Journal of the Korean Statistical Society (2000), **29**: 1, pp 95–108

A Bayes Rule for Determining the Number of Common Factors in Oblique Factor Model

Hea-Jung Kim[1]

Abstract

Consider the oblique factor model $X = \Lambda f + \varepsilon$, with defining relation $\Sigma = \Lambda\Phi\Lambda' + \Psi$. This paper is concerned with suggesting an optimal Bayes criterion for determining the number of factors in the model, i.e. dimension of the vector f. The use of marginal likelihood as a method for calculating posterior probability of each model with given dimension is developed under a generalized conjugate prior. Then based on an appropriate loss function, a Bayes rule is developed by use of the posterior probabilities. It is shown that the approach is straightforward to specify distributionally and to implement computationally, with output readily adopted for constructing required criterion.

1. Introduction

The most widely used multivariate statistical model in the social and behavioral sciences involves linear structural relations among observed and latent variables. Linear structural equation models can be described as a class of models in which a p-variate observation X_i on subject i is presumed to be generated as $X_i = A\zeta_i$, $i = 1, \ldots, N$, where the matrix A is a function of a basic vector of parameters and the underlying k ($k \geq p$) generating variables ζ may represent measured, latent, or residual random or fixed variables (cf. Yuan and Bentler, 1997). Examples of such models include path analysis, factor analysis, errors-in-variable models, and simultaneous equations (see, e.g., Bollen and Long 1993; Byrne 1994; Hoyle 1995; and Kim 1998). Among them factor analysis attempts to simplify complex and diverse relationships that exist among a set

[1]Department of Statistics, Dongguk University, Seoul, 100-715, Korea

of observed variables by uncovering common dimensions or factors that link together the seemingly uncorrelated variables. To obtain the factor analysis model, let $A = (\Lambda, I)$ and $\zeta_i = (f_i', \varepsilon_i')'$. The linear latent variable structure becomes

$$X_i = \Lambda_m f_i + \varepsilon_i, \quad m < p, \quad i = 1, \ldots, N, \tag{1.1}$$

where Λ_m denotes a $p \times m$ matrix of constants called the factor loading matrix, f_i denotes $m \times 1$ factor score vector for subject i having $E(f_i) = 0$ and $\mathrm{Var}(f_i) = \Phi_m$. In the model ε_i's are assumed to be mutually uncorrelated and normally distributed as $N_p(0, \Psi_m)$ where Ψ_m is a diagonal matrix. The model (1.1) with $m \times 1$ factor vectors will be denoted by M_m in the sequel. Several authors suggested Bayesian analysis of the model under various prior assumption of the model, see, e.g., Kaufman and Press (1973; 1976), Wong (1980), Lee (1981), Euverman and Vermulst (1983), Shigemasu (1986) and Press and Shigemasu (1989). However, before using their methods, we need to choose the number of common factors to be included in the model, i.e. determination of m. Several criteria have been proposed for choosing m, the number of factors. Among them following five criteria are frequently used in practice (cf. Rencher 1995): (i) Scree test; (ii) Eigen value criterion; (iii) Communality criterion; (iv) Likelihood ratio criterion; (v) AIC by Akaike (1987). But these criteria may be criticized in that the first three criteria are informal and heuristically based criteria, while the last two criteria are formal but it is based upon asymptotic result and does not have a closed form expression.

Purpose of the present paper is to develop yet another criterion which gives formal, exact and a closed form criterion for determining the number of factors. The development involves calculation of posterior probability of each possible oblique factor model with given value of m via a Bayesian test criterion, called Bayes factor, and suggests a procedure to determine optimal value of m. The paper is constructed so that the basic model we are adopting is set out in Section 2. Section 3 develops the basic factor analysis model, introduces prior distributions on the parameters and calculates the joint posterior density of the parameters. Finally we find the marginal likelihood of the model. The procedures for calculating the Bayes factor and the posterior probability of each possible factor analysis model are discussed in Section 4, and it concludes the section by suggesting a Bayes rule that leads to the new criterion. Section 5 illustrates performance of the criterion on a real data example. Section 6 includes some concluding remarks.

