Multivariate Regression and Outliers: A MCMC approach

Wolfgang Polasek

Institute of Advanced Studies, Vienna
Stumpergasse 56, 1060 Vienna, Austria
Email: Polasek@ihs.ac.at
October 1, 2002

Abstract

Assuming a normal-Wishart modeling framework we show how to find outliers in a multivariate regression system. We assume an informative prior and show how the posterior distribution can be simulated in multivariate blocks by a Gibbs sampling algorithm. The number of outliers is determined using the usual marginal likelihood and the posterior marginal likelihood criteria. The approach is demonstrated for the language data set of Fuller (1987).

Keywords: Multivariate outliers, Gibbs sampling, marginal likelihood.

1 Introduction

Multivariate regression models are a frequent tool in social sciences (e.g., economics, psychology and sociology) for explaining a multivariate data matrix by a common set of "independent" variables or regressors. This is particularly useful if one encounters many variables (repeated measures) which can be related by some common variables. There are only a few suggestions how to find Multivariate outliers, and here we demonstrate a computational Bayesian approach.

The question arises as to whether or not the new technique of Monte Carlo Markov Chain (MCMC) methods can be used to detect multivariate outliers. The question of distributional assumptions is also important. Using the Gibbs sampling approach of Verdinelli and Wasserman (1991) for multivariate location shift outlier models, we show how to derive the marginal likelihoods for the Gibbs sampling algorithm.

Marginal likelihoods are used in Bayesian tests and we can use them for model selection. Thus, the marginal likelihood can be used to determine the nand multivariate extension of the outlier approach of Verdinelli and Wasserman (1991). In section four we analyse the language data set of Fuller (1987) and describe inferences for possible (location shift) outliers. Section five concludes and in the appendix we have listed computational details for the full conditional distribution of the Gibbs sampler and the components of the marginal likelihood.

2 Multivariate Regression and Outliers

The multivariate regression model is given by

$$\underset{(n\times p)}{\mathbf{Y}} = \underset{(n\times K)}{\mathbf{X}} \underset{(K\times p)}{\mathbf{B}} + \underset{(n\times p)}{\mathbf{E}}.$$

Let $\mathbf{A}$ be a $n \times p$ location shift parameter matrix and $\mathbf{D}_\vartheta = diag(\vartheta_1, \ldots, \vartheta_n)$ an $n \times n$ indicator matrix for multivariate outliers. We can then formulate the model by assuming a normal distribution

$$\mathbf{Y} \sim \mathcal{N}_{n\times p}[\mathbf{XB} + \mathbf{D}_\vartheta \mathbf{A}, \mathbf{\Psi} \otimes \mathbf{I}_n].$$

The prior information can be compactly formulated as

$$\begin{aligned} \mathbf{B} &\sim \mathcal{N}_{K\times p}[\mathbf{B}_*, \mathbf{G}_* \otimes \mathbf{H}_*], \\ \mathbf{\Psi} &\sim \mathcal{W}_p[\mathbf{\Psi}_*, n_*], \\ \mathbf{A} &\sim \mathcal{N}_{n\times p}[\mathbf{A}_*, \mathbf{P}_* \otimes \mathbf{I}_n], \\ \vartheta_i &\sim Ber[\varepsilon_{i*}], \qquad i = 1, \ldots, n, \end{aligned}$$

where $Ber[\varepsilon_{i*}]$ denotes the Bernoulli distribution and ε_{i*} is the prior probability that observation i is an outlier.
The joint distribution of the data $\mathbf{Y}$ and the parameter $\theta = (\mathbf{B}, \mathbf{\Psi}, \mathbf{A}, \mathbf{D}_\vartheta)$ is

$$\begin{aligned} p(\mathbf{Y}, \theta) = {} & \mathcal{N}_{n\times p}[\mathbf{Y} \mid \mathbf{XB} + \mathbf{D}_\vartheta \mathbf{A}, \mathbf{\Psi} \otimes \mathbf{I}_n] \cdot \mathcal{N}_{\mathbf{K}\times\mathbf{p}}[\mathbf{B} | \mathbf{B}_*, \mathbf{G}_* \otimes \mathbf{H}_*] \cdot \\ & \cdot \mathcal{W}_p[\mathbf{\Psi}^{-1} \mid \mathbf{\Psi}_*, n_*] \cdot \mathcal{N}_{n\times p}[\mathbf{A} \mid \mathbf{A}_*, \mathbf{P}_* \otimes \mathbf{I}_n] \cdot \sum_{i=1}^{n} Ber(\vartheta_i \mid \varepsilon_{i*}). \end{aligned}$$

The full conditional distributions are
a) For the matrix regression coefficients $\mathbf{B}$:

$$p(\mathbf{B} \mid \mathbf{Y}, \theta^c) = \mathcal{N}_{n\times p}[\mathbf{B}_{**}, \mathbf{C}_{**}]$$

a multivariate normal distribution with the parameters

$$
\begin{aligned}
\mathbf{C}_{**}^{-1} &= \mathbf{G}_{*}^{-1} \otimes \mathbf{H}_{*}^{-1} + \mathbf{\Psi}^{-1} \otimes \mathbf{X}'\mathbf{X}, \\
\text{vec } \mathbf{B}_{**} &= \mathbf{C}_{**}[\text{vec } (\mathbf{G}_{*}^{-1}\mathbf{B}_{*}\mathbf{H}_{*}^{-1} + \mathbf{X}'(\mathbf{Y} - \mathbf{D}_{\vartheta}\mathbf{A})\mathbf{\Psi}^{-1})].
\end{aligned}
$$

b) For the covariance matrix $\mathbf{\Psi}$:

$$p(\mathbf{\Psi}^{-1} \mid \mathbf{Y}, \theta^c) = \mathcal{W}_p[\mathbf{\Psi}_{**}, n_{**} = n_* + n]$$

a p-dimensional Wishart distribution with scale parameter

$$\mathbf{\Psi}_{**} = \mathbf{\Psi}_* + (\mathbf{Y} - \mathbf{XB} - \mathbf{D}_{\vartheta}\mathbf{A})(\mathbf{Y} - \mathbf{XB} - \mathbf{D}_{\vartheta}\mathbf{A})'.$$

c) For the level shift matrix $\mathbf{A}$:

$$p(\mathbf{A} \mid \mathbf{Y}, \theta^c) = \mathcal{N}_{n \times n}[\mathbf{A}_{**}, \mathbf{G}_{**}]$$

a multivariate normal distribution with the parameters

$$
\begin{aligned}
\mathbf{G}_{**}^{-1} &= \mathbf{P}_{*}^{-1} \otimes \mathbf{I}_n + \mathbf{\Psi}^{-1} \otimes \mathbf{D}_{\vartheta}'\mathbf{D}_{\vartheta}, \\
\text{vec } \mathbf{A}_{**} &= \mathbf{G}_{**}[\text{vec } (\mathbf{A}_*\mathbf{P}_{*}^{-1} + \mathbf{D}_{\vartheta}(\mathbf{Y} - \mathbf{XB})\mathbf{\Psi}^{-1})].
\end{aligned}
$$

