PSYCHOMETRIKA—VOL. 43, NO. 1.
MARCH, 1978
NOTES AND COMMENTS

# ON BAYESIAN ESTIMATION IN UNRESTRICTED FACTOR ANALYSIS

RAYMOND F. KOOPMAN

SIMON FRASER UNIVERSITY

It is shown that the common and unique variance estimates produced by Martin & McDonald's Bayesian estimation procedure for the unrestricted common factor model have a predictable sum which is always greater than the maximum likelihood estimate of the total variance. This fact is used to justify a suggested simple alternative method of specifying the Bayesian parameters required by the procedure.

Key words: communalities, uniquenesses, common variance, unique variance, Heywood cases, unbiased estimates, missing data.

Martin & McDonald [1975] have proposed a Bayesian procedure for the unrestricted common factor model. There is, however, a certain inconsistency between their equations and one of their numerical examples. Although they do not comment on it, the matrix $F$ of estimated factor loadings given in their Table 5(b) and the diagonal matrix $U^2$ of estimated unique variances given in the corresponding column of their Table 4(a) have the property that

$$H^2 + U^2 = \operatorname{diag} A \tag{1}$$

within roundoff error, where

$$H^2 = \operatorname{diag} FF' \tag{2}$$

contains the estimates of the common variances, and $A$ is the maximum likelihood estimate of the total covariance matrix of the variables. From this one might be led to the conclusion that the Bayesian estimates of the communalities and uniquenesses sum to the maximum likelihood estimates of the total variances. Such a conclusion would be wrong.

The proof of this is as follows. For any positive diagonal matrix $U^2$, let $D$ and $E$ be diagonal matrices of the $r$ largest and $n - r$ smallest eigenvalues, respectively, of $U^{-1}AU^{-1}$, and let $V$ and $W$ be the corresponding matrices of column-orthonormal eigenvectors of $U^{-1}AU^{-1}$, so that

$$U^{-1}AU^{-1} = VDV' + WEW'. \tag{3}$$

Then it is well known that, for any fixed $U^2$, the $F$ which conditionally minimizes $\phi(F, U^2)$ is given by

$$F = UV(D - I)^{1/2}. \tag{4}$$

Since the Bayesian procedure, as well as the maximum likelihood procedure, seeks such a conditional minimum, we may substitute from (4) into (2), yielding

$$h_{jj}^2 + u_{jj}^2 = a_{jj} - u_{jj}^2 \sum_{q=1}^{n-r} (e_q - 1) w_{jq}^2. \tag{5}$$

However, minimizing the Bayesian function $\phi^*(U^2)$ requires solving for $U^2$ such that Martin & McDonald's expression (20) is zero, which in turn requires (cf. Clarke, 1970, equation 6)

$$\sum_q \frac{(1 - e_q)w_{jq}^2}{u_{jj}^2} - \frac{\alpha_j}{u_{jj}^4} = 0 \tag{6}$$

Requests for reprints should be sent to R. F. Koopman, Psychology Dept., Simon Fraser University, Burnaby, B.C., Canada V5A 1S6.

for all $j$. It then follows upon substitution into (5) that

$$h_{jj}^2 + u_{jj}^2 = a_{jj} + \alpha_j. \tag{7}$$

Thus in the case at point the sums of the common and unique variance estimates should be 1.014 rather than 1.000. Although the difference between the two values is small, it is definitely nonzero and should not be attributed to roundoff error. R. P. McDonald has confirmed (private communication) that there was indeed a mistake in the computer program that supplied the numerical examples, and that the corrected program gives results in agreement with the above conclusion.

This being the case, some comment is in order regarding the presentation of results when a factoring of the data "in standard score form" is desired. This is usually what is implied when the only data input to the factoring routine is a sample correlation matrix, say $R$. In such cases it is not only permissible but mandatory to rescale the results so that $H^2 + U^2 = I$. This is in no sense "cheating". It is merely a recognition of the fact that $R$ is a rescaling of $A$ and that, since the factoring procedure is scale-invariant, we may interpret a factoring of $R$ as an arbitrary rescaling of a factoring of $A$. Had we actually factored $A$ in the first place and then been asked to report the results in units of the total standard deviations, there would be no question but that we should rescale so that $H^2 + U^2 = I$. The fact that the factoring was performed using $R$ instead of $A$ is therefore irrelevant.

However, the fact of (7) does allow us to make the following simplifying suggestion regarding the problem of arriving at a "reasonable" specification of the $\alpha_j$: always take $\alpha_j = a_{jj}/(N - 1)$. The justification for this is that (7) guarantees that the estimated total variance will then equal $a_{jj}N/(N - 1)$, which is the familiar unbiased estimate of the total variance and which is surely always "reasonable". Should this result in a value of $\phi$ which is excessive, we may conclude that some aspect of the model is inappropriate for the data at hand. The typical, but not necessary, conclusion in such a case would be that the number of factors should be increased.

In making the above suggestion, it is assumed that the data matrix from which $A$ is obtained has no missing entries. Should there be any data missing, then the following modification is suggested. Form $A$ in the usual way (e.g., as in Martin & McDonald's equation 5), substituting for each missing observation the mean of the non-missing observations on that variable. Then take $\alpha_j = a_{jj}(N - N_j + 1)/(N_j - 1)$, where $N_j$ is the number of non-missing observations on variable $j$. This will give unbiased total variance estimates and will reduce to the desired form when there are no missing observations.

## REFERENCES

Clarke, M. R. B. A rapidly convergent method for maximum-likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology*, 1970, *23*, 43–52.

Martin, J. K., & McDonald, R. P. Bayesian estimation in unrestricted factor analysis: a treatment for Heywood cases. *Psychometrika*, 1975, *40*, 505–517.

*Manuscript received 4/20/76*
*First revision received 12/15/76*
*Final version received 8/8/77*