2. Bayesian Approach for Model Selection

Suppose we have data $(X_1, \ldots, X_N) \equiv \mathbf{X}'$ assumed to have arisen under one of all possible models $M_1, \ldots, M_{p-1}$ having probability densities $P(\mathbf{X}|\Lambda_m, F_m, \Phi_m, \Psi_m)$ under M_m, $m = 1, \ldots, p-1$, where $F'_m = (f_1, \ldots, f_N)$ and m denotes the dimension of factor score vectors (i.e. the number of common factors) in (1.1). Given a prior distribution $\pi(\Lambda_m, F_m, \Phi_m, \Psi_m|M_m)$ for the parameters of each model, together with prior probability p_m of each model being true, the data produce the posterior probability of M_m being true as

$$P(M_m|\mathbf{X}) = \frac{P(\mathbf{X}|M_m)p_m}{\sum_{j=1}^{p-1} P(\mathbf{X}|M_j)p_j}, \quad m = 1, \ldots, p-1, \tag{2.1}$$

where the densities $P(\mathbf{X}|M_m)$ are obtained by integrating over the parameter spaces, so that

$$P(\mathbf{X}|M_m) = \int P(\mathbf{X}|\Lambda_m, F_m, \Phi_m, \Psi_m, M_m)\pi(\Lambda_m, F_m, \Phi_m, \Psi_m|M_m)\partial\Lambda_m\partial F_m\partial\Phi_m\partial\Psi_m. \tag{2.2}$$

The above equation is called the marginal likelihood of $\mathbf{X}$ under M_m. Using the marginal likelihood, Jeffrey's (1961) defined the Bayes factor for comparing M_m against $M_{m'}$, $m \neq m'$, such that

$$B_{mm'} = \frac{P(\mathbf{X}|M_m)}{P(\mathbf{X}|M_{m'})}. \tag{2.3}$$

The Bayes factor can be viewed as the ratio of the posterior odds of M_m to its prior odds, regardless of the value of the prior odds, and hence it is the weighted likelihood ratio of M_m to $M_{m'}$ giving a comparative support of the data for the two models (cf. Kass and Raftery, 1995; Kim, 1999). The posterior probability (2.1) that M_m is true is then expressed in terms of the Bayes factors

$$P(M_m|\mathbf{X}) = \left(\sum_{m'=1}^{p-1} \frac{p_{m'}}{p_m} B_{m'm} \right)^{-1}, \quad m = 1, \ldots, p-1, \tag{2.4}$$

where $B_{mm'} = 1/B_{m'm}$. The posterior model probabilities given by (2.4) will lead to a criterion for determining the number of common factors (i.e. determining m) in (1.1).

3. Marginal Likelihood of Oblique Factor Analysis Model

To recapitulate, define p-variate observation vectors, $(X_1, \ldots, X_N) \equiv \mathbf{X}'$, on N subjects. The means are assumed to have been subtracted out, so that $E(\mathbf{X}) = 0$. Under M_m, if we suppress the subscripts in Λ_m, F_m, Ψ_m and Φ_m, the factor analysis model can be written in matrix notation as

$$X_i = \Lambda f_i + \varepsilon_i,\ i = 1, \ldots, N,\ \ m < p, \tag{3.1}$$

where Λ denotes a $p \times m$ matrix of constants called the factor loading matrix; f_i denotes $m \times 1$ factor score vector for subject i; $F' \equiv (f_1, \ldots, f_N)$. The ε_i's are assumed to be mutually uncorrelated and normally distributed as $N_p(0, \Psi)$, $\Psi > 0$. If we assume (Λ, F, Ψ) are unobserved and fixed quantities, X_i is normal with mean Λf_i and covariance matrix Ψ. and hence the likelihood for (Λ, F, Ψ) is

$$P(\mathbf{X}|\Lambda, F, \Psi) = (2\pi)^{-Np/2}|\Psi|^{-N/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\Psi^{-1}(\mathbf{X} - F\Lambda')'(\mathbf{X} - F\Lambda')]\right\}. \tag{3.2}$$

