Behaviormetrika
1989, No. 25, 15-33

COVARIANCE STRUCTURE ANALYSIS OF CONTINUOUSLY CHANGING POPULATIONS

Haruhiko Ogasawara*

A model, in which the means and the variance-covariance matrix of observed variables change with an external variable, is proposed. This is an extension of the analysis of covariance structures in several populations. Assuming that the observed variables, given the value of an external variable, have a multivariate normal distribution, the maximum likelihood estimates of the parameters in the model can be obtained by the Fisher's scoring method. The model with a constant variance-covariance matrix, the model with constant correlations, the model of a single common factor and the model of oblique multiple factors with constant factor loadings are disucussed for the model of the variance-covariance matrix. Finally, examples of intelligence test scores are provided, where the external variable is age.

1. Introduction

The analysis of covarince structures with several common parameters over different populations has been developed by Jöreskog (1971) and Sörbom (1974). This method is based on the assumption that the sample covariance matrix of each population has the independent Wishart distribution. McGaw & Jöreskog (1971) applied the method to a situation where the covariance structures were supposed to vary in several populations having different socio-economic statuses.

In Jöreskog's (1971) original literature, the mean structure was not paid attention, but, Sörbom (1974) treated the expected values of factor scores in each population as the parameters to be estimated by the maximum likelihood method. This model is called structured means model and the estimation procedure of the model is covered by the program LISREL (Jöreskog & Sörbom, 1981 ; Sörbom, 1981). A similar model in item response theory is found in Mislevy (1987), in which the mean of the examinee parameters, which represent a latent factor, is determined by auxiliary examinee variables.

In the possible applications of the models of several populations in social sciences, external variables which define subpopulations are socio-economic status, years of education, age and the like.

In this case, our interest is in the situation in which the parameters change continuously with the external variables. The problem of development and decay of intelligence is an example of this case , where the population changes with age. In this paper, we develop a model in which the variance-covariance matrix changes continuously as a function of an external variable. The model is a natural extension of the analysis of covariance structures in several populations.

2. General model specification and estimation

The models which will be proposed by this paper may be taken as the multivariate

Key Words and Phrases ; covariance structure analysis, maximum likelihood factor analysis, multiple populations, multivariate regression analysis, Fisher's scoring method, structured means
* Railway Technical Research Institute, 2-8-38, Hikari-cho, Kokubunji, Tokyo, 185 Japan

versions of Ogasawara's (1986) model. He developed a model in which s, a score, was normally distributed according to $N(\mu(y), \sigma^2(y))$, conditional on an external variable y. And $\mu(y)$ and $\sigma^2(y)$ indicate some specific functions of y. He applied the model to the situation in which the conditional distribution of adult intelligence scores depends on subjects' ages. Further, Ogasawara (1988) proposed a model in which the logarithm of a test score, given an external variable y, was distributed according to $N(\mu(y), \sigma^2(y))$.

Before specifying our models we explain the notation used here. Let the number of dependent variables be p and $\boldsymbol{s}_i$ be the column vector of these variables for the i-th individual, $(i=1, \cdots, N)$, that is,

$$\boldsymbol{s}_i=(s_{1i}, s_{2i}, \cdots, s_{pi})',$$

where s_{ki} denotes the value of k-th variable for the i-th individual.

Suppose the distribution of $\boldsymbol{s}_i$ depends on the observed value of y_i, and hence $\boldsymbol{s}_i \sim N(\boldsymbol{\mu}_i, \Sigma_i)$, where $\boldsymbol{\mu}_i$ and Σ_i denote the values of $\boldsymbol{\mu}$ and Σ when $y=y_i$, which are the functions of y. These functions, $\boldsymbol{\mu}$ and Σ, are specified by the vectors of parameters, $\boldsymbol{\theta}_\mu$ and $\boldsymbol{\theta}_\sigma$, respectively. Further, we assume that $\boldsymbol{s}_i(i=1, \cdots, N)$ are mutually independent, that is, we consider only the cross-sectional data instead of the longitudinal one for the study of change.

In the case of the analysis of covariance structures in several populations, $y_i(i=1, \cdots, N)$ are grouped and we can use tha likelihood of several independent Wishart distributions. However, in our case we cannot use this likelihood, since y changes continuously and the minimum size of the sample in which y takes the same value, y_i is generally one. Thus, we have to consider the likelihood of the multivariate normal distribution.