For each observation the posterior mean can be calculated by breaking up the system estimate as

$$\mathbf{G}_{**i}^{-1} = \mathbf{P}_{*}^{-1} + {\vartheta_i}^2\mathbf{\Psi}^{-1}, \quad i = 1, \ldots, n,$$

$$\mathbf{a}_{**i} = \mathbf{G}_{**i}[\mathbf{P}_{*}^{-1}\mathbf{a}_{*i} + \vartheta_i\mathbf{\Psi}^{-1}(\mathbf{y}_i - \mathbf{Bx}_i)].$$

d) For the indicator variables ϑ_i:

$$Pr(\vartheta_i \mid \mathbf{Y}, \theta^c) = Ber[\varepsilon_{i**} = \frac{c_i}{c_i + d_i}], \qquad i = 1, \ldots, n,$$

a Bernoulli distribution with the components obtained via Bayes theorem, i.e.,

$$
\begin{aligned}
c_i &= \mathcal{N}_p[\mathbf{y}_i \mid \mathbf{x}_i\mathbf{B} + \mathbf{a}_i, \mathbf{\Psi}] \cdot \varepsilon_{i*}, \\
d_i &= \mathcal{N}_p[\mathbf{y}_i \mid \mathbf{x}_i\mathbf{B}, \mathbf{\Psi}] \cdot (1 - \varepsilon_{i*}), \qquad i = 1, \ldots, n,
\end{aligned}
$$

where $\mathbf{x}_i$ is the i-th row of $\mathbf{X}$ and $\mathbf{a}_i$ is the i-th row of $\mathbf{A}$.

3 The computation of the marginal likelihood

This appendix calculates the marginal likelihood of the factor analysis model (see section 2.1) using the method of Chib (1995).
1. Use the Gibbs run of length J of the estimation procedure to estimate the ordinate at the location $\hat{\mathbf{Z}}$:

$$
\begin{aligned}
p(\hat{\mathbf{Z}} \mid \mathbf{Y}) &= \int \mathcal{N}[\hat{\mathbf{Z}} \mid \mathbf{Z}_{**}, \mathbf{I}_n \otimes \mathbf{D}_Z] \times p(\boldsymbol{\Lambda}, \boldsymbol{\Psi}^{-1}, \boldsymbol{\Phi}^{-1} \mid \mathbf{Y}) d\boldsymbol{\Lambda} d\boldsymbol{\Psi}^{-1} d\boldsymbol{\Phi}^{-1} \\
&= \frac{1}{J}\sum_{j=1}^{J} \mathcal{N}[\hat{\mathbf{Z}} \mid \mathbf{Z}_{**}^{(j)}, \mathbf{I}_n \otimes \mathbf{D}_Z^{(j)}] = \frac{1}{J}\sum_{j=1}^{J}\prod_{i=1}^{n} \mathcal{N}[\hat{\mathbf{z}}_i | \mathbf{z}_{**i}^{(j)}, \mathbf{D}_Z^{(j)}], \qquad (1)
\end{aligned}
$$

with $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ and the following moments of the multivariate normal distribution

$$
\begin{aligned}
\mathbf{Z}_{**}^{(j)} &= (\boldsymbol{\Phi}_j^{-1} + \boldsymbol{\Lambda}_j' \boldsymbol{\Psi}_j^{-1} \boldsymbol{\Lambda}_j)^{-1} \boldsymbol{\Lambda}_j' \boldsymbol{\Psi}_j^{-1} \mathbf{Y}, \\
\mathbf{D}_Z^{(j)} &= (\boldsymbol{\Phi}_j^{-1} + \boldsymbol{\Lambda}_j' \boldsymbol{\Psi}_j^{-1} \boldsymbol{\Lambda}_j)^{-1}.
\end{aligned}
$$

2. Reduce the Gibbs run to estimate the second component of the posterior ordinate

$$
\begin{aligned}
p(\hat{\boldsymbol{\Lambda}} \mid \hat{\mathbf{Z}}, \mathbf{Y}) &= \int \mathcal{N}[\hat{\boldsymbol{\Lambda}} \mid \boldsymbol{\Lambda}_{**}, \mathbf{C}_{**}] \times p(\boldsymbol{\Psi}^{-1}, \boldsymbol{\Phi}^{-1} \mid \mathbf{Y}) \; d\boldsymbol{\Psi}^{-1} d\boldsymbol{\Phi}^{-1} \\
&= \frac{1}{J}\sum_{j=1}^{J} \mathcal{N}[\hat{\boldsymbol{\Lambda}} \mid \boldsymbol{\Lambda}_{**}^{(j)}, \mathbf{C}_{**}^{(j)}] \qquad (2)
\end{aligned}
$$

with the reduced parameters

$$
\boldsymbol{\Lambda}_{**}^{(j)} = \mathbf{C}_{**}^{(j)}[\text{vec } (\mathbf{G}_*^{-1} \boldsymbol{\Lambda}_* \mathbf{H}_*^{-1} + \boldsymbol{\Psi}_j^{-1} \mathbf{Y} \hat{\mathbf{Z}})]
$$

and

$$
\mathbf{C}_{**}^{(j)} = (\mathbf{H}_*^{-1} \otimes \mathbf{G}_*^{-1} + \hat{\mathbf{Z}}' \hat{\mathbf{Z}} \otimes \boldsymbol{\Psi}_j^{-1})^{-1}, \quad j = 1, \ldots, J.
$$

Note that this implies that the remaining two conditional distributions have to be estimated by

$$
\boldsymbol{\Psi}_{**} = \boldsymbol{\Psi}_* + (\mathbf{Y} - \hat{\boldsymbol{\Lambda}} \hat{\mathbf{Z}})(\mathbf{Y} - \hat{\boldsymbol{\Lambda}} \hat{\mathbf{Z}})' \qquad (3)
$$

and

$$\mathbf{\Phi}_{**} = \mathbf{\Phi}_* + \hat{\mathbf{Z}}\hat{\mathbf{Z}}'. \tag{4}$$