3.1. Prior Specifications

The usual assumption about the oblique factor score vectors (f_i) is that they are independent and normally distributed , namely $N_m(0, \Phi)$, which is equivalent to a prior assumption about the oblique factor parameters, Φ. Now, since Φ is parameter of the oblique factors, one must introduce a prior distribution for its analysis. Thus the prior information is introduced in two stages: First, the conditional prior distribution of f_i given Φ is

$$f_i|\Phi \overset{iid}{\sim} N_m(0, \Phi),\ \ i = 1, \ldots, N, \tag{3.3}$$

which is the usual assumption about the oblique random factors, and we use a generalized natural conjugate family (see Press and Shigemasu (1989)) of prior distributions for (Λ, Ψ) so that

$$P(\Lambda|\Psi) = (2\pi)^{-mp/2}\frac{|H|^{p/2}}{|\Psi|^{m/2}}\exp\left\{-\frac{1}{2}\mathrm{tr}[(\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)'\Psi^{-1}]\right\}, \tag{3.4}$$

$$P(\Psi) = c_0(\nu, p)|\Psi|^{-\nu/2}|B|^{(\nu-p-1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\Psi^{-1}B]\right\}, \tag{3.5}$$

with B a diagonal matrix and $H > 0$ a symmetric matrix, where, for integer values α and β,

$$c_0^{-1}(\alpha,\beta) = 2^{(\alpha-\beta-1)\beta/2}\Gamma_\beta(\theta),$$

where $\theta = \alpha - \beta - 1$ and

$$\Gamma_\beta(\theta) = \pi^{\beta(\beta-1)/4}\prod_{j=1}^{\beta}\Gamma\left(\theta - \frac{j-1}{2}\right)$$

is the β dimensional gamma function. Secondly, assume secondary parameter Φ is Inverted Wishart with scale parameter H and degrees of freedom κ so that

$$\Phi \sim W^{-1}(H, m, \kappa), \quad \kappa > 2m, \tag{3.6}$$

As will be seen, this form of the prior specification produces analytically tractable marginal likelihood for $\mathbf{X}$.

Thus, Ψ follows an Inverted 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 assessed. Note that $E(F'F) \propto H$ leads to the scale parameter in (3.4) and $E(\Psi|B)$ is a diagonal to represent traditional views of the factor model containing "common" and "specific" factors. Also note that if $\Lambda \equiv (\lambda_1', \dots, \lambda_m')'$, then $\text{var}(\Lambda|\Psi) = \Psi \otimes H^{-1}$, and $\text{cov}[(\lambda_i, \lambda_j)|\Psi] = \Psi_{ij}H^{-1}$. In case the model is for the orthogonal factors, we may take $H = n_0 I_m$, for preassigned scalar $n_0 = \kappa - 2m - 2$, so that $\text{var}(f_i) = I_m$. These interpretations of the hyperparameters will simplify assessment.

3.2. Marginal Likelihood of X

Combining (3.2) through (3.6), we obtain the joint posterior density of the parameters

$$P(\Lambda, F, \Phi, \Psi|\mathbf{X}, M_m) \propto C|\Psi|^{-(N+m+\nu)/2}\exp\left\{-\frac{1}{2}\text{tr}[\Psi^{-1}G]\right\}P(F, \Phi), \tag{3.7}$$

where

$$\begin{aligned}
C &= c_0(\nu, p)(2\pi)^{-(m+N)p/2}|H|^{p/2}|B|^{(\nu-p-1)/2},\\
P(F,\Phi) &= (2\pi)^{-mN/2}|\Phi|^{-N/2}\exp\left\{-\frac{1}{2}\text{tr}[\Phi^{-1}F'F]\right\}\\
&\times\ c_0(\kappa, m)|H|^{(\kappa-m-1)/2}|\Phi|^{-\kappa/2}\exp\left\{-\frac{1}{2}tr[\Phi^{-1}H]\right\},\\
G &= (\mathbf{X} - F\Lambda')'(\mathbf{X} - F\Lambda') + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B.
\end{aligned}$$

