Bayesian Inference in Factor Analysis – Revised

S. James Press and K. Shigemasu

Technical Report No. 243

UNIVERSITY of CALIFORNIA
Riverside

Department of Statistics

Bayesian Inference in Factor Analysis – Revised[1]

S. James Press[2]
K. Shigemasu[3]

ABSTRACT We propose a new method for analyzing factor analysis models using a Bayesian approach. Normal theory is used for the sampling distribution, and we adopt a model with a full disturbance covariance matrix. Using vague and natural conjugate priors for the parameters, we find that the marginal posterior distribution of the factor scores is approximately a matrix T-distribution, in large samples. This explicit result permits simple interval estimation and hypothesis testing of the factor scores. Explicit point and interval estimators of the factor score elements, in large samples, are obtained as means as means of the respective marginal posterior distributions. Factor loadings are estimated as joint modes (with factor scores), or alternatively as means or modes of the distribution of the factor loadings conditional upon the estimated factor scores. Disturbance variances and covariances are estimated conditional upon the estimated factor scores and factor loadings. This revision includes the correction of some typographical errors and some revised computations, plus an appendix that provides some intermediate results.

1 Introduction

This paper proposes a new method for analyzing factor analysis models using a Bayesian point of view. We use normal theory for the sampling distribution, and vague and natural conjugate theory for the prior distributions. We adopt a general disturbance covariance matrix whose prior mean is diagonal. We show that in large samples, for a variety of prior distributions for the factor scores (including vague and normal priors), the marginal posterior distribution of the scores is approximately matrix T. As a result, we are able to make both point and interval estimates of the factor scores, thereby improving upon earlier research. For comparison, we give below a brief review of earlier research in this area.

[1] *Acknowledgements* The authors would like to acknowledge the assistance of Masanori Ickikawa in computing the original numerical results in Section 6, and the assistance of Daniel Bryant Rowe for discovering several typographical errors in the original manuscript, recalculating the numerical results, providing the Appendix that has been included in this revision, and typing this manuscript. We are also grateful to an anonymous referee for helpful comments on related research on which we report in Section 1.

[2] University of California, Riverside
[3] Tokyo Institute of Technology

An early formulation (see Press 1972, 1982) of a Bayesian factor analysis model used the Wishart distribution for the sample covariance matrix, and a vague prior distribution for the parameters. Only implicit numerical solutions could be obtained from this model. Kaufman and Press (1973a, b) proposed a new formulation of that model in terms of a model with more factors than observations (a characteristic of the model shared with that of Guttman, 1953), but the prior on most of the factors was centered at zero. This work was developed further in Kaufman and Press (1976), who showed that the posterior distribution of the factor loading matrix was truncated multivariate normal in large samples.

Martin and McDonald (1975) approached the factor analysis problem looking for a Bayesian solution to Heywood cases. They adopted a diagonal disturbance covariance matrix and used a Jeffreys type vague prior for the elements. They proposed finding posterior joint modal estimators of the factor loading and disturbance covariance matrices, and obtained an implicit numerical solution. A point estimate of the factor loading matrix was also obtained.

Wong (1980) addressed the factor analysis problem from the empirical Bayes point of view, adopting normal priors for the factor loadings. He suggesed use of the EM algorithm (see Dempster, Laird, and Rubin, 1977) to find a posterior mode for the factor loading matrix, but an explicit algorithm was not obtained.

Lee (1981) adopted a higherarchial Bayesian approach to confirmatory factor analysis, starting from the assumption that the free parameters in the factor loading matrix were exchangeable and normally distributed. The disturbance covariance matrix was assumed to be diagonal. Joint modal estimates were found of the factor loading matrix, the variances of the disturbances, and the covariance matrix of the factors. This modal solution was implicit and numerical. A point estimate of the factor loading matrix was obtained.

Euverman and Vermulst (1983) studied the Bayesian Factor analysis model with a diagonal disturbance covariance matrix, and a preassigned number of factors. A numerical computer routine was described for implicitly finding the posterior joint mode of the factor loadings and error variances.

Mayekawa (1985) studied the Bayesian factor analysis problem examining factor scores as well as factor loadings and error (specific) variances. The factor loadings were assumed to to be normal, a priori. The author used the EM algorithm to find point estimates of the parameters as marginal modes of the posterior distributions. Unfortunately, however, there were no proofs about convergence of the EM algorithm used.

Shigemasu (1986) used a natural conjugate prior Bayesian approach to the factor analysis model and found implicit numerical solutions for the factor loading matrix and specific variances.

Akaike (1987) suggested that the AIC criterion could be used to select the appropriate number of factors to use in a Bayesian model (see also Press, 1982). He was motivated by the desire to deal with the problem of frequent occurrence of improper solutions in maximum likelihood factor analysis caused by overparameterization of the model. The introduction of prior information in our model directly addresses this issue. By minimizing the AIC in this, and other factor analysis models, results can be used to test hypothesis about the appropriate number of factors.

In contrast to the earlier work on Bayesian factor analysis which focused upon point estimation, in this paper we also develop methods for obtaining large sample interval estimators of factor scores, factor loadings, and specific variances. Consequently, standard Bayesian hypothesis testing methods (see, e.g. Press, 1982; 1989) can be used for testing hypothesis about all the fundamental quantities (apart from the number of factors) in the model. Because we develop exact (large sample) posterior distributions for these quantities, level curves, or contours, of the posterior distributions can be studied for sensitivity around the point estimators by examining the steepness and shape of the level curves. (Earlier research in which only point estimators were proposed has not suggested simple methods for studying estimator sensitivity.) Finally, our development yields *explicit* analytical results for the distributions of the quantities of interest (as well as some general implicit solutions), whereas most earlier work focused only on *implicit* numerical solutions of the matrix equations.

This paper is constructed so that the basic model we are adopting is set out in Section 2. The procedures for estimating factor scores and loadings, and disturbance variances and covariances, are discussed in Sections 3, 4, and 5, respectively. The paper concludes in Section 6 with a numerical illustration of the procedures. In this revision of the paper we have included an appendix that supplies some intermediate results.

2 Model

In this section we develop the basic factor analysis model. We first define the likelihood function. Then we introduce prior distributins on the parameters and calculate the joint posterior density of the parameters. Finally we find the marginal posterior densities for the parameters.

LIKELIHOOD FUNCTION

Define p-variate observation vectors, $(x_1, \ldots, x_N) \equiv X'$ on N subjects. The means are assumed to have been subtracted out, so that $E(X') = 0$. The prime denotes transposed matrix. The traditional factor analysis model is

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

for $j = 1, \ldots, N$, where Λ denotes a matrix of constants called the factor loading matrix; f_j denotes the factor score vector for subject j; $F' \equiv (f_1, \ldots, f_N)$. The ϵ_j's are assumed to be mutually uncorrelated and normally distributed as $N(0, \Psi)$, for Ψ a symmetric positive definite matrix, i.e. $\Psi > 0$. Note that Ψ is not assumed to be diagonal (but note from (2.5b) that $E(\Psi)$ is diagonal).

We assume that (Λ, F, Ψ) are unobserved and fixed quantities, and we assume that we can write the probability law of x_j as

$$\mathcal{L}(x_j|\Lambda, f_j, \Psi) = N(\Lambda f_j, \Psi), \tag{2.2}$$

where $\mathcal{L}(\cdot)$ denotes probability law. Equivalently, if "$\propto$" denotes proportionality, the likelihood for (Λ, F, Ψ) is

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

We will use $p(\cdot)$ generically to denote "density"; the p's will be distinguished by their arguments. This should not cause confusion. The proportionality constant in (2.3) is numerical, depending only on (p, N) and not upon (Λ, F, Ψ).

