Bayesian Factor Analysis By Gibbs Sampling and Iterated Conditional Modes

Daniel B. Rowe and S. James Press

Key Words: Press-Shigemasu, scores, loadings.

Abstract

Press and Shigemasu proposed a Bayesian factor analysis model in which factor scores, factor loadings, and disturbance variances and covariances were estimated in closed form using a large sample approximation for one of the terms in the posterior distribution. This paper shows that by using Gibbs sampling or Iterated Conditional Modes approaches to estimation instead of the large sample approximation, we can obtain improved point estimators in small samples.

1 Introduction

The Bayesian factor analysis model of Press and Shigemasu, 1989 (we will refer to it as PS89) was revised in 1997 to correct some typographical errors. The model adopted the classical normal sampling model, but assumed that the disturbance covariance matrix was a full positive definite matrix. One of the prior assumptions, however, was that the expected value of the disturbance covariance matrix was diagonal. Natural conjugate prior distributions were assumed for the unknown matrices. A large sample approximation was made for one of the terms in the joint posterior density with the result that the marginal posterior distribution of the factor scores was found to be approximately matrix T. The factor loading matrix was estimated conditional on the factor scores, and the disturbance covariance matrix was estimated conditional on estimates of the factor scores and the factor loadings.

In this paper, we adopt the same model as in PS89, but we estimate the unknown quantities by means of Gibbs sampling and by Iterated Conditional Modes (ICM), see Lindley and Smith, 1972. For both approaches we will require conditional posterior distributions for the factor score vectors given the factor loadings, the disturbance covariance matrix, and the data; the conditional posterior distribution of the factor loadings given the factor score vectors, the disturbance covariance matrix, and the data; the conditional posterior distribution of the disturbance covariance matrix given the factor score vectors, the factor loadings, and the data. All three can be found explicitly. Gibbs marginal and ICM conditional estimators may then readily be found from the conditional posterior distributions.

The plan of the paper is to review the model and to adopt prior distributions in Section 2. We outline the estimation procedures in Section 3, and obtain conditional posterior densities in Section 4. Section 5 discusses the numerical Gibbs sampling and ICM estimation of the unknown quantities of a Bayesian factor analysis in an example examined in PS89 and estimated there by other means.

2 Model

2.1 Likelihood Function

The p-variate observation vectors are, $(x_1, \ldots, x_N) \equiv X'$ on N subjects. We will assume that they have mean zero, so that $E(X') = 0$. The prime denotes transposition. The Bayesian factor analysis model of PS89 is:

$$\underset{(p\times 1)}{(x_j|\Lambda, f_j)} = \underset{(p\times m)}{\Lambda} \; \underset{(m\times 1)}{f_j} + \underset{(p\times 1)}{\epsilon_j} \quad , \quad m < p, \tag{2.1}$$

for $j = 1, \ldots, N$, where Λ is a matrix of constants called the factor loading matrix; f_j is the factor score vector for subject j; $F' \equiv (f_1, \ldots, f_N)$, and the ϵ_j's are assumed to be mutually uncorrelated and normally distributed $N(0, \Psi)$ variables.

In the traditional model, Ψ is taken to be a diagonal matrix so that we can readily distinguish common and specific factors; in the the PS89 model, Ψ is taken to be a general symmetric, positive definite covariance matrix with the property of being diagonal on the average, i.e., $E(\Psi) =$ a diagonal matrix.

We assume that (Λ, F, Ψ) are unobservable but fixed quantities, and that we can write the distribution of each x_j as

$$p(x_j|\Lambda, f_j, \Psi) = (2\pi)^{-\frac{p}{2}}|\Psi|^{-\frac{1}{2}}e^{-\frac{1}{2}(x_j-\Lambda f_j)'\Psi^{-1}(x_j-\Lambda f_j)}. \tag{2.2}$$

If we denote proportionality by "$\propto$" then, the likelihood for (Λ, F, Ψ) is

$$p(X|\Lambda, F, \Psi) \propto |\Psi|^{-\frac{N}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(X-F\Lambda')'(X-F\Lambda')}. \tag{2.3}$$

The notation $p(\cdot)$ will denote "density"; densities will be distinguished by their arguments. The proportionality constant in (2.3) depends only on

(p, N) and not on (Λ, F, Ψ).

2.2 Priors

We adopt the same natural conjugate prior distributions (Press, 1982, p. 253–254) as in PS89, so that the joint prior density is:

$$p(\Lambda, F, \Psi) \propto p(\Lambda|\Psi)p(\Psi)p(F), \tag{2.4}$$

where

$$p(\Lambda|\Psi) \propto |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)'}, \tag{2.5}$$

$$p(\Psi) \propto |\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2} tr \Psi^{-1} B}, \tag{2.6}$$

$$p(F) \propto e^{-\frac{1}{2} tr F'F} \tag{2.7}$$

with $H, B, \Psi > 0$ and B a diagonal matrix. That, is Λ conditional on Ψ has elements which are jointly normally distributed, and hyperparameters (Λ_0, H) are to be assesssed. The matrix Ψ^{-1} follows a Wishart distribution, with hyperparameters (ν, B) which are to be assessed. We are assuming that $E(\Psi|B)$ and $E(\Psi)$ are diagonal, in order to represent traditional views of the factor model containing "common" and "specific" factors.

The joint normal distribution for $(\Lambda|\Psi)$ comes from writing $\Lambda' \equiv (\lambda_1, \ldots, \lambda_p)$, as $\lambda \equiv vec(\Lambda') = (\lambda_1', \ldots, \lambda_p')'$; then $\text{var}(\lambda|\Psi) = \Psi \otimes H^{-1}$, which can be written as a matrix normal distribution (Kotz and Johnson, 1985, p. 326–333). Also, as in PS89, we will take $H = h_0 I$, for some preassigned scalar h_0. Assessment of the hyperparameters is simplified by these interpretations.

2.3 Joint Posterior

Using Bayes rule, we combine (2.3)–(2.5), to get the joint posterior density of the parameters

$$p(F, \Lambda, \Psi|X) \quad \propto \quad e^{-\frac{1}{2}trF'F}|\Psi|^{-\frac{(N+m+\nu)}{2}}e^{-\frac{1}{2}tr\Psi^{-1}U} \tag{2.8}$$

where $U = (X - F\Lambda')'(X - F\Lambda') + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B$.

3 Estimation

3.1 Gibbs Sampling

For our Gibbs sampling approach (Geman & Geman, 1984 and Gelfand & Smith, 1990) to estimation we draw random samples from the posterior conditional distribution for each of the parameters conditional on the fixed values of all the other parameters and the data X. Let $p(\theta|X)$ be the posterior distribution of the parameters where θ is the set of parameters and X is the data. If θ is partitioned by $\theta = (\theta_1, \theta_2, \ldots, \theta_J)$ into J groups of parameters, we begin with an initial value $\bar{\theta}^{(0)} = (\bar{\theta}_1^{(0)}, \bar{\theta}_2^{(0)}, \ldots, \bar{\theta}_J^{(0)})$ and at the i^{th} iteration define $\bar{\theta}^{(i+1)} = (\bar{\theta}_1^{(i+1)}, \bar{\theta}_2^{(i+1)}, \ldots, \bar{\theta}_J^{(i+1)})$ by

$$\bar{\theta}_1^{(i+1)} = \text{a random sample from } p(\bar{\theta}_1|\bar{\theta}_2^{(i)}, \bar{\theta}_3^{(i)}, \ldots, \bar{\theta}_J^{(i)}, X) \tag{3.1}$$

$$\bar{\theta}_2^{(i+1)} = \text{a random sample from } p(\bar{\theta}_2|\bar{\theta}_1^{(i+1)}, \bar{\theta}_3^{(i)}, \ldots, \bar{\theta}_J^{(i)}, X) \tag{3.2}$$

$$\vdots$$

$$\bar{\theta}_J^{(i+1)} = \text{a random sample from } p(\bar{\theta}_J|\bar{\theta}_2^{(i+1)}, \bar{\theta}_3^{(i+1)}, \ldots, \bar{\theta}_{J-1}^{(i+1)}, X) \tag{3.3}$$

that is, at each step drawing a random sample from the conditional posterior distribution. To apply this method we need to determine the posterior conditionals of θ_j, conditional on the fixed values of all the other elements of θ and X from $p(\theta|X)$.