Lemma 3.1 (Press and Shigemasu (1989)). For G, given in (3.7)

$$G = R_F + (\Lambda - \Lambda_F)Q_F(\Lambda - \Lambda_F)', \tag{3.8}$$

where

$$\begin{aligned} Q_F &\equiv H + F'F, \\ R_F &\equiv \mathbf{X}'\mathbf{X} + B + \Lambda_0 H \Lambda_0' - (\mathbf{X}'F + \Lambda_0 H)Q_F^{-1}(\mathbf{X}'F + \Lambda_0 H)', \\ \Lambda_F &\equiv (\mathbf{X}'F + \Lambda_0 H)(H + F'F)^{-1}. \end{aligned}$$

Lemma 3.2. For Q_F and R_F in Lemma 3.1,

$$|Q_F||R_F| = |W||A + (F - \hat{F})'(I_N - \mathbf{X}W^{-1}\mathbf{X}')(F - \hat{F})|, \tag{3.9}$$

where

$$\begin{aligned} W &\equiv \mathbf{X}'\mathbf{X} + B + \Lambda_0 H \Lambda_0', \\ A &\equiv H - H'\Lambda_0' W^{-1}\Lambda_0 H - (H'\Lambda_0' W^{-1}\mathbf{X}')(I_N - \mathbf{X}W^{-1}\mathbf{X}')^{-1}(H'\Lambda_0' W^{-1}\mathbf{X}')', \\ \hat{F} &\equiv (I_N - \mathbf{X}W^{-1}\mathbf{X}')^{-1}\mathbf{X}W^{-1}\Lambda_0 H. \end{aligned}$$

Proof. Under the same notations,

$$\begin{aligned} |Q_F||R_F||W|^{-1} &= |Q_F||I_p - (H'\Lambda_0' + F'\mathbf{X})'(H + F'F)^{-1}(H'\Lambda_0' + F'\mathbf{X})W^{-1}| \\ &= |Q_F - (H'\Lambda_0' + F'\mathbf{X})W^{-1}(H'\Lambda_0' + F'\mathbf{X})'| \\ &= |A + (F - \hat{F})'(I_N - \mathbf{X}W^{-1}\mathbf{X}')(F - \hat{F})|. \end{aligned}$$

Multiplying $|W|$ on both sides of the equation, we have the result.

Theorem 3.1. Given the model (3.1), say M_m having m common factors, if we set $\kappa = m + \nu - 2p$, the marginal likelihood of $\mathbf{X}$ under M_m is given by

$$P(\mathbf{X}|M_m) = \Delta_m |\mathbf{X}'\mathbf{X} + B + \Lambda_0 H \Lambda_0'|^{-(\gamma - m)/2} |I_N - \mathbf{X}W^{-1}\mathbf{X}'|^{-m/2}, \tag{3.10}$$

where

$$\begin{aligned} \Delta_m &= \pi^{-Np/2}|H|^{(\gamma-N-m)/2}|B|^{(\gamma-N-m)/2}|A|^{-(\gamma-N-m)/2} \\ &\times \frac{\Gamma_p\left(\frac{\gamma-m}{2}\right)}{\Gamma_p\left(\frac{\gamma-N-m}{2}\right)} \frac{\Gamma_m\left(\frac{\gamma-N-m}{2}\right)}{\Gamma_m\left(\frac{\gamma-m}{2}\right)} \frac{\Gamma_m\left(\frac{\gamma-m-p}{2}\right)}{\Gamma_m\left(\frac{\gamma-N-m-p}{2}\right)}, \\ \gamma &= N + m + \nu - p - 1, \end{aligned}$$

Proof. The marginal likelihood is obtained by integrating the right hand side of (3.7) over the parameter spaces, so that

$$P(\mathbf{X}|M_m) = C\int P(F,\Phi)|\Psi|^{-(N+m+\nu)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\Psi^{-1}G]\right\}\partial\Psi\partial\Lambda\partial\Phi\partial F.$$