PRIORS

We use a generalized natural conjugate family (see Press, 1982) of prior distributions for (Λ, F). We take as prior density for the unobservables (to represent our state of uncertainty)

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

where

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

$$p(\Psi) \propto |\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2} tr \Psi^{-1} B}, \tag{2.5b}$$

with B a diagonal matrix, and $H > 0$. (Choices for the prior density $p(F)$ will be discussed in Section 3.) Thus, Ψ^{-1} follows a Wishart distribution, (ν, B) are hyperparameters to be assessed; Λ conditional on Ψ has elements which are jointly normally distributed, and (Λ_0, H) are hyperparameters to be assesssed. Note that $E(\Psi|B)$ is diagonal, to represent traditional views of the factor model containing "common" and "specific" factors. Also, note that if $\Lambda \equiv (\lambda_1, \ldots, \lambda_m)$, $\lambda \equiv vec(\Lambda) = (\lambda_1', \ldots, \lambda_m')'$, then $\text{var}(\lambda|\Psi) = H^{-1} \otimes \Psi$, $\text{var}(\lambda) = H^{-1} \otimes (E\Psi)$, and $\text{cov}[(\lambda_i, \lambda_j)|\Psi] = H_{ij}^{-1}\Psi$. Moreover, we will often take $H = n_0 I$, for some preassigned scalar n_0. These interpretations of the hyperparameters will simplify assessment.

JOINT POSTERIOR

Combining (2.3)-(2.5), the joint posterior density of the parameters becomes

$$p(\Lambda, F, \Psi|X) \propto p(F)|\Psi|^{-\frac{(N+m+\nu)}{2}} e^{-\frac{1}{2}tr\Psi^{-1}G}, \tag{2.6}$$

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

MARGINAL POSTERIORS

Integrating with respect to Ψ, and using properties of the Inverted Wishart density, gives the marginal posterior density of (Λ, F):

$$p(\Lambda, F|X) \propto p(F)|G|^{-\frac{(N+m+\nu-p-1)}{2}}. \tag{2.7}$$

We next want to integrate (2.7) with respect to Λ, to obtain the marginal posterior density of F. We accomplish this by factoring G into a form which makes it transparent that in terms of Λ, the density is proportional to a matrix T-density. Thus, completing the square in Λ in the G function defined in (2.6), (2.7) may be rewritten as

$$p(\Lambda, F|X) \propto \frac{p(F)}{|R_F + (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)'|^{\frac{\gamma}{2}}}. \tag{2.8}$$

where

$$\begin{aligned} Q_F &= H + F'F, \\ R_F &= X'X + B + \Lambda_0 H\Lambda_0' - (X'F + \Lambda_0 H)Q_F^{-1}(X'F + \Lambda_0 H)', && (2.9) \\ \Lambda_F &= (X'F + \Lambda_0 H)(H + F'F)^{-1} && (2.10) \\ \gamma &= N + m + \nu - p - 1 && (2.11) \end{aligned}$$

(2.8) is readily integrated with respect to Λ (by using the normalizing constant of a matrix T-distribution) to give the marginal posterior density of F,

$$p(F|X) \propto \frac{p(F)}{|R_F|^{\frac{\gamma-m}{2}}|Q_F|^{\frac{p}{2}}}. \tag{2.12}$$

After some algebra, the marginal posterior density of F in (2.12) may be written in the form

$$p(F|X) \propto \frac{p(F)|H + F'F|^{\frac{\gamma-m-p}{2}}}{|A + (F - \hat{F})'(I_N - XW^{-1}X')(F - \hat{F})|^{\frac{\gamma-m}{2}}}. \tag{2.13}$$

where

$$\hat{F} \equiv (I_N - XW^{-1}X')^{-1}XW^{-1}\Lambda_0 H$$

$$
\begin{aligned}
&= (I_N - X(X'X - W)^{-1}X')XW^{-1}\Lambda_0 H \qquad (2.14)\\
W &\equiv X'X + B + \Lambda_0 H \Lambda_0', \qquad (2.15)\\
A &\equiv H - (\Lambda_0 H)'W^{-1}\Lambda_0 H\\
&\quad -(XW^{-1}\Lambda_0 H)'(I_N - XW^{-1}X')^{-1}(XW^{-1}\Lambda_0 H)\\
&\equiv H - (\Lambda_0 H)'W^{-1}\Lambda_0 H\\
&\quad -(XW^{-1}\Lambda_0 H)'(I_N - X(X'X - W)^{-1}X')(XW^{-1}\Lambda_0 H) \qquad (2.16)
\end{aligned}
$$

Note 1. In (2.14) and (2.16) the second representations of $\hat{F}$ and A are more convenient for numerical computation than the first ones, because we need only invert a matrix of order p, instead of one of order N.

Note 2. In (2.15) the quantity $\frac{X'X}{N}$ is the sample covariance matrix of the observed data (since the data are assumed to have mean zero). If the data are scaled to have variance of unity, $\frac{X'X}{N}$ denotes the data correlation matrix.

Note 3. In (2.16), $H = H'$, but we have left H' to preserve the symmetry of the representation.

3 Estimation of Factor Scores

Now examine (2.13). There are several cases of immediate interest. We take factor scores of subjects to be independent, a priori, so we can think of $p(F)$ as $p(f_1), p(f_2), \ldots, p(f_N)$.

HISTORICAL DATA ASSESSMENT OF F

Suppose that, on the basis of historical data that is very similar to the current data set, we can assess $p(F)$. We can then evaluate $p(F|X)$ numerically from (2.13) to construct point estimators, and we can make interval estimates from the cdf of $p(F|X)$.

VAGUE PRIOR ESTIMATION OF F

Suppose instead that we are uninformed about F, a priori, and we accordingly adopt a vague prior

$$p(F) \propto \text{constant}. \qquad (3.1)$$

Then (2.13) becomes

$$p(F|X) \propto \frac{|H + F'F|^{\frac{\gamma-m-p}{2}}}{|A + (F - \hat{F})'(I_N - XW^{-1}X')(F - \hat{F})|^{\frac{\gamma-m}{2}}}. \qquad (3.2)$$

Again, interval estimates of F can be made numerically from the cdf of $p(F|X)$, and point estimates can also be obtained numerically from (3.2). Such numerical evaluations are treated in Press and Davies (1987).

LARGE SAMPLE ESTIMATION OF F

We note that

$$\frac{F'F}{N} = \frac{1}{N}\sum_{j=1}^{N} f_j f_j'.$$

If we assume (without loss of generality) that $E(f_j) = 0$, $\mathrm{var}(f_j) = I_m$, then for large N, by the law of large numbers,

$$\frac{F'F}{N} \approx I_m.$$

Thus, for large N, $|H + F'F| \approx |H + NI_m|$, a term which can be incorporated into the proportionality constant in (2.13), because it no longer depends on F. (2.13) may now be rewritten, for large N, as

$$p(F|X) \propto \frac{p(F)}{|A + (F - \hat{F})'(I_N - XW^{-1}X')(F - \hat{F})|^{\frac{\gamma - m}{2}}},$$

where $\hat{F}$ is defined by (2.14).

Suppose $p(F) \propto$ constant. Then, $(F|X)$ follows a matrix T-distribution with density

$$p(F|X) \propto |A + (F - \hat{F})'(I_N - XW^{-1}X')(F - \hat{F})|^{-\frac{(\gamma - m)}{2}}. \tag{3.3}$$

Alternatively, suppose $\mathcal{L}(f_j) = N(0, I_m)$, and the f_j's are mutually independent. Then,

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

For large N, since $F'F \approx NI_m$, $p(F)$ can be incorporated into the proportionality constant in (2.13) to yield (3.3). The same argument applies to any prior density for F which can depent upon $F'F$.

In summary, we conclude that for large N, and for a wide variety of important priors for F (a vague prior, or for any prior which depends on F only through $F'F$), the marginal posterior density of F, given the observed data vectors, is approximately matrix T, as given in (3.3), centered at $\hat{F}$. In particular, $E(F|X) \approx \hat{F}$, for large N.