We will have $\bar{\theta}^{(1)}, \bar{\theta}^{(2)}, \bar{\theta}^{(s)}, \ldots, \bar{\theta}^{(s+1)}, \ldots, \bar{\theta}^{(s+t)}$. The first s random samples called the "burn in" are discarded and the remaining t samples are kept. Our marginal posterior mean and modal estimators of the parameters are $\bar{\theta} = (\bar{\theta}_1, \ldots, \bar{\theta}_J)$ where

$$\bar{\theta}_j = \frac{1}{t} \sum_{k=1}^{t} \bar{\theta}_j^{(s+k)}, \quad j = 1, \ldots, J. \tag{3.4}$$

3.2 Iterated Conditional Modes (ICM)

Iterated Conditional Modes (Lindley and Smith, 1972) is a deterministic optimization method that finds the joint posterior modal estimators of $p(\theta|X)$ where θ denotes the vector of parameters, and X denotes the data.

Assume that $\theta = (\theta_1, \theta_2)$ where θ_1 and θ_2 are scalars and the posterior density of θ is $p(\theta_1, \theta_2|X)$. We have a surface in 3-Dimensional space. We have θ_1 along one axis and θ_2 along the other with $p(\theta_1, \theta_2|X)$ being the height of the surface or hill.

We want to find the top of the hill which is the same as finding the peak or maximum of the function $p(\theta_1, \theta_2|X)$ with respect to both θ_1 and θ_2. Well we find the maximum of a surface by differentiating with respect to each variable (direction).

The maximum of the function $p(\theta_1, \theta_2|X)$ satisfies

$$\frac{\partial}{\partial\theta_1} p(\theta_1, \theta_2|X) \Big|_{\theta_1=\tilde{\theta}_1} = \frac{\partial}{\partial\theta_2} p(\theta_1, \theta_2) \Big|_{\theta_2=\tilde{\theta}_2} = 0, \tag{3.5}$$

which is the same as

$$\frac{\partial}{\partial \theta_1} p(\theta_1|\theta_2, X) p(\theta_2|X) \Big|_{\theta_1=\tilde{\theta}_1} = \frac{\partial}{\partial \theta_2} p(\theta_2|\theta_1, X) p(\theta_1|X) \Big|_{\theta_2=\tilde{\theta}_2} = 0 \quad (3.6)$$

or

$$p(\theta_2|X) \frac{\partial}{\partial \theta_1} p(\theta_1|\theta_2, X) \Big|_{\theta_1=\tilde{\theta}_1} = p(\theta_1|X) \frac{\partial}{\partial \theta_2} p(\theta_2|\theta_1, X) \Big|_{\theta_2=\tilde{\theta}_2} = 0 \quad (3.7)$$

assuming that $p(\theta_1|X) \neq 0$ and $p(\theta_2|X) \neq 0$.

We can obtain the posterior conditionals (functions) $p(\theta_1|\theta_2, X)$ and $p(\theta_2|\theta_1, X)$ along with their respective modes (maximum) $\tilde{\theta}_1 = \tilde{\theta}_1(\theta_2, X)$ and $\tilde{\theta}_2 = \tilde{\theta}_2(\theta_1, X)$.

We have the maximum of θ_1, $\tilde{\theta}_1$ for a given value of (conditional on) θ_2, and the maximum of θ_2, $\tilde{\theta}_2$ for a given value of (conditional on) θ_1.

The optimization procedure consists of

(1) Selecting an initial value for θ_2; call it $\tilde{\theta}_2^{(0)}$.

(2) Calculate the modal (maximal) value of $p(\theta_1|\tilde{\theta}_2^{(0)}, X)$, $\tilde{\theta}_1^{(1)}$.

(3) Calculate the modal (maximal) value of $p(\theta_2|\tilde{\theta}_1^{(1)}, X)$, $\tilde{\theta}_2^{(1)}$.

(4) Continue to calculate the remainder of the sequence $\tilde{\theta}_1^{(1)}, \tilde{\theta}_2^{(1)}, \tilde{\theta}_1^{(2)}, \tilde{\theta}_2^{(2)}, \ldots$ until convergence is reached.

This method always converges to a mode or maximum. If the posterior conditional distributions are not unimodal, we may converge to a local maximum and not the global maximum. If the posterior conditionals are unimodal, then we will always converge to the global maximum.

This method can be generalized to more than two parameters. According to O'Hagan (1994), If θ is partitioned by $\theta = (\theta_1, \theta_2, \ldots, \theta_J)$ into J groups of parameters, we begin with a starting point $\tilde{\theta}^{(0)} = (\tilde{\theta}_1^{(0)}, \tilde{\theta}_2^{(0)}, \ldots, \tilde{\theta}_J^{(0)})$ and at the i^{th} iteration define $\tilde{\theta}^{(i+1)}$ by

$$\tilde{\theta}_1^{(i+1)} = \tilde{\theta}_1(\tilde{\theta}_2^{(i)}, \tilde{\theta}_3^{(i)}, \ldots, \tilde{\theta}_J^{(i)}) \tag{3.8}$$

$$\tilde{\theta}_2^{(i+1)} = \tilde{\theta}_2(\tilde{\theta}_1^{(i+1)}, \tilde{\theta}_3^{(i)}, \ldots, \tilde{\theta}_J^{(i)}) \tag{3.9}$$

$$\vdots$$

$$\tilde{\theta}_J^{(i+1)} = \tilde{\theta}_1(\tilde{\theta}_2^{(i+1)}, \tilde{\theta}_3^{(i+1)}, \ldots, \tilde{\theta}_{J-1}^{(i+1)}) \tag{3.10}$$

at each step computing the maximum or mode. To apply this method we need to determine the functions $\tilde{\theta}_j$ which give the maximum of $p(\theta|X)$ with respect to $\tilde{\theta}_j$, conditional on the fixed values of all the other elements of θ.

This is the form of ICM we will be implementing.

3.3 Comparison With Other Methods

The estimation procedure proposed in PS89 was obtained in closed form so that no iterations are required. In PS89, a large sample approximation was made for one of the terms in the joint posterior density with the result that the marginal posterior distribution of the factor scores was found to be asymptotically matrix T. The factor loading matrix was estimated conditional on the factor scores, and the disturbance covariance matrix was estimated conditional on estimates of the factor scores and the factor loadings.

Marginal posterior point and interval estimates were obtained for the factor scores; as well as point and interval estimates for the factor loadings conditional on the mean and modal value for the factor scores; and point and interval estimates for the disturbance covariance matrix conditional on the marginal mean and modal value for the factor scores and the conditional mean and modal value for the factor loadings given the marginal mean and modal value for the factor scores.

It is of interest to compare the PS89 results with those obtained from a Gibbs sampler as well as for ICM. In addition to being another estimation procedure to compare to, there are several advantages to ICM in this model.

(1) We will show that each of the posterior conditional distributions is unimodal, so that the joint maxima ICM finds are joint global maxima. The reason one would use a stochastic procedure like Gibbs sampling over a deterministic procedure like ICM is to eliminate the possibility of coverging to a local joint mode when the posterior distribution is multimodal.

