KM 21(1986)
pag 19 - 36

CONFIRMATORY BAYESIAN FACTOR ANALYSIS [1)]

Teije J. Euverman [2)]
Ad A. Vermulst [3)]
Charles Lewis [4)]

Summary

Confirmatory factor analysis is considered from a Bayesian viewpoint with prior information based on substantive theoretical knowledge. A reparameterized factor model is used for the parameters of which prior densities may be specified using an interactive computer program. The posterior density may then be optimized using an iterative algorithm to obtain Bayesian modal estimates. A numerical example from the literature is presented for illustrative purposes.

Keywords: CONFIRMATORY FACTOR ANALYSIS; BAYESIAN STATISTICS; INFORMATIVE PRIOR DENSITIES; POSTERIOR DENSITIES; INTERACTIVE COMPUTER PROGRAM; OPTIMIZATION COMPUTER PROGRAM

The authors are indebted to W. Molenaar and J.M.F. ten Berge for their valuable suggestions for improvement.

---

1) Naar aanleiding van een voordracht op 12 april 1985 door A.Vermulst op de statistische dagen van de VVS.

2) Rijksuniversiteit Utrecht, vakgroep Psychonomie, sektie methoden, Heidelberglaan 1, 3508 TC Utrecht. Tel. 030-534594.
3) Katholieke Universiteit Nijmegen, vakgroep WHEP, Erasmusplein 1, 6500 HD Nijmegen. Tel. 080-516283.
4) Rijksuniversiteit Groningen, vakgroep Statistiek en Meettheorie der F.S.W., Oude Boteringestraat 23, 9712 GC Groningen. Tel. 050-636188.

1. *Introduction*

In the past few decades, computational tools such as LISREL (Jöreskog & Sörbom, 1984) have become available which allow investigators to test hypotheses regarding the covariance structure of sets of variables. Within the context of factor analysis, this development has had the positive effect of allowing the emphasis to shift from exploratory to confirmatory studies, where substantive theory is given a formal role in the analysis.

In this paper, we discuss two modifications of the standard approach (neither of which is novel, although the combination has not been previously considered) which are intended to further support the practice of confirmatory factor analysis. The first modification is a reparameterization of the usual model, with factor-variable correlations replacing raw factor regression weights as the primary description of the relations between factors and observable variables. The "structure vs pattern" issue is an old one in factor analysis (see, for instance, Harman, 1967), but it is our experience that investigators in the social sciences are more capable of expressing hypotheses about relationships in terms of correlations than in terms of regression coefficients.

The second modification of the standard approach which we consider is the replacement of a sampling theory interpretation of statistical inference by a Bayesian one. (See Mayekawa, 1985, for a review of other Bayesian treatments of factor analysis.) An advantage of this approach for confirmatory analysis is that it replaces the sampling theory idea of hypothesis <u>testing</u> by that of hypothesis <u>revision</u>, based on Bayes' theorem. Thus substantive theory should provide the basis for a prior distribution of the parameters of the factor model. This prior is combined with sample information, resulting in a posterior distribution which can then be compared with the prior to see where the greatest changes have taken place.

In section 2 the reparameterized factor model will be presented. Prior densities for the new parameters will be discussed in section 3, and in section 4 procedures for eliciting prior information will be described. The authors have written an interactive computer program, BAYFAC-I (Bayesian Factor Analysis-Interactive), which incorporates these procedures. In section 5 the posterior density is given, for which the joint modal estimates will be regarded as providing Bayesian estimates. These may be obtained by means of an iterative program, BAYFAC-O (Bayesian Factor Analysis-Optimization), written by the authors. An example is given in section 6. The present article ends with a discussion in section 7.

2. *The reparameterized factor model*

The basic factor model is

$$\underline{x} = \Lambda\underline{f} + \underline{u} + \underline{\mu} \;, \tag{1}$$

where $\underline{x}$ is a random vector of p variables, Λ is a p*k matrix of factor loadings, $\underline{f}$ is a random vector of k common factors, $\underline{u}$ is a random vector of p unique factors an $\underline{\mu}$ is a vector of means with p constant values. The usually stated assumptions are

$$\begin{bmatrix} \underline{f} \\ \\ \underline{u} \end{bmatrix} \sim N_{k+p}\left\{ \begin{bmatrix} \underline{0} \\ \\ \underline{0} \end{bmatrix} \;;\; \left[\begin{array}{c:c} \Phi & 0 \\ \hdashline 0 & \Psi \end{array}\right] \right\} \tag{2}$$

where Φ is a k*k positive definite factor correlation matrix and Ψ a p*p diagonal matrix with positive unique variances.
Then, defining Σ as $\Lambda\Phi\Lambda' + \Psi$,

$$\underline{x} \sim N_p(\underline{\mu},\Sigma) \;. \tag{3}$$

The relevant factor parameters are contained in the matrices Λ, Φ and Ψ.