3. The third component of the posterior ordinate can be estimated without a Gibbs run because

$$\begin{aligned} p(\hat{\mathbf{\Psi}}^{-1} \mid \hat{\mathbf{\Lambda}}, \hat{\mathbf{Z}}, \mathbf{Y}) &= \int \mathcal{W}_p[\hat{\mathbf{\Psi}}^{-1} \mid \ \hat{\mathbf{\Lambda}}, \hat{\mathbf{Z}}, \mathbf{Y}] d\mathbf{\Phi}^{-1} \\ &= \mathcal{W}_p\ [\hat{\mathbf{\Psi}}^{-1} \mid \mathbf{\Psi}_{**}, n_{**}], \end{aligned} \tag{5}$$

which does not depend on the remaining $\mathbf{\Phi}$ parameters and with $\mathbf{\Psi}_{**}$ given as in (**??**).
4. The last component of the posterior ordinate can be also calculated without a Gibbs run:

$$p(\hat{\mathbf{\Phi}}^{-1} \mid \hat{\mathbf{\Psi}}^{-1}, \hat{\mathbf{\Lambda}}, \hat{\mathbf{Z}}, \mathbf{Y}) = \mathcal{W}\ [\hat{\mathbf{\Phi}}^{-1} \mid \mathbf{\Phi}_{**}, \nu_{**}], \tag{6}$$

where $\mathbf{\Phi}_{**}$ is given in (**??**).

3.1 Selecting the number of factors

In order to choose between factor analysis models with different numbers of factors (in short: factor analysis of rank r), we use the Bayes factor and the marginal likelihood as model choice criteria (see also section 3.2). For equal prior probabilities the Bayes factor is given by the ratio of marginal likelihoods

$$B = \frac{p(\mathbf{Y}|\text{rank} = r)}{p(\mathbf{Y}|\text{rank} = s)} \tag{7}$$

and can be used to choose between factor analysis models with rank r or s. The marginal likelihood for a certain rank specification is calculated by

$$p(\mathbf{Y}) = \frac{p(\mathbf{Y} \mid \hat{\theta})\ p(\hat{\theta})}{p(\hat{\theta} \mid \mathbf{Y})}. \tag{8}$$

For the denominator in (**??**) we need the following posterior ordinate decomposition for an appropriate chosen point $\hat{\theta} = (\hat{\mathbf{Z}}, \hat{\mathbf{\Lambda}}, \hat{\mathbf{\Psi}}, \hat{\mathbf{\Phi}})$:

$$p(\hat{\theta} \mid \mathbf{Y}) = p(\hat{\mathbf{Z}} \mid \mathbf{Y}) \times p(\hat{\mathbf{\Lambda}} \mid \hat{\mathbf{Z}}, \mathbf{Y}) \times p(\hat{\mathbf{\Psi}}^{-1} \mid \hat{\mathbf{\Lambda}}, \hat{\mathbf{Z}}, \mathbf{Y}) \times p(\hat{\mathbf{\Phi}}^{-1} \mid \hat{\mathbf{\Psi}}^{-1}, \hat{\mathbf{\Lambda}}, \hat{\mathbf{Z}}, \mathbf{Y}). \tag{9}$$

We show in Appendix B how the four components in (**??**) can be computed by additional steps in the Gibbs algorithm. Following Chib (1995), the three components on the right hand side of (**??**), which depend on $\hat{\theta}$, are computed in the following way:
The likelihood ordinate in (**??**) is calculated by

$$p(\mathbf{Y} \mid \hat{\theta}) = \mathcal{N}_{p\times n}[\mathbf{Y} \mid \hat{\mathbf{\Lambda}}\hat{\mathbf{Z}}, \mathbf{I}_n \otimes \hat{\mathbf{\Psi}}] \tag{10}$$

and the ordinate of the prior density is given by

$$p(\hat{\theta}) = \mathcal{W}[\hat{\mathbf{\Psi}}^{-1} \mid \mathbf{\Psi}_*, n_*] \cdot \mathcal{W}[\hat{\mathbf{\Phi}}^{-1} \mid \mathbf{\Phi}_*, \nu_*] \cdot \mathcal{N}[\hat{\mathbf{Z}}|\mathbf{0}, \mathbf{I}_n \otimes \hat{\mathbf{\Phi}}] \cdot \mathcal{N}[\hat{\mathbf{\Lambda}}|\mathbf{\Lambda}_*, \mathbf{H}_* \otimes \mathbf{G}_*]. \tag{11}$$

Alternatively we can compute the logarithm of the marginal likelihood (**??**) by

$$log\ p(\mathbf{Y}) = log\ p(\mathbf{Y} \mid \hat{\theta}) + log\ p(\hat{\theta}) - log\ p(\hat{\theta} \mid \mathbf{Y}). \tag{12}$$

This (log) marginal likelihood can be calculated for models with different number of factors $k = 1, \ldots, p$ and we choose the model with the highest marginal likelihood.

4 Factor analysis with outliers

Consider the factor analysis model under the normal distribution as in (**??**):

$$\mathbf{y}_i \sim \mathcal{N}_p[\mathbf{\Lambda}\mathbf{z}_i, \mathbf{\Psi}], \quad i = 1, \ldots, n,$$

we now expose this model to additive outliers. The univariate location shift outlier model was first analyzed by the Gibbs sampler in Verdinelli and Wasserman (1991) . The multivariate location shift outlier model is formulated with the n indicator variables $\vartheta_1, \ldots, \vartheta_n$

$$\begin{aligned} f(\mathbf{y}_i \mid \vartheta_i) &= (1-\varepsilon_*)\ \mathcal{N}_p\ [\mathbf{y}_i \mid \mathbf{\Lambda}\mathbf{z}_i, \mathbf{\Psi}] \\ &\quad + \ \varepsilon_*\ \mathcal{N}_p\ [\mathbf{y}_i \mid \mathbf{a}_i + \mathbf{\Lambda}\mathbf{z}_i, \mathbf{\Psi}] \end{aligned} \tag{13}$$

where $\mathbf{y}_i$ is the i-th column of $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$, and $\mathbf{a}_i$ is the i-th column of $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$. $\mathbf{D}_\vartheta = diag(\vartheta_1, \ldots, \vartheta_n)$ is a $n \times n$ indicator matrix for the (multivariate) outliers. We assume that each indicator is distributed as a Bernoulli random variable with parameter ε_*, the prior probability that the

i-th observation is an outlier.
The factor analysis model with outliers can be written as

$$\mathbf{Y} \;\sim\; \mathcal{N}_{p\times n}[\mathbf{\Lambda Z} + \mathbf{A D}_\vartheta, \mathbf{I}_n \otimes \mathbf{\Psi}] \tag{14}$$

