DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

INCORPORATING PRIOR KNOWLEDGE REGARDING THE MEAN IN BAYESIAN FACTOR ANALYSIS

Daniel B. Rowe

SOCIAL SCIENCE WORKING PAPER 1097

July 2000

Incorporating Prior Knowledge Regarding the Mean in Bayesian Factor Analysis

Daniel B. Rowe

Abstract

In the Bayesian factor analysis model (Press & Shigemasu, 1989), available knowledge regarding the model parameters is incorporated in the form of prior distributions. This has the added consequence of eliminating the ambiguity of rotation found in the traditional factor analysis model. In the model presented by Press and Shigemasu, a vague prior distribution was implicitly specified for the population mean. The sample size was assumed to be large enough to estimate the overall population mean by the sample mean. In this paper, available prior knowledge regarding the population mean is incorporated into the inferences in the form of a prior distribution. The population mean is estimated along with the other parameters by both Gibbs sampling and Iterated Conditional Modes.

1 Introduction

A factor analysis is performed to explain the relationship among a set of observed variables in terms of a smaller number of unobserved variables or latent factors which underlie the observations. This smaller number of variables can be used to find a meaningful structure in the observed variables. This structure will aid in the interpretation and explanation of the process that has generated the observations.

In the Bayesian factor analysis model first proposed by Press & Shigemasu, 1989 (henceforth PS89) the classical normal sampling model was assumed, but the disturbance covariance matrix was assumed to be a full positive definite matrix. One of the prior assumptions, however, was that the expected value of the disturbance covariance matrix was diagonal in order to represent traditional views of the factor model containing "common" and "specific" factors. Natural conjugate prior distributions were specified for the unknown matrices.

Bayesian statistical methods not only incorporate available prior information either from substantive experts or previous data, but allow the the knowledge regarding the parameter values to accumulate as subsequent data is acquired. In the non-Bayesian Factor Analysis model, the factor loading matrix is determinate up to an orthogonal rotation. Typically after a non-Bayesian Factor Analysis is performed, an orthogonal rotation is performed on the factor loading matrix according to one of many subjective criteria. This is not the case in Bayesian Factor Analysis. The rotation is automatically found.

In PS89, the model parameters were estimated by marginalization and conditional estimation with the use of a large sample approximation. The need for this large sample approximation was alleviated in Rowe & Press, 1998 (henceforth RP98) by estimating the model parameters exactly by both Gibbs sampling (Geman & Geman, 1984 and Gelfand & Smith, 1990) and iterated conditional modes (Lindley & Smith, 1972 and O'Hagen, 1994).

The marginalization and conditional estimation procedure can not be used when available prior information regarding the population mean is incorporated into the inferences because none of the marginal posterior distributions may be found in a convenient closed form. Thus, Gibbs sampling and iterated conditional modal estimates are computed.

In this paper, the same model as in PS89 and RP98 is adopted, with the explicit inclusion of the population mean. Parameters are estimated exactly by Gibbs sampling and by Iterated Conditional Modes (ICM). For both approaches conditional posterior distributions for each of the parameters given the other parameters and the data are needed. All four can be found explicitly. Gibbs posterior marginal mean and ICM posterior joint modal 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. Present the conditional posterior distributions along with the Gibbs sampling and ICM algorithms in Section 3. In Section 4 an example is detailed, and estimates from Gibbs sampling and ICM estimation methods are presented.

2 Model

2.1 Likelihood Function

The Bayesian factor analysis model is:

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

for $j = 1, \ldots, n$, where x_j is the j^{th} observation, μ is the overall population mean, Λ is a matrix of constants called the factor loading matrix; f_j is the factor score vector for subject j; 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 common and specific factors can be readily distinguished. In the the current 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.

It is assumed that the distribution of each x_j can be written as

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

If proportionality is denoted by "$\propto$" and the Kroneker product by $\otimes$ then, the likelihood for (μ, Λ, F, Ψ) is

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

where the p-variate observation vectors on n subjects are $X' = (x_1, \ldots, x_n)$, the factor scores are $F' = (f_1, \ldots, f_n)$, and the errors of observation are $E' = (\epsilon_1, \ldots, \epsilon_n)$. The notation $p(\cdot)$ will generically denote a distribution which is distinguished by its argument. The proportionality constant in (2.3) depends only on (p, n) and not on (μ, Λ, F, Ψ).

