{1}{100} \sum_{d=1}^{100} \|\hat{\boldsymbol{\theta}}^{(d)} - \boldsymbol{\theta}_0\|^2, \tag{19}$$

where $\boldsymbol{\theta}_0$ are true values of $\boldsymbol{\theta}$, i.e. $\boldsymbol{\theta}_0 = (0.5, 0.6, 0.7, 0.7, 0.6, 0.1, 0.7, 0.5, 0.5)'$, and $\hat{\boldsymbol{\theta}}^{(d)}$ are the estimates of parameters for d-th dataset. From Table 1, we can see that each procedure becomes better in terms of minimizing the MSE as N increase. For each N, we obtained

Table 1: Frequency of improper solutions (Frequency), mean squared error for parameters (MSE), mean value of hyper-parameters ($\overline{\lambda_1}$ and $\overline{\lambda_2}$), and the mean value of the GBIC ($\overline{\text{GBIC}}$)

	procedure	Frequency	MSE ($\times 10^1$)	$\overline{\lambda_1}$ ($\times 10^3$)	$\overline{\lambda_2}$ ($\times 10^3$)	$\overline{\text{GBIC}}$ ($\times 10^{-3}$)
	MLE	39	6.688	—	—	—
$N = 100$	PMLE_1	0	2.267	3.880	—	1.077
	PMLE_2	0	2.516	4.112	3.836	1.077
	MLE	24	4.223	—	—	—
$N = 150$	PMLE_1	0	1.485	2.588	—	1.612
	PMLE_2	0	1.563	2.922	2.615	1.612
	MLE	36	2.109	—	—	—
$N = 200$	PMLE_1	0	0.869	1.928	—	2.139
	PMLE_2	0	0.879	2.420	1.808	2.139

improper solutions several times for maximum likelihood procedure, whereas our proposed method prevented the occurrence of improper solutions for all datasets. Moreover, the MSE of PMLE_1 is much smaller than that of MLE, which means the Bayesian approach yields more stable estimates than maximum likelihood technique. In addition, the values of $\overline{\lambda_1}$, $\overline{\lambda_2}$ for PMLE_2 are almost the same, and they are also similar to the value of $\overline{\lambda_1}$ for PMLE_1. Thus, it seems that PMLE_1 and PMLE_2 selected almost the same models.

Tables 2 and 3 show the mean squared error (MSE) for each parameter. When we compare the MSE of PMLE_1 with that of MLE, a large difference occurred in σ_η and σ_{y_1}. Regarding the σ_{y_1}, we obtained improper solutions several times since the true value of σ_{y_1} is relatively small compared with other parameters of error variances. Consequently, the maximum likelihood estimate of σ_{y_1} is very unstable. On the other hand, the proposed procedures PMLE_1 or PMLE_2 produced more stable estimates than MLE. For error variance σ_η, the proposed methods also provide much better estimates than MLE.

As a result, our proposed method prevents the occurrence of improper solutions and also yields stable estimates. Also, the result of PMLE_1 is very similar to that of PMLE_2. In structural equation models, the observable variables and exogenous variables are usually normalized, and thus they may have similar values, which may cause the similarity of results between PMLE_1 and PMLE_2.

Table 2: Mean squared error (MSE) for each parameter of coefficients

	procedure	γ ($\times 10^3$)	a_{y_2} ($\times 10^2$)	a_{x_1} ($\times 10^2$)	a_{x_2} ($\times 10^2$)
	MLE	8.738	5.479	1.435	1.255
$N = 100$	PMLE_1	8.813	5.011	1.320	1.138
	PMLE_2	9.555	5.985	1.381	1.178
	MLE	8.182	2.796	0.849	0.871
$N = 150$	PMLE_1	8.671	2.091	0.905	0.936
	PMLE_2	8.873	2.192	0.941	0.964
	MLE	4.649	1.848	0.446	0.682
$N = 200$	PMLE_1	4.666	1.005	0.503	0.712
	PMLE_2	4.688	0.960	0.533	0.756