with the prior information given compactly as

$$\begin{aligned}
\mathbf{Z} &\sim \mathcal{N}_{K\times n}[\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Phi}]_*, \\
\mathbf{\Lambda} &\sim \mathcal{N}_{p\times K}[\mathbf{\Lambda}_*, \mathbf{H}_* \otimes \mathbf{G}_*], \\
\mathbf{\Psi} &\sim \mathcal{W}_p[\mathbf{\Psi}_*, n_*], \quad \mathbf{\Phi} \sim \mathcal{W}_k[\mathbf{\Phi}_*, \nu_*], \\
\mathbf{A} &\sim \mathcal{N}_{p\times n}[\mathbf{A}_*, \mathbf{I}_n \otimes \mathbf{P}_*], \\
\vartheta_i &\sim Ber[\varepsilon_*], \quad i = 1, \ldots, n,
\end{aligned} \tag{15}$$

where $\mathbf{A}_* : p \times K$ and $\mathbf{P}_* : p \times p$ are a-priori known parameter matrices for the location and variances of outliers and $Ber[\varepsilon_*]$ denotes a Bernoulli distributed random variable with known success probability ε_*. The full conditional distributions of the Gibbs sampler for the factor analysis model with outliers are derived in Appendix C.

4.1 The marginal likelihood for factor analysis with outliers

Using the approach of Chib (1995) we will evaluate the marginal likelihood at the point

$$\hat{\theta}_1 = (\hat{\theta}_0, \hat{\mathbf{A}} = \mathbf{0}, \hat{\mathbf{D}}_\vartheta = \mathbf{0}) \tag{16}$$

where $\hat{\theta}_0 = (\hat{\mathbf{Z}}, \hat{\mathbf{\Lambda}}, \hat{\mathbf{\Psi}}, \hat{\mathbf{\Phi}})$ is the same point as for the factor analysis without outliers. Now we have the following factorization

$$p(\hat{\theta} \mid \mathbf{Y}) \;=\; p(\hat{\mathbf{D}}_\vartheta \mid \mathbf{Y}) \cdot p(\hat{\mathbf{A}} \mid \hat{\mathbf{D}}_\vartheta, \mathbf{Y}) \cdot p(\hat{\theta}_0 \mid \hat{\mathbf{A}}, \hat{\mathbf{D}}_\vartheta, \mathbf{Y}) \tag{17}$$

and

$$p(\hat{\theta}) \;=\; p(\hat{\theta}_0)\mathcal{N}[\mathbf{A}_*, \mathbf{I}_n \otimes \mathbf{P}_*] \prod_{i=1}^{n} Ber(\varepsilon_*). \tag{18}$$

1. Use the Gibbs run of J sample points of the 'factor analysis with outliers' program to calculate the ordinate:

$$
\begin{aligned}
p(\hat{\mathbf{D}}_\vartheta \mid \mathbf{Y}) &= \int \prod_{i=1}^{n} Ber(\varepsilon_{i**}) p(\theta \mid \mathbf{Y}) d\theta \\
&= \frac{1}{J} \sum_{j=1}^{J} \prod_{i=1}^{n} Ber(\varepsilon_{i**}^{(j)}) \qquad (19)
\end{aligned}
$$

where the parameter of the i-th posterior density of the Bernoulli distribution is given by

$$
\begin{aligned}
\varepsilon_{i**}^{(j)} &= \frac{c_i^{(j)}}{c_i^{(j)} + d_i^{(j)}}, \\
c_i^{(j)} &= \mathcal{N}_p[\mathbf{y}_i \mid \mathbf{a}^{(j)} \vartheta_i^{(j)} + \mathbf{\Lambda}_j \mathbf{z}_i^{(j)}],
\end{aligned}
$$

with

$$
d_i^{(j)} = \mathcal{N}_p[\mathbf{y}_i \mid \mathbf{\Lambda}^{(j)} \mathbf{z}_i^{(j)}].
$$

2. The ordinate for the second component can be obtained without a Gibbs sampling output by

$$
p(\hat{\mathbf{A}} \mid \hat{\mathbf{D}}_\vartheta = \mathbf{0}, \mathbf{Y}) = \mathcal{N}_{p \times n}[\hat{\mathbf{A}} \mid \mathbf{A}_{**}, \mathbf{I}_n \otimes \mathbf{G}_{**}] = \prod_{i=1}^{n} \mathcal{N}[\hat{\mathbf{a}}_i | \mathbf{a}_{i**}, \mathbf{G}_{**}]. \qquad (20)
$$

It can be seen from the f.c.d. for $\mathbf{A}$ that for $\hat{\mathbf{D}}_\vartheta = \mathbf{0}$ the conditional distribution equals the prior distribution:

$$
\mathbf{G}_{**} = \mathbf{I}_n \otimes \mathbf{P}_* \qquad \text{and} \qquad vec\ \mathbf{A}_{**} = vec\ \mathbf{A}_*.
$$

3. Finally we can obtain the ordinate of the third factor $p(\hat{\theta}_0 \mid \hat{\mathbf{A}}, \hat{\mathbf{D}}_\vartheta, \mathbf{Y})$ in (**??**) by the marginal likelihood output of the factor analysis program without outliers. This follows from the fact that the reduced Gibbs run at this step of the factor analysis program with outliers is equivalent to a factor analysis program without outliers.
The log marginal likelihood is now computed as

$$
\log p(\mathbf{Y}) = log\ p(\mathbf{Y} \mid \hat{\theta}_1) + log\ p(\hat{\theta}_1) - log\ p(\hat{\theta} \mid \mathbf{Y}) \qquad (21)
$$

where the likelihood part is given by

$$p(\mathbf{Y} \mid \hat{\theta}_1) = \mathcal{N}[\mathbf{Y} \mid \hat{\mathbf{\Lambda}}\hat{\mathbf{Z}}, \mathbf{I}_n \otimes \hat{\mathbf{\Psi}}], \tag{22}$$

which is the same value as for the factor analysis without outliers, since $\hat{\mathbf{D}}_\vartheta = \mathbf{0}$.
Note that formula (**??**) is a simplification since the components for the location shifts $\hat{\mathbf{A}}$ in (**??**) cancel out.