2.2 Priors

The same prior distributions are adopted as in PS89 and again in RP98 with the exception of an additional natural conjugate normal distribution for the mean. The joint prior distribution is:

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

where

$$p(\mu) \propto |\Psi|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mu-\mu_0)'(\psi_0\Psi)^{-1}(\mu-\mu_0)}, \tag{2.5}$$

$$p(\Lambda|\Psi) \propto |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)'}, \tag{2.6}$$

$$p(\Psi) \propto |\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2}tr\Psi^{-1}B}, \tag{2.7}$$

$$p(F) \propto e^{-\frac{1}{2}trF'F} \tag{2.8}$$

with $H, B, \Psi > 0$ and B a diagonal matrix. A natural conjugate normal distribution is specified to quantify prior knowledgeregarding the population mean where μ_0 and ψ_0 are hyperparameters to be assessed. The matrix Λ conditional on Ψ has elements which are jointly normally distributed, and hyperparameters (Λ_0, H) are to be assesssed. The matrix Ψ follows an Inverted Wishart distribution, with hyperparameters (ν, B) which are to be assessed. It is specified that $E(\Psi)$ is 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 and RP98, $H = h_0 I$, for some preassigned scalar h_0 to simplify hyperparameter assessment.

2.3 Joint Posterior

Using Bayes rule, combine (2.3)–(2.8), to get the joint posterior density of the parameters

$$p(\mu, F, \Lambda, \Psi|X) \propto e^{-\frac{1}{2}tr F'F}|\Psi|^{-\frac{(n+m+\nu+1)}{2}} e^{-\frac{1}{2}tr\Psi^{-1}U} \tag{2.9}$$

where

$$\begin{aligned} U &= (\mu - \mu_0)\psi_0^{-1}(\mu - \mu_0)' + (X - e_n \otimes \mu' - F\Lambda')'(X - e_n \otimes \mu' - F\Lambda') + \\ &\quad (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B. \end{aligned}$$

3 Estimation

As stated earlier, marginal mean and joint modal posterior estimates are found by the Gibbs sampling and iterated conditional modes algorithms. For both the conditional posterior distributions are needed.

3.1 Conditional Posterior Densities

The four posterior conditional distributions are as follows.

$$\begin{aligned} p(\mu|\Lambda, F, X, X) &\propto p(\mu)p(X|\mu, F, \Lambda, \Psi) \\ &\propto |\Psi|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mu-\mu_0)'(\psi_0\Psi)^{-1}(\mu-\mu_0)} \end{aligned}$$

$$
\begin{aligned}
&\cdot \quad |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(X - e_n \otimes \mu' - F\Lambda')'(X - e_n \otimes \mu' - F\Lambda')} \\
&\propto \quad e^{-\frac{1}{2}(\mu - \tilde{\mu})' \left(\frac{n\psi_0}{1+n\psi_0}\Psi\right)^{-1}(\mu - \tilde{\mu})} \qquad (3.1)
\end{aligned}
$$

where $\tilde{\mu} = [\frac{1}{1+n\psi_0}\mu_0 + \frac{n\psi_0}{1+n\psi_0}(\bar{x} - \Lambda\bar{f})]$.

$$
\begin{aligned}
p(\Lambda|\mu, F, \Psi, X) &\propto p(\Lambda|\Psi)p(X|\mu, F, \Lambda, \Psi) \\
&\propto |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)'} \\
&\cdot \quad |\Psi|^{-\frac{N}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(X - e_n \otimes \mu' - F\Lambda')'(X - e_n \otimes \mu' - F\Lambda')} \\
&\propto e^{-\frac{1}{2} tr \Psi^{-1}(\Lambda - \tilde{\Lambda})(H + F'F)(\Lambda - \tilde{\Lambda})'} \qquad (3.2)
\end{aligned}
$$