Table 3: Mean squared error (MSE) for each parameter of error variances

	procedure	σ_η ($\times 10^1$)	σ_{y_1} ($\times 10^1$)	σ_{y_2} ($\times 10^2$)	σ_{x_1} ($\times 10^2$)	σ_{x_2} ($\times 10^2$)
	MLE	2.545	2.463	2.047	3.182	2.523
$N = 100$	PMLE_1	0.292	0.411	2.465	2.668	2.154
	PMLE_2	0.361	0.459	2.418	2.806	2.243
	MLE	1.691	1.551	1.069	1.662	1.754
$N = 150$	PMLE_1	0.257	0.285	1.152	1.677	1.807
	PMLE_2	0.288	0.298	1.147	1.755	1.872
	MLE	0.780	0.686	0.708	0.861	1.409
$N = 200$	PMLE_1	0.135	0.154	0.739	0.940	1.435
	PMLE_2	0.139	0.144	0.703	0.997	1.540

5 Real data example

We applied the proposed modeling procedure to the slump of personal consumption dataset (Toyoda, 1998). This data set was surveyed from 10/06/1998 to 15/07/1998. During this long period, Japan was suffering a slump. The aim of this analysis is to find out the cause of the recession by conducting a causality analysis. This dataset consists of $N = 405$ samples and $p = 21$ observable variables $X_1, \ldots, X_{21}$, and Toyoda (1998) considered 8 latent variables $F_1, \ldots, F_8$. The 8 latent variables and corresponding observable variables are given in Appendix B.

We made a hypothetic model based on Toyoda (1998), which is given in Figure 2. First, the model of Figure 2 is estimated by maximum likelihood procedure, which is given in Figure 3. This procedure produced improper solutions since the error variance for X_{15} was -0.046. The standard error of error variance for X_{15} was 0.649, and thus the 95% confidence interval includes 0. This means the cause of improper solutions might be sampling fluctuation (see, e.g., van Driel, 1978; Chen *et al.*, 2001).

In order to prevent the occurrence of improper solutions, we applied the proposed procedures (PMLE_1 and PMLE_2) to this dataset. The estimates in PMLE_1 and PMLE_2 were respectively given in Figure 4 and Figure 5. The result of GFI, AGFI and GBIC and corresponding hyper-parameters λ_1 and λ_2 for each procedure is also given in Table 4.

From Figure 4, the penalized maximum likelihood estimate of error variance for X_{15} was positive, which means the proposed procedure prevented the occurrence of improper solutions. Additionally, Figure 4 and 5 indicate the estimates in PMLE_2 and PMLE_1 are very similar. Moreover, the result of GFI and AGFI are also almost the same. This means it is not necessary to assume that error variances for observable variables and those for latent variables have different hyper-parameters.

6 Concluding and remarks

In the structural equation modeling, the maximum likelihood estimates of error variances can turn out be zero or negative. In order to overcome this difficulty, the Bayesian approach is employed by specifying a prior distribution for error variances. Crucial issues in this modeling procedure include the choice of hyper-parameters in the prior distribution. We derived a model selection criterion from a Bayesian point of view to select these

Table 4: The result of GFI, AGFI and GBIC and corresponding hyper-parameters λ_1 and λ_2 for MLE, PMLE_1 and PMLE_2 for the hypothetic model given in Figure 2.

	MLE	PMLE_1	PMLE_2
GFI	0.9037	0.9038	0.9038
AGFI	0.8736	0.8738	0.8738
λ_1	—	0.0014	0.0006
λ_2	—	—	0.0017
GBIC	—	23494	23492