4.2 Model selection with Bayes factors

Posterior odds are used in Bayesian analysis to choose between two or more different models for the same data set. The basic formula for choosing between models M_1 and M_2 is

$$\begin{aligned} \text{posterior odds} &= \text{Bayes factor} \cdot \text{prior odds} \\ &\text{or} \\ \frac{p(M_1|\mathbf{Y})}{p(M_2|\mathbf{Y})} &= B \cdot \frac{p(M_1)}{p(M_2)}, \end{aligned} \tag{23}$$

where $p(M_1|\mathbf{Y})$ and $p(M_2|\mathbf{Y})$ are the posterior probabilities for models M_1 and M_2, respectively. $p(M_1)$ and $p(M_2)$ are the prior probabilities for models M_1 and M_2, and, in the simplest case, they are set to be equal. Thus, in these cases the posterior odds are equal to the Bayes factor, which is defined as the ratio of marginal likelihoods

$$B = \frac{p(M_1|\mathbf{Y})}{p(M_2|\mathbf{Y})} = \frac{\int p(\mathbf{Y}, \theta_1)d\theta_1}{\int p(\mathbf{Y}, \theta_2)d\theta_2},$$

where θ_1 and θ_2 are the parameters for models M_1 and M_2, respectively. If $B > 1$ we choose model M_1 and if $B < 1$ we choose model M_2. Therefore the model with the largest marginal likelihood will be chosen using simple Bayes factors. For example: The Bayes factor for testing the factor analysis model with outliers against no outliers is:

$$B = \frac{p(\mathbf{Y}|\text{outliers})}{p(\mathbf{Y}|\text{no outliers})}.$$

Clearly, this approach can be combined with the Bayes factor in (**??**).

5 Example

The main purpose of factor analysis is the reduction of dimensionality. Therefore it will generally be difficult to come up with precise prior information. While classical factor analysis achieves identification of the model by non-stochastic restrictions, Bayesian factor analysis can be thought of as a stochastic weakening of these assumptions. Therefore we suggest using the eigenvalue decomposition of covariance matrices as a data-based prior distribution. This approach seems to be plausible since the goal is to explore the effects of outliers. The rationale of this specification is as follows: even if outliers are present, we expect the factors and the factor loadings to be in the "neighborhood" of the original model.

We use the language data in Fuller (1987, page 154) as an example for identifying outliers in a factor analysis. This data consists of 100 observations with eight items, three items related to the essay, three items related to the language used, and two items related to the writing style.

We are interested to see if there are outliers in the 8-dimensional data set which might affect the factor analysis. The assignment of the prior parameters was based on the classical (least squares) approach to factor analysis in the following way:

1) We calculated the eigenvalue decomposition:

$$\begin{aligned} \mathbf{YY}' &= \mathbf{U\Gamma U}', \\ \mathbf{U} &= (\mathbf{u}_1, \ldots, \mathbf{u}_p), \\ \mathbf{\Gamma} &= diag(\gamma_1, \ldots, \gamma_p). \end{aligned}$$

2) If we choose $K = 2$ factors, then we use the first two eigenvectors of $\mathbf{U}$ together with the largest two eigenvalues in $\mathbf{\Gamma}$:

$$\begin{aligned} \mathbf{U}_k &= (\mathbf{u}_1, \mathbf{u}_2), \\ \mathbf{\Gamma}_k &= \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix}. \end{aligned}$$

3) The prior parameters are now

$$\begin{aligned} \mathbf{\Lambda}_* &= \mathbf{YU}_k, \\ \mathbf{H}_* \otimes \mathbf{G}_* &= \mathbf{\Gamma}_k \otimes \mathbf{I}_p, \\ \mathbf{\Phi}_* &= \mathbf{I}_k, \\ \mathbf{\Psi}_* &= diag(var(\mathbf{y}_1), \ldots, var(\mathbf{y}_p))/10. \end{aligned}$$

4) The prior distribution for the outlier parameters in model (14) consists of two parts. The first part is the set of parameters which is identical to the factor analysis model without outliers in section two. For the prior distribution of the location shifts we have assumed $\mathbf{A}_* = \mathbf{0}$ and $var(\mathbf{a}_i) = \mathbf{P}_* = diag(var(\mathbf{y}))$. Finally, we assume that the residual variances of the factor model are about one tenth of the variances of the observed variables. The value of the prior information of the Wishart distribution is $n_* = \nu_* = 1$, i.e. $1/100$ in terms of the sample size $n = 100$.
Convergence of the Gibbs sampler was achieved quite quickly for the present specification. The convergence was monitored by diagnostic measures proposed in the CODA package of Best et al. (1995) written in S-plus which uses the Gelman and Rubin (1992) and the Raftery and Lewis (1992) statistics. A good introduction to the theory and practice of MCMC modeling can be found in Gilks et al. (1990). Only the last 100 simulations of the MCMC sequence were used to calculate the mean and variances of the posterior distribution. Note that convergence could not be achieved for arbitrary prior distributions for factors and factor loadings. Quick and satisfactory results were only obtained for the proposed data-based prior distribution. This approach also guarantees an insensitive calculation of the marginal likelihood. Priors in the vicinity of the data-based prior will produce identical results for model selection via the marginal likelihood criterion and only little variation for Bayes factors.
The eigenvalues of the data covariance matrix are

$$diag\boldsymbol{\Gamma} = (207.339, 15.079, 13.848, 10.185, 10.026, 9.400, 8.038, 6.572),$$

and the cumulated percentages are

$$p_j = \frac{\sum_{i=1}^{j} \gamma_i}{\sum_{i=1}^{8} \gamma_i}, \quad j = 1, \ldots, 8,$$

the results are given by Table **??**.
Table **??** lists the (ordinary) marginal likelihood and the posterior marginal likelihood (see Appendix D) for different number of factors. Two factors are chosen by the maximum marginal likelihood criterion for the model with and without outliers. The Bayes factor is clearly in favor of the factor analysis model with outliers.
Table **??** is the summary of the important result from our factor analysis

j	1	2	3	4	5	6	7	8
p_j	0.739	0.793	0.842	0.879	0.914	0.948	0.977	1.000

Table 1: Cumulated percentages

model with outliers. The first column is the observation number and the second column is the probability of its being an outlier. The third to the tenth columns contain the outlier shifts and the standard deviations of the outliers in parentheses.
Table **??** shows the row estimates of the location shift matrix **A** for which the posterior probability parameter ε_{i**} (the probability of being an outlier) is larger than 1/2. The prior probability that observation i is an outlier is assumed to be $\varepsilon_{i*} = 0.1$. The standard deviations of the location shifts are printed in parentheses. Those location shifts a_{ij} which are larger than the standard deviation are in bold font. It is interesting to note that all five outlier points have location shifts which are shifted by more than one standard deviation in exactly one of the eight variables. This shows that the grading process of the language papers was quite independent with respect to these eight judgment categories. No outlier point shows location shifts in *two* or more variables jointly. Note that the standard deviations of the location shifts varies quite a lot across the outliers. There seems to be no obvious relation to the posterior probability of being an outlier. The size of the location shifts are not too large but make sense if they are compared to the original data.
Figure **??** shows the eight row vectors of the factor loadings matrix $\mathbf{\Lambda}$ for the model with and without outliers. The eight row vectors of the factor loadings for the model without outliers are more spread out than the factor loadings for the factor model with the outliers removed.
Figure **??** shows the posterior distribution of the location shifts by parallel box plots. The posterior means ε_{i**} are plotted in Figure **??** and are interpreted as posterior probabilities of being an outlier. This leads us to the following stochastic outlier analysis: While the posterior means for three observations are above 60%, two more observations just make it above the 50% line. Three (or almost four) additional observations are above 40% while the other observations are certainly not candidates for outlier locations. We