LARGE SAMPLE ESTIMATION OF f_j

Since $(F|X)$ is approximately distributed as matrix T, $(f_j|X)$ is distributed as multivariate t (see, Theorem 6.2.4, in Press, 1982, p. 140), where f_j is the j^{th} column of F', with density given by

$$p(f_N|X) \propto \left(P_{22.1}^{-1} + (f_N - \hat{f}_N)'A^{-1}(f_N - \hat{f}_N)\right)^{-\frac{(\gamma - m - N + 1)}{2}}, \tag{3.4}$$

where $P_{22.1} = P_{22} - P_{21}P_{11}^{-1}P_{12}$ is obtained from

$$P \equiv I_N - XW^{-1}X' = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \qquad P_{11} : (N-1) \times (N-1).$$

By reindexing the subjects, (3.4) gives the posterior density of the factor score vector for any of the N subjects. (3.4) can readily be placed into canonical form

$$p(f_N|X) \propto \left(1 + (f_N - \hat{f}_N)' \left(\frac{A}{P_{22.1}}\right)^{-1} (f_N - \hat{f}_N)\right)^{-\frac{(\delta+m)}{2}}, \qquad (3.5)$$

where $\delta \equiv \nu - p - m$.

LARGE SAMPLE ESTIMATION OF THE ELEMENTS OF f_j.

Now suppose we wish to make posterior probability statements about a particular element of $f_N \equiv (f_{kN})$, $k = 1, \ldots, m$, say f_{1N}. We use the posterior density of a Student t-variate obtained as the marginal of the multivariate t-density in (3.5) (see e.g., Press (1982), p. 137). It is given by

$$p(f_{1N}|X) \propto \left(\delta + \left(\frac{f_{1N} - \hat{f}_{1N}}{\sigma_1}\right)^2\right)^{-\frac{(\delta+1)}{2}}, \qquad (3.6)$$

where σ_1^2 is the $(1,1)$ element of

$$\frac{A}{\delta P_{22.1}} = \begin{pmatrix} \sigma_1^2 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

$\hat{f}_{1N}$ is of course the $(1,N)$ element of $\hat{F}'$. From (3.6) we can make posterior probability statements about any factor score for any subject; i.e., we can obtain credibility (confidence) intervals for any factor score. For example,

$$\left(\frac{f_{1N} - \hat{f}_{1N}}{\sigma_1}\right)\Bigg| X \sim t_\delta.$$

4 Estimation of the Factor Loadings Matrix

We now return to the joint posterior density of (Λ, F), given in (2.8). One method of estimating Λ would be to integrate F out of (2.8) to obtain the marginal posterior density of Λ. Then, some measure of location of the distribution could be used as a point estimator of Λ. Unfortunately, while the integration can be carried out, the resulting marginal density is extremely complicated, and it does not seem possible to obtain a mean or mode of the distribution for any realistic prior densities for F, except numerically. The result is

$$p(\Lambda|X) \propto |P_\Lambda|^{-\frac{\gamma}{2}} |\Lambda' P_\Lambda^{-1} \Lambda|^{-\frac{N}{2}} |Z|^{-\frac{(\gamma-m)}{2}},$$

where

$$P_\Lambda \equiv B + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)',$$

and

$$Z \equiv I_N + XP_\Lambda^{-1}X' - (XP_\Lambda^{-1}\Lambda)(\Lambda' P_\Lambda^{-1}\Lambda)^{-1}(XP_\Lambda^{-1}\Lambda)'$$

Since this distribution is so complicated, we will alternatively estimate Λ for given $F = \hat{F}$. First note from (2.8) that

$$p(\Lambda|F, X) \propto |R_F + (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)'|^{-\frac{\gamma}{2}}. \tag{4.1}$$

That is, the conditional distribution of Λ for prespecified F is matrix T. Our point estimator of Λ is $E(\Lambda|\hat{F}, X)$, or

$$\hat{\Lambda} \equiv \Lambda_{\hat{F}} = (X'\hat{F} + \Lambda_0 H)(H + \hat{F}'\hat{F})^{-1} \tag{4.2}$$

Any scalar element of Λ, conditional on $(\hat{F}, X)$, follows a general Student t-distribution, analogous to the general univariate marginal Student t-density in (3.6), which corresponds to the matrix T-density in (3.3).

We note also that $\hat{\Lambda}$ in (4.2) is both a mean and a modal estimator of the joint distribution of $(\Lambda, F|X)$ under a vague prior for F (see (2.8) and (2.10)). This follows from the unimodality and the symmetry of the density in (2.8). Thus, in this case, $(\hat{F}, \hat{\Lambda})$ is a joint modal estimator of (F, Λ).

5 Estimation of the Disturbance Covariance Matrix

The disturbance covariance matrix, Ψ, is estimated conditional upon $(F, \Lambda) = (\hat{F}, \hat{\Lambda})$. The joint posterior density of $(\Psi, \Lambda, F|X)$ is given in (2.6). The conditional density of $(\Psi|\Lambda, F, X)$ is obtained by dividing (2.6) by (2.7) and setting $G = \hat{G}$ ($\hat{G}$ depends only on the data X, and the hyperparameters H, B, and Λ_0). The result is

$$p(\Psi|\hat{\Lambda}, \hat{F}, X) \propto \frac{e^{-\frac{1}{2}tr\Psi^{-1}\hat{G}}}{|\Psi|^{\frac{(N+m+\nu)}{2}}} \tag{5.1}$$

where

$$\hat{G} = (X - \hat{F}\hat{\Lambda}')'(X - \hat{F}\hat{\Lambda}') + (\hat{\Lambda} - \Lambda_0)H(\hat{\Lambda} - \Lambda_0)' + B. \tag{5.2}$$

That is, the posterior conditional distribution of Ψ given $(\hat{F}, \hat{\Lambda}, X)$ is Inverted Wishart. A point estimator of Ψ is given by $\hat{\Psi} = E(\Psi|\hat{\Lambda}, \hat{F}, X)$. Equation (5.2.4), page 119 in Press (1982) gives

$$\hat{\Psi} = \frac{\hat{G}}{N + m + \nu - 2p - 2}, \tag{5.3}$$

with $\hat{G}$ given in (5.2).

6 Example

We have extracted some data from an illustrative example in Kendall (1980), and have analyzed this data from the Bayesian viewpoint using our model. 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 subset of factors that explain the variation observed in the scores? If so, then each applicant 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 after subtracting it out.)

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

Table 2: Correlation matrix of variables 1 through 15.

	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

Now we postulate a model with 4 factors. This choice is based upon our having carried out a principal components analysis and our having found that 4 factors accounted for 81.5% of the variance. This is therefore our first guess, a conclusion that might be modified if we were to do hypothesis testing to see how well a four factor model fit the data. Based upon underlying theory we constructed the prior factor loading matrix

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

The hyperparameter H was assessed as $H = 10I_4$. The prior distribution for Ψ was assessed with $B = 0.2I_{15}$, and $\nu = 33$. Note that when our observational data is augmented by proper prior information, as in this example, the identification-of-parameters problem of classical factor analysis disappears. The factor scores, factor loadings, and disturbance variances and covariances may now be estimated from (2.14), (4.2), and (5.3), respectively. Results are given in Tables 3, 4, and 5.

Note that since we used standardized scores, the elements in Table 2 may be interpred as correlations. It may also be noted from Table 5 that most of the off diagonal elements of the estimated disturbance matrix Ψ are very small relative to the diagonal elements (the variances). That is, Ψ is approximately diagonal. Tables 3, 4, and 5 give the Bayesian point estimates for (F, Λ, Ψ). We next obtain two-tailed 95% credibility intervals for the 48^{th} subject's factor scores, and for the last (15^{th}) row of the factor loading matrix.

The factor scores for subject 48 are given in the last row of the matrix in Table 3 as

$$(-2.156, 2.038, -2.529, -0.751).$$

Now calculate the two-tailed credibility interval at the 95% level from (3.6) and find the intervals

$$\begin{array}{l} [-3.160, -1.152]\,, \\ [-0.578, \quad 4.649]\,, \\ [-4.058, -1.001]\,, \\ [-2.617, \quad 1.115]\,. \end{array}$$

The factor loadings for row 15 of the factor loading matrix are obtained from Table 4 as

$$(0.128, -0.015, 0.667, 0.011).$$

Now calculate 95% two-tailed credibility intervals from the marginals of (4.1), just as we obtained the result in (3.6) from (3.3). Results for the last row factor loadings are

$$\begin{array}{l} [-0.219, \ 0.475]\,, \\ [-0.295, \ 0.266]\,, \\ [\ \ 0.341, \ 1.012]\,, \\ [-0.314, \ 0.337]\,. \end{array}$$

Hypothesis about the elements of (F, Λ, Ψ) may be tested using the associated marginal posterior densities. These are quite simple, being Student t, Student t for given F, and Inverted Wishart given F, and Λ, respectively. For example, note that the credibility intervals for the first, second and fourth factor loadings corresponding to the last row of Table 4 include the origin. A commonly used Bayesian hypothesis testing procedure suggests that we should therefore conclude that we cannot reject the hypothesis that these three factor loadings are zero.