Integrating with respect Ψ, and using properties of the Inverted Wishart density gives the marginal density of (Λ, F), and then arrange the remaining integral with respect to Λ. This can be accomplished by factoring G (using Lemma 3.1) into a form which makes it transparent that in terms of Λ, the density is proportional to a matrix T-density

$$P(\mathbf{X}|M_m) = C_1\int \frac{P(F,\Phi)}{|R_F + (\Lambda-\Lambda_F)Q_F(\Lambda-\Lambda_F)'|^{\gamma/2}}\partial\Lambda\partial\Phi\partial F,$$

where $C_1 = C/c_0(\gamma+p+1,p)$. This is readily integrated with respect to Λ by using the normalizing constant of the matrix T-distribution (cf. Dickey 1967) and applying Lemma 3.2 to give the joint posterior density of Φ and F. Then integrating the joint posterior density with respect to Φ, using the normalizing constant of the Inverted Wishart distribution, we have, under the condition that $\kappa = m+\nu-2p$,

$$\begin{aligned} P(\mathbf{X}|M_m) &= \pi^{-N(p+m)/2}|B|^{(\gamma-N-m)/2}|H|^{-(\gamma-N-m)/2}|\mathbf{X}'\mathbf{X}+B+\Lambda_0 H\Lambda_0'|^{-(\gamma-m)/2} \\ &\times \frac{\Gamma_p\left(\frac{\gamma-m}{2}\right)}{\Gamma_p\left(\frac{\gamma-N-m}{2}\right)}\frac{\Gamma_m\left(\frac{\gamma-m-p}{2}\right)}{\Gamma_m\left(\frac{\gamma-N-m-p}{2}\right)} \\ &\times \int |A+(F-\hat{F})'(I_N-\mathbf{X}W^{-1}\mathbf{X}')(F-\hat{F})|^{(\gamma-m)/2}\partial F. \end{aligned}$$

Integrating it with respect to F using the normalizing constants of the matrix T-distribution, we have the result.

Note 1. $|I_N - \mathbf{X}W^{-1}\mathbf{X}'|$ and $(I_N - \mathbf{X}W^{-1}\mathbf{X}')^{-1}$ are equivalent to $|I_p - W^{-1}\mathbf{X}'\mathbf{X}|$ and $(I_N - \mathbf{X}(\mathbf{X}'\mathbf{X}-W)^{-1}\mathbf{X}')$, respectively. The latter representations are more convenient for numerical calculations than the former, because we need only invert matrices of order p, instead of those of order N.

4. New Criterion

The marginal likelihood is a summary of the evidence provided by a data set represented by a statistical model (i.e. Factor analysis model). When several alternative factor analysis models M_m's ($m = 1, \ldots, p-1$) are being considered, their marginal likelihoods yield posterior probability of each model. Suppose that p-dimensional oblique factor analysis models, $M_1, M_2, \ldots, M_{p-1}$, are being considered. Each of $M_1, \ldots, M_{p-1}$ is compared in turn with M_m, yielding Bayes factors $B_{1m}, \ldots, B_{(p-1)m}$, $B_{mm} \equiv 1$. Then the posterior probability of M_m is given by (2.4). Thus Theorem 3.1 and (2.4) yields the following posterior probability of M_m.

Lemma 4.1. Restoring the subscripts for (Λ, F, Ψ) and corresponding hyperparameters (Λ_0, H, B) and taking all the prior odds $p_m/p_{m'}$ equal to 1, for $m' = 1, \ldots, p-1$, and $1 \le m \le p-1$, we have the posterior probability of M_m, $P(M_m|\mathbf{X}) =$

$$\left(\sum_{m'=1}^{p-1} \frac{\Delta_{m'} |\mathbf{X}'\mathbf{X} + B_{m'} + \Lambda_{m'0} H_{m'} \Lambda'_{m'0}|^{-(\gamma-m')/2} |I_N - \mathbf{X} W_{m'}^{-1} \mathbf{X}'|^{-m'/2}}{\Delta_m |\mathbf{X}'\mathbf{X} + B_m + \Lambda_{m0} H_m \Lambda'_{m0}|^{-(\gamma-m)/2} |I_N - \mathbf{X} W_m^{-1} \mathbf{X}'|^{-m/2}} \right)^{-1}, \tag{4.1}$$

where $(\Lambda_{m'0}, H_{m'}, B_{m'})$ and (Λ_{m0}, H_m, B_m) denote corresponding hyperparameters of $(\Lambda_{m'}, F_{m'}, \Psi_{m'})$ in $M_{m'}$ and (Λ_m, F_m, Ψ_m) in M_m, respectively.