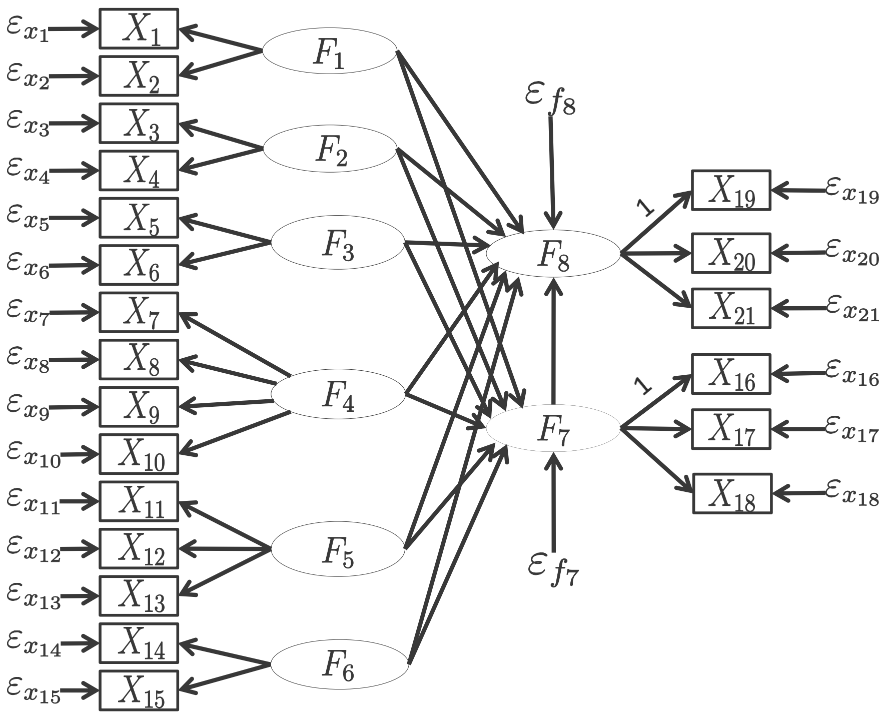

Figure 2: Hypothetic model for slump of personal consumption data

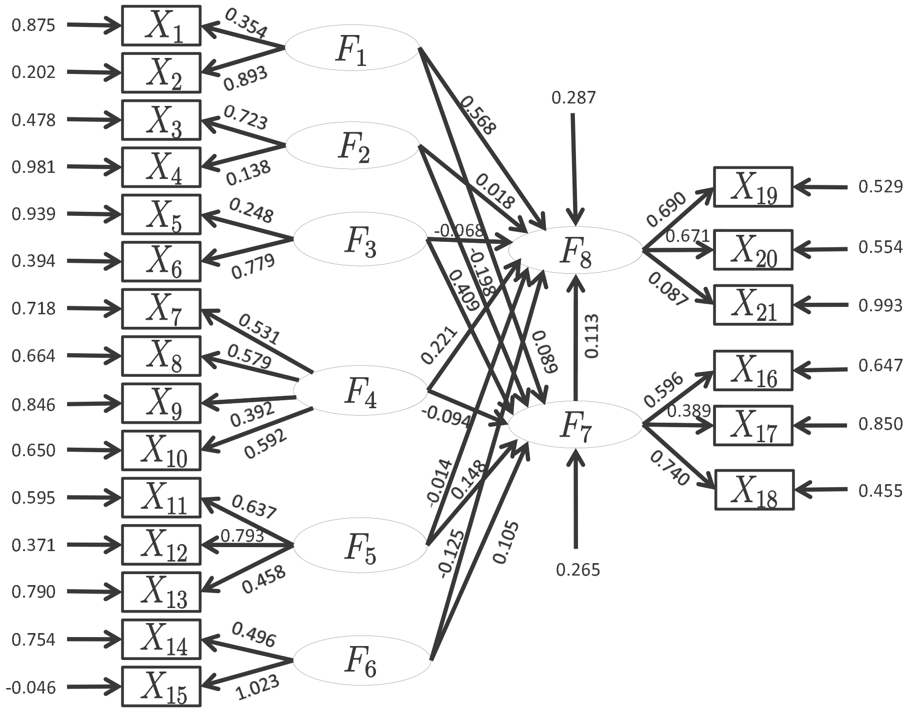

Figure 3: Maximum likelihood estimates

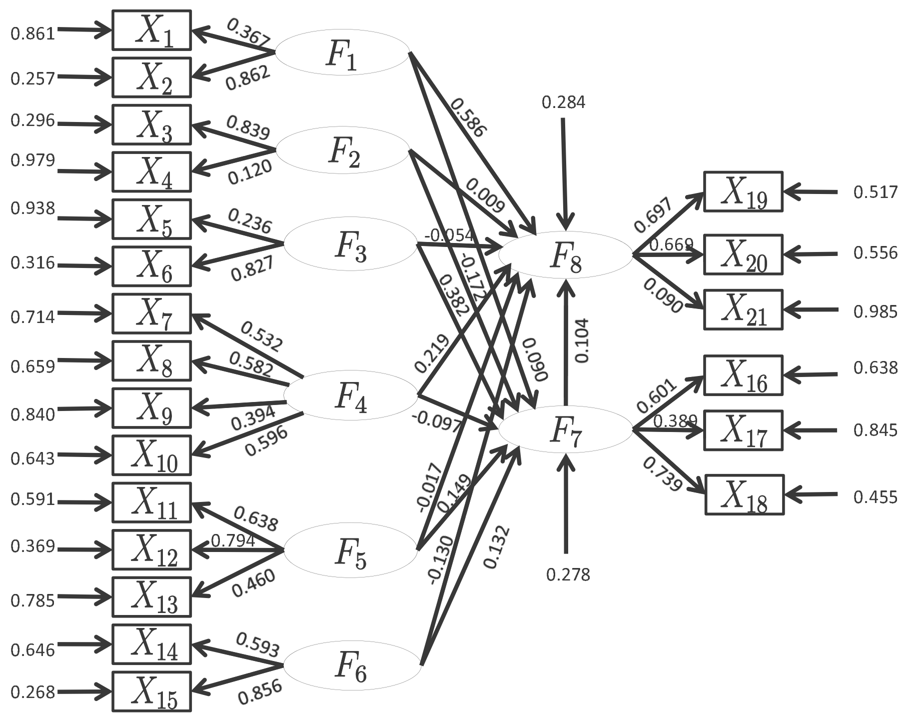

Figure 4: PMLE_1

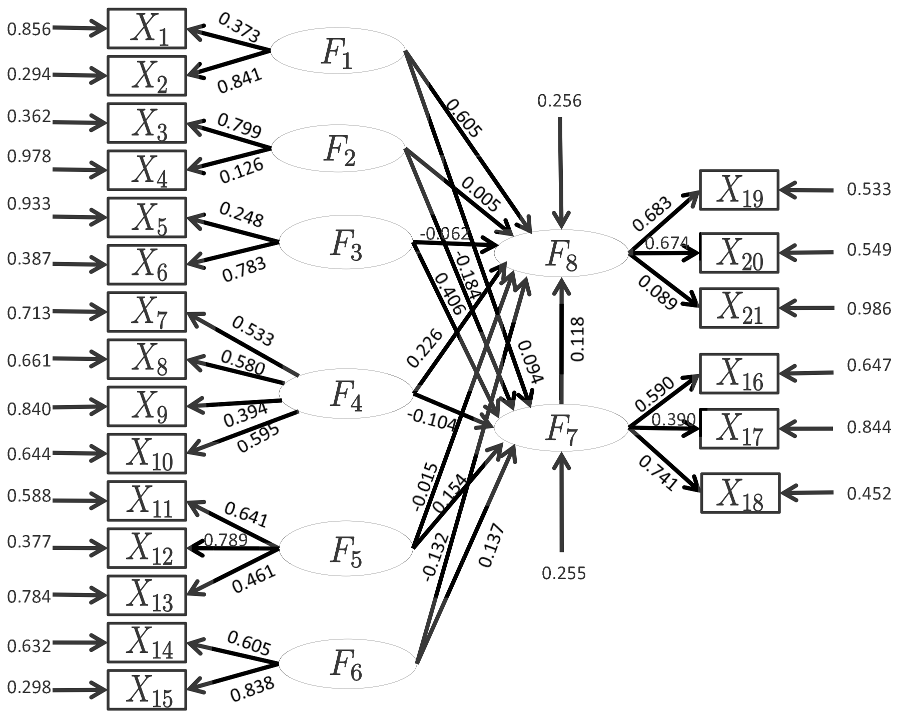

Figure 5: PMLE_2

parameters. In order to reduce a computational load of the proposed modeling procedure, we considered two kinds of restrictions: all error variances have the same hyper-parameter (PMLE_1), and the error variances for observable variables and those for latent variables have different hyper-parameters (PMLE_2). The proposed procedure was applied to artificial data, and we found that our method prevents the occurrence of improper solutions and provides stable estimates. Our modeling strategy is applied to the slump of personal consumption data, and also prevented the occurrence of improper solutions. For both artificial and real data, PMLE_1 and PMLE_2 yield almost the same result. Therefore, from a practical point of view, the PMLE_1 may be preferable since it does not need computation time compared with PMLE_2.

The structural equation modeling is usually used to investigate the linear relationships among observed variables and latent variable. However, models that have nonlinear structure are often encountered in social and behavioral sciences (see, e.g., Lee and Zhu, 2003). As a future research topic, it is interest to propose a selection procedure of hyper-parameters for nonlinear structural equation modeling.