(2) Both ICM and Gibbs sampling do not require the exact form of the conditionals, that is they do not require the proportionality constant but Gibbs requires the exact form otherwise while ICM only requires the mode. We know the kernel of the conditionals, which are well known distributions with modes known in closed form. For ICM one simply has to cycle through the posterior conditional modes.

(3) Convergence is not uncertain with ICM as it is with Gibbs sampling with small numbers of iterations. With Gibbs sampling there is always some uncertainty as to whether the sequence has effectively converged, and hence uncertainty about how accurate the estimates are (O'Hagan

p.241), without taking additional samples from the posterior.

Although, both ICM and Gibbs sampling have slow convergence when the correlations between the partitioned parameters is close to ± 1 we can often group the parameters so that the correlations between groups is small.

However, Gibbs sampling can yield marginal posterior point estimates as well as marginal posterior interval estimates (see Geman & Geman, 1984 and Gelfand & Smith, 1990) while ICM can yield joint posterior point estimates and conditional posterior interval estimates. Of course in large samples the PS89 estimates are correct and have the advantage of being easy to compute as well as having associated distributional theory available.

3.4 Convergence

For ICM, we define convergence to mean that each of the elements of the estimator for the factor scores changes by less than 0.001 after an additional thousand iterations. For Gibbs sampling, we suggest standard diagnostic methods which include several parallel runs of differing lengths and burn-in periods.

4 Conditional Posterior Densities

The three conditional posterior densities required are found as follows.

$$\begin{aligned} p(\Lambda|F,\Psi,X) &\propto p(F,\Lambda,\Psi)p(X|F,\Lambda,\Psi) \qquad (4.1) \\ &= p(\Lambda|\Psi)p(\Psi)p(F)p(X|F,\Lambda,\Psi) \end{aligned}$$

$$
\begin{aligned}
&\propto \quad |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)'} \\
&\cdot \quad |\Psi|^{-\frac{N}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(X-F\Lambda')'(X-F\Lambda')} \\
&\propto \quad e^{-\frac{1}{2} tr \Psi^{-1}[(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)'+(X-F\Lambda')'(X-F\Lambda')]}
\end{aligned}
$$

which after some algebra becomes

$$
p(\Lambda|F,\Psi,X) \quad \propto \quad e^{-\frac{1}{2} tr \Psi^{-1}(\Lambda-\tilde{\Lambda})(H+F'F)(\Lambda-\tilde{\Lambda})'} \tag{4.2}
$$

where $\tilde{\Lambda} = (X'F + \Lambda_0 H)(H + F'F)^{-1}$.

$$
\begin{aligned}
p(\Psi|F,\Lambda,X) \quad &\propto \quad p(\Psi)p(\Lambda|\Psi)p(F)p(X|F,\Lambda,\Psi) &(4.3)\\
&\propto \quad |\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2} tr \Psi^{-1} B} |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)'} \\
&\cdot \quad |\Psi|^{-\frac{N}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(X-F\Lambda')'(X-F\Lambda')} \\
&\propto \quad |\Psi|^{-\frac{(N+m+\nu)}{2}} e^{-\frac{1}{2} tr \Psi^{-1} U} &(4.4)
\end{aligned}
$$

where $U = (X - F\Lambda')(X - F\Lambda')' + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B$.

$$
\begin{aligned}
p(F|\Lambda,\Psi,X) \quad &\propto \quad p(\Psi)p(\Lambda|\Psi)p(F)p(X|F,\Lambda,\Psi) &(4.5)\\
&\propto \quad e^{-\frac{1}{2} tr F'F} |\Psi|^{-\frac{N}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(X-F\Lambda')'(X-F\Lambda')} \\
&\propto \quad e^{-\frac{1}{2} tr F'F} e^{-\frac{1}{2} tr (X-F\Lambda')\Psi^{-1}(X-F\Lambda')'}
\end{aligned}
$$

which after some algebra can be written as

$$
p(F|\Lambda,\Psi,X) \quad \propto \quad e^{-\frac{1}{2} tr (F-\tilde{F})(I_m+\Lambda'\Psi^{-1}\Lambda)(F-\tilde{F})'} \tag{4.6}
$$

where $\tilde{F} \equiv X\Psi^{-1}\Lambda(I_m + \Lambda'\Psi^{-1}\Lambda)^{-1}$.

The modes of these conditional distributions are $\tilde{F}$, $\tilde{\Lambda}$ (as defined above), and

$$\tilde{\Psi} \;=\; \frac{U}{N+m+\nu}, \tag{4.7}$$

respectively.

The Gibbs Sampling Algorithm

For Gibbs estimation of the posterior, we start with initial values for F and Ψ say $\bar{F}_{(0)}$ and $\bar{\Psi}_{(0)}$. Then cycle through

$$\begin{aligned}
\bar{\Lambda}_{(i+1)} &\equiv \text{ a random sample from } p(\Lambda|\bar{F}_{(i)},\bar{\Psi}_{(i)},X) \\
\bar{\Psi}_{(i+1)} &\equiv \text{ a random sample from } p(\Psi|\bar{F}_{(i)},\bar{\Lambda}_{(i+1)},X) \\
\bar{F}_{(i+1)} &\equiv \text{ a random sample from } p(F|\bar{\Lambda}_{(i+1)},\bar{\Psi}_{(i+1)},X)
\end{aligned}$$

and we have $(\bar{\Lambda}_{(1)},\bar{\Psi}_{(1)},\bar{F}_{(1)}),\ \ldots,(\bar{\Lambda}_{(s)},\bar{\Psi}_{(s)},\bar{F}_{(s)}),\ (\bar{\Lambda}_{(s+1)},\bar{\Psi}_{(s+1)},\bar{F}_{(s+1)})$, …, $(\bar{\Lambda}_{(s+t)}\ ,\bar{\Psi}_{(s+t)},\bar{F}_{(s+t)})$. The first s random samples are discarded and the remaining t samples are kept. The means of the remaining random samples

$$\bar{F} = \frac{1}{t}\sum_{k=1}^{t}\bar{F}_{(s+k)}$$

$$\bar{\Lambda} = \frac{1}{t}\sum_{k=1}^{t}\bar{\Lambda}_{(s+k)}$$

$$\bar{\Psi} = \frac{1}{t}\sum_{k=1}^{t}\bar{\Psi}_{(s+k)}$$

are used as the posterior estimates of the parameters.

The ICM Algorithm

For Iterated Conditional Modes estimatiom of the posterior, we start with an initial value for $\tilde{F}$, say $\tilde{F}_{(0)}$ and then cycle through

$$\begin{aligned}
\tilde{\Lambda}_{(i+1)} &\equiv (X'\tilde{F}_{(i)} + \Lambda_0 H)(H + \tilde{F}'_{(i)}\tilde{F}_{(i)})^{-1} \\
\tilde{\Psi}_{(i+1)} &\equiv \frac{(X - \tilde{F}_{(i)}\tilde{\Lambda}'_{(i+1)})'(X - \tilde{F}_{(i)}\tilde{\Lambda}'_{(i+1)}) + (\tilde{\Lambda}_{(i+1)} - \Lambda_0)H(\tilde{\Lambda}_{(i+1)} - \Lambda_0)' + B}{N + m + \nu} \\
\tilde{F}_{(i+1)} &\equiv X\tilde{\Psi}^{-1}_{(i+1)}\tilde{\Lambda}_{(i+1)}(I_m + \tilde{\Lambda}'_{(i+1)}\tilde{\Psi}^{-1}_{(i+1)}\tilde{\Lambda}_{(i+1)})^{-1}.
\end{aligned}$$

until convergence is reached with the joint posterior modal estimator $(\tilde{F}, \tilde{\Lambda}, \tilde{\Psi})$.