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

$$
\begin{aligned}
p(\Psi|\mu, F, \Lambda, X) &\propto p(\Psi)p(\Lambda|\Psi)p(X|\mu, F, \Lambda, \Psi) \\
&\propto |\Psi|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mu - \mu_0)'(\psi_0 \Psi)^{-1}(\mu - \mu_0)} \\
&\cdot \quad |\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2} tr \Psi^{-1} B} |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)'} \\
&\cdot \quad |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(X - e_n \otimes \mu' - F\Lambda')'(X - e_n \otimes \mu' - F\Lambda')} \\
&\propto |\Psi|^{-\frac{(n+m+\nu+1)}{2}} e^{-\frac{1}{2} tr \Psi^{-1} U} \qquad (3.3)
\end{aligned}
$$

$$
\begin{aligned}
U \quad = \quad & (\mu - \mu_0)\psi_0^{-1}(\mu - \mu_0)' + (X - e_n \otimes \mu' - F\Lambda')'(X - e_n \otimes \mu' - F\Lambda') + \\
& (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B.
\end{aligned}
$$

$$
\begin{aligned}
p(F|\mu, \Lambda, \Psi, X) &\propto p(F)p(X|\mu, F, \Lambda, \Psi) \\
&\propto e^{-\frac{1}{2} tr F'F} |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(X - e_n \otimes \mu' - F\Lambda')'(X - e_n \otimes \mu' - F\Lambda')} \\
&\propto e^{-\frac{1}{2} tr (F - \tilde{F})(I_m + \Lambda'\Psi^{-1}\Lambda)(F - \tilde{F})'} \qquad (3.4)
\end{aligned}
$$

where $\tilde{F} = (X - e_n \otimes \mu')\Psi^{-1}\Lambda(I_m + \Lambda'\Psi^{-1}\Lambda)^{-1}$.

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

$$\tilde{\Psi} = \frac{U}{n+m+\nu+1}, \tag{3.5}$$

respectively.

3.2 The Gibbs Sampling Algorithm

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

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

The first s random samples are discarded and the remaining t samples are kept. The means of the remaining random samples

$$\bar{\mu} = \frac{1}{t}\sum_{l=1}^{t}\bar{\mu}_{(s+k)} \quad \bar{F} = \frac{1}{t}\sum_{k=1}^{t}\bar{F}_{(s+k)} \quad \bar{\Lambda} = \frac{1}{t}\sum_{k=1}^{t}\bar{\Lambda}_{(s+k)} \quad \bar{\Psi} = \frac{1}{t}\sum_{k=1}^{t}\bar{\Psi}_{(s+k)}.$$

are the marginal posterior mean estimates of the parameters.

3.3 The ICM Algorithm

For iterated conditional modes estimatiom of the posterior, start with an initial value for $\tilde{\mu}$ and $\tilde{F}$, say $\tilde{\mu}_{(0)}$ and $\tilde{F}_{(0)}$ and then cycle through

$$\tilde{\Lambda}_{(i+1)} = [(X - e_n \otimes \tilde{\mu}'_{(i)})'\tilde{F}_{(i)} + \Lambda_0 H)(H + \tilde{F}'_{(i)}\tilde{F}_{(i)})^{-1}$$

$$
\begin{aligned}
\tilde{\Psi}_{(i+1)} &= \{[(X - e_n \otimes \tilde{\mu}'_{(i)}) - \tilde{F}_{(i)}\tilde{\Lambda}'_{(i+1)}]'[(X - e_n \otimes \tilde{\mu}'_{(i)}) - \tilde{F}_{(i)}\tilde{\Lambda}'_{(i+1)}] + \\
&\quad (\tilde{\Lambda}_{(i+1)} - \Lambda_0)H(\tilde{\Lambda}_{(i+1)} - \Lambda_0)' + B\}/(n + m + \nu) \\
\tilde{F}_{(i+1)} &= (X - e_n \otimes \tilde{\mu}'_{(i)})\tilde{\Psi}^{-1}_{(i+1)}\tilde{\Lambda}_{(i+1)}(I_m + \tilde{\Lambda}'_{(i+1)}\tilde{\Psi}^{-1}_{(i+1)}\tilde{\Lambda}_{(i+1)})^{-1} \\
\tilde{\mu}_{(i+1)} &= \left[\frac{1}{1 + n\psi_0}\mu_0 + \frac{n\psi_0}{1 + n\psi_0}(\bar{x} - \tilde{\Lambda}_{(i+1)}\bar{\tilde{f}}_{(i+1)})\right].
\end{aligned}
$$

until convergence is reached with the joint posterior modal estimator $(\tilde{\mu}, \tilde{F}, \tilde{\Lambda}, \tilde{\Psi})$. The mean of the factor score vectors has been denoted by $\bar{\tilde{f}}$.