Appendix A: Derivation of $J_{\boldsymbol{\lambda}}(\boldsymbol{\theta})$

This appendix derives the $J_{\boldsymbol{\lambda}}(\boldsymbol{\theta})$ included in the second differential of penalized log-likelihood function. First, we define a function F of $\boldsymbol{\theta}$:

$$F = \log|\Sigma| + \mathrm{tr}(\Sigma^{-1}S). \tag{A1}$$

The relationship between F and second differential of log-likelihood function in (5) can be obtained as follows:

$$\frac{\partial^2 \log f(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N|\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j} = -\frac{N}{2}\frac{\partial^2 F}{\partial\theta_i\partial\theta_j} \quad (i, j = 1, \cdots, k). \tag{A2}$$

Hence, if we derive a second differential of F, the second differential of log-likelihood function can be obtained. It is known that the second differential of F is given by (Lee and Jennrich, 1979):

$$\frac{\partial^2 F}{\partial\theta_i\partial\theta_j} = \mathrm{tr}\Sigma^{-1}\dot{\Sigma}_i\Sigma^{-1}\dot{\Sigma}_j + \mathrm{tr}\Sigma^{-1}(\Sigma - S)\Sigma^{-1}(\ddot{\Sigma}_{ij} - 2\dot{\Sigma}_i\Sigma^{-1}\dot{\Sigma}_j), \tag{A3}$$

where

$$\dot{\Sigma}_j = \frac{\partial\Sigma}{\partial\theta_j}, \tag{A4}$$

$$\ddot{\Sigma}_{ij} = \frac{\partial^2\Sigma}{\partial\theta_i\partial\theta_j}. \tag{A5}$$

For structural equation modeling, Equation (A4) and (A5) are given by

$$\frac{\partial\Sigma}{\partial a_{\alpha\beta}} = GT\Delta_{\alpha\beta}T\Sigma_{\boldsymbol{u}}T'G' + GT\Sigma_{\boldsymbol{u}}T'\Delta_{\beta\alpha}T'G', \tag{A6}$$

$$\frac{\partial\Sigma}{\partial\sigma^u_{wx}} = GT\Delta_{wx}T'G', \tag{A7}$$

$$\begin{aligned}\frac{\partial^2\Sigma}{\partial a_{\gamma\delta}\partial a_{\alpha\beta}} &= GT(\Delta_{\gamma\delta}T\Delta_{\alpha\beta}T\Sigma_{\boldsymbol{u}} + \Delta_{\alpha\beta}T\Delta_{\gamma\delta}T\Sigma_{\boldsymbol{u}} + \Delta_{\alpha\beta}T\Sigma_{\boldsymbol{u}}T'\Delta_{\delta\gamma})T'G' \\ &\quad + GT(\Delta_{\gamma\delta}T\Sigma_{\boldsymbol{u}}T'\Delta_{\beta\alpha} + \Sigma_{\boldsymbol{u}}T'\Delta_{\delta\gamma}T'\Delta_{\beta\alpha} + \Sigma_{\boldsymbol{u}}T'\Delta_{\beta\alpha}T'\Delta_{\delta\gamma})T'G',\end{aligned} \tag{A8}$$

$$\frac{\partial^2\Sigma}{\partial\sigma^u_{wx}\partial a_{\alpha\beta}} = GT(\Delta_{\alpha\beta}T\Delta_{wx} + \Delta_{wx}T'\Delta_{\beta\alpha})T'G', \tag{A9}$$

$$\frac{\partial^2\Sigma}{\partial\sigma^u_{yz}\partial\sigma^u_{wx}} = \boldsymbol{O}_{p\times p}, \tag{A10}$$

where Δ_{ij} is a matrix with one on (i, j)-th element and zeros elsewhere.

To derive $J_{\boldsymbol{\lambda}}(\boldsymbol{\theta})$, we need to obtain the second differential of the logarithm of prior distribution $\log\pi(\boldsymbol{\theta})$ regarding error variances, which is given by

$$\frac{\partial^2 \log\pi(\boldsymbol{\theta})}{\partial(\sigma^u_{v_i})^2} = \frac{2}{(\sigma^u_{v_i})^2} - \frac{2N\lambda_i}{(\sigma^u_{v_i})^3}, \quad (i = 1, \cdots, q_v). \tag{A11}$$

The second differential for other parameters is zero.