conclude that checking for outliers can be important for factor analysis with data sets which are exposed to possibly aberrant observations.

6 Summary

The paper shows that a sophisticated multivariate model of factor analysis can be combined quite successfully with a multivariate approach to outlier analysis. Under the usual assumption of a normal-Wishart distribution and a location-shift outlier model with Bernoulli-distributed indicator variables, all the full conditional distributions of the Gibbs sampler are given in closed form. The method is demonstrated with the Fuller (1987) language data and is tested for the presence of outliers. We have also shown that the problem of estimating the number of factors can be solved by using the concept of marginal likelihoods. The marginal likelihood can be computed as an additional step in the Gibbs sampling algorithm. With our example we could demonstrate that the marginal likelihood (and posterior marginal likelihood) criteria pick the same number of factors. The present approach uses a data-based prior distribution which yields a fast convergence in the MCMC algorithm. Recently, Shera and Ibrahim (1998) have suggested using historical data for a prior distribution which could also be used in the context of outliers. Further research in this area will show if the model choice in factor analysis can be analysed successfully or more efficiently by different MCMC methods or different outlier modeling approaches.
Acknowlegement: The author wants to thank two unknown referees for their comments. The program can be found in the program package BASEL under
http://www.unibas.ch/iso/basel.

Appendix

Appendix A demonstrates that the multivariate regression model with outliers is a special case of the factor analysis with outlier model.

A The Gibbs sampler for the factor analysis with outliers

The joint distribution of the parameters $\theta = (\mathbf{\Lambda}, \mathbf{Z}, \mathbf{\Psi}, \mathbf{\Phi}, \mathbf{A}, \vartheta)$ and the data $\mathbf{Y}$ is found from the distributions in (??) and (??):

$$
\begin{aligned}
p(\mathbf{Y}, \theta) \; = \; & \mathcal{N}_{p\times n}[\mathbf{Y} \mid \mathbf{A}\mathbf{D}_\vartheta + \mathbf{\Lambda}\mathbf{Z}, \mathbf{I}_n \otimes \mathbf{\Psi}] \times \mathcal{W}_p[\mathbf{\Psi}^{-1} \mid \mathbf{\Psi}_*, n_*] \\
& \times \mathcal{N}_{K\times n}[\mathbf{Z} \mid \mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Phi}] \times \mathcal{W}_K[\mathbf{\Phi}^{-1} \mid \mathbf{\Phi}_*, \nu_*] \\
& \times \mathcal{N}_{p\times K}[\mathbf{\Lambda} \mid \mathbf{\Lambda}_*, \mathbf{H}_* \otimes \mathbf{G}_*] \times \mathcal{N}_{p\times K}[\mathbf{A} \mid \mathbf{A}_*, \mathbf{I}_K \otimes \mathbf{P}_*] \\
& \times \prod_{i=1}^{n} Ber[\vartheta_i, \varepsilon_*].
\end{aligned}
$$

The full conditional distributions are:
a) For the latent factors $\mathbf{Z}$:

$$
\begin{aligned}
p(\mathbf{Z} \mid \mathbf{Y}, \theta^{\mathbf{c}}) \; \propto \; & \exp\{-\frac{1}{2} tr(\mathbf{Y} - \mathbf{A}\mathbf{D}_\vartheta - \mathbf{\Lambda}\mathbf{Z})' \mathbf{\Psi}^{-\mathbf{1}} (\mathbf{Y} - \mathbf{A}\mathbf{D}_\vartheta - \mathbf{\Lambda}\mathbf{Z})\} \\
& \times \exp\{-\frac{1}{2} \mathbf{Z}' \mathbf{\Phi}^{-1} \mathbf{Z}\} \\
= \; & \mathcal{N}_{K\times n}[\mathbf{Z}_{**}, \mathbf{I}_n \otimes \mathbf{D}_Z]. \qquad (24)
\end{aligned}
$$

Because vec $\mathbf{\Lambda}\mathbf{Z} = (\mathbf{I}_n \otimes \mathbf{\Lambda})$ vec $\mathbf{Z}$, we find for the parameters

$$
\mathbf{D}_Z^{-1} \; = \; \mathbf{\Phi}^{-1} + \mathbf{\Lambda}' \mathbf{\Psi}^{-\mathbf{1}} \mathbf{\Lambda},
$$

and

$$
\mathbf{Z}_{**} = \mathbf{D}_Z \mathbf{\Lambda}' \mathbf{\Psi}^{-\mathbf{1}} (\mathbf{Y} - \mathbf{A}\mathbf{D}_\vartheta).
$$

b) For the matrix of factor loadings $\mathbf{\Lambda}$:

$$
\begin{aligned}
p(\mathbf{\Lambda} \mid \mathbf{Y}, \theta^{c}) \; \propto \; & \exp\{-\frac{1}{2} tr(\mathbf{Y} - \mathbf{A}\mathbf{D}_\vartheta - \mathbf{\Lambda}\mathbf{Z})' \mathbf{\Psi}^{-1} (\mathbf{Y} - \mathbf{A}\mathbf{D}_\vartheta - \mathbf{\Lambda}\mathbf{Z})\} \\
& \times \exp\{-\frac{1}{2} tr \mathbf{H}_*^{-1} (\mathbf{\Lambda} - \mathbf{\Lambda}_*)' \mathbf{G}_*^{-1} (\mathbf{\Lambda} - \mathbf{\Lambda}_*)\} \\
= \; & \mathcal{N}_{p\times K}[\mathbf{\Lambda}_{**}, \mathbf{C}_{**}] \qquad (25)
\end{aligned}
$$