The decomposition of Σ will now be considered. If $(Diag(\Sigma))^{\frac{1}{2}}$ is defined as Δ, then Σ can be written as follows:

$$\Sigma = \Delta\Delta^{-1}(\Lambda\Phi\Lambda' + \Psi)\Delta^{-1}\Delta = \Delta(\Delta^{-1}\Lambda\Phi\Lambda'\Delta^{-1} + \Delta^{-1}\Psi\Delta^{-1})\Delta \;, \tag{4}$$

where the last matrix in parentheses is the p*p correlation matrix of the variables. Because $Diag(\Delta^{-1}\Lambda\Phi\Lambda'\Delta^{-1} + \Delta^{-1}\Psi\Delta^{-1}) = I$ and $Diag(\Delta^{-1}\Lambda\Phi\Lambda'\Delta^{-1})$ and $\Delta^{-1}\Psi\Delta^{-1}$ are positive definite, the diagonal elements of $\Delta^{-1}\Lambda\Phi\Lambda'\Delta^{-1}$ and $\Delta^{-1}\Psi\Delta^{-1}$ are between zero and one. The diagonal matrix $\Delta^{-1}\Psi\Delta^{-1}$ will be denoted by Ψ^{*} and refered to as the matrix with proportions unique variances. The correlations between the variables and factors can be expressed with the matrices Δ,Λ and Φ. This can be seen as follows:

$$COV(\underline{x},\underline{f}) \equiv E((\underline{x} - \underline{\mu})\underline{f}') = E((\Lambda\underline{f} + \underline{u})\underline{f}') = E(\Lambda\underline{f}\underline{f}') = \Lambda\Phi \quad , \tag{5}$$

where (2) is used. So the p*k matrix $\Delta^{-1}\Lambda\Phi$ contains as its ijth element the correlation between variable $i(i=1,\ldots,p)$ and factor $j(j=1,\ldots,k)$. If $\Lambda^{-1}\Lambda\Phi$ is denoted by Ω, then (4) can be written as follows:

$$\Sigma = \Delta(\Omega\Phi^{-1}\Omega' + \Psi^{*})\Delta = \Delta(\Omega\Phi^{-1}\Omega' + I\text{-}Diag(\Omega\Phi^{-1}\Omega'))\Delta \quad , \tag{6}$$

because $Diag(\Omega\Phi^{-1}\Omega' + \Psi^{*}) = I$. The covariance matrix Σ is now expressed in terms of the standard deviations of the variables (Δ), the variable-factor correlations (Ω) and the factor-factor correlations (Φ). In the following section prior densities will be specified for the elements in Δ, Ω and Φ.

3. *Prior specified densities*

Arranging the diagonal elements $\delta_i(i=1,\ldots,p)$ of Δ in a p-dimensional random vector $\underline{\delta}$, it is assumed that its prior density is the product of p inverse chi densities, i.e.

$$p(\underline{\delta}) \propto \prod_{i=1}^{p} \delta_i^{-(\nu_i+1)} \; exp[-\tfrac{1}{2}\, \frac{\nu_i c_i^2}{\delta_i^2}] \;,\; 0 < \delta_i < \infty \quad . \tag{7}$$

In (7) c_i an ν_i respectively are the "scale factor" and the number of degrees of freedom, associated with δ_i. The inverse chi distribution is the natural conjugate for the standard deviation in univariate normal Bayesian analysis (see, for instance, Novick & Jackson, 1974, p. 195). The prior modal value of δ_i, say $\bar{\delta}_i$, and ν_i are determined in BAYFAC-I, through which c_i can be obtained using the expression for $\bar{\delta}_i$:

$$\bar{\delta}_i = c_i\left(\frac{\nu_i}{\nu_i+1}\right)^{\frac{1}{2}} \quad . \tag{8}$$

Arranging the elements of Ω row-wise in a pk-dimensional vector $\underline{\omega}$ and, similarly, the elements of the strict upper triangle of Φ in a $\frac{1}{2}k(k-1)$-dimensional vector $\underline{\phi}$ it is assumed, for reasons discussed below, that

$$\underline{\omega} \sim N(\bar{\underline{\omega}},\Pi) \quad \text{(a)} \qquad \text{and} \qquad \underline{\phi} \sim N(\bar{\underline{\phi}},Z) \quad \text{(b)} \;, \tag{9}$$

where Π and Z are positive definite diagonal matrices. It is furthermore assumed that the distributions in (7) and (9) are independent. In BAYFAC-I values are assigned to the parameters of the normal densities in (9).