Appendix B: Description of real data

The 8 latent variables and corresponding observable variables for the slump of personal consumption dataset are given in the following:

- F_1: Changes in income
  — X_1: Increase and decrease in income
  — X_2: Consciousness of joy and sorrow for life

- F_2: Fears of a recession
  — X_3: Sense for the state of the economy
  — X_4: Prospect of the state of the economy

- F_3: Expectation for decrease in price
  — X_5: Prospect of decrease in price
  — X_6: The number of products that have a prospect of decrease in price

- F_4: Saturation of consumption
  — X_7: Getting away from from shopping
  — X_8: Overmuch fullness
  — X_9: Getting tired of shopping
  — X_{10}: Lack of fascinating products

- F_5: Prospect of future of society
  — X_{11}: Society of guarantee for position
  — X_{12}: Society of increase and decrease in income
  — X_{13}: Society of guarantee for old people

- F_6: Self-searching for life
  — X_{14}: Self-searching for luxury
  — X_{15}: Self-searching for shopping

- F_7: Buyer motivate
  — X_{16}: The number of goods that have appetites
  — X_{17}: A great desire to buy
  — X_{18}: The number of goods that have appetites if the price is down

- F_8: Buyer behavior
  — X_{19}: Limits on spending
  — X_{20}: The number of goods that limit on spending
  — X_{21}: The rate of realization for purchasing

For detail of the explanation of each variables, we refer to Toyoda (1998).

References

[1] Akaike, H. (1987): Factor analysis and AIC. *Psychometrika*, **52**, 317–332.

[2] Anderson, J. C. and Gerbing, D. W . (1984): The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, **49**, 155–173.

[3] Bentler, P. M. and Stein, J. A. (1992): Structural equation models in medical research. *Stat. Methods Med. Res.*, **1**, 159–181.

[4] Bentler, P. M. and Weeks, D. G. (1980): Linear structural equations with latent variables. *Psychometrika*, **45**, 289–308.

[5] Bock, R. D. and Bargmann, R. E. (1966): Analysis of covariance structures. *Psychometrika*, **31**, 507–534.

[6] Boomsma, A. (1985): Nonconvergence, improper solutions, and starting values in lisrel maximum likelihood estimation. *Psychometrika*, **50**, 229–242.

[7] Chen, F., Bollen, K. A., Paxton, P., Curran, P. J. and Kirby, J. B. (2001): Improper solutions in structural equation models, causes, consequences, and strategies. *Soc. Methods Res.*, **29**, 468–508.

[8] Chib, S. (1995): Marginal likelihood from the Gibbs output. *J. Amer. Statist. Assoc.*, **90**, 1313–1321.

[9] Chib, S. and Jeliazkov, I. (2001): Marginal likelihood from the Metropolis-Hastings output. *J. Amer. Statist. Assoc.*, **96**, 280–281.

[10] Diciccio, T., Kass, R., Raftery, A. and Wasserman, L. (1997): Computing Bayes factors by combining simulation and asymptotic approximations. *J. Am. Statist. Ass.*, **92**, 903–915.

[11] Flora, D. B. and Curran, P. J. (2004): An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol. Methods*, **9**, 466–491.

[12] Gerbing, D. W. and Anderson J. C. (1987): Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. *Psychometrika*, **52**, 99–111.

[13] Hirose, K., Kawano, S., Konishi, S. and Ichikawa, M. (2008): Bayesian factor analysis and model selection. Preprint, MHF2008-2, Kyushu University.

[14] Jöreskog, K. G. (1970): A general method for analysis of covariance structures. *Biometrika*, **57**, 239–251.

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

[16] Kano, Y. (1998): Improper solutions in exploratory factor analysis: Causes and treatments, in: A. Rizzi, M. Vichi and H. Bock (Eds.). *Advances in Data Sciences and Classification*, Springer, Berlin, 375–382.

[17] Konishi, S., Ando, T. and Imoto, S. (2004): Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika*, **91**, 27–43.

[18] Konishi, S. and Kitagawa, G. (2008): *Information Criteria and Statistical Modeling.* New York: Springer.

[19] Lee, S. Y. and Jennrich, R. I. (1979): A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika*, **44**, 99–113.

[20] Lee, S. Y. and Zhu, H. T. (2003): Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *Br. J. Math. Stat. Psychol.*, **53**, 209–232.

[21] Liu, B., de la Fuente, A. and Hoeschele, I. (2008): Gene network inference via structural equation modeling in genetical genomics experiments. *Genetics*, **178**, 1763–1776.

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

[23] McArdle, J. J. (1980): Causal modeling applied to psychonomic systems simulation. *Behav. Res. Methods Instrum.*, **12**, 193–209.

[24] McArdle, J. J. and McDonald, R. P. (1984): Some algebraic properties of the reticular action model for moment structures. *Br. J. Math. Stat. Psychol.*, **37**, 234–251.

[25] Pugesek, B. H., Tomer, A. and von Eye, A. (2003): *Structural Equation Modeling Applications in Ecological and Evolutionary Biology.* New York: Cambridge University Press.