4 Example

In this section the ICM and the Gibbs Sampler procedures for estimating the parameters of the Bayesian factor analysis model are used and the resulting estimators are presented. The data is extracted from an example in Kendall 1980, p.53. The problem as stated in PS89 and again in RP98 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.

Note that the initial values for the ICM and Gibbs sampling estimation procedures have little effect on the final result, because for ICM there are unimodal posterior conditional distributions so the algorithm is sure to converge to the joint mode, and for Gibbs sampling, there is a burn-in period. The initial values for $\tilde{\mu}$ and $\tilde{F}$ are chosen to be $\tilde{\mu}_{(0)} = \bar{x}$ and $\tilde{F}_{(0)} = \hat{F}$, the estimators of PS89. This choice of the initial values hastens convergence.

The same underlying structure is postulated 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

$$\Lambda_0' = \begin{pmatrix} 0 & 0 & 0 & 0 & .7 & .7 & 0 & .7 & 0 & .7 & .7 & .7 & .7 & 0 & 0 \\ 0 & 0 & .7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7 & 0 & 0 & 0 & 0 & 0 & .7 \\ 0 & 0 & 0 & .7 & 0 & 0 & .7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

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 prior population mean is assessed as $\mu_0 = 7.5e_{15}$ and the hyperparameter determining prior variability in the population mean as $\psi_0 = 1/5$. The 15 dimensional unit vector has been denoted by e_{15}. The population mean, factor scores, factor loadings, and disturbance covariance matrix may now be estimated. It was found that a burn in period of $5,000$ samples worked well, so then the next $25,000$ samples were taken for the Gibbs estimates.

Table 2 displays the Gibbs sampling and ICM estimates of the population mean along with the prior and sample means.

Table 2: Gibbs Sampling and ICM estimates of the mean.

p	Gibbs Mean	ICM Mean	Sample Mean	Prior Mean
1	6.1441	6.1397	6.0000	7.5000
2	7.1219	7.1223	7.0833	7.5000
3	7.1306	7.1343	7.0833	7.5000
4	6.2737	6.2731	6.1458	7.5000
5	6.9984	6.9891	6.9375	7.5000
6	6.4458	6.4424	6.3333	7.5000
7	7.9919	7.9905	8.0417	7.5000
8	5.0484	5.0456	4.7917	7.5000
9	4.5391	4.5386	4.2292	7.5000
10	5.5215	5.5179	5.3125	7.5000
11	6.1255	6.1210	5.9792	7.5000
12	6.3671	6.3679	6.2500	7.5000
13	5.8585	5.8595	5.6875	7.5000
14	5.7449	5.7407	5.5625	7.5000
15	6.1058	6.1026	5.9583	7.5000

Table 3 displays the Gibbs sampling and ICM estimates of the factor loadings. For enhanced interpretability, the rows of the factor loading matrices have been rearranged. It is seen that factor 1 loads heavily for variables 5, 6, 8, 10, 11, 12, and 13; factor 2

Table 3: Gibbs (left) and ICM (right) Estimates of Factor Loadings.