with the parameters

$$
\mathbf{C}_{**}^{-1} = \mathbf{H}_*^{-1} \otimes \mathbf{G}_*^{-1} + \mathbf{Z}\mathbf{Z}' \otimes \mathbf{\Psi}^{-1}
$$

and

$$\text{vec } \mathbf{\Lambda}_{**} = \mathbf{C}_{**}[\text{vec } (\mathbf{G}_*^{-1}\mathbf{\Lambda}_*\mathbf{H}_*^{-1} + \mathbf{\Psi}^{-1}(\mathbf{Y} - \mathbf{A}\mathbf{D}_\vartheta)\mathbf{Z})]$$

since also vec $\mathbf{\Lambda Z} = (\mathbf{Z}' \otimes \mathbf{I}_p)$ vec $\mathbf{\Lambda}$ and

$$\text{vec } \mathbf{\Lambda}_{**} = \mathbf{C}_{**}[(\mathbf{H}_*^{-1} \otimes \mathbf{G}_*^{-1}) \text{ vec } \mathbf{\Lambda}_* + (\mathbf{Z} \otimes \mathbf{\Psi}^{-1})\text{vec } (\mathbf{Y} - \mathbf{A}\mathbf{D}_\vartheta)].$$

c) For the covariance matrix $\mathbf{\Psi}$:

$$p(\mathbf{\Psi}^{-1} \mid \mathbf{Y}, \theta^c) = \mathcal{W}_p[\mathbf{\Psi}_{**}, n_{**} = n_* + n] \tag{26}$$

with

$$\mathbf{\Psi}_{**} = \mathbf{\Psi}_* + (\mathbf{Y} - \mathbf{A}\mathbf{D}_\vartheta - \mathbf{\Lambda Z})(\mathbf{Y} - \mathbf{A}\mathbf{D}_\vartheta - \mathbf{\Lambda Z})'$$

because the residual matrix is $\mathbf{E} = \mathbf{Y} - \mathbf{A}\mathbf{D}_\vartheta - \mathbf{\Lambda Z}$, and the f.c.d. is proportional to

$$\begin{aligned} p(\mathbf{\Psi}^{-1} \mid \mathbf{Y}, \theta^c) \;\; &\propto \;\; |\mathbf{\Psi}^{-1}|^{-\frac{n}{2}} \exp\{-\frac{1}{2} tr \mathbf{E}'\mathbf{\Psi}^{-1}\mathbf{E}\} \\ & \quad\;\; |\mathbf{\Psi}^{-1}|^{\frac{n_* - p - 1}{2}} \exp\{-\frac{1}{2} tr \mathbf{\Psi}^{-1}\mathbf{\Psi}_*\}. \end{aligned}$$

d) For the covariance matrix $\mathbf{\Phi}$:

$$\begin{aligned} p(\mathbf{\Phi}^{-1} \mid \mathbf{Y}, \theta^c) \;\; &\propto \;\; |\mathbf{\Phi}^{-1}|^{\frac{n}{2}} \exp\{-\frac{1}{2} tr \mathbf{Z}'\mathbf{\Phi}^{-1}\mathbf{Z}\} && (27) \\ & \quad\;\; |\mathbf{\Phi}^{-1}|^{\frac{\nu_* - K - 1}{2}} \exp\{-\frac{1}{2} tr \mathbf{\Phi}^{-1}\mathbf{\Phi}_*\} && (28) \\ &= \;\; \mathcal{W}_K[\mathbf{\Phi}_{**}, \nu_{**} = \nu_* + n] && (29) \end{aligned}$$

we find as f.c.d. a Wishart distribution with scale matrix

$$\mathbf{\Phi}_{**} = \mathbf{\Phi}_* + \mathbf{Z}\mathbf{Z}'.$$

e) For the level shift matrix $\mathbf{A}$:

$$\begin{aligned} p(\mathbf{A} \mid \mathbf{Y}, \theta^c) \;\; &\propto \;\; \exp\{-\frac{1}{2} tr \mathbf{E}'\mathbf{\Psi}^{-1}\mathbf{E}\} && (30) \\ & \quad\;\; \times \exp\{-\frac{1}{2} tr (\mathbf{A} - \mathbf{A}_*)'\mathbf{P}_*^{-1}(\mathbf{A} - \mathbf{A}_*)\} && (31) \\ &= \;\; \mathcal{N}_{p \times n}[\mathbf{A}_{**}, \tilde{\mathbf{G}}_{**}]. && (32) \end{aligned}$$

Because of the vectorization

$$\text{vec } \mathbf{A}\mathbf{D}_\vartheta = (\mathbf{D}'_\vartheta \otimes \mathbf{I}_n) \text{ vec } \mathbf{A},$$

we have for the parameters of the multivariate normal distribution

$$\begin{aligned} \tilde{\mathbf{G}}_{**}^{-1} &= \mathbf{I}_n \otimes \mathbf{P}_*^{-1} + \mathbf{D}_\vartheta \mathbf{D}'_\vartheta \otimes \mathbf{\Psi}^{-1}, \\ \text{vec } \mathbf{A}_{**} &= \tilde{\mathbf{G}}_{**}[(\mathbf{I}_n \otimes \mathbf{P}_*^{-1}) \text{ vec } \mathbf{A}_* + (\mathbf{D}_\vartheta \otimes \mathbf{\Psi}^{-1}) \text{ vec } (\mathbf{Y} - \mathbf{\Lambda}\mathbf{Z})] \\ &= \tilde{\mathbf{G}}_{**}[\text{vec } (\mathbf{P}_*^{-1}\mathbf{A}_* + \mathbf{\Psi}^{-1}(\mathbf{Y} - \mathbf{\Lambda}\mathbf{Z})\mathbf{D}'_\vartheta)]. \end{aligned}$$