Proof. Direct application of the marginal likelihood in Theorem 3.1 to (2.4) gives the result.

Taking all the prior odds $p_m/p_{m'}$ equal to 1 is natural choice, but other values may be used to reflect prior information about the relative plausibility of competing models. In the present paper our aim is focused on the selection of a single best model in the presence of a set of rival models. Lemma 4.1 gives an optimal rule for selecting the best fitting factor analysis model. Suppose we use the marginal $P(\mathbf{X}|M_{m'})$, $m' = 1, \ldots, p-1$, for selecting the factor analysis model, and suppose $p_{m'}$ is the prior probability of $M_{m'}$ being the true model, then the risk incurred in choosing M_m as the best fitting factor analysis model is

$$Risk(M_m|\mathbf{X}) = \frac{\sum_{m'=1}^{p-1} L(M_{m'}, M_m) P(\mathbf{X}|M_{m'}) p_{m'}}{\sum_{m'=1}^{p-1} P(\mathbf{X}|M_{m'}) p_{m'}}, \tag{4.2}$$

where $L(M_{m'}, M_m)$ is the cost or loss associated with the model selection error. Let assume the special but commonly used loss function;

$$L(M_{m'}, M_m) = 1 - \delta(M_{m'}, M_m), \tag{4.3}$$

where $\delta(M_{m'}, M_m) = 1$ if $M_{m'} = M_m$, otherwise it is zero. Then we have the following theorem.

Theorem 4.1. A Bayes rule for choosing the number of common factors in the factor analysis model is to choose m (i.e. M_m) if

$$P(M_m|\mathbf{X}) = Max \;\; P(M_{m'}|\mathbf{X}), \;\; m' = 1, \dots, p-1. \tag{4.4}$$

Proof. The risk incurred in choosing the best fitting factor analysis model is (4.2). This can be minimized by choosing M_m that minimizes the numerator in (4.2). Minimizing the numerator, we have $P(\mathbf{X}|M_m)p_m = Max \;\; P(\mathbf{X}|M_{m'})p_{m'}, \;\; m' = 1, \dots, p-1$. Applying Bayes theorem to this gives (4.4).

The rule resulting from choosing M_m to minimize $Risk(M_m|\mathbf{X})$ in (4.2) is known as the Bayes rule, and it achieves minimal choice risk among all possible models based on the posterior probabilities $P(M_{m'}|\mathbf{X})$.

5. An Illustrative Example

We have extracted the data reported in Dillon and Goldstein (1984), and have used it to illustrate the new criterion for determining optimal number of factors. Table 1 gives raw data on 10 characteristics for 14 selected nations. The characteristics are

X_1: GNP per Capita($)	X_2: Trade (Millions of $)
X_3: Power (Rank)	X_4: Stability
X_5: Freedom of Group Opposition	X_6: Foreign Conflict
X_7: Agreement with U.S. in U.N.	X_8: Defence Budget (Millions of $)
X_9: GNP for Defense (%)	X_{10}: Acceptance of International Law.