p	1	2	3	4	1	2	3	4
5	0.7756	0.0041	-0.1428	0.0278	0.8033	-0.0860	-0.0772	0.0397
6	0.7420	-0.0470	-0.0075	0.1163	0.7284	-0.0571	0.0545	0.1271
8	0.7214	-0.0300	-0.0302	-0.0232	0.7292	-0.0937	0.0455	-0.0110
10	0.6169	-0.0170	0.0723	0.0554	0.6151	-0.0611	0.1288	0.0699
11	0.7465	-0.0411	-0.0393	-0.0211	0.7512	-0.0941	0.0264	-0.0194
12	0.6941	-0.0421	0.1305	0.1585	0.6657	-0.0034	0.1799	0.1698
13	0.6160	-0.0086	0.1547	0.2457	0.5814	0.0619	0.1779	0.2459
3	0.0091	0.5782	0.2157	0.0515	0.0135	0.7225	0.0430	0.0343
1	-0.0031	0.0149	0.6896	0.1048	-0.0359	-0.1283	0.8418	0.1140
9	-0.0938	-0.0193	0.6986	0.0291	-0.1160	0.0376	0.7165	0.0349
15	0.1695	-0.1107	0.6851	0.0865	0.1322	-0.0885	0.7722	0.1023
4	0.0381	-0.0587	0.0507	0.7379	-0.0254	-0.0435	0.1065	0.7365
7	0.1040	-0.0049	-0.1157	0.7125	0.0791	-0.0101	-0.0867	0.7272
2	0.2902	-0.0755	0.1332	0.2032	0.2619	-0.0248	0.1801	0.2039
14	0.1882	-0.1238	-0.0365	0.3065	0.1693	-0.2881	0.1053	0.3486

heavily on variable 3; factor 3 heavily on variables 1, 9, and 15; while factor 4 loads heavily on variables 4 and 7. These factors in terms of the original

Table 4: Gibbs (left) and ICM (right) Estimates of the Factor Scores.

Person	1	2	3	4	1	2	3	4
1	0.7118	-2.4892	0.2169	-0.5717	0.7386	-3.4983	0.3604	-0.4227
2	1.0629	-0.8825	0.5688	0.3767	1.0961	-1.6503	0.7898	0.5002
3	0.9115	-1.9654	0.5230	-0.1073	0.9591	-2.8608	0.6610	0.0954
4	-0.5177	-0.1194	0.2060	-0.0164	-0.4647	0.6155	0.0076	0.0387
5	-0.9215	-0.7615	0.1675	0.7036	-1.0261	0.3567	-0.0795	0.6712
6	-0.0208	-0.1685	0.6613	0.3878	0.0992	-0.0778	0.6247	0.4926
7	0.3003	0.1877	0.8164	0.1918	0.2878	0.3211	0.8012	0.1890
8	0.5029	0.9491	0.7432	0.0914	0.5244	1.0105	0.7535	0.1663
9	0.2968	-0.4315	1.1360	0.2556	0.2720	-0.3026	1.0959	0.2785
10	2.4248	1.4154	0.7934	-1.3036	2.4728	2.0816	0.7462	-1.2395
11	1.7095	1.1739	-0.0693	-2.6849	1.7562	2.0222	-0.2271	-2.8669
12	2.1232	1.1713	0.3322	-0.6580	2.1414	2.0767	0.2633	-0.6632
13	-0.9458	-0.1660	-0.1732	1.2082	-1.0368	0.3823	-0.2421	1.1192
14	-0.5658	-0.1231	0.6098	0.9378	-0.6606	0.4722	0.5439	0.7392
15	-0.8618	-0.6758	0.1394	0.8358	-0.8589	0.5047	-0.1319	0.7978