5 Example

In this section we illustrate the use of ICM and Gibbs Sampler procedures for estimating the parameters of the Bayesian factor analysis model, and we compare the resulting estimators with those obtained in PS89. The data is extracted from an example in Kendall 1980, p.53. The problem as stated in PS89 is the following.

> There are 48 applicants for a certain job, and they have been scored on 15 variables regarding their acceptability. They are:

(1)	Form of letter application	(9)	Experience
(2)	Appearance	(10)	Drive
(3)	Academic ability	(11)	Ambition
(4)	Likeabiliy	(12)	Grasp
(5)	Self-confidence	(13)	Potential
(6)	Lucidity	(14)	Keenness to join
(7)	Honesty	(15)	Suitability
(8)	Salesmanship		

The raw scores of the applicants on these 15 variables, measured on the same scale, are presented in Table 1. The question is, Is there an underlying

Table 1: Raw scores of 48 applicants scaled on 15 variables.

Person	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	6	7	2	5	8	7	8	8	3	8	9	7	5	7	10
2	9	10	5	8	10	9	9	10	5	9	9	8	8	8	10
3	7	8	3	6	9	8	9	7	4	9	9	8	6	8	10
4	5	6	8	5	6	5	9	2	8	4	5	8	7	6	5
5	6	8	8	8	4	5	9	2	8	5	5	8	8	7	7
6	7	7	7	6	8	7	10	5	9	6	5	8	6	6	6
7	9	9	8	8	8	8	8	8	10	8	10	8	9	8	10
8	9	9	9	8	9	9	8	8	10	9	10	9	9	9	10
9	9	9	7	8	8	8	8	5	9	8	9	8	8	8	10
10	4	7	10	2	10	10	7	10	3	10	10	10	9	3	10
11	4	7	10	0	10	8	3	9	5	9	10	8	10	2	5
12	4	7	10	4	10	10	7	8	2	8	8	10	10	3	7
13	6	9	8	10	5	4	9	4	4	4	5	4	7	6	8
14	8	9	8	9	6	3	8	2	5	2	6	6	7	5	6
15	4	8	8	7	5	4	10	2	7	5	3	6	6	4	6
16	6	9	6	7	8	9	8	9	8	8	7	6	8	6	10
17	8	7	7	7	9	5	8	6	6	7	8	6	6	7	8
18	6	8	8	4	8	8	6	4	3	3	6	7	2	6	4
19	6	7	8	4	7	8	5	4	4	2	6	8	3	5	4
20	4	8	7	8	8	9	10	5	2	6	7	9	8	8	9
21	3	8	6	8	8	8	10	5	3	6	7	8	8	5	8
22	9	8	7	8	9	10	10	10	3	10	8	10	8	10	8
23	7	10	7	9	9	9	10	10	3	9	9	10	9	10	8
24	9	8	7	10	8	10	10	10	2	9	7	9	9	10	8
25	6	9	7	7	4	5	9	3	2	4	4	4	4	5	4
26	7	8	7	8	5	4	8	2	3	4	5	6	5	5	6
27	2	10	7	9	8	9	10	5	3	5	6	7	6	4	5
28	6	3	5	3	5	3	5	0	0	3	3	0	0	5	0
29	4	3	4	3	3	0	0	0	0	4	4	0	0	5	0
30	4	6	5	6	9	4	10	3	1	3	3	2	2	7	3
31	5	5	4	7	8	4	10	3	2	5	5	3	4	8	3
32	3	3	5	7	7	9	10	3	2	5	3	7	5	5	2
33	2	3	5	7	7	9	10	3	2	2	3	6	4	5	2
34	3	4	6	4	3	3	8	1	1	3	3	3	2	5	2
35	6	7	4	3	3	0	9	0	1	0	2	3	1	5	3
36	9	8	5	5	6	6	8	2	2	2	4	5	6	6	3
37	4	9	6	4	10	8	8	9	1	3	9	7	5	3	2
38	4	9	6	6	9	9	7	9	1	2	10	8	5	5	2
39	10	6	9	10	9	10	10	10	10	10	8	10	10	10	10
40	10	6	9	10	9	10	10	10	10	10	10	10	10	10	10
41	10	7	8	0	2	1	2	0	10	2	0	3	0	0	10
42	10	3	8	0	1	1	0	0	10	0	0	0	0	0	10
43	3	4	9	8	2	4	5	3	6	2	1	3	3	3	8
44	7	7	7	6	9	8	8	6	8	8	10	8	8	6	5
45	9	6	10	9	7	7	10	2	1	5	5	7	8	4	5
46	9	8	10	10	7	9	10	3	1	5	7	9	9	4	4
47	0	7	10	3	5	0	10	0	0	2	2	0	0	0	0
48	0	6	10	1	5	0	10	0	0	2	2	0	0	0	0

subset of factors that explain the variation observed in the scores? If so, then the applicants could be compared more easily. The correlation matrix for the 15 variables is given in Table 2. (Note: We assume the sample size of 48 is large enough to estimate the mean well enough for it to be ignored.)

Table 2: Correlation Matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.000	0.239	0.044	0.306	0.092	0.229	-0.107	0.269	0.548	0.346	0.285	0.338	0.367	0.467	0.586
2		1.000	0.123	0.380	0.431	0.376	0.354	0.477	0.141	0.341	0.550	0.506	0.507	0.284	0.384
3			1.000	0.002	0.001	0.080	-0.030	0.046	0.266	0.094	0.044	0.198	0.290	-0.323	0.140
4				1.000	0.302	0.489	0.645	0.347	0.141	0.393	0.347	0.503	0.606	0.685	0.327
5					1.000	0.802	0.410	0.816	0.015	0.704	0.842	0.721	0.672	0.482	0.250
6						1.000	0.360	0.823	0.155	0.700	0.758	0.890	0.785	0.533	0.420
7							1.000	0.231	-0.156	0.280	0.215	0.386	0.416	0.448	0.003
8								1.000	0.233	0.811	0.860	0.766	0.735	0.549	0.548
9									1.000	0.337	0.195	0.299	0.348	0.215	0.693
10										1.000	0.780	0.714	0.788	0.613	0.623
11											1.000	0.784	0.769	0.547	0.435
12												1.000	0.876	0.549	0.528
13													1.000	0.539	0.574
14														1.000	0.396
15															1.000

Note that the initial values for the ICM and Gibbs sampling estimation procedures have little effect on the final result, because for ICM we have unimodal posterior conditional distributions so we are sure to converge to the mode quickly, and for Gibbs sampling, we have a burn-in period. We choose the initial value for $\tilde{F}$ to be $\tilde{F}_{(0)} = \hat{F}$, the estimator of PS89. This choice of the initial value hastens convergence. (For ICM, we have used other initial values such as every element of $\tilde{F}_{(0)}$ being either $+10$ or -10 and converged to the same values but it took many more iterations.)

We postulate the same underlying structure as in as PS89, a model with 4 factors. This choice is based upon PS89 having carried out a principal components analysis and having found that 4 factors accounted for 81.5% of the variance. Based upon underlying theory they constructed the prior factor loading matrix

$$\Lambda_0 = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \end{array} \begin{pmatrix} 0 & 0 & .7 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & .7 & 0 & 0 \\ 0 & 0 & 0 & .7 \\ .7 & 0 & 0 & 0 \\ .7 & 0 & 0 & 0 \\ 0 & 0 & 0 & .7 \\ .7 & 0 & 0 & 0 \\ 0 & 0 & .7 & 0 \\ .7 & 0 & 0 & 0 \\ .7 & 0 & 0 & 0 \\ .7 & 0 & 0 & 0 \\ .7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & .7 & 0 \end{pmatrix}$$

In PS89, the hyperparameter H was assessed as $H = 10I_4$, B was assessed as $B = 0.2I_{15}$, and ν was assessed as $\nu = 33$. The factor scores, factor loadings, and disturbance variances and covariances may now be estimated. We found that a burn in period of $50,000$ samples would work well, so we used the next $200,000$ samples for our Gibbs estimates. Tables 3, 4, and 5 present comparisons of ICM, Gibbs, and PS89 modal estimators of the factor scores, factor loadings, and disturbance variances and covariances, respectively.