It should be noted with regard to (9b) that when an orthogonal model is assumed a priori, it is in fact supposed that all probability mass for the elements in $\underline{\phi}$ is fixed at zero values. It is then implicitly believed that sample information cannot weaken this situation. In this particular case the factor model is considered conditionally on $\Phi = I$, an option which is available in BAYFAC-I.

Finally, for $\underline{\mu}$ the improper uniform density, assumed independent of (7) and (9) is used:

$$p(\underline{\mu}) \propto c \quad , \tag{10}$$

where c is a constant (see Press, 1972, p. 71). Obviously it is possible that a user has knowledge regarding this parameter but, for reasons of computational and conceptual simplicity, this knowledge is not used for estimating the factor parameters.

Independent normal priors have been chosen in (9) primarily as an aid to the researcher who must provide information about $\underline{\omega}$ and $\underline{\phi}$. They clearly must be considered approximations to the "true" priors, given the restricted range of the correlation coefficient and the necessity of positive semidefinite correlation matrices. Regarding the first of these points, one might consider applying the standard Fisher-Z transformation to the elements of $\underline{\omega}$ and $\underline{\phi}$ and assuming prior normality for the transformed values. While this is a line which the authors intend to pursue as part of a more general study of possible priors, the choice of working with untransformed correlations was made here with the goal in mind of keeping interpretational problems for the researcher to a minimum. The normal density is the most familiar one to many researchers and one whose two distributional parameters are conceptually meaningful. An obvious advantage of using a symmetric density is the equality of the mean, median and mode, eliminating an indeterminacy in the choice of which characteristic is the best representation of the prior beliefs. Moreover, it is often plausible that an investigator has some best value in mind for a correlation, around which a symmetric interval is constructed.

The assumption of joint prior independence among all parameters has also been made primarily for practical reasons. First it is unreasonable to demand that an investigator, in addition to providing information about the marginal distributions of all parameters in the model under consideration, also describe all dependencies among these parameters. Second, in applying Bayes' theorem, the prior is combined with a likelihood function which will, among other things, introduce meaningful dependencies in the resulting posterior.

4. Assessment and correction of the prior information

The program BAYFAC-I is developed to assist in the assessment of the relevant prior information with regard to the prior densities. The ultimate result will be an external file with the characteristics of these densities, which later can be linked to a file with sample data by using BAYFAC-O, with which posterior estimates can be obtained. It should be emphasized that, as in all Bayesian analysis, the prior must be specified before the current sample data have been examined. Otherwise, the basic assumption of independence of prior and sample information will almost certainly be violated, and Bayes' theorem will no longer be appropiate for combining the two sources.

The prior information will be of a more-or-less subjective nature, depending on the degree of knowledge and the kind of information which an investigator uses. The inexperienced user is the presupposed conversational partner in BAYFAC-I. A more experienced user or one who uses more-or-less objective information from previous analyses may skip parts of the interactive interrogative sequence by entering the required information directly. The relevant prior information which is obtained by BAYFAC-I are values for $\bar{\delta}_i$, ν_i (see section 4.1) and for the elements in $\bar{\underline{\omega}}$, $\bar{\underline{\phi}}$, Π and Z (see section 4.2). In the subsequent sections a concise description of the basic underlying reasoning will be given.

4.1 Determination of $\bar{\delta}_i$ and ν_i in (7)

To be able to give judgments about $\bar{\delta}_i$, knowledge is necessary regarding the spread of the variable x_i. Starting from assumption (3), the investigator is asked to imagine a normally distributed variable and is then requested to answer questions through which a prior modal estimate is obtained.
Based on robustness considerations, a slight deviation from normality is acceptable in BAYFAC-I. However, if an investigator has variables for which a normal density is completely inappropriate, even as an approximation, this Bayesian Factor Analysis is not recommended.

The following basic idea is relevant. Given the positive quantile z of the standard normal density associated with a fixed HDR-percentage, the interval length is 2z and the standard deviation δ in a normal density with interval length ℓ is then

$$\delta = \frac{\ell}{2z} \quad . \tag{11}$$

So, if estimates are available of ℓ and of the HDR-percentage (translated to

a value of z), $\bar{\delta}$ can be calculated.
In BAYFAC-I, this idea is used several times. The way in which modal estimates are obtained for δ_i is dependent on the kind of the variable x_i in question:

Type 1 variables are defined as scales with 9 or fewer (discrete) response categories (e.g. scores on test items in survey questionnaires). For this kind of variable the number of categories is the most informative quantity.
Type 2 variables are defined as scales with more than 9 (discrete) response categories or which are continuous. In this case an investigator must be able to specify certain percentiles in the density.