Table 3: Bayes estimates of factor scores.

Person	1	2	3	4
1	0.728	-3.548	0.405	-0.301
2	1.476	-1.454	1.225	0.735
3	1.020	-2.850	0.726	0.231
4	-0.288	0.640	0.226	-0.021
5	-0.324	0.640	0.691	0.735
6	0.263	-0.058	0.868	0.510
7	1.188	0.640	1.942	0.456
8	1.475	1.338	1.942	0.456
9	0.876	-0.058	1.799	0.456
10	1.880	2.035	0.050	-1.336
11	1.550	2.035	-0.382	-2.957
12	1.547	2.035	-0.525	-0.832
13	-0.590	0.640	0.261	1.239
14	-0.623	0.640	0.472	0.708
15	-0.708	0.640	0.049	0.762
16	0.903	-0.756	1.123	0.204
17	0.422	-0.058	0.903	0.204
18	-0.195	0.640	-0.458	-1.111
19	-0.217	0.640	-0.315	-1.390
20	0.729	-0.058	-0.237	1.014
21	0.597	-0.756	-0.415	1.014
22	1.591	-0.058	0.650	1.014
23	1.591	-0.058	0.295	1.266
24	1.364	-0.058	0.506	1.518
25	-0.932	-0.058	-0.602	0.483
26	-0.702	-0.058	0.007	0.456
27	0.326	-0.058	-1.024	1.266
28	-1.822	-1.454	-1.464	-1.642
29	-2.047	-2.152	-1.819	-3.038
30	-0.976	-1.454	-1.244	0.510
31	-0.586	-2.152	-0.923	0.762
32	-0.151	-1.454	-1.422	0.762
33	-0.492	-1.454	-1.600	0.762
34	-1.601	-0.756	-1.566	-0.553
35	-2.198	-2.152	-0.889	-0.526
36	-0.699	-1.454	-0.213	-0.301
37	0.674	-0.756	-1.388	-0.553
38	0.722	-0.756	-1.388	-0.327
39	1.720	1.338	2.120	1.518
40	1.859	1.338	2.120	1.518
41	-2.283	0.640	2.120	-3.236
41	-2.709	0.640	2.120	-3.794
43	-1.647	1.338	0.015	-0.381
44	1.090	-0.058	0.581	-0.048
45	-0.007	2.035	-0.069	1.266
46	0.521	2.035	-0.212	1.518
47	-2.156	2.035	-2.529	-0.247
48	-2.156	2.035	-2.529	-0.751

Table 4: Bayes estimates of factor loadings.

	1	2	3	4
1	-0.045	-0.065	0.711	0.028
2	0.241	0.046	0.094	0.175
3	0.000	0.726	0.000	0.000
4	-0.010	-0.010	0.152	0.703
5	0.775	-0.051	-0.192	-0.029
6	0.719	-0.011	-0.058	0.035
7	0.012	0.010	-0.152	0.722
8	0.754	-0.049	0.016	-0.091
9	-0.081	0.080	0.737	-0.040
10	0.653	-0.020	0.116	-0.021
11	0.771	-0.046	-0.028	-0.100
12	0.659	0.057	0.052	0.066
13	0.592	0.120	0.094	0.141
14	0.244	-0.263	0.221	0.311
15	0.128	-0.015	0.677	0.011

Table 5: Bayes estimates of the disturbance covariance matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.278	0.008	0.000	0.035	0.015	0.009	-0.035	-0.026	-0.148	-0.035	0.029	0.008	0.000	0.082	-0.127
2		0.650	0.000	-0.025	0.019	-0.077	0.026	0.035	-0.058	-0.099	0.109	0.016	-0.002	-0.061	0.050
3			0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	-0.001	0.000
4				0.132	-0.055	0.026	-0.128	0.002	-0.024	-0.015	0.018	-0.001	0.026	0.049	-0.010
5					0.150	-0.017	0.055	0.001	0.013	-0.031	0.015	-0.052	-0.061	-0.018	-0.029
6						0.148	-0.026	-0.011	-0.002	-0.077	-0.071	0.055	-0.022	-0.025	-0.008
7							0.132	-0.002	0.024	0.015	-0.018	0.001	-0.025	-0.048	0.010
8								0.126	-0.011	0.001	-0.003	-0.055	-0.054	0.007	0.038
9									0.199	-0.004	-0.003	0.005	0.002	-0.026	-0.047
10										0.203	-0.018	-0.077	0.003	0.052	0.039
11											0.128	-0.033	-0.014	0.021	-0.026
12												0.141	0.025	-0.012	-0.012
13													0.125	-0.023	-0.002
14														0.308	-0.055
15															0.177

REFERENCES

Akaiki, H.(1987). "Factor analysis and AIC. " *Psychometrika*, Vol 52, No 3, 317–32.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum likelihood from incomplete Data via the EM algorithm", Jour Roy Statist Soc (B), 39, 1–38.

Euverman, T.J. and Vermulst, A.A. (1983). *Bayesian Factor Analysis.* Department of Social Sciences, State University of Groningin, the Netherlands, Nov 3, 1983.

Guttman, L. (1953). "Image theory for the structure of quantitative variates." *Psychometrika*, Vol 18, No 4, 277–96.

Kaufman, G.M. and Press, S. J.(1973a). "Bayesian factor analysis." Report No. 73t-25, *Bull. Inst. of Math Statist.*, Vol 2, No 2, Issue No. 7, March, 1973.

Kaufman, G.M. and Press, S. J. (1973b). "Bayesian Factor analysis." Report No. 7322 Center for Mathematical Studies in Business and Economics, University of Chicago, April, 1973.

Kaufman, G.M. and Press, S. J. (1976), "Bayesian Factor analysis." Working Paper No. 413, Department of Commerce and Business Administration, University of British Columbia, Vancouver, B.C. Canada, September,1983.

Kendall, M. (1980), *Multivariate Analysis.* Second edition, Charles Griffin Pub, 34.

Lee, S. Y. (1981), "A Bayesian approach to confirmatory Factor analysis." *Psychometrika*, Vol 46, No 2, June, 1981, 153–59.

Martin, J. K. and McDonald, R. P.(1975). "Bayesian estimation in unrestricted Factor analysis: A treatment for Heywood cases.", *Psychometrika*, Vol 40, No 4, 505–17.

Mayekawa, S. (1985). "Bayesian Factor analysis", ONR Tech Report No 85-3, School of Education, The University of Iowa, Iowa City, Iowa.

S. James Press (1972). *Applied Multivariate Analysis.* New York: Holt, Rinehart, and Winston, Inc.

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

S. James Press (1989). *Bayesian Statistics: Principles Models, and Applications* John Wiley and Sons, New York.

S. James Press and W. Davies (1987). "Asymptotics for the ratio of multiple t-densities." in *Contributions to the Theory and Applications of Statistics*, ed. Alan Gelfand, New York, Academic Press, 155–77.

Shigemasu, K. (1986)."A Bayesian estimation procedure with informative prior in factor analysis model." *Tokyo Institute of Technology*, Jinmon-Ronso, No 11, 67–77.

Wong, G. (1980). "A Bayesian approach to factor analysis." Program in Statistical Technical Report, No 80-9, Educational Testing Service, Princeton, N.J., August, 1980.

Appendix

by Daniel B. Rowe

This appendix provides some of the intermediate calculations, and the numerical results, to correct some errors in the original manuscript.