16	0.2543	-1.2285	0.3074	-0.0612	0.2844	-0.9899	0.2669	-0.0209
17	0.1121	0.4814	0.4789	0.0591	0.1995	-0.0897	0.5699	0.0914
18	0.0915	1.1476	-0.2172	-1.1720	0.1242	0.6791	-0.0693	-1.1090
19	0.1061	0.6828	-0.0098	-1.2754	0.0723	0.6791	0.0464	-1.3217
20	0.7054	-0.4150	-0.1679	0.7158	0.7039	-0.1506	-0.1656	0.9484
21	0.7763	-1.4665	-0.0304	0.9502	0.7767	-0.7999	-0.1541	1.0241
22	1.0731	0.7744	-0.1638	0.5748	1.0872	-0.2114	0.1304	0.8161
23	0.8510	0.1328	-0.7256	0.7601	0.8436	-0.3711	-0.5334	0.9652
24	0.7597	0.5218	-0.4502	1.1360	0.7052	-0.2705	-0.1665	1.3109
25	-0.8415	-0.2442	-0.5539	0.5583	-0.9449	-0.1766	-0.5408	0.4813
26	-0.5011	-0.4622	0.2649	0.6505	-0.6210	-0.1442	0.2242	0.5177
27	0.4279	-1.0006	-0.7710	1.2120	0.4237	-0.1963	-0.8976	1.2136
28	-1.3879	0.2809	-1.1130	-1.3723	-1.3537	-1.0852	-0.9699	-1.4632
29	-2.2051	-1.1689	-2.2388	-2.6562	-2.3286	-1.9992	-2.2926	-2.9928
30	-0.8779	0.1096	-1.1420	0.2445	-0.6340	-1.2144	-0.9348	0.4856
31	-0.7858	-0.6550	-1.1889	0.5420	-0.6209	-1.9597	-1.0462	0.7136
32	0.1889	-1.0450	-0.8955	0.8578	0.2507	-1.1550	-0.9339	0.9648
33	-0.1146	-0.9462	-1.0788	0.7934	-0.0191	-1.1266	-1.0871	0.9326
34	-1.3932	-0.3138	-1.3336	-0.4273	-1.4112	-0.5686	-1.3885	-0.4289
35	-1.6393	-1.3486	-0.2466	-0.4018	-1.6450	-1.9979	-0.1553	-0.3739
36	-0.2604	-0.8560	0.2881	-0.1372	-0.3051	-1.3903	0.4503	-0.1532
37	1.2467	-0.1587	-0.8276	-0.5021	1.3349	-0.6570	-0.6548	-0.5308
38	0.9098	-0.2763	-1.4017	-0.3756	0.9033	-0.7818	-1.2231	-0.4233
39	0.5700	1.3850	0.6538	1.1015	0.6292	1.1159	0.7557	1.2655
40	0.7462	1.4721	0.6649	1.1149	0.7889	1.1237	0.7611	1.2498
41	-1.6427	-0.3268	2.9197	-2.6276	-1.7645	0.6502	2.7348	-2.9693
42	-2.1441	0.1538	2.6863	-3.0716	-2.2710	0.8198	2.5185	-3.4880
43	-1.9302	-0.0159	-0.3647	-0.1285	-2.0652	1.2588	-0.6411	-0.3608
44	0.7434	-0.0751	0.2238	-0.0654	0.7971	-0.1032	0.1625	-0.1094
45	0.8373	1.8021	1.0493	1.6923	0.7500	2.1625	1.0969	1.5753
46	1.3061	1.3896	0.8043	1.9544	1.1330	2.0221	0.8323	1.7723
47	-1.2781	1.7912	-1.2774	0.0682	-1.1442	2.2629	-1.4356	-0.0206
48	-1.1482	2.0438	-1.1055	-0.4417	-0.9740	2.3570	-1.2440	-0.4703