For type 1 variables, the interval length ℓ is defined as being equal to the number of categories, assuming that these categories represent equidistant values on the scale with unit spacing. If the user has entered the number of categories, four symmetric histograms are displayed. These histograms are constructed from normal HDRs with varying percentages (and thus standard deviations). A high percentage (e.g. 99.68, so z = 2.95) delivers a highly peaked histogram with a small standard deviation (0.17ℓ). A low percentage (e.g. 88.18, so z = 1.56) delivers a flat distribution with a larger standard deviation (0.32ℓ). Studying these histograms and confronting them with the user's prior expectations about the spread of the variable can give an idea about the value of the modal estimate of δ.
On request four non-normal histograms will be shown, together with their standard deviations (Two histograms with a different degree of skewness, one rectangular histogram and one bimodal histogram). This option is in conflict with the normality demand but it is assumed that a slight deviation from normality will not seriously disturb the final results. Default modal estimates, based on pragmatic choices, are available. As an option, the investigator may directly enter a value for $\bar{\delta}_i$.

It is assumed that for type 2 variables the user should be able to specify some percentiles in the normal density of the variable in question. By such a specification a user provides information about a value for ℓ (the difference between two percentiles) and thus an estimate can be obtained of δ using (11). If a user is able and willing to specify more percentiles, more estimates of δ will be the result. Finally the user chooses a specific value for $\bar{\delta}$ which maximally fits his/her prior beliefs.

After a modal estimate $\bar{\delta}_i$ has been obtained, a value for ν_i must be established. This number of degrees of freedom can be seen as a measure of certainty of the user about the value of the population standard deviation. The user is asked to express his/her (un)certainty about δ_i by entering a lower point δ_{ℓ_i} and an upper point δ_{u_i} of likely values around $\bar{\delta}_i$. The two bounds are considered to be the limits of a 95% HDR in the inverse chi density of $\bar{\delta}_i$. Together with $\bar{\delta}_i$ two inverse chi densities are fit, one based on δ_{ℓ_i} and $\bar{\delta}_i$, the second based on δ_{u_i} and $\bar{\delta}_i$. For both densities the number of degrees of freedom is computed by means of an approximation procedure in BAYFAC-I. The mean of the two calculated values for the degrees of freedom is shown to the user, as well as the associated 95% HDR around $\bar{\delta}_i$. The user may accept this interval and thus ν_i or may choose a new interval if the proposed one deviates too much from prior expectations. A default value of $\nu_i = 12$ together with the corresponding 95% interval is also shown. The final result will be a value for ν_i. From $\bar{\delta}_i$ and ν_i, c_i is computed by using (8), and (7) is completely specified.

4.2 Determination of $\bar{\underline{\omega}}$, $\bar{\underline{\phi}}$ and Π, Z in (9)

An interval of likely values for a population correlation coefficient is obtained in BAYFAC-I by requesting to enter a lower (ρ_ℓ) and an upper (ρ_u) bound. From the boundary points, the mode and variance in each normal density are determined by considering the interval as a 95% HDR. The mode of the prior density for the population correlation is fitted by

$$\bar{\rho} = \frac{\rho_\ell + \rho_u}{2} \tag{12}$$

and the standard deviation by

$$\bar{\sigma}_\rho = \frac{\rho_u - \rho_\ell}{3.92} , \tag{13}$$

where relation (11) is again recognized. In (12) and (13) ρ may either be an element from $\underline{\omega}$ or from $\underline{\phi}$.

After specification and personal correction of the densities in (9), it should be ideally true that $p(\underline{\phi})$ is such that the probability of a negative smallest eigenvalue of the corresponding Φ-matrix is zero and that $p(\underline{\phi})$ and $p(\underline{\omega})$ are such that the probability of a proportion unique variance outside (0 , 1) is zero. By means of Monte Carlo procedures which are implemented in BAYFAC-I an attempt is made to approach these ideals, recognizing that exactly zero probabilities are not possible as a consequence of using normal densities. The procedure will now be briefly outlined.

In order to examine the probability of a negative smallest eigenvalue of Φ in the density $p(\underline{\phi})$, a random sample of ϕ-vectors is drawn (based on $p(\underline{\phi})$) and the smallest eigenvalue (say ξ_1) of each corresponding Φ-matrix is computed. If there is an indication that the probability of an inadmissible ϕ-matrix is not negligible, a new density $p(\underline{\phi})$ is constructed. To adjust the density, two groups of ϕ-vectors are formed, one with the smallest ξ_1-values, and the other with the greatest ξ_1-values. The mean ϕ-vectors of the groups are compared element by element. The modal value for an element will be adjusted if its mean in one subgroup deviates considerably from its mean in the other. The standard deviations of these elements remain unchanged.

Say that n modal values are adjusted. The direction of the change is toward the mean of the subgroup with the greatest ξ_1-values. The magnitude of the change is positively related to the observed proportion of inadmissible Φ-matrices in the sample and to the difference between the two subgroup means, and negatively related to the magnitude of n.