Due to different measurement units, we standardized the data for the model selection. The hyperparameter H_m was assessed as $H_m = I_m + .1(J(m,m) - I_m)$, where $J(m,m)$ is a $m \times m$ matrix with all elements equal to one. We took $\nu_m = 2p + m + 3$ to reflect minimal prior knowledge but to permit $E\Psi_m$ and $E\Phi_m$ $(m = 1, \dots, 9)$ to exist. The prior distribution of Ψ_m and Λ_m were assessed with $B_m = .5(m+1)I_{10}$ and $\Lambda_{m0} = (0.5/(m(1 + .1(m-1))))^{1/2}J(p, m)$, so that $E[\Lambda]E[\Phi]E[\Lambda]' + E[\Psi] = \Lambda_{m0}H\Lambda_{m0}' + B_m/(m+1)$ has a correlation matrix pattern. This assessment is consistent with the defining relation of the oblique factor model. Our program using SAS/IML calculates values of the marginal likelihood and posterior probability of each model M_m. Those are noted in Table 2. It points out that the new criterion attains maximum posterior probability $(*)$ for

TABLE I. Sample Data

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Brazil	91	2,729	7	0	2	0	69.1	148	2.8	0
Bruma	51	407	4	0	1	0	-9.5	74	6.9	0
China	58	349	11	0	0	1	41.7	3,054	8.7	0
Egypt	134	928	5	1	1	1	-15.4	158	6.0	1
India	70	2,689	10	0	2	0	-28.6	410	1.9	1
Indonesia	129	1,601	8	0	1	0	-21.4	267	6.7	0
Israel	515	415	2	1	2	1	42.9	33	2.7	1
Jordan	70	83	1	0	1	1	8.3	29	25.7	0
Netherlands	707	5,395	6	1	2	0	52.3	468	6.1	1
Poland	468	1,852	9	0	0	1	-41.7	220	1.5	0
U.S.S.R.	746	6,530	13	1	0	1	-41.7	34,000	20.4	0
U.K.	998	18,667	12	1	2	1	69.0	3,934	7.8	0
U.S.	2,334	26,836	14	1	2	1	100	40,641	12.2	1

four factors; so we should conclude that a four common factor model M_4 is best fitted to the data. In our present example, the same progam with $H_m = I_m + .05(J(m,m) - I_m)$ and $H_m = I_m + .2(J(m,m) - I_m)$ yield posterior probabilities of M_4 as .8587 and .7984, respectively. This notes that, in selecting the best fitted model, the suggested method is insensitive with respect to small difference in assessment of values in H_m. For the reference, AIC's for models with different numbers of factors are also calculated via SAS/STAT PROC FACTOR (cf. Ray 1982) and listed in the table. In this numerical example, AIC presents the value calculated from the maximum likelihood factor method (for more than 7 common factors minimum eigen value criterion stopped calculating AIC). More than 5 common factor analysis generates the message, "Warning: Two many factors for a unique solution." The parameter in the model exceeds the number of elements in the correlation matrix from which they can be estimated, so an infinite number of different perfect solution can be estimated. The degrees of freedom for the chi-square test (LRT) are nonpositive for $m > 5$, so that probability levels cannot be computed for $m(> 5)$ factor models. The probability levels for the chi-square test are .0001 for the hypothesis of no common factors, .0806 for one common factor, and .3783 for two common factors. Akaike's information criterion attains their minimum values at two common factors, so there is little doubt that M_2 is appropriate for the data. However, Dillon and Goldstein (1984, p. 95) notes that four common factors are appropriate for the data set. This confirms the statement that AIC and the chi-square test tend to include common factors that are statistically significant but inconsequential for practice (cf. Ray 1982, p. 328). Thus, the new criterion can be viewed as a good surrogate for the impractical classical criteria in a sense that the criterion gives the same conclusion with that of Dillon and Goldstein. Also note that when our observational data is augmented by proper prior information, as in this example, the identification-of-parameter problem of the classical criteria disappears.

TABLE 2. Selection of Common Factors
("-" denotes that corresponding value is not available)

Number of Factors	$P(\mathbf{X}\|M_m) \times 10^{84}$	$P(M_m\|\mathbf{X})$	AIC	LRT(p-value)
1	0.0019	.0000	4.5893	.0806
2	29.474	.0042	-4.1528	.3782
3	1023.271	.1464	-0.6433	.4178
4	5938.575	.8493*	-0.9026	.5297
5	0.0005	.0000	-0.0375	.5189
6	0.0002	.0000	2.3697	-
7	0.0001	.0000	8.0011	-
8	0.0001	.0000	-	-
9	0.0001	.0000	-	-

6. Concluding Remarks

We have suggested a criterion for selecting the number of common factors in oblique factor analysis. The development is pertaining to deriving the posterior probability of each factor analysis model with given number of factors m and constructing a Bayes rule for optimal selection of m. The appeal of this criterion is that it is not only optimal but a closed form which has not been available yet, and hence it is easy to apply for choosing the number of common factors to be included in the oblique factor model. Moreover it is free from the identification-of-parameter problem of the classical criteria (LRT and AIC) and provides probabilistic considerations for selecting promising number of common factors so that one may compare the factor extractions among the highly plausible models.