Table 3:

	Gibbs Mean and Modal Estimates of Factor Scores, $\bar{F}$.				ICM Modal Estimates of Factor Scores, $\tilde{F}$.				PS89 Modal Estimates of Factor Scores, $\hat{F}$.			
j	1	2	3	4	1	2	3	4	1	2	3	4
1	0.747	-3.280	0.258	-0.542	0.748	-3.508	0.345	-0.463	0.728	-3.548	0.405	-0.301
2	1.345	-1.468	0.917	0.218	1.426	-1.537	1.086	0.463	1.476	-1.454	1.225	0.735
3	0.964	-2.667	0.521	-0.130	1.002	-2.846	0.635	0.031	1.020	-2.850	0.726	0.231
4	0.319	0.491	0.345	0.172	-0.315	0.624	0.280	0.113	-0.288	0.640	0.226	-0.021
5	-0.517	0.190	0.620	0.717	-0.431	0.465	0.701	0.765	-0.324	0.640	0.691	0.735
6	0.258	-0.088	0.936	0.562	0.256	-0.047	0.882	0.561	0.263	-0.058	0.868	0.510
7	1.029	0.390	1.728	0.164	1.112	0.489	1.873	0.297	1.188	0.640	1.942	0.456
8	1.290	1.117	1.701	0.095	1.385	1.190	1.858	0.268	1.475	1.338	1.942	0.456
9	0.711	-0.266	1.594	0.162	0.796	-0.189	1.730	0.309	0.876	-0.058	1.799	0.456
10	1.982	1.981	-0.025	-1.333	1.949	2.046	0.008	-1.359	1.880	2.035	0.050	-1.336
11	1.678	1.871	-0.304	-2.722	1.622	2.024	-0.352	-2.866	1.550	2.035	-0.382	-2.957
12	1.606	1.904	-0.528	-0.745	1.586	2.035	-0.537	-0.781	1.547	2.035	-0.525	-0.832
13	-0.726	0.322	0.072	1.051	-0.654	0.490	0.209	1.148	-0.590	0.640	0.261	1.239
14	-0.723	0.337	0.396	0.640	-0.669	0.527	0.447	0.692	-0.623	0.640	0.472	0.708
15	-0.750	0.271	0.056	0.878	-0.732	0.541	0.064	0.850	-0.708	0.640	0.049	0.762
16	0.828	-0.972	0.919	-0.030	0.874	-0.870	1.047	0.069	0.903	-0.756	1.123	0.204
17	0.415	0.071	0.851	0.090	0.420	-0.046	0.883	0.124	0.422	-0.058	0.903	0.204
18	-0.161	0.788	-0.509	-1.278	-0.167	0.692	-0.520	-1.205	-0.195	0.640	-0.458	-1.111
19	-0.168	0.698	-0.269	-1.357	-0.189	0.680	-0.317	-1.381	-0.217	0.640	-0.315	-1.390
20	0.592	-0.144	-0.479	0.671	0.670	-0.122	-0.341	0.861	0.729	-0.058	-0.237	1.014
21	0.558	-0.921	-0.556	0.887	0.589	-0.810	-0.478	0.958	0.597	-0.756	-0.415	1.014
22	1.425	0.057	0.450	0.597	1.508	-0.097	0.572	0.816	1.591	-0.058	0.650	1.014
23	1.354	-0.186	-0.074	0.659	1.483	-0.200	0.145	0.968	1.591	-0.058	0.295	1.266
24	1.156	-0.029	0.297	1.119	1.255	-0.134	0.441	1.335	1.364	-0.058	0.506	1.518
25	-0.988	-0.213	-0.669	0.383	-0.957	-0.123	-0.629	0.446	-0.932	-0.058	-0.602	0.483
26	-0.762	-0.219	-0.026	0.447	-0.733	-0.125	0.010	0.462	-0.702	-0.058	0.007	0.456
27	0.288	-0.351	-1.220	1.047	0.326	-0.153	-1.130	1.160	0.326	-0.058	-1.024	1.266
28	-1.606	-0.842	-1.076	-1.189	-1.735	-1.206	-1.302	-1.438	-1.822	-1.454	-1.464	-1.642
29	-1.894	-1.786	-1.490	-2.517	-1.999	-2.039	-1.617	-2.841	-2.047	-2.152	-1.819	-3.038
30	-0.870	-0.961	-1.189	0.427	-0.920	-1.272	-1.270	0.466	-0.976	-1.454	-1.244	0.510
31	-0.530	-1.693	-0.811	0.786	-0.568	-2.001	-0.890	0.772	-0.586	-2.152	-0.923	0.762
32	-0.024	-1.138	-1.090	1.205	-0.110	-1.277	-1.284	1.011	-0.151	-1.454	-1.422	0.762
33	-0.334	-1.085	-1.296	1.155	-0.428	-1.252	-1.492	0.976	-0.492	-1.454	-1.600	0.762
34	-1.485	-0.514	-1.316	-0.191	-1.559	-0.637	-1.447	-0.373	-1.601	-0.756	-1.566	-0.553
35	-2.141	-1.968	-0.789	-0.466	-2.169	-2.065	-0.872	-0.474	-2.198	-2.152	-0.889	-0.526
36	-0.747	-1.379	-0.157	-0.348	-0.726	-1.416	-0.216	-0.276	-0.699	-1.454	-0.213	-0.301
37	0.819	-0.622	-1.437	-0.693	0.773	-0.677	-1.480	-0.648	0.674	-0.756	-1.388	-0.553
38	0.772	-0.680	-1.513	-0.574	0.770	-0.737	-1.490	-0.484	0.722	-0.756	-1.388	-0.327
39	1.560	1.319	2.067	1.396	1.623	1.273	2.135	1.466	1.720	1.338	2.120	1.518
40	1.714	1.345	2.054	1.381	1.773	1.278	2.128	1.445	1.859	1.338	2.120	1.518
41	-2.142	0.399	2.326	-2.779	-2.217	0.616	2.230	-3.019	-2.283	0.640	2.120	-3.236
42	-2.453	0.684	2.561	-2.972	-2.602	0.735	2.354	-3.421	-2.709	0.640	2.120	-3.794
43	-1.574	1.134	0.192	0.115	-1.629	1.293	0.160	-0.166	-1.647	1.338	0.015	-0.381
44	1.094	-0.072	0.630	0.002	1.088	-0.063	0.600	-0.029	1.090	-0.058	0.581	-0.048
45	0.001	2.039	0.102	1.503	-0.013	2.084	0.006	1.432	-0.007	2.035	-0.069	1.266
46	0.455	1.838	-0.160	1.591	0.486	1.997	-0.192	1.605	0.521	2.035	-0.212	1.518
47	-1.886	2.092	-2.340	0.083	-2.009	2.156	-2.498	-0.092	-2.156	2.035	-2.529	-0.247
48	-1.841	2.207	-2.269	-0.340	-1.989	2.213	-2.476	-0.551	-2.156	2.035	-2.529	-0.751

Table 4:

Gibbs Mean and Modal Estimates of Factor Loadings, $\bar{\Lambda}$. ICM Modal Estimates of Factor Loadings, $\tilde{\Lambda}$. PS89 Modal Estimates of Factor Loadings, $\hat{\Lambda}$.