After the adjustments, the n new modal estimates are presented to the user if the indications are that the probability of an inadmissible Φ-matrix is negligible. (Otherwise the estimates are further adjusted.) The user may accept the new values or make new adjustments. Obviously the ultimate result must be a density with an indicated low probability of an inadmissible Φ-matrix. Experience has shown that extreme values of the modes and standard deviations for the elements in $\underline{\phi}$ are likely to produce a non-negligible proportion of inadmissible Φ-matrices. In particular, the combination of extreme high and low values for the modes of different elements of $\underline{\phi}$ is apt to cause difficulties.

To check the admissibility of the marginal densities of the diagonal elements of Ψ^* (the proportions unique variance), a similar procedure is followed. A random sample of admissible ϕ-vectors is combined with a random sample of vectors of one row in Ω. (Each row corresponds to one variable and thus to one diagonal element in Ψ^*.) This results in a set of sampled values for the proportion unique variance of a variable. If there is an indication of a non-negligible probability of an improper value, an adjustment procedure is applied, resulting in one or more adapted modal estimates of the correlations in the given row of Ω. The procedure is repeated for every row of Ω, the results are checked, and the user is given the opportunity to make further adjustments, which are also checked. Again the experience is that extreme values for the modes and standard deviations of the elements in $\underline{\omega}$ are likely to produce improper sampled proportions unique variances.

5. *The posterior density*

The joint posterior density of all parameters in the factor analysis model is proportional to the product of the likelihood and the prior density $p(\underline{\delta},\underline{\omega},\underline{\phi},\underline{\mu})$. Knowing that $\underline{x}$ is distributed according to (3), the likelihood function for a random sample of N vectors $\underline{x}_i (i=1,\ldots,N)$ with sample mean $\bar{\underline{x}}$ and sample dispersion matrix S, defined as $N^{-1}\sum_{i=1}^{N}(\underline{x}_i\underline{x}_i' - N\bar{\underline{x}}\bar{\underline{x}}')$ is then as follows:

$$L(\Delta,\Omega,\Phi,\underline{\mu}|S,\bar{\underline{x}}) =$$

$$2\pi^{-\frac{1}{2}N}|\Sigma|^{-\frac{1}{2}N} \; exp[-\tfrac{1}{2}N tr(S\Sigma^{-1} + \Sigma^{-1}(\bar{\underline{x}} - \underline{\mu})(\bar{\underline{x}} - \underline{\mu})')] \quad . \tag{14}$$

Then the posterior distribution is given as

$$p(\underline{\delta},\underline{\omega},\underline{\phi},\underline{\mu}|S,\bar{\underline{x}}) \propto L(\Delta,\Omega,\Phi,\underline{\mu}|S,\bar{\underline{x}})\,p(\underline{\delta},\underline{\omega},\underline{\phi},\underline{\mu}) \quad , \tag{15}$$

where

$$p(\underline{\delta},\underline{\omega},\underline{\phi},\underline{\mu}) \propto \prod_{i=1}^{p} \delta_i^{-(\nu_i+1)} \; exp[-\tfrac{1}{2}\frac{\nu_i c_i^2}{\delta_i^2}]$$

$$* \; exp[-\tfrac{1}{2}tr(\Pi^{-1}(\underline{\omega} - \bar{\underline{\omega}})(\underline{\omega} - \bar{\underline{\omega}})')]$$

$$* \; exp[-\tfrac{1}{2}tr(Z^{-1}(\underline{\phi} - \bar{\underline{\phi}})(\underline{\phi} - \bar{\underline{\phi}})')] \quad . \tag{16}$$

In BAYFAC-O, the file created by BAYFAC-I, with all relevant prior estimates, is linked with a file containing the sample information.

Due to the intractability of the posterior density, derivation of interesting distributional characteristics (e.g. the posterior standard deviations) is too complex a matter. For a large number of observations the posterior density can be approximated by a multivariate normal distribution with mean vector equal to the vector with the generalized maximum likelihood estimators (the largest joint posterior mode) and covariance matrix equal to a generalization of the observed Fisher Information matrix evaluated at these generalized maximum likelihood estimators (Berger, 1985). However, large sample reasoning logically implies a minor attributed importance to the prior distribution which is not generally satisfactory to the present authors. Joint modal Bayesian estimates can however be obtained by employing an optimization routine. The optimization of (15) is performed by means of a quasi

Newton algorithm from the NAG-library. Computational details are considered in Euverman & Vermulst (1983).