A Intermediate Calculations

A.1 F

The likelihood is a matrix normal distribution.

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

From the matrix normal prior on Λ and the Inverted Wishart on Ψ we get the full prior

$$p(\Lambda, F, \Psi) \propto p(F)|\Psi|^{-\frac{(m+\nu)}{2}} e^{-\frac{1}{2}tr\Psi^{-1}[(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)'+B]}$$

Combining these by Bayes rule we get the full posterior

$$p(\Lambda, F, \Psi|X) \propto p(F)|\Psi|^{-\frac{(N+m+\nu)}{2}} e^{-\frac{1}{2}tr\Psi^{-1}G}, \tag{2.6}$$

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

We now integrate the full posterior distribution with respect to Ψ to obtain the joint marginal posterior distribution $p(\Lambda, F|X)$. We do this by using the properties of the Inverted Wishart distribution (see. p. 117, Press 1982).

$$p(\Lambda, F|X) \propto \frac{p(F)}{|G|^{\frac{(N+m+\nu-p-1)}{2}}}. \tag{2.7}$$

After some matrix algebra on G (see section on Algebra), we obtain

$$p(\Lambda, F|X) \propto \frac{p(F)}{|R_F + (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)'|^{\frac{\gamma}{2}}}. \tag{2.8}$$

where

$$\begin{aligned}
Q_F &\equiv H + F'F, \\
R_F &\equiv X'X + B + \Lambda_0 H\Lambda_0' - (X'F + \Lambda_0 H)Q_F^{-1}(X'F + \Lambda_0 H)', \quad (2.9) \\
\Lambda_F &\equiv (X'F + \Lambda_0 H)(H + F'F)^{-1} \quad (2.10) \\
\gamma &\equiv N + m + \nu - p - 1 \quad (2.11)
\end{aligned}$$

We can now obtain the marginal posterior density $p(F|X)$. We do this by integrating $p(\Lambda, F|X)$ with respect to Λ. We do this by using the properties of the matrix T-distribution (see. p. 138, Press 1982).

$$p(F|X) \propto \frac{p(F)}{|R_F|^{\frac{(\gamma-m)}{2}}|Q_F|^{\frac{p}{2}}}. \tag{2.12}$$

Substituting the expressions represented by R_F and letting

$$W \equiv X'X + B + \Lambda_0 H \Lambda_0'$$

we get

$$p(F|X) \propto \frac{p(F)|Q_F|^{\frac{-p}{2}}}{|W - (X'F + \Lambda_0 H)Q_F^{-1}(X'F + \Lambda_0 H)'|^{\frac{(\gamma-m)}{2}}}.$$

Now use the alternative representation of the Matrix T-distribution (see. p. 139, Press 1982).

$$p(F|X) \propto \frac{|W|^{-\frac{(\gamma-2m)}{2}}|Q_F|^{\frac{N}{2}}}{|W|^{-\frac{(\gamma-2m)}{2}}|Q_F|^{\frac{N}{2}}} \frac{p(F)|Q_F|^{\frac{-p}{2}}}{|W + (X'F + \Lambda_0 H)[-Q_F]^{-1}(X'F + \Lambda_0 H)'|^{\frac{(\gamma-m)}{2}}}.$$

$$p(F|X) \propto |Q_F|^{\frac{N}{2}} p(F)|Q_F|^{\frac{-p}{2}} \frac{|Q_F|^{\frac{(\gamma-m-N)}{2}}|W|^{\frac{m}{2}}}{|Q_F - (X'F + \Lambda_0 H)'W^{-1}(X'F + \Lambda_0 H)|^{\frac{(\gamma-m)}{2}}}.$$

$$p(F|X) \propto \frac{p(F)|H + F'F|^{\frac{(\gamma-m-p)}{2}}}{|Q_F - (X'F + \Lambda_0 H)'W^{-1}(X'F + \Lambda_0 H)|^{\frac{(\gamma-m)}{2}}}.$$

After some matrix algebra on

$$M_F \equiv Q_F - (X'F + \Lambda_0 H)'W^{-1}(X'F + \Lambda_0 H)$$

(see section on Algebra), the marginal posterior density of F may be written in the form

$$p(F|X) \propto \frac{p(F)|H + F'F|^{\frac{(\gamma-m-p)}{2}}}{|A + (F - \hat{F})'P(F - \hat{F})|^{\frac{(\gamma-m)}{2}}}. \tag{2.13}$$

where

$$\begin{aligned} P &\equiv (I_N - XW^{-1}X') \\ A &\equiv H - (\Lambda_0 H)'W^{-1}(\Lambda_0 H) \\ &\quad -(XW^{-1}\Lambda_0 H)'P^{-1}(XW^{-1}\Lambda_0 H) \\ \hat{F} &\equiv P^{-1}(XW^{-1}\Lambda_0 H) \end{aligned}$$

Now, using the large sample approximation $F'F \approx \frac{I}{N}$ and either a vague or normal prior on F, we obtain

$$p(F|X) \propto \frac{1}{|A + (F - \hat{F})'P(F - \hat{F})|^{\frac{(\gamma - m)}{2}}}.$$

We can convert this to the usual form of the matrix T-distribution (see p.138, Press 1982).

$$p(F|X) \propto \frac{1}{|P^{-1} + (F - \hat{F})A^{-1}(F - \hat{F})'|^{\frac{(\gamma - m)}{2}}}.$$

And since the mean and mode of this matrix T-distribution is $\hat{F}$ we will use $\hat{F}$ as a point estimator for F.

We can obtain the marginal posterior distribution for any row of F. We will obtain the marginal posterior distribution the last row of F namely f'_N (see Theorem 6.2.4, p. 140, Press 1982). It is

$$p(f_N|X) \propto \left(P_{22.1}^{-1} + (f_N - \hat{f}_N)'A^{-1}(f_N - \hat{f}_N)\right)^{-\frac{(\gamma - m - N + 1)}{2}},$$

and $P_{22.1} = P_{22} - P_{21}P_{11}^{-1}P_{12}$ is obtained from

$$P \equiv I_N - XW^{-1}X' = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \qquad P_{11} : (N-1) \times (N-1).$$

The above distribution can be placed in the canonical form

$$p(f_N|X) \propto \left(1 + (f_N - \hat{f}_N)'\left(\frac{A}{P_{22.1}}\right)^{-1}(f_N - \hat{f}_N)\right)^{-\frac{(\delta + m)}{2}}, \qquad (3.5)$$

where $\delta \equiv \nu - p - m$. The marginal posterior distribution of any of the elements say the k^{th} (see p.137, Press 1982) is

$$p(f_{kN}|X) \propto \left(\delta + \left(\frac{f_{kN} - \hat{f}_{kN}}{\sigma_k}\right)^2\right)^{-\frac{(\delta + 1)}{2}}, \qquad (3.6)$$

where σ_k^2 is the (k,k) element of $\Sigma \equiv \frac{A}{\delta P_{22.1}}$, and $\hat{f}_{kN}$ is of course the (k, N) element of $\hat{F}'$. From the above we can make posterior probability statements about any factor score for any subject; i.e., we can obtain credibility (confidence) intervals for any factor score.

For example, a 95% credibility interval for f_{kN} is

$$\hat{f}_{kN} \pm \sigma_k t_\delta(0.975)$$

where $t_\delta(0.975)$ is the 97.5^{th} percentile of a t-distribution with δ degrees of freedom.