	Gibbs 1	Gibbs 2	Gibbs 3	Gibbs 4	ICM 1	ICM 2	ICM 3	ICM 4	PS89 1	PS89 2	PS89 3	PS89 4
1	0.016	-0.061	0.706	0.036	-0.021	-0.068	0.708	0.037	-0.045	-0.065	0.711	0.028
2	0.296	-0.015	0.049	0.099	0.281	0.020	0.073	0.141	0.241	0.046	0.094	0.175
3	0.008	0.738	0.052	0.020	0.012	0.734	0.029	0.014	0.000	0.726	0.000	0.000
4	0.102	-0.060	0.156	0.722	0.038	-0.038	0.164	0.721	-0.010	-0.010	0.152	0.703
5	0.779	-0.038	-0.159	-0.004	0.780	-0.045	-0.181	-0.019	0.775	-0.051	-0.192	-0.029
6	0.733	-0.018	-0.014	0.085	0.722	-0.012	-0.035	0.061	0.719	-0.011	-0.058	0.035
7	0.113	-0.025	-0.142	0.728	0.062	-0.007	-0.148	0.735	0.012	0.010	-0.152	0.722
8	0.765	-0.049	0.037	-0.068	0.761	-0.051	0.028	-0.084	0.754	-0.049	0.016	-0.091
9	-0.020	0.059	0.745	-0.005	-0.057	0.076	0.739	-0.022	-0.081	0.080	0.737	-0.040
10	0.672	-0.023	0.148	0.024	0.658	-0.024	0.137	0.000	0.653	-0.020	0.116	-0.021
11	0.774	-0.051	-0.008	-0.083	0.775	-0.050	-0.016	-0.096	0.771	-0.046	-0.028	-0.100
12	0.684	0.032	0.092	0.108	0.668	0.050	0.074	0.092	0.659	0.057	0.052	0.066
13	0.629	0.085	0.138	0.192	0.605	0.108	0.122	0.173	0.592	0.120	0.094	0.141
14	0.303	-0.271	0.205	0.302	0.267	-0.279	0.221	0.312	0.244	-0.263	0.221	0.311
15	0.201	-0.042	0.662	0.020	0.162	-0.026	0.671	0.015	0.128	-0.015	0.677	0.011

The ICM, Gibbs, and PS89 Estimators are now compared. We are using the modal values of the quantities instead of the means. In PS89 they use mean estimators, but they are also equal to the modal values for the factor scores and loadings while the disturbance covariance matrix differs by a multiplicative constant because the mean and the mode of an inverted Wishart distribution differs by a multiplicative constant. They use the mean

$$\hat{\Psi} = \frac{(X - F\Lambda')(X - F\Lambda')' + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B}{N + m + \nu - 2p - 2}$$

while we are using the mode

$$\hat{\Psi}_{mode} = \frac{(X - F\Lambda')(X - F\Lambda')' + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B}{N + m + \nu}.$$

ICM finds joint posterior modes and Gibbs sampling finds marginal posterior means which are equal to the marginal posterior modal values because of the large sample size.

Table 5:
Gibbs Mean and Modal Estimate of the Disturbance Covariance Matrix, $\bar{\Psi}$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.201	0.052	-.015	0.044	0.014	0.006	-.010	-.006	-.096	-.018	0.031	0.010	0.005	0.082	-.061
2		0.484	0.051	0.062	0.016	-.033	0.076	0.037	0.006	-.042	0.084	0.046	0.045	0.000	0.084
3			0.043	0.009	-.017	-.002	-.003	-.003	0.015	-.005	0.002	0.016	0.025	-.026	0.014
4				0.125	-.027	0.019	-.055	0.017	-.017	-.005	0.031	0.011	0.028	0.073	0.020
5					0.100	-.010	0.038	-.001	0.001	-.019	0.012	-.032	-.038	0.007	-.024
6						0.104	-.024	-.002	-.011	-.049	-.038	0.045	-.009	-.004	-.007
7							0.122	0.002	0.008	0.005	-.007	0.002	-.017	0.002	0.018
8								0.090	-.004	0.008	0.005	-.027	-.024	0.026	0.034
9									0.144	-.006	0.003	0.003	0.002	-.018	-.011
10										0.144	-.001	-.043	0.010	0.052	0.033
11											0.096	-.008	0.008	0.035	-.009
12												0.113	0.035	0.007	0.001
13													0.104	-.004	0.011
14														0.264	-.015
15															0.157

ICM Modal Estimate of the Disturbance Covariance Matrix, $\tilde{\Psi}$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.173	0.025	-.003	0.022	0.015	0.006	-.017	-.012	-.095	-.023	0.023	0.006	-.001	0.055	-.075
2		0.426	0.019	0.017	0.008	-.046	0.039	0.023	-.017	-.056	0.068	0.020	0.013	-.022	0.052
3			0.004	0.005	-.006	-.002	-.003	0.001	0.001	0.000	0.003	0.003	0.006	0.000	0.005
4				0.085	-.028	0.014	-.076	0.010	-.021	-.009	0.021	0.000	0.015	0.039	0.000
5					0.086	-.013	0.032	-.005	0.009	-.020	0.005	-.034	-.038	-.003	-.018
6						0.091	-.023	-.007	-.005	-.048	-.043	0.035	-.013	-.011	-.008
7							0.083	0.001	0.013	0.007	-.008	-.004	-.021	-.019	0.013
8								0.075	-.006	0.001	-.004	-.032	-.030	0.012	0.024
9									0.121	-.007	0.001	0.001	-.002	-.019	-.027
10										0.127	-.009	-.046	0.004	0.036	0.022
11											0.079	-.017	-.003	0.021	-.014
12												0.091	0.019	-.003	-.007
13													0.081	-.011	-.001
14														0.204	-.030
15															0.118

PS89 Modal Estimate of the Disturbance Covariance Matrix, $\hat{\Psi}_{mode}$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.174	0.005	0.000	0.022	0.009	0.006	-.022	-.016	-.092	-.022	0.018	0.005	0.000	0.051	-.079
2		0.405	0.000	-.015	0.012	-.048	0.016	0.022	-.036	-.062	0.068	0.010	-.001	-.038	0.031
3			0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-.001	0.000
4				0.082	-.034	0.016	-.080	0.001	-.015	-.010	0.011	-.001	0.016	0.031	-.006
5					0.093	-.011	0.034	0.000	0.008	-.019	0.009	-.033	-.038	-.011	-.018
6						0.092	-.016	-.007	-.001	-.048	-.044	0.034	-.014	-.015	-.005
7							0.082	-.001	0.015	0.009	-.011	0.001	-.016	-.030	0.006
8								0.078	-.007	0.000	-.002	-.035	-.033	0.004	0.024
9									0.124	-.002	-.002	0.003	0.002	-.016	-.029
10										0.127	-.011	-.048	0.002	0.032	0.024
11											0.080	-.021	-.009	0.013	-.016
12												0.088	0.016	-.008	-.007
13													0.078	-.015	-.001
14														0.192	-.034
15															0.111

PS89 approximated $\frac{F'F}{N}$ by I_m for large N, for some prior distributions. So we will address the question of how good was this approximation in this example where $N = 48$.