[26] Schwarz, G. (1978): Estimating the dimension of a model. *Ann. Statist.* **6**, 461–464.

[27] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002): Bayesian measures of model complexity and fit (with discussion). *J. Roy. Statist. Soc. Ser. B*, **64**, 583-639.

[28] Toyoda, H. (1998): *Covariance Structure Analysis [Case Examples] — Structural Equation Modeling— (in Japanese).* Kitaohji-shobo Publishing Co., Ltd.

[29] Xiong, M., Li, J. and Fang, X. (2004): Identification of genetic networks. *Genetics*, **166**, 1037–1052.

[30] van Driel, O. P. (1978): On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika*, **43**, 225–243.

List of MI Preprint Series, Kyushu University

The Global COE Program
Math-for-Industry Education & Research Hub

MI

MI2008-1 Takahiro ITO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Abstract collision systems simulated by cellular automata

MI2008-2 Eiji ONODERA
The intial value problem for a third-order dispersive flow into compact almost Hermitian manifolds

MI2008-3 Hiroaki KIDO
On isosceles sets in the 4-dimensional Euclidean space

MI2008-4 Hirofumi NOTSU
Numerical computations of cavity flow problems by a pressure stabilized characteristic-curve finite element scheme

MI2008-5 Yoshiyasu OZEKI
Torsion points of abelian varieties with values in nfinite extensions over a p-adic field

MI2008-6 Yoshiyuki TOMIYAMA
Lifting Galois representations over arbitrary number fields

MI2008-7 Takehiro HIROTSU & Setsuo TANIGUCHI
The random walk model revisited

MI2008-8 Silvia GANDY, Masaaki KANNO, Hirokazu ANAI & Kazuhiro YOKOYAMA
Optimizing a particular real root of a polynomial by a special cylindrical algebraic decomposition

MI2008-9 Kazufumi KIMOTO, Sho MATSUMOTO & Masato WAKAYAMA
Alpha-determinant cyclic modules and Jacobi polynomials

MI2008-10 Sangyeol LEE & Hiroki MASUDA
Jarque-Bera Normality Test for the Driving Lévy Process of a Discretely Observed Univariate SDE

MI2008-11 Hiroyuki CHIHARA & Eiji ONODERA
A third order dispersive flow for closed curves into almost Hermitian manifolds

MI2008-12 Takehiko KINOSHITA, Kouji HASHIMOTO and Mitsuhiro T. NAKAO
On the L^2 a priori error estimates to the finite element solution of elliptic problems with singular adjoint operator

MI2008-13 Jacques FARAUT and Masato WAKAYAMA
Hermitian symmetric spaces of tube type and multivariate Meixner-Pollaczek polynomials

MI2008-14 Takashi NAKAMURA
Riemann zeta-values, Euler polynomials and the best constant of Sobolev inequality

MI2008-15 Takashi NAKAMURA
Some topics related to Hurwitz-Lerch zeta functions

MI2009-1 Yasuhide FUKUMOTO
Global time evolution of viscous vortex rings

MI2009-2 Hidetoshi MATSUI & Sadanori KONISHI
Regularized functional regression modeling for functional response and predictors

MI2009-3 Hidetoshi MATSUI & Sadanori KONISHI
Variable selection for functional regression model via the L_1 regularization

MI2009-4 Shuichi KAWANO & Sadanori KONISHI
Nonlinear logistic discrimination via regularized Gaussian basis expansions

MI2009-5 Toshiro HIRANOUCHI & Yuichiro TAGUCHII
Flat modules and Groebner bases over truncated discrete valuation rings

MI2009-6 Kenji KAJIWARA & Yasuhiro OHTA
Bilinearization and Casorati determinant solutions to non-autonomous 1+1 dimensional discrete soliton equations

MI2009-7 Yoshiyuki KAGEI
Asymptotic behavior of solutions of the compressible Navier-Stokes equation around the plane Couette flow

MI2009-8 Shohei TATEISHI, Hidetoshi MATSUI & Sadanori KONISHI
Nonlinear regression modeling via the lasso-type regularization

MI2009-9 Takeshi TAKAISHI & Masato KIMURA
Phase field model for mode III crack growth in two dimensional elasticity

MI2009-10 Shingo SAITO
Generalisation of Mack's formula for claims reserving with arbitrary exponents for the variance assumption

MI2009-11 Kenji KAJIWARA, Masanobu KANEKO, Atsushi NOBE & Teruhisa TSUDA
Ultradiscretization of a solvable two-dimensional chaotic map associated with the Hesse cubic curve