The posterior probabilities of m-factor ($m = 1, \dots, p-1$) models are calculated by use of corresponding marginal likelihoods. To derive the marginal likelihoods we use a natural conjugate priors for the parameters that uses a generalized natural conjugate family (cf. Press 1982) of prior distributions for (Λ, Ψ). Once we get the best fitted model having optimal m, we may proceed further analysis for the selected model such as point and interval estimate of factor scores, factor loadings and specific variances. A study pertaining to those analyses is straightforward and is left as a future study of interest.

References

[**1**] Ray, A. A. (Editor) (1982). *SAS User's Guide: Statistics*, 1982 edition, SAS Instititute Inc., North Carolina.

[**2**] Akaike, H. (1987). Factor analysis and AIC, *Psychometrika*, Vol. 52, No. 3, 317-332.

[**3**] Bollen, K. A. and Long, J. S. (1993). *Testing Structural Equation Model*, Newbury Park, CA:Sage.

[**4**] Byrne, B. M. (1994). *Structural Equation Modeling with EQS and EQS/Windows*, Thousand Oaks, CA:Sage.

[**5**] Dickey, J. M. (1967). Multivariate generalization of the multivariate t-distribution and the inverted multivariate t-distribution, *Annals of Mathematical Statistics*, Vol. 38, 511-518.

[**6**] Dillon, W. R. and Goldstein, M. (1984). *Multivariate Analysis: Methods and Applications*, Wiley, New York.

[**7**] Jeffreys, H. (1961). *Theory of Probability*, 3rd ed., Oxford, U. K.:Oxford University Press.

[**8**] Euverman, T. J. and Vermulst, A. A. (1983). Bayesian factor analysis, *Department of Social Science, State University of Gorningen*, the Netherlands.

[**9**] Hoyle, R. (1995). *Structural Equation Modeling: Concepts, Issues and Applications*, Thousand Oaks, CA:Sage.

[**10**] Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the American Statistical Association*, Vol. 90, 773-795.

[11] Kaufman, G. M. and Press, S. J. (1973). Bayesian factor analysis, Report No. 7322, *Bulletin of Institute of Mathematical Statistics*, Vol. 2, Issue No. 7.

[12] 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.

[13] Kim, H. J. (1998). Variable selection in linear random effects models for normal data, *Journal of the Korean Statistical Society*, Vol. 27, NO. 4, 407-420.

[14] Kim, H. J. (1999). A method for testing nested point null hypotheses using multiple Bayes factor, *Annals of the Institute of Statistical Mathematics*, Vol. 51, NO. 3, 585-602.

[15] Lee, S. Y. (1981). A Bayesian approach to confirmatory factor analysis, *Psychometrika*, Vol. 46, No. 2, 153-159.

[16] Press, S. J. (1982). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods for Inference*, Robert E. Krieger Publishing Co., Malabar, Florida.

[17] Press, S. J. and Shigemasu, K. (1989). Bayesian inference in factor analysis, *Contributions to Probability and Statistics: Assays in Honor of Ingram Olkin*, Eds. Gleser, L. J., Perlman, M. D., Press, S. J., and Sampson, A. R., Springer-Verlag, New York, 271-287.

[18] Rencher, A. C. (1995). *Methods of Multivariate Analysis*, John Wiley and Sons, INC., New York.

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

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

[21] Yuan, H. E. and Bentler, P. M. (1997). Mean and covariance structure analysis: theoretical and practical improvements, *Journal of the American Statistical Association*, Vol. 92, 767-774.