If the optimization is successful, posterior modal estimates of Δ, Ω and Φ, say $\hat{\Delta}$, $\hat{\Omega}$, and $\hat{\Phi}$, are obtained. By using these estimates, estimates of Λ, Ψ and Ψ^* can also be obtained. The estimates $\hat{\Delta}$, $\hat{\Omega}$ and $\hat{\Phi}$ are unique, i.e. any rotation will destroy the validity of the prior density. The estimate $\hat{\Delta}$ is necessarily positive definite as a consequence of the definition of the inverse chi density used in the prior. Optimization of (15) should be performed such that the elements δ_i are restricted to the admissible parameter space. In the algorithm this is solved by restarting the optimization with a smaller stepsize in the iterations, if the standard stepsize would produce a negative value for δ_i. The estimate $\hat{\Phi}$ necessarily is non-singular, but it need not be positive definite. However, having defined a sound prior density $p(\underline{\phi})$, this is not likely to be a problem. The estimate $\hat{\Omega}$ should be well interpretable, i.e. having values between -1 and +1. The corresponding estimate $\hat{\Sigma}$ is likely to be positive definite. Eigenvalues of $\hat{\Phi}$ and $\hat{\Sigma}$ are printed in BAYFAC-O.

6. *Numerical example*

It is rather difficult to use an example from the literature in the present Bayesian context, as the question of whose prior information should be used wil necessarily arise. The authors have used an existing example: The Wechsler Preschool and Primary Scale of Intelligence (WPPSI) One of the authors has studied extensively the manual of the WPPSI to make himself familiar with the construction and meaning of the scale in order to collect sufficient prior knowledge regarding a factor analysis. The WPPSI is a scale designed in the USA consisting "of a battery of subtests, each of which when treated separately may be considered as measuring a different ability, and when combined into a composite score as a measure of overall or global intellectual capacity" (see Wechsler, 1967, pp. 1-2). The subtests are split up into a verbal and a performance cluster in the following way:

I	II
Verbal	Performance
1 Information	6 Animal House
2 Vocabulary	7 Picture Completion
3 Similarities	8 Mazes
4 Comprehension	9 Geometric Design
5 Arithmetic	10 Block Design

A supplementary 11th subtest Sentences is usually excluded from the battery. The prior information, which is primarily based on substantive considerations in the WPPSI-manual is used to analyze the results of the investigation of Yule, Berger, Butter, Newham and Tizard (1969).

It is important that the prior data are independent of the sample results. The design of Yule et al. is taken into account in specifying the prior information, which is set up with regard to children aged 5½.

The prior values for δ_i and ν_i are developed from the following information. In the Wechsler manual the raw scores are converted to scores with a mean of 10 and a standard deviation of 3. In additional samples described in the manual it was found that the standard deviation was generally below 3, probably due to the fact that the scaling procedure of the WPPSI is based on an extensively stratified sample, whereas the additional samples were less diversified. This is also true for the sample of Yule et al. The $\bar{\delta}_i$'s were entered directly in BAYFAC-I without using the interrogative question procedures. In an interactive way the ν_i were determined. The results are given in Table 1.

The prior choices for $\bar{\underline{\omega}}$, $\bar{\underline{\phi}}$ and Π, Z are based on the following considerations. Basic to the definition of the two clusters in the diagram of the preceeding page is the rough distinction in the kind of answers expected from the children and the tangibility of the materials involved, the latter typically regarded as indicative for performance tests. However the particular performance tests do require a certain level of verbal comprehension, i.e. the two clusters or factors cannot be regarded as uncorrelated. In subtests of both kinds, counting activities are involved. The mentioned distinction is thus not clearcut. A kind of numerical factor may then be defined but it should be possible to place this factor into the two-dimensional space spanned by the factors Verbal and Performance. It will generally be more determined by performance activities than by verbal comprehension. With this in mind, prior intervals for the elements of Ω were entered in BAYFAC-I, and are shown in Table 1. It can be seen that the correlations for Information, Arithmetic and Picture Completion are less discriminating between the two factors than the others. The test Information consists of expected verbal answers, but in a number of the questions counting activities are involved as well.

Arithmetic as a whole is a counting test and in some of the answers performance activities are required of the child. Picture Completion requires short verbal answers together with the use of concrete pictures. The values assigned to the distributional parameters in the prior densities were such that no correction procedures were necessary in BAYFAC-I.

Table 1. The prior information for WPPSI

	$\bar{\delta}_i$	ν_i	correlation intervals					
			Verbal		Perfor-mance		Verbal x Performance	
1. Information	2.7	61	.60	.90	.30	.70	.30	.70
2. Vocabulary	2.9	35	.60	.90	.20	.60		
3. Similarities	2.8	45	.60	.90	.20	.60		
4. Comprehension	2.6	44	.60	.90	.20	.60		
5. Arithmetic	2.6	44	.20	.60	.30	.70		
6. Animal House	2.8	66	.20	.60	.60	.90		
7. Picture Completion	2.8	66	.30	.70	.60	.90		
8. Mazes	2.8	66	.20	.60	.60	.90		
9. Geometric Design	2.6	44	.20	.60	.60	.90		
10. Block design	2.8	66	.20	.60	.60	.90		

The sample correlation matrix and standard deviations for these ten tests based on 150 observations of Yule et al. (1969) and the prior estimates are the input for BAYFAC-O, where the corresponding posterior density (15) is maximized to obtain posterior modal estimates of the elements in $\bar{\underline{\delta}}$, $\bar{\underline{\omega}}$, and $\bar{\underline{\phi}}$ (see Tables 2 and 3).