variables are factor 1: Self-confidence, Lucidity, Salesmanship, Drive, Ambition, Grasp, Potential; factor 2: Academic ability; factor 3: Form of letter application, Experience, Suitability; and factor 4: Likeabiliy, Honesty. These factors may be loosely interpreted as factor 1 being personality, factor 2 being academic ability, factor 3 being position match, and factor 4 being charisma.

In Table 4, the Gibbs sampling and ICM estimates of the factor scores are presented. Note the similarity of most of the values but there are some differences. An employer may now decide on a criteria to select a person. For example, if the employer wished to hire a person that is "very" hard working with a "good" academic record and a "fair" match for the position, person 10 might be selected.

Table 5: Gibbs (top) and ICM (bottom) Estimates of the Disturbance Covariance Matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.4865	.0905	-.0888	.2558	.1106	.1796	-.1064	.3684	.3654	.3265	.2751	.1873	.2462	.4795	.2100
2		.5536	.0725	.0756	.0515	-.0012	.0140	.1544	.1164	.0527	.1656	.0769	.0932	.1017	.1272
3			.3348	.0330	-.0603	.0168	-.0504	.0347	.1332	.0333	.0094	.0685	.1215	-.1068	.0349
4				.2752	.0734	.1815	-.1140	.3256	.3701	.2800	.2238	.1776	.2438	.3343	.2478
5					.1702	.0689	.0227	.1702	.2090	.1444	.1314	.0479	.0733	.1940	.1024
6						.2351	-.0704	.2780	.3460	.2121	.1379	.1800	.1823	.2711	.1904
7							.1696	-.0983	-.1160	-.0723	-.0798	-.0590	-.1017	-.0583	-.0619
8								.6753	.7040	.5236	.3779	.2740	.3832	.5188	.4425
9									.9918	.6222	.4567	.3826	.5131	.5737	.4591
10										.6228	.3394	.2373	.3857	.5070	.4084
11											.3526	.1832	.2651	.3745	.2615
12												.2601	.2415	.2970	.2011
13													.3998	.3494	.2827
14														.7282	.3909
15															.4286
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.3069	.0471	.0427	.1831	.0697	.1270	-.0941	.2699	.2574	.2393	.2008	.1358	.1969	.3140	.1039
2		.4172	.0435	.0572	.0292	-.0049	.0004	.1158	.0809	.0321	.1296	.0592	.0751	.0602	.0915
3			.0174	.0458	.0133	.0330	-.0194	.0764	.0892	.0648	.0531	.0433	.0598	.0628	.0566
4				.2070	.0485	.1443	-.1199	.2565	.2831	.2133	.1812	.1410	.2001	.2269	.1816
5					.0978	.0303	-.0048	.1099	.1514	.0886	.0799	.0177	.0444	.1201	.0546
6						.1838	-.0782	.2112	.2657	.1540	.0995	.1407	.1445	.1941	.1386
7							.0925	-.0960	-.0999	-.0773	-.0750	-.0677	-.0976	-.0856	-.0689
8								.5321	.5534	.4058	.2887	.2105	.3053	.3864	.3389
9									.8100	.4914	.3547	.2974	.4067	.4369	.3671
10										.4891	.2568	.1782	.3060	.3746	.3181
11											.2706	.1416	.2144	.2731	.1939
12												.2072	.1929	.2227	.1534
13													.3227	.2784	.2299
14														.4848	.2625
15															.3334

Table 5 displays the Gibbs sampling and ICM estimates of the disturbance covariance matrix.

5 Conclusion

A Bayesian statistical model was detailed in which available prior information either from substantive experts or previous experiments is incorporated. An added feature of the Bayesian factor analysis model is that the there is no need to rotate the factor loading matrix. The rotation is automatically found. Available prior information regarding the population mean was incorporated along with the other parameters through a prior distribution. By incorporating prior knowledge regarding the mean, estimation of it and the other parameters may be improved.

References

[1] A. E. Gelfand and A. F. M Smith. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

[2] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, 6:721–741, 1984.

[3] K. Hayashi. *The Press-Shigemasu Bayesian Factor Analysis Model With Estimated Hyperparameters.* PhD thesis, Department of Psychology, University of North Carolina, Chapel Hill, NC 27599, 1997.

[4] M. Kendall. *Multivariate Analysis*. Charles Griffin & Company LTD, London, second edition, 1980.

[5] D. V. Lindley and A. F. M. Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society B*, 34(1), 1972.

[6] A. O'Hagen. *Kendalls' Advanced Theory of Statistics, Volume 2B Bayesian Inference*. John Wiley and Sons Inc., New York, 1994.

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

[8] S. J. Press. *Bayesian Statistics: Principles, Models, and Applications*. John Wiley and Sons, New York, 1989.

[9] S. J. Press and K. Shigemasu. Bayesian inference in factor analysis. In *Contributions to Probability and Statistics*, chapter 15. Springer-Verlag, 1989.

[10] S. J. Press and K. Shigemasu. Bayesian inference in factor analysis-Revised. Technical Report No. 243, Department of Statistics, University of California, Riverside, CA 92521, May 1997.

[11] D. B. Rowe. *Correlated Bayesian Factor Analysis*. PhD thesis, Department of Statistics, University of California, Riverside, CA 92521, December 1998.

[12] D. B. Rowe. On estimating the mean in Bayesian factor analysis. Technical Report 1096, Division of Humanities and Social Sciences, Caltech, Pasadena, CA 91125, July 2000.

[13] D. B. Rowe and S. J. Press. Gibbs sampling and hill climbing in Bayesian factor analysis. Technical Report No. 255, Department of Statistics, University of California, Riverside, CA 92521, May 1998.

[14] D. M. Shera. *Bayesian Factor Analysis.* PhD thesis, Harvard University, School of Public Health, Boston, MA 02115, March 2000.