Tables 3, 4, and 5 display trios of entire matrices of values. It is not simple to arrive at quantitative conclusions about how well these estimated matrices in each trio compare, since they contain so many values. So to assist us in such comparisons we have adopted several distinct scalar performance measures that can be used. Accordingly, we have evaluated:

$$\left|\frac{F'F}{N}\right|, \quad \left[tr\left(\frac{F'F}{N}\right)\right]^{\frac{1}{2}}, \text{ and } \quad \left|\frac{F'F}{N} - I_m\right|,$$

for the ICM, Gibbs, and PS89 estimators. (Note that $[tr(\frac{F'F}{N})]^{\frac{1}{2}}$ denotes the norm of the $\frac{F'F}{N}$ matrix.) We have also differenced the matrix estimators pairwise, for the factor scores, factor loadings, and disturbance covariance matrices, to form:

$$\Delta F_{IP} = F_I - F_P, \quad \Delta F_{GP} = F_G - F_P,$$

$$\Delta\Lambda_{IP} = \Lambda_I - \Lambda_P, \quad \Delta\Lambda_{GP} = \Lambda_G - \Lambda_P, \quad \text{and}$$

$$\Delta\Psi_{IP} = \Psi_I - \Psi_P, \Delta\Psi_{GP} = \Psi_G - \Psi_P,$$

where I, G, and P denote the modal estimators by ICM, Gibbs, and PS89 respectively.

Moreover, we have computed the scalar measures of the differenced matrices:

$$|(\Delta F')(\Delta F)|^{\frac{1}{2}}, \quad |(\Delta\Lambda')(\Delta\Lambda)|^{\frac{1}{2}}, \quad |(\Delta\Psi')(\Delta\Psi)|^{\frac{1}{2}},$$

and have compared their numerical values. All these comparisons are displayed in Tables 6 and 7. The $\frac{F'F}{N}$ matrices themselves are given below for the three types of estimators. Complete difference matrices are given in the Appendix.

Gibbs Matrix
$$\frac{\bar{F}'\bar{F}}{N} = \begin{pmatrix} 1.399 & 0.162 & 0.372 & 0.335 \\ 0.162 & 1.561 & 0.206 & -0.167 \\ 0.372 & 0.206 & 1.479 & 0.056 \\ 0.335 & -0.167 & 0.056 & 1.349 \end{pmatrix},$$

ICM Matrix $$\frac{\tilde{F}'\tilde{F}}{N} = \begin{pmatrix} 1.547 & 0.146 & 0.533 & 0.560 \\ 0.146 & 1.805 & 0.220 & -0.030 \\ 0.533 & 0.220 & 1.433 & 0.061 \\ 0.560 & -0.030 & 0.061 & 1.438 \end{pmatrix},$$

PS89 Matrix $$\frac{\hat{F}'\hat{F}}{N} = \begin{pmatrix} 1.638 & 0.210 & 0.656 & 0.750 \\ 0.210 & 1.884 & 0.290 & -0.028 \\ 0.656 & 0.290 & 1.469 & 0.169 \\ 0.750 & -0.028 & 0.169 & 1.612 \end{pmatrix}.$$

Table 6:
Measures of Quality Of The PS89 Approximation.

Performance Measures	Gibbs Estimation	ICM Estimation	PS89 Estimation	Ideal Values
$\left\lvert \frac{F'F}{N} \right\rvert$	3.650	4.166	4.495	1
$\left[tr\left(\frac{F'F}{N}\right)\right]^{\frac{1}{2}}$	2.406	2.495	2.172	2
$\left\lvert \frac{F'F}{N} - I_m \right\rvert$	-.013	-.084	-.133	0

We note from inspection of Table 6 that the Gibbs sampling estimator of $\frac{F'F}{N}$ is closer to the ideal values for two of the three measures of performance which is better than either of the other two estimators. Since the matrix $\frac{F'F}{N}$ represents the sample covariance for the factor scores, and they have been assumed to have an identity covariance matrix, for large N, the determinant should be the determinant of an identity covariance matrix, which is 1.000. That's what is meant by the ideal value in the last column of the table. Similarly for the norm of this matrix in the middle row (since the square root of the trace of the identity matrix of of order 4 is 2).

Table 7:

Difference Measures Of Quality Of The PS89 Approximation.

Performance Measures	Gibbs-PS89 Estimation	ICM-PS89 Estimation	Ideal Values
$\|(\Delta F')(\Delta F)\|^{\frac{1}{2}}$	0.280	0.028	0
$\|(\Delta\Lambda')(\Delta\Lambda)\|^{\frac{1}{2}}$	8.85×10^{-5}	4.68×10^{-6}	0
$\|(\Delta\Psi')(\Delta\Psi)\|^{\frac{1}{2}}$	4.93×10^{-27}	3.05×10^{-51}	0

The difference measures in the last two rows of Table 7 clearly show that the factor loadings and disturbance variance and covariance estimates of PS89 are very good relative to either the Gibbs or ICM evaluations. The entries in the first row show that for the factor scores, the deviation of the PS89 estimates from the Gibbs sampling results (0.364) can be appreciable (for the small sample size of $N = 48$, and the small ratio of sample size to dimension of $\frac{48}{15}$.

It should be noted that while the Gibbs sampling and ICM procedures can claim to be exact marginal and joint posterior modal estimates respectively, the estimating procedures of PS89 are computationally faster, since they don't involve iterative sampling from the posterior distribution or optimization. We therefore conclude that at least in this example, the PS89 method appears to yield very good estimates for the factor loadings and the disturbance variances and covariances, even in small samples, but that if the factor loadings are of strong concern, a Gibbs sampler, or ICM method of estimation, should be used in small samples; in large samples, the PS89 procedure will be quite sufficient.

A Appendix

Table A1: Difference of Factor Score Estimates, ΔF_{GP} and ΔF_{IP}.

Person	1	2	3	4		1	2	3	4
1	0.019	0.268	-0.147	-0.241		0.020	0.040	-0.060	-0.162
2	-0.131	-0.014	-0.308	-0.517		-0.050	-0.083	-0.138	-0.272
3	-0.056	0.183	-0.205	-0.360		-0.018	0.003	-0.092	-0.199
4	-0.030	-0.148	0.119	0.193		-0.027	-0.015	0.054	0.134
5	-0.193	-0.450	-0.071	-0.018		-0.107	-0.175	0.010	0.030
6	-0.005	-0.030	0.068	0.052		-0.008	0.011	0.014	0.051
7	-0.159	-0.250	-0.214	-0.292		-0.076	-0.151	-0.069	-0.159
8	-0.185	-0.221	-0.242	-0.361		-0.089	-0.148	-0.084	-0.188
9	-0.164	-0.208	-0.205	-0.293		-0.080	-0.130	-0.069	-0.147
10	0.103	-0.055	-0.074	0.003		0.070	0.010	-0.042	-0.023
11	0.127	-0.165	0.078	0.235		0.071	-0.011	0.031	0.091
12	0.059	-0.131	-0.003	0.087		0.039	0.000	-0.012	0.050
13	-0.137	-0.318	-0.189	-0.188		-0.064	-0.150	-0.052	-0.092
14	-0.100	-0.303	-0.075	-0.068		-0.046	-0.112	-0.025	-0.016
15	-0.043	-0.368	0.008	0.116		-0.024	-0.098	0.015	0.088
16	-0.075	-0.216	-0.203	-0.234		-0.029	-0.114	-0.075	-0.135
17	-0.007	0.129	-0.052	-0.113		-0.002	0.012	-0.020	-0.079
18	0.034	0.149	-0.051	-0.168		0.028	0.052	-0.061	-0.094
19	0.049	0.059	0.045	0.033		0.029	0.040	-0.003	0.009
20	-0.137	-0.086	-0.242	-0.343		-0.059	-0.064	-0.103	-0.153
21	-0.039	-0.165	-0.141	-0.127		-0.008	-0.054	-0.062	-0.056
22	-0.165	0.115	-0.200	-0.417		-0.083	-0.038	-0.078	-0.198
23	-0.237	-0.128	-0.369	-0.607		-0.108	-0.142	-0.150	-0.298
24	-0.208	0.029	-0.210	-0.399		-0.108	-0.076	-0.066	-0.184
25	-0.056	-0.155	-0.068	-0.100		-0.025	-0.065	-0.028	-0.037
26	-0.061	-0.161	-0.033	-0.009		-0.031	-0.066	0.003	0.006
27	-0.038	-0.293	-0.195	-0.219		-0.001	-0.095	-0.106	-0.106
28	0.216	0.612	0.388	0.453		0.087	0.248	0.162	0.205
29	0.153	0.366	0.329	0.520		0.048	0.113	0.202	0.197
30	0.106	0.493	0.055	-0.083		0.056	0.182	-0.026	-0.044
31	0.056	0.459	0.112	0.024		0.018	0.151	0.033	0.010
32	0.127	0.316	0.332	0.443		0.042	0.177	0.138	0.249
33	0.158	0.369	0.304	0.393		0.064	0.202	0.108	0.214
34	0.116	0.242	0.249	0.362		0.042	0.119	0.119	0.180
35	0.057	0.183	0.100	0.060		0.029	0.086	0.017	0.052
36	-0.048	0.075	0.056	-0.048		-0.027	0.038	-0.003	0.025
37	0.146	0.134	-0.049	-0.140		0.100	0.079	-0.092	-0.096
38	0.050	0.076	-0.125	-0.247		0.048	0.019	-0.102	-0.157
39	-0.160	-0.019	-0.052	-0.122		-0.096	-0.064	0.015	-0.052
40	-0.145	0.007	-0.066	-0.138		-0.086	-0.060	0.009	-0.073
41	0.141	-0.241	0.206	0.457		0.066	-0.024	0.110	0.217
42	0.257	0.045	0.441	0.822		0.107	0.095	0.234	0.373
43	0.074	-0.204	0.176	0.497		0.018	-0.044	0.145	0.216
44	0.004	-0.014	0.049	0.050		-0.001	-0.005	0.020	0.019
45	0.008	0.003	0.171	0.237		-0.006	0.048	0.074	0.165
46	-0.066	-0.197	0.053	0.073		-0.035	-0.038	0.020	0.087
47	0.270	0.056	0.189	0.330		0.147	0.121	0.031	0.155
48	0.315	0.171	0.260	0.411		0.167	0.178	0.054	0.199