The absolute average discrepancy (AD) between the prior modal and sample standard deviations is equal to .223, between the prior and posterior modal values equal to .173, and between the sample and posterior values equal to .062, a rough indication that, regarding these standard deviations, the sample relatively had a larger weight in determining the posterior estimates. It should be noted that the posterior modal standard deviations of Mazes and Geometric Design are not between the prior modal and sample standard deviations.

By inspecting the prior and posterior agreement in the modal values of the elements in Ω in Table 3 it can be quickly seen that there is a high degree of comparability. Tucker's coefficient of congruence (T_c), being an overall measure of proportionality (equality of factor interpretation), and AD-values are shown in Table 3.

Similarities shows a drop of .11 in the correlation with the verbal factor. A possible explanation for this observation is that this subtest has 10 questions in which concrete applications of familiar oblects are described. Striking is also the rise of .20 in the correlation of Arithmetic with the verbal factor. Arithmetic perhaps requires more verbal comprehension than was expected a priori. The drop of .12 in the correlation of Animal House with Performance is hard to explain. The subtest requires color association, perhaps being less a performance act than was expected initially. The relatively high proportions unique variance in Table 2 may be an indication of the restricted value of the two-factor solution for this variable. The drop in the correlation between Geometric Design and Performance is such that it does not lie in the interval specified a priori (indicated with * in Table 3). A possible explanation is found in Yule et al. (1969): "The values for Geometric Design are almost considerably lower. Since this was the test which departed most from the American mean and since its scoring involves a great deal of judgment, it is possible that the scoring is highly unreliable and that the marks have been too lenient" (p.10). In Table 2 it is shown that the proportion unique variance for Geometric Design is high, indicating that this variable can hardly to be assumed to be explained by the two common factors.

It may be interesting to compare the posterior results with other factor analyses performed on the WPPSI scale.
In Table 4 results are printed of a factor analysis performed by Hollenbeck and Kaufman (1973). This analysis was conducted on the original Wechsler sample. Hollenbeck and Kaufman performed a principal factor analysis followed by a Biquartimin (oblique) rotation. The similarity of the structure is self-evident regarding the factor Verbal, but for the factor Performance the relatively high correlation with Information should be noted, perhaps due to the difference between the English and American samples. Hollenbeck and Kaufman conclude that the WPPSI scale can be regarded as separable into two distinct factors, thereby noticing that the subtest Arithmetic is an exception. They note: "This dual loading may have resulted from the fact that the Arithmetic test is the only test on the Verbal Scale that uses tangible materials (blocs, card with pictures) and the only one that has items requiring the child to 'perform' -i.e., count or point to the right picture - rather than to 'verbalize'" (p.44).

As a general conclusion it is stated that interpretable results were obtained by the Bayesian approach in accordance with results found in the literature, and the rotational problem has been avoided by using an interpretable prior specification. Prior thinking about correlation coefficients furthermore facilitates the interpretation of the factor structure a posteriori.

Table 2. Results for the standard deviations and the proportions unique variances

	standard deviations			prop. unique variances	
	prior $\bar{\delta}_i$	sample s_i	posterior $\hat{\delta}_i$	prior	posterior
1. Information	2.70	2.98	2.93	.42	.40
2. Vocabulary	2.90	3.15	3.11	.44	.38
3. Similarities	2.80	2.40	2.52	.44	.57
4. Comprehension	2.60	2.51	2.52	.44	.41
5. Arithmetic	2.60	2.66	2.64	.72	.53
6. Animal House	2.80	2.63	2.74	.44	.60
7. Picture Completion	2.80	2.62	2.72	.42	.47
8. Mazes	2.80	2.68	2.91	.44	.44
9. Geometric Design	2.60	3.08	3.09	.44	.66
10. Block Design	2.80	2.54	2.65	.44	.30

Table 3. Results for the correlation structure and factor intercorrelation

	verbal		performance		verbal x performance	
	prior	post.	prior	post.	prior	post.
1. Information	.75	.77	.50	.44	.50	.47
2. Vocabulary	.75	.78	.40	.46		
3. Similarities	.75	.64	.40	.44		
4. Comprehension	.75	.76	.40	.45		
5. Arithmetic	.40	.60	.50	.58		
6. Animal House	.40	.39	.75	.63		
7. Picture Completion	.50	.48	.75	.71		
8. Mazes	.40	.35	.75	.75		
9. Geometric Design	.40	.36	.75	.57*		
10. Block Design	.40	.47	.75	.83		
T_c		.990		.991		
AD		.056		.071		