MI2009-12 Tetsu MASUDA
Hypergeometric τ-functions of the q-Painlevé system of type $E_8^{(1)}$

MI2009-13 Hidenao IWANE, Hitoshi YANAMI, Hirokazu ANAI & Kazuhiro YOKOYAMA
A Practical Implementation of a Symbolic-Numeric Cylindrical Algebraic Decomposition for Quantifier Elimination

MI2009-14 Yasunori MAEKAWA
On Gaussian decay estimates of solutions to some linear elliptic equations and its applications

MI2009-15 Yuya ISHIHARA & Yoshiyuki KAGEI
Large time behavior of the semigroup on L^p spaces associated with the linearized compressible Navier-Stokes equation in a cylindrical domain

MI2009-16 Chikashi ARITA, Atsuo KUNIBA, Kazumitsu SAKAI & Tsuyoshi SAWABE
Spectrum in multi-species asymmetric simple exclusion process on a ring

MI2009-17 Masato WAKAYAMA & Keitaro YAMAMOTO
Non-linear algebraic differential equations satisfied by certain family of elliptic functions

MI2009-18 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of an Elliptical Flow Subjected to a Coriolis Force

MI2009-19 Mitsunori KAYANO & Sadanori KONISHI
Sparse functional principal component analysis via regularized basis expansions and its application

MI2009-20 Shuichi KAWANO & Sadanori KONISHI
Semi-supervised logistic discrimination via regularized Gaussian basis expansions

MI2009-21 Hiroshi YOSHIDA, Yoshihiro MIWA & Masanobu KANEKO
Elliptic curves and Fibonacci numbers arising from Lindenmayer system with symbolic computations

MI2009-22 Eiji ONODERA
A remark on the global existence of a third order dispersive flow into locally Hermitian symmetric spaces

MI2009-23 Stjepan LUGOMER & Yasuhide FUKUMOTO
Generation of ribbons, helicoids and complex scherk surface in laser-matter Interactions

MI2009-24 Yu KAWAKAMI
Recent progress in value distribution of the hyperbolic Gauss map

MI2009-25 Takehiko KINOSHITA & Mitsuhiro T. NAKAO
On very accurate enclosure of the optimal constant in the a priori error estimates for H_0^2-projection

MI2009-26 Manabu YOSHIDA
Ramification of local fields and Fontaine's property (Pm)

MI2009-27 Yu KAWAKAMI
Value distribution of the hyperbolic Gauss maps for flat fronts in hyperbolic three-space

MI2009-28 Masahisa TABATA
Numerical simulation of fluid movement in an hourglass by an energy-stable finite element scheme

MI2009-29 Yoshiyuki KAGEI & Yasunori MAEKAWA
Asymptotic behaviors of solutions to evolution equations in the presence of translation and scaling invariance

MI2009-30 Yoshiyuki KAGEI & Yasunori MAEKAWA
On asymptotic behaviors of solutions to parabolic systems modelling chemotaxis

MI2009-31 Masato WAKAYAMA & Yoshinori YAMASAKI
Hecke's zeros and higher depth determinants

MI2009-32 Olivier PIRONNEAU & Masahisa TABATA
Stability and convergence of a Galerkin-characteristics finite element scheme of lumped mass type

MI2009-33 Chikashi ARITA
Queueing process with excluded-volume effect

MI2009-34 Kenji KAJIWARA, Nobutaka NAKAZONO & Teruhisa TSUDA
Projective reduction of the discrete Painlevé system of type$(A_2 + A_1)^{(1)}$

MI2009-35 Yosuke MIZUYAMA, Takamasa SHINDE, Masahisa TABATA & Daisuke TAGAMI
Finite element computation for scattering problems of micro-hologram using DtN map

MI2009-36 Reiichiro KAWAI & Hiroki MASUDA
Exact simulation of finite variation tempered stable Ornstein-Uhlenbeck processes

MI2009-37 Hiroki MASUDA
On statistical aspects in calibrating a geometric skewed stable asset price model

MI2010-1 Hiroki MASUDA
Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes

MI2010-2 Reiichiro KAWAI & Hiroki MASUDA
Infinite variation tempered stable Ornstein-Uhlenbeck processes with discrete observations

MI2010-3 Kei HIROSE, Shuichi KAWANO, Daisuke MIIKE & Sadanori KONISHI
Hyper-parameter selection in Bayesian structural equation models