We can calculate the posterior mean for each observation as follows

$$\tilde{\mathbf{G}}_{**}^{-1} = \mathbf{P}_*^{-1} + \vartheta_i^2 \mathbf{\Psi}^{-1},$$

$$\mathbf{a}_{**i} = \tilde{\mathbf{G}}_{**}[\mathbf{P}_*^{-1} + \vartheta_i \mathbf{\Psi}^{-1}(\mathbf{y}_i - \mathbf{\Lambda}\mathbf{z}_i)].$$

f) For the indicator variables ϑ_i the f.c.d. is

$$p(\vartheta_i \mid \mathbf{Y}, \theta^c) = Ber\left[\varepsilon_{i**} = \frac{c_i}{c_i + d_i}\right], \quad i = 1, \ldots, n, \tag{33}$$

we obtain again a Bernoulli distribution for each observation by a straightforward application of the Bayes theorem with the components

$$\begin{aligned} c_i &= Pr(\mathbf{y}_i \mid \vartheta_i = 1, \mathbf{Y}, \theta^c)\varepsilon_* \\ &= \mathcal{N}_p[\mathbf{y}_i \mid \mathbf{a}_i + \mathbf{\Lambda}\mathbf{z}_i, \mathbf{\Psi}]\varepsilon_* \end{aligned}$$

and

$$\begin{aligned} d_i &= Pr(\mathbf{y}_i \mid \vartheta_i = 0, \mathbf{Y}, \theta^c)(1 - \varepsilon_*) \\ &= \mathcal{N}_p[\mathbf{y}_i \mid \mathbf{\Lambda}\mathbf{z}_i, \mathbf{\Psi}](1 - \varepsilon_*), \quad i = 1, \ldots, n, \end{aligned}$$

where $\mathbf{z}_i$ is the i-th column of $\mathbf{Z}$, ϑ_i is the i-th column of $\mathbf{D}_\vartheta$, and $\mathcal{N}_p$ is the p-dimensional normal density function.
The posterior probability that observation i is an outlier is estimated from the MCMC output for the Bernoulli parameter ε_{i**}:

$$\bar{\varepsilon}_{i**} = \frac{1}{J}\sum_{j=1}^{J} \varepsilon_{i**}^{(j)}.$$

B The posterior marginal likelihood

Since the computation of the marginal likelihoods needs considerable computation time, we propose the calculation of the posterior marginal likelihood (see Aitkin 1991) from the MCMC output for large data sets.

1) The factor analysis model:

The posterior marginal likelihood ($PoML$) of the factor analysis model of rank K can be estimated from the MCMC output of size R in following way:

$$\begin{aligned} PoML &= \frac{1}{R}\sum_{j=1}^{R} \mathcal{N}[\mathbf{Y}|\mathbf{\Lambda}_{(j)}\mathbf{Z}_{(j)}, \mathbf{I}_n \otimes \mathbf{\Psi}_{(j)}] \cdot \mathcal{N}[\mathbf{Z}_{(j)}|\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Phi}_{(j)}] \cdot \\ & \mathcal{N}[\mathbf{\Lambda}_{(j)}|\mathbf{\Lambda}_*, \mathbf{H}_* \otimes \mathbf{G}_*] \cdot \mathcal{W}[\mathbf{\Psi}_{(j)}^{-1}|\mathbf{\Psi}_*, n_*] \cdot \mathcal{W}[\mathbf{\Phi}_{(j)}^{-1}|\mathbf{\Phi}_*, \nu_*]. \end{aligned}$$

2) The factor analysis model with outliers:

The $PoML$ for the factor analysis model with outliers is given by

$$\begin{aligned} PoML &= \frac{1}{R}\sum_{j=1}^{R} \mathcal{N}[\mathbf{Y}|\mathbf{\Lambda}_{(j)}\mathbf{Z}_{(j)} + \mathbf{A}_{(j)}\mathbf{D}_{\theta}^{(j)}, \mathbf{I}_n \otimes \mathbf{\Psi}_{(j)}] \cdot \mathcal{N}[\mathbf{Z}_{(j)}|\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Phi}_{(j)}] \cdot \\ & \cdot\mathcal{N}[\mathbf{\Lambda}_{(j)}|\mathbf{\Lambda}_*, \mathbf{H}_* \otimes \mathbf{G}_*] \cdot \mathcal{N}[\mathbf{A}_{(j)}|\mathbf{A}_*, \mathbf{I}_n \otimes \mathbf{P}_*] \cdot \\ & \cdot\mathcal{W}[\mathbf{\Psi}_{(j)}^{-1}|\mathbf{\Psi}_*, n_*] \cdot \mathcal{W}[\mathbf{\Phi}_{(j)}^{-1}|\mathbf{\Phi}_*, \nu_*] \cdot \prod_{i=1}^{n} Ber(\theta_i^{(j)}|\varepsilon_{i*}). \end{aligned}$$

We have shown in several simulation runs that the $PoML$ criterion choses always the same model as the ordinary marginal likelihood criterion. This result can be expected since we use a data-based prior. Therefore we can recommend the $PoML$ criterion for model selection for large data sets, since it is faster to compute.

C References

Aitkin, M. (1991): Posterior Bayes Factors, *J. R. Statistical Society*, 53, 111-142.

Akaike, H. (1987): Factor analysis of AIC, *Psychometrika 52*, 317-332.

Barnett, V. and Lewis, T. (1984): *Outliers in Statistical Data*, Wiley, Chichester.

Best, N.G.; Cowles, M.K. and Vines, K. (1995): CODA: Convergence diagnosis and output analysis software for Gibbs sampling output, Version 3.0, Technical report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.

Chib, S. (1995): Marginal likelihood from the Gibbs output. *JASA, 90*, 1313-1321.

Fuller, W.A. (1987): *Measurement Error Models*, John Wiley & Sons, NY.

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

Gelman, A. and Rubin, D.B. (1992): Inference from iterative simulation using multiple sequences (with discussion), *Stat. Science, 7*, 457-511.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1990): Markov Chain Monte Carlo in Practice, Chapman and Hall, London.

Kitagawa, G. and Akaike, H. (1992): A Quasi Bayesian Approach to Outlier Detection, *Ann. Inst. Stat. Mathematics, 34*, 95-104.

Magnus, J.R. and Neudecker, H. (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley, Chichester.

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

Pettit, L.I. and Smith, A.F.M. (1985): Outliers and Influential Observations in Linear Models, in J.M. Bernardo et al. (eds.) *Bayesian Statistics, 2*, 473-474, Elsevier Publ. co.

Polasek, W. (1994): Gibbs sampling in B-VAR models with latent variables. WWZ-discussion papers Nr. 9415, University of Basel.

Polasek, W. (ed.) (1998): The BASEL package, ISO-WWZ, University of Basel, mimeo.
ftp://iso.iso.unibas.ch/pub/basel

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

Raftery, A.E. and Lewis, S.M. (1992): One long run with diagnostics: implementation strategies for Markov chain Monte Carlo, *Stat. Science, 7*, 493-497.

Reyment, R. and Joereskog, K.G. (1996): *Applied Factor Analysis in the Natural Sciences*, Cambridge University Press.

Shera D.M. and Ibrahim, J.G. (1998): *Prior Elicitation and Computation for Bayesian Factor Analysis*, mimeo, Harvard School of Public Health.

Verdinelli, I. and Wasserman, L. (1991): Bayesian analysis of outlier problems using the Gibbs sampler, *Statistics and Computing 1991-1*, 105-117.