* = Posterior mode falling outside the interval given a priori

able 4. Posterior Structure and the Corresponding One of Hollenbeck and Kaufman

	Correlation-structure verbal post.	verbal H&K	performance post.	performance H&K	verbal x performance post.	verbal x performance H&K
1. Information	.77	.78	.44	.56	.47	.57
2. Vocabulary	.78	.72	.46	.48		
3. Similarities	.64	.63	.44	.32		
4. Comprehension	.76	.76	.45	.49		
5. Arithmetic	.60	.71	.58	.64		
6. Animal House	.39	.46	.63	.64		
7. Picture Completion	.48	.51	.71	.60		
8. Mazes	.35	.47	.75	.65		
9. Geometric Design	.36	.47	.57	.75		
10. Block Design	.47	.51	.83	.74		
T_c	.994		.987			
AD	.107		.085			

7. *Discussion*

The present Bayesian approach is different from previous ones (e.g. Lee, 1981) in the sense that the assignment of values to distributional parameters of the prior densities is performed within a substantive theoretical framework. The authors do not claim that a Bayesian approach as such is always preferable to, for instance, a maximum likelihood treatment of the confirmatory factor analysis problem. Bayesian results such as those presented here should be carefully interpreted if one is used to thinking in the classical statistical way. It is always true that the posterior results are influenced by the prior specifications. In reporting Bayesian investigations a researcher should be conscientious in reporting the lines of reasoning used in specifying the prior information. It is possible to create posterior results such that the prior beliefs are confirmed, for example, by using a small sample size. Generalizing propositions based on posterior results should in fact lead to new investigations where prediction and testing is possible.

However as is well known, in factor analysis interpretational problems are often encountered. A Bayesian approach is useful in avoiding rotational indeterminacy or negative variance estimates by using prior densities which are supported theoretically. Compared to the maximum likelihood treatment of confirmatory factor analysis in which prior <u>values</u> are assigned to certain factor parameters, a Bayesian approach is more flexible in that prior densities are incorporated.

In conclusion, it may be noted that the authors are currently working on a revised form of the joint prior described in section 3 which is intended to provide meaningful dependencies among the parameters, thus replacing the assumption of prior independence now made in BAYFAC-I. The approach which is adopted stems from work by, for instance, Steiger (1980) on normal approximations to the sampling distribution of the sample correlation matrix, and has the advantage that the covariances among parameters can be written in terms of the modal values for the parameters, thus avoiding the necessity for additional specifications by the researcher.

References

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York, Berlin, Heidelberg, Tokyo. Sec. edition.

Euverman, T.J. & Vermulst, A.A. (1983). *Bayesian Factor Analysis.* Doctoral dissertation, University of Nijmegen. Krips Repro, Meppel.

Harman, H.H. (1976). *Modern Factor Analysis.* The University of Chicago Press, Chicago, London.

Hollenbeck, G.P. & Kaufman, A.S. (1973). Factor Analysis of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI). *Journal of Clinical Psychology,* 29, 41-45.

Jöreskog, K.G. & Sörbom, D. (1984). LISREL VI. *Analysis of linear structural relationships by maximum likelihood, instrumental variables and least squares methods.* User's guide. Uppsala: University of Uppsala, Department of Statistics.

Lee, S.Y. (1981). A Bayesian Approach to Confirmatory Factor Analysis. *Psychometrika,* 46, 153-160.

Mayekawa, S. (1985). *Bayesian Factor Analysis.* ONR Technical Report 85-3. CADA Research Group. The University of Iowa.

Novick, M.R. & Jackson, P.H. (1974). *Satistical Methods for Educational and Psychological Research.* McGraw-Hill, New York.

Press, S.J. (1978). *Applied Multivariate Analysis.* New York: Holt, Rinehart and Winston.

Steiger, J.H. (1980). Tests for comparing the elements of a correlation matrix. *Psychological Bulletin,* 87, 245-251.

Vermulst, A.A. & Euverman, T.J. (1983). *Inspection and Adjustment of Φ and Ψ^* in BAYFAC-I*. Internal Report. Institute for Educational Sciences, University of Nijmegen.

Wechsler, D. (1967). *Manual for the Preschool and Primary Scale of Intelligence*. New York: The Psychological Corporation.

Yule, W., Berger, M., Butler, S., Newham, V. & Tizard, J. (1969). The WPPSI: An Empirical Evaluation with a British Sample. *The British Journal of Educational Psychology*, 39, 1-13.