A.2 Λ

We can find the posterior marginal distribution $p(\Lambda|X)$. We start with equation 2.7. After performing some algebra on G (see section on Algebra), the joint marginal posterior density of Λ, F may be written in the form

$$p(\Lambda, F|X) \propto \frac{p(F)}{|P_\Lambda + (F\Lambda' - X)' I_N (F\Lambda' - X)|^{\frac{\gamma}{2}}}.$$

We can rewrite this using the alternative representation of the Matrix T-distribution (see. p. 139, Press 1982).

$$p(\Lambda, F|X) \propto \frac{1}{|P_\Lambda|^{\frac{(\gamma-N)}{2}} |I_N|^{\frac{m}{2}}} \frac{|P_\Lambda|^{\frac{(\gamma-N)}{2}} |I_N|^{\frac{m}{2}} p(F)}{|P_\Lambda + (F\Lambda' - X)' I_N (F\Lambda' - X)|^{\frac{\gamma}{2}}}.$$

$$p(\Lambda, F|X) \propto \frac{1}{|P_\Lambda|^{\frac{(\gamma-N)}{2}} |I_N|^{\frac{m}{2}}} \frac{1}{|P_\Lambda|^{\frac{N}{2}} |I_N|^{\frac{(\gamma-m)}{2}}} \frac{p(F)}{|I_N + (F\Lambda' - X) P_\Lambda^{-1} (F\Lambda' - X)'|^{\frac{\gamma}{2}}}.$$

$$p(\Lambda, F|X) \propto \frac{p(F)|P_\Lambda|^{-\frac{\gamma}{2}}}{|I_N + (F\Lambda' - X) P_\Lambda^{-1} (F\Lambda' - X)'|^{\frac{\gamma}{2}}}.$$

After some matrix algebra on

$$K = I_N + (F\Lambda' - X) P_\Lambda^{-1} (F\Lambda' - X)'$$

(see section on Algebra), we obtain

$$p(\Lambda, F|X) \propto \frac{p(F)|P_\Lambda|^{-\frac{\gamma}{2}}}{|Z + (F - F_0)(\Lambda' P_\Lambda^{-1} \Lambda)(F - F_0)'|^{\frac{\gamma}{2}}}$$

where

$$\begin{aligned} F_0 &\equiv X P_\Lambda^{-1} \Lambda (\Lambda' P_\Lambda^{-1} \Lambda)^{-1} \\ Z &= I_N + X P_\Lambda^{-1} X' - (X P_\Lambda^{-1} \Lambda)(\Lambda' P_\Lambda^{-1} \Lambda)^{-1} (X P_\Lambda^{-1} \Lambda)' \end{aligned}$$

We can now obtain the marginal posterior density $p(\Lambda|X)$. We do this by integrating $p(\Lambda, F|X)$ with respect to F. Assuming vague or normal prior on F and using the properties of the matrix T-distribution (see. p. 138, Press 1982). The result is

$$p(\Lambda|X) \propto |P_\Lambda|^{-\frac{\gamma}{2}} |\Lambda' P_\Lambda^{-1} \Lambda|^{-\frac{N}{2}} |Z|^{-\frac{(\gamma-m)}{2}}.$$

This posterior marginal distribution is very complicated. We will alternatively, estimate Λ for given $F = \hat{F}$. Note, that

$$p(\Lambda|F, X) = \frac{p(\Lambda, F|X)}{p(F|X)}.$$

Using the appropriate expressions, we obtain

$$p(\Lambda|F, X) \propto \frac{1}{|R_F + (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)'|^{\frac{\gamma}{2}}} \tag{4.1}$$

where $\Lambda_F = (X'F + \Lambda_0 H)(H + F'F)^{-1}$ as before.

If we then let $V = R_F^{-1}$, $E = Q_F^{-1}$, and $F = \hat{F}$ we get

$$p(\Lambda|\hat{F}, X) \propto \frac{1}{|V^{-1} + (\Lambda - \Lambda_{\hat{F}})E^{-1}(\Lambda - \Lambda_{\hat{F}})'|^{\frac{\gamma}{2}}}$$

where

$$\Lambda_{\hat{F}} = (X'\hat{F} + \Lambda_0 H)(H + \hat{F}'\hat{F})^{-1}. \tag{4.2}$$

And since the mean and modal estimators of this matrix T-distribution is $\hat{\Lambda} = \Lambda_{\hat{F}}$ we will use $\hat{\Lambda}$ as a point estimator for Λ .

The marginal distribution of any row of Λ can then be found. In particular, the marginal posterior distribution of λ_p, the last row of Λ is (see Theorem 6.2.4, p. 140, Press 1982)

$$p(\lambda_p|\hat{F}, X) \propto \frac{1}{|V_{22.1}^{-1} + (\lambda_p - \hat{\lambda}_p)E^{-1}(\lambda_p - \hat{\lambda}_p)'|^{\frac{\gamma}{2}}}$$

and $V_{22.1} = V_{22} - V_{21}V_{11}^{-1}V_{12}$ is obtained from

$$V \equiv R_F^{-1} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}, \qquad V_{11} : (p-1) \times (p-1).$$

The above distribution can be placed in the canonical form

$$p(\lambda_p|\hat{F}, X) \propto \left(1 + (\lambda_p - \hat{\lambda}_p)\left(\frac{E}{V_{22.1}}\right)^{-1}(\lambda_p - \hat{\lambda}_p)'\right)^{-\frac{(\eta+m)}{2}},$$

where $\eta \equiv N + \nu - 2p$. The marginal posterior distribution of any of the elements say the k^{th} (see p.137, Press 1982) is

$$p(\lambda_{pk}|X) \propto \left(\eta + \left(\frac{\lambda_{pk} - \hat{\lambda}_{pk}}{\theta_k}\right)^2\right)^{-\frac{(\eta+1)}{2}},$$

where θ_k^2 is the (k, k) element of $\Theta \equiv \frac{E}{\eta V_{22.1}}$, and $\hat{\lambda}_{pk}$ is of course the (p, k) element of $\hat{\Lambda}$. From the above we can make posterior probability statements about any factor score for any subject; i.e., we can obtain credibility (confidence) intervals for any factor score.

For example, a 95% credibility interval for λ_{pk} is

$$\hat{\lambda}_{pk} \pm \theta_k t_\eta(0.975)$$

where $t_\eta(0.975)$ is the 97.5^{th} percentile of a t-distribution with η degrees of freedom.

A.3 Ψ

The posterior disturbance covariance matrix is also estimated conditional on $(\Lambda, F) = (\hat{\Lambda}, \hat{F})$. Note, that

$$p(\Psi|\Lambda, F, X) = \frac{p(\Psi, \Lambda, F|X)}{p(\Lambda, F|X)}.$$

Using the appropriate expressions, we obtain

$$p(\Psi|\Lambda, F, X) \propto |\Psi|^{-\frac{(N+m+\nu)}{2}} e^{-\frac{1}{2}\Psi^{-1}G}$$

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

A point estimator for Ψ is given by (see Theorem 5.2.2, p. 119, Press 1982)

$$\hat{\Psi} = E(\Psi|\Lambda, F, X) = \frac{G}{N + m + \nu - 2p - 2} \tag{5.1}$$

Using $(\Lambda, F) = (\hat{\Lambda}, \hat{F})$

$$\hat{\Psi} = E(\Psi|\hat{\Lambda}, \hat{F}, X) = \frac{\hat{G}}{N + m + \nu - 2p - 2} \tag{5.3}$$

where

$$\hat{G} = (X - \hat{F}\hat{\Lambda}')'(X - \hat{F}\hat{\Lambda}') + (\hat{\Lambda} - \Lambda_0)H(\hat{\Lambda} - \Lambda_0)' + B. \tag{5.2}$$

A.4 Algebra

Here is the required matrix algebra on G, M_F, and K.