Table A2: Difference of Factor Loading Estimates, $\Delta\Lambda_{GP}$ and $\Delta\Lambda_{IP}$.

	1	2	3	4		1	2	3	4
1	0.062	0.003	-0.005	0.008		0.024	-0.003	-0.003	0.009
2	0.056	-0.061	-0.045	-0.076		0.041	-0.026	-0.022	-0.034
3	0.008	0.012	0.052	0.020		0.012	0.008	0.029	0.014
4	0.112	-0.051	0.004	0.019		0.048	-0.028	0.012	0.018
5	0.004	0.013	0.033	0.024		0.005	0.005	0.011	0.009
6	0.013	-0.008	0.044	0.050		0.003	-0.001	0.022	0.026
7	0.101	-0.034	0.010	0.006		0.050	-0.016	0.004	0.013
8	0.010	0.000	0.020	0.022		0.007	-0.002	0.011	0.006
9	0.061	-0.021	0.008	0.035		0.024	-0.004	0.002	0.018
10	0.019	-0.002	0.032	0.045		0.005	-0.004	0.021	0.021
11	0.003	-0.006	0.020	0.017		0.004	-0.004	0.011	0.005
12	0.025	-0.025	0.040	0.041		0.009	-0.008	0.022	0.026
13	0.037	-0.034	0.044	0.051		0.013	-0.012	0.028	0.032
14	0.059	-0.008	-0.015	-0.009		0.022	-0.016	0.000	0.001
15	0.074	-0.027	-0.015	0.009		0.034	-0.011	-0.006	0.004

Table A3: Difference of Disturbance Covariance Estimates, $\Delta\Psi_{GP}$ and $\Delta\Psi_{IP}$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.027	0.046	-.015	0.022	0.004	0.000	0.012	0.010	-.004	0.004	0.014	0.006	0.004	0.030	0.018
2		0.079	0.051	0.078	0.004	0.015	0.060	0.015	0.042	0.019	0.017	0.036	0.047	0.039	0.053
3			0.041	0.009	-.017	-.002	-.003	-.003	0.015	-.005	0.003	0.016	0.024	-.025	0.014
4				0.043	0.007	0.003	0.025	0.016	-.002	0.004	0.020	0.012	0.012	0.042	0.026
5					0.006	0.001	0.004	-.001	-.007	0.000	0.003	0.001	0.000	0.018	-.007
6						0.012	-.008	0.005	-.010	0.000	0.006	0.010	0.005	0.012	-.003
7							0.040	0.003	-.007	-.004	0.005	0.001	-.002	0.032	0.012
8								0.012	0.003	0.008	0.007	0.008	0.009	0.021	0.010
9									0.020	-.004	0.005	0.000	0.001	-.002	0.019
10										0.018	0.010	0.005	0.009	0.019	0.008
11											0.016	0.013	0.016	0.022	0.007
12												0.024	0.019	0.014	0.009
13													0.026	0.011	0.012
14														0.072	0.020
15															0.047

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-.001	0.020	-.003	0.000	0.005	0.000	0.005	0.005	-.003	-.001	0.005	0.001	-.001	0.003	0.004
2		0.021	0.019	0.032	-.004	0.002	0.023	0.001	0.019	0.005	0.000	0.010	0.015	0.016	0.021
3			0.002	0.005	-.006	-.002	-.003	0.001	0.001	0.000	0.003	0.003	0.005	0.001	0.005
4				0.003	0.006	-.002	0.004	0.009	-.006	0.000	0.010	0.001	-.001	0.009	0.006
5					-.008	-.003	-.002	-.005	0.001	-.001	-.004	-.001	0.000	0.009	0.000
6						-.001	-.007	0.000	-.004	0.000	0.001	0.001	0.000	0.005	-.003
7							0.001	0.003	-.002	-.003	0.003	-.005	-.006	0.010	0.007
8								-.003	0.002	0.000	-.002	0.002	0.004	0.008	0.001
9									-.003	-.004	0.003	-.002	-.004	-.003	0.002
10										0.000	0.002	0.002	0.002	0.004	-.003
11											-.001	0.004	0.006	0.008	0.001
12												0.003	0.004	0.005	0.001
13													0.003	0.004	0.001
14														0.012	0.004
15															0.007

References

Gelfand, A. E. and Smith, A. F. M. (1990). "Sampling based approaches to calculating marginal densities." *J. Amer. Stat. Assoc.* 85, 398-409.

Geman, S. and Geman, D. (1984). "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images." *IEEE Trans. on pattern analysis and machine intelligence* 6, 721-741.

Kendall, M.(1980). *Multivariate Analysis* Charles Griffin & Company LTD, London, Second Edition

Kotz, S. and Johnson, N. (1985), Editors. *Encyclopedia of Statistical Science*, Volume 5. John Wiley and Sons, Inc., New York, pages 326-333.

Lindley, D. V. and Smith, A. F. M. (1972). "Bayes estimates for the linear model", *Journal of the Royal Statistical Society B*, Volume 34, No. 1.

O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics*, Volume 2B Bayesian Inference. John Wiley and Sons, Inc., New York.

Press, S. J. and Shigemasu, K. (1989). "Bayesian inference in factor analysis". In *Contributions to Probability and Statistics*, Chapter 15, Springer-Verlag.

Press, S. J. and Shigemasu, K. (1997). "Bayesian inference in factor analysis-Revised." Technical Report No. 243, Department of Statistics, University of California, Riverside.

Press, S. J. (1982). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference.* Robert E. Krieger Publishing Company, Malabar, Florida.