A.4.1 Algebra on G for $p(F|X)$

$$\begin{aligned}
G &= (X - F\Lambda')'(X - F\Lambda') + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B \\
&= (X' - \Lambda F')(X - F\Lambda') + (\Lambda H - \Lambda_0 H)(\Lambda' - \Lambda_0') + B \\
&= X'(X - F\Lambda') - \Lambda F'(X - F\Lambda') + \\
&\quad \Lambda H(\Lambda' - \Lambda_0') - \Lambda_0 H(\Lambda' - \Lambda_0') + B \\
&= X'X - X'F\Lambda' - \Lambda F'X + \Lambda F'F\Lambda' + \\
&\quad \Lambda H\Lambda' - \Lambda H\Lambda_0' - \Lambda_0 H\Lambda' + \Lambda_0 H\Lambda_0' + B \\
&= X'X + \Lambda_0 H\Lambda_0' + B + \Lambda H\Lambda' + \Lambda F'F\Lambda' - \Lambda F'X - \Lambda H\Lambda_0' \\
&\quad -X'F\Lambda' - \Lambda_0 H\Lambda'
\end{aligned}$$

$$
\begin{aligned}
&= W + \Lambda[(H + F'F)\Lambda' - F'X - H\Lambda_0'] - X'F\Lambda' - \Lambda_0 H\Lambda' \\
&= W + \Lambda Q_F[\Lambda' - Q_F^{-1}(F'X + H\Lambda_0')] - X'F\Lambda' - \Lambda_0 H\Lambda' \\
&= W + \Lambda Q_F[\Lambda - (X'F + \Lambda_0 H)Q_F^{-1}]' - X'F\Lambda' - \Lambda_0 H\Lambda' \\
&= W + \Lambda Q_F(\Lambda - \Lambda_F)' - X'F\Lambda' - \Lambda_0 H\Lambda' \\
&= W + (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)' + \Lambda_F Q_F(\Lambda - \Lambda_F)' \\
&\quad -X'F\Lambda' - \Lambda_0' H\Lambda' \\
&= W + (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)' + (X'F + \Lambda_0 H)Q_F^{-1}Q_F\Lambda' - \\
&\quad \Lambda_F Q_F\Lambda_F' - X'F\Lambda' - \Lambda_0 H\Lambda' \\
&= W + (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)' + X'F\Lambda' - \Lambda_0 H\Lambda' - \\
&\quad \Lambda_F Q_F\Lambda_F' - X'F\Lambda' - \Lambda_0 H\Lambda' \\
&= W + (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)' - \Lambda_F Q_F\Lambda_F' \\
&= W + (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)' - \\
&\quad (X'F + \Lambda_0 H)Q_F^{-1}Q_F Q_F^{-1}(F'X + H\Lambda_0') \\
&= W + (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)' - \\
&\quad (X'F + \Lambda_0 H)Q_F^{-1}(F'X + H\Lambda_0') \\
&= W - (X'F + \Lambda_0 H)Q_F^{-1}(X'F + \Lambda_0 H)' + \\
&\quad (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)' \\
&= R_F + (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)'
\end{aligned}
$$

A.4.2 Algebra on M_F

$$
\begin{aligned}
M_F &= H + F'F - (X'F + \Lambda_0 H)'W^{-1}(X'F + \Lambda_0 H) \\
&= H + F'F - (F'XW^{-1} + H\Lambda_0' HW^{-1})(X'F + \Lambda_0 H) \\
&= H + F'F - F'XW^{-1}(X'F + \Lambda_0 H) - \\
&\quad H\Lambda_0' W^{-1}(X'F + \Lambda_0 H) \\
&= H + F'F - F'XW^{-1}X'F - F'XW^{-1}\Lambda_0 H - \\
&\quad H\Lambda_0' W^{-1}X'F - H\Lambda_0' W^{-1}\Lambda_0 H \\
&= H - H\Lambda_0' W^{-1}\Lambda_0 H - H\Lambda_0' W^{-1}X'F + \\
&\quad F'[(I - XW^{-1}X')F - XW^{-1}\Lambda_0 H] \\
&= H - H\Lambda_0' W^{-1}\Lambda_0 H - H\Lambda_0' W^{-1}X'F + \\
&\quad F'(PF - XW^{-1}\Lambda_0 H) \\
&= H - H\Lambda_0' W^{-1}\Lambda_0 H - H\Lambda_0' W^{-1}X'F + \\
&\quad F'P(F - P^{-1}XW^{-1}\Lambda_0 H) \\
&= H - H\Lambda_0' W^{-1}\Lambda_0 H - H\Lambda_0' W^{-1}X'F + F'P(F - \hat{F}) \\
&= H - H\Lambda_0' W^{-1}\Lambda_0 H - H\Lambda_0' W^{-1}X'F + \\
&\quad (F - \hat{F})'P(F - \hat{F}) + \hat{F}'P(F - \hat{F})
\end{aligned}
$$

$$
\begin{aligned}
&= H - H\Lambda_0' W^{-1}\Lambda_0 H - H\Lambda_0' W^{-1} X' F + \\
&\quad (F-\hat{F})' P (F-\hat{F}) + \hat{F}' P F - \hat{F}' P \hat{F} \\
&= H - H\Lambda_0' W^{-1}\Lambda_0 H + (F-\hat{F})' P (F-\hat{F}) - H\Lambda_0' W^{-1} X' F + \\
&\quad H\Lambda_0' W^{-1} X' P^{-1} P F - \hat{F}' P P^{-1} X W^{-1} \Lambda_0 H \\
&= H - H\Lambda_0' W^{-1}\Lambda_0 H + (F-\hat{F})' P (F-\hat{F}) - H\Lambda_0' W^{-1} X' F + \\
&\quad H\Lambda_0' W^{-1} X' F - H\Lambda_0' W^{-1} X' P^{-1} P P^{-1} X W^{-1} \Lambda_0 H \\
&= H - H\Lambda_0' W^{-1}\Lambda_0 H + (F-\hat{F})' P (F-\hat{F}) - \\
&\quad (X W^{-1}\Lambda_0 H)' P^{-1} (X W^{-1} \Lambda_0 H) \\
&= A + (F-\hat{F})' P (F-\hat{F})
\end{aligned}
$$

A.4.3 Algebra on G for $p(\Lambda|X)$

$$
\begin{aligned}
G &= (X - F\Lambda')'(X - F\Lambda') + (\Lambda - \Lambda_0) H (\Lambda - \Lambda_0)' + B \\
&= (X' - \Lambda F')(X - F\Lambda') + (\Lambda H - \Lambda_0 H)(\Lambda' - \Lambda_0') + B \\
&= X'(X - F\Lambda') - \Lambda F'(X - F\Lambda') + \Lambda H (\Lambda' - \Lambda_0') - \\
&\quad \Lambda_0 H (\Lambda' - \Lambda_0') + B \\
&= X'X - X'F\Lambda' - \Lambda F' X + \Lambda F' F \Lambda' + \\
&\quad \Lambda H \Lambda' - \Lambda H \Lambda_0' - \Lambda_0 H \Lambda' + \Lambda_0 H \Lambda_0' + B \\
&= (F\Lambda')'(F\Lambda') - (F\Lambda')' X - X(F\Lambda') + X'X + \\
&\quad \Lambda H \Lambda' - \Lambda H \Lambda_0' - \Lambda_0 H \Lambda' + \Lambda_0 H \Lambda_0' + B \\
&= (F\Lambda')'[F\Lambda' - X] - X'(F\Lambda') + X'X + \\
&\quad \Lambda H \Lambda' - \Lambda H \Lambda_0' - \Lambda_0 H \Lambda' + \Lambda_0 H \Lambda_0' + B \\
&= (F\Lambda' - X)'(F\Lambda' - X) + X'[(F\Lambda') - X] - X'(F\Lambda') + \\
&\quad X'X + \Lambda H \Lambda' - \Lambda H \Lambda_0' - \Lambda_0 H \Lambda' + \Lambda_0 H \Lambda_0' + B \\
&= (F\Lambda' - X)'(F\Lambda' - X) + X'F\Lambda' - X'X - X'(F\Lambda') + \\
&\quad X'X + \Lambda H \Lambda' - \Lambda H \Lambda_0' - \Lambda_0 H \Lambda' + \Lambda_0 H \Lambda_0' + B \\
&= (F\Lambda' - X)'(F\Lambda' - X) + \Lambda H \Lambda' - \Lambda H \Lambda_0' - \\
&\quad \Lambda_0 H \Lambda' + \Lambda_0 H \Lambda_0' + B \\
&= (F\Lambda' - X)'(F\Lambda' - X) + \Lambda H (\Lambda' - \Lambda_0') - \\
&\quad \Lambda_0 H \Lambda' + \Lambda_0 H \Lambda_0' + B \\
&= (F\Lambda' - X)'(F\Lambda' - X) + (\Lambda - \Lambda_0) H (\Lambda - \Lambda_0)' + B \\
&= (F\Lambda' - X)'(F\Lambda' - X) + P_\Lambda
\end{aligned}
$$

A.4.4 Algebra on K

$$
K = I_N + (F\Lambda' - X) P_\Lambda^{-1} (F\Lambda' - X)'
$$

$$
\begin{aligned}
&= I_N + (F\Lambda' P_\Lambda^{-1} - X P_\Lambda^{-1})(\Lambda F' - X') \\
&= I_N + F\Lambda' P_\Lambda^{-1}(\Lambda F' - X') - X P_\Lambda^{-1}(\Lambda F' - X') \\
&= I_N + F(\Lambda' P_\Lambda^{-1}\Lambda)F' - F\Lambda' P_\Lambda^{-1} X' - X P_\Lambda^{-1}\Lambda F' + X P_\Lambda^{-1} X' \\
&= I_N + X P_\Lambda^{-1} X' + F[(\Lambda' P_\Lambda^{-1}\Lambda)F' - \Lambda' P_\Lambda^{-1} X'] - X P_\Lambda^{-1}\Lambda F' \\
&= I_N + X P_\Lambda^{-1} X' + F(\Lambda' P_\Lambda^{-1}\Lambda)[F' - (\Lambda' P_\Lambda^{-1}\Lambda)^{-1}\Lambda' P_\Lambda^{-1} X'] - X P_\Lambda^{-1}\Lambda F' \\
&= I_N + X P_\Lambda^{-1} X' + F(\Lambda' P_\Lambda^{-1}\Lambda)(F - F_0)' - X P_\Lambda^{-1}\Lambda F' \\
&= I_N + X P_\Lambda^{-1} X' + (F - F_0)(\Lambda' P_\Lambda^{-1}\Lambda)(F - F_0)' - X P_\Lambda^{-1}\Lambda F' + \\
&\quad F_0(\Lambda' P_\Lambda^{-1}\Lambda)(F - F_0)' \\
&= I_N + X P_\Lambda^{-1} X' + (F - F_0)(\Lambda' P_\Lambda^{-1}\Lambda)(F - F_0)' - X P_\Lambda^{-1}\Lambda F' + \\
&\quad F_0(\Lambda' P_\Lambda^{-1}\Lambda)F' - F_0(\Lambda' P_\Lambda^{-1}\Lambda)F_0' \\
&= I_N + X P_\Lambda^{-1} X' + (F - F_0)(\Lambda' P_\Lambda^{-1}\Lambda)(F - F_0)' - X P_\Lambda^{-1}\Lambda F' + \\
&\quad X P_\Lambda^{-1}\Lambda(\Lambda' P_\Lambda^{-1}\Lambda)^{-1}(\Lambda' P_\Lambda^{-1}\Lambda)F' - \\
&\quad X P_\Lambda^{-1}\Lambda(\Lambda' P_\Lambda^{-1}\Lambda)^{-1}(\Lambda' P_\Lambda^{-1}\Lambda)(\Lambda' P_\Lambda^{-1}\Lambda)^{-1}\Lambda' P_\Lambda^{-1} X' \\
&= I_N + X P_\Lambda^{-1} X' + (F - F_0)(\Lambda' P_\Lambda^{-1}\Lambda)(F - F_0)' - X P_\Lambda^{-1}\Lambda F' + \\
&\quad X P_\Lambda^{-1}\Lambda F' - X P_\Lambda^{-1}\Lambda(\Lambda' P_\Lambda^{-1}\Lambda)^{-1}\Lambda' P_\Lambda^{-1} X' \\
&= I_N + X P_\Lambda^{-1} X' + (F - F_0)(\Lambda' P_\Lambda^{-1}\Lambda)(F - F_0) - \\
&\quad (X P_\Lambda^{-1}\Lambda)(\Lambda' P_\Lambda^{-1}\Lambda)^{-1}(X P_\Lambda^{-1}\Lambda)' \\
&= Z + (F - F_0)(\Lambda' P_\Lambda^{-1}\Lambda)(F - F_0)
\end{aligned}
$$

B Numerical Results

The means for the 15 variables are

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
6.000	7.083	7.083	6.146	6.938	6.333	8.042	4.792	4.229	5.312	5.979	6.250	5.688	5.562	5.958

The variances for the 15 variables are

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
7.000	3.785	3.868	7.708	5.725	9.764	6.290	11.748	10.718	8.507	8.437	9.021	9.923	6.913	10.665

The correlation matrix, Table 2 is computed after standardizing the data.

The results of a principal components analysis of Table 2 which uses the Data from Table 1 are shown below.
Percent variance explained by each principal component

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
50.0%	13.7%	9.7%	8.0%	4.9%	3.3%	2.4%	2.1%	1.7%	1.3%	1.0%	0.6%	0.6%	0.4%	0.2%

Using the data from Table 1, and the appropriate prior information we obtain estimators of $\hat{F}$, $\hat{\Lambda}$, and $\hat{\Psi}$ which are as given in Tables 3, 4, and 5 from equations (2.14), (4.2), and (5.3).

Interval Estimates for Column N of F'

$\gamma = 69$

$$A = \begin{pmatrix} 0.058 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.392 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.134 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.200 \end{pmatrix}$$

$P_{22.1} = 0.0189$

$$\frac{A}{P_{22.1}} = \begin{pmatrix} 3.07 & 0.00 & 0.00 & 0.00 \\ 0.00 & 20.78 & 0.00 & 0.00 \\ 0.00 & 0.00 & 7.11 & 0.00 \\ 0.00 & 0.00 & 0.00 & 10.60 \end{pmatrix}$$

$\delta = 14$

$$\Sigma = \begin{pmatrix} 0.219 & 0.000 & 0.000 & 0.000 \\ 0.000 & 1.484 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.508 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.757 \end{pmatrix}$$

Square roots of diagonal elements of Σ, i.e. $\sqrt{\sigma_k^2}$, are:
0.468 1.218 0.713 0.870

$t_\delta(\alpha/2) = 2.145$

Credibility intervals are:
$f_{kN} \pm \sigma_k * t_\delta(\alpha/2)$

95% Credibility intervals for Row N of F

Upper CI of Row N of F:	-1.152	4.649	-1.001	1.115
Row N of $\hat{F}$:	-2.156	2.035	-2.529	-0.751
Lower CI of Row N of F:	-3.160	-0.578	-4.058	-2.617

Interval Estimates for Row p of Λ

$$E = \begin{pmatrix} 0.016 & -0.001 & -0.005 & -0.006 \\ -0.001 & 0.010 & -0.001 & 0.001 \\ -0.005 & -0.001 & 0.015 & 0.001 \\ -0.006 & 0.001 & 0.001 & 0.014 \end{pmatrix}$$

$V_{22.1} = 0.0103$

$$\frac{E}{V_{22.1}} = \begin{pmatrix} 1.524 & -0.088 & -0.523 & -0.581 \\ -0.088 & 0.993 & -0.144 & 0.065 \\ -0.523 & -0.144 & 1.424 & 0.081 \\ -0.581 & 0.065 & 0.081 & 1.340 \end{pmatrix}$$

$\eta = 51$

$$\Theta = \begin{pmatrix} 0.02988 & -0.00172 & -0.01025 & -0.01139 \\ -0.00172 & 0.01948 & -0.00282 & 0.00127 \\ -0.01025 & -0.00282 & 0.02792 & 0.00159 \\ -0.01139 & 0.00127 & 0.00159 & 0.02628 \end{pmatrix}$$

Square roots of diagonal elements of Θ, ie $\sqrt{\theta_k^2}$, are:

0.173 0.140 0.167 0.162

$t_\eta(\alpha/2) = 2.008$

Credibility intervals are:

$\lambda_{Nk} \pm \theta_k * t_\eta(\alpha/2)$

95% Credibility intervals for Row p of Λ

Upper CI of Row p of Λ:	0.475	0.266	1.012	0.337
Row p of $\hat{\Lambda}$:	0.128	-0.015	0.677	0.011
Lower CI of Row p of Λ:	-0.219	-0.295	0.341	-0.314

Large Sample Diagnostic

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