The likelihood of $\boldsymbol{\theta}_\mu$ and $\boldsymbol{\theta}_\sigma$, given $s_1, s_2, \cdots, s_N$ and $y_1, y_2, \cdots, y_N$, is

$$L(\boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\sigma \mid s_1, s_2, \cdots, s_N, y_1, y_2, \cdots, y_N)$$
$$=\prod_{i=1}^{N}(2\pi)^{-p/2}\,|\Sigma_i|^{-1/2}\exp\,\{-\frac{1}{2}(\boldsymbol{s}_i-\boldsymbol{\mu}_i)'\Sigma_i^{-1}(\boldsymbol{s}_i-\boldsymbol{\mu}_i)\}. \tag{1}$$

Actual forms of $\boldsymbol{\mu}_i$ and Σ_i should be specified according to each application. Hence, in this section we deal only with our model in a general form.

The negative of the logarithm of (1) is

$$f=-\log L=\frac{Np}{2}\log(2\pi)+\frac{1}{2}\sum_{i=1}^{N}\log|\Sigma_i|+\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{s}_i-\boldsymbol{\mu}_i)'\Sigma_i^{-1}(\boldsymbol{s}_i-\boldsymbol{\mu}_i). \tag{2}$$

We minimize (2) with respect to $\boldsymbol{\theta}_\mu$ and $\boldsymbol{\theta}_\sigma$, which is equivalent to maximizing (1), using the Fisher's scoring method (the Gauss-Newton method; Lee & Jennrich, 1979).

Next, we consider the method of testing the goodness-of-fit of the model. In the case of grouped data, we can use the likelihood ratio test of the model against the unconstrained model. This follows from the fact that

$$F=\sum_{i=1}^{g}\frac{N_i}{N}\{\log|\hat{\Sigma}_i S_i^{-1}|+\mathrm{tr}\,(S_i\hat{\Sigma}_i^{-1})\}-p \tag{3}$$

has the asymptotic x^2-distribution with $d.f.=\frac{1}{2}gp(p+1)-t$, where g is the number of groups, N_i the sample size, t the total number of independent parameters, $\hat{\Sigma}_i$ the fitted covariance matrix and S_i the sample covariance matrix in the i-th group.

In the case of non-grouped data, where populations are changing continuously, there is no appropriate statistic for assessing the goodness-of-fit of the model. But, the asymptotic standard errors and the AIC (Akaike, 1976, 1987) may be used for assessing and comparing models.

3. Various models

3.1 Constant variance-covariance matrix model (multivariate regression model)

In this section we deal with a model in which the expected values of observed variables vary with an external variable, but the variance-covariance matrix is constant. Here, we do not consider the structure of the matrix. Therefore,

$$\Sigma_i = \Sigma \quad (i=1, \cdots, N). \tag{4}$$

Equation (4) is often assumed in the case of multivariate regression analysis.

Now, we consider the following model.

$$\boldsymbol{s}_i = B\boldsymbol{y}_i + \boldsymbol{\varepsilon}_i, \tag{5}$$

where $\boldsymbol{\varepsilon}_i \sim^{i.i.d.} N(\boldsymbol{0}, \Sigma)$ and $\boldsymbol{y}_i' = [1, y_{1i}, \cdots, y_{qi}]$. $y_{1i}, \cdots, y_{qi}$ are the values of q external variables for the i-th individual. The matrix B is the $p \times (q+1)$ matrix of regression coefficients. In this case $\boldsymbol{\mu}_i = B\boldsymbol{y}_i$ and, $\boldsymbol{\theta}_\mu$ and $\boldsymbol{\theta}_\sigma$ consist of B and the unique elements of Σ. The number of independent parameters is $p(q+1)+p(p+1)/2$. When a polynomial of degree q is chosen, we simply replace $\boldsymbol{y}_i'$ by $[1, y_i, y_i^2, \cdots, y_i^q]$, others unchanged.

Let β_{kl} be the (k, l)-th element of B. Then the partial derivative of f is

$$\begin{aligned}\frac{\partial f}{\partial \beta_{kl}} &= -\sum_{i=1}^{N} \boldsymbol{y}_i' I_{lk} \Sigma^{-1}(\boldsymbol{s}_i - B\boldsymbol{y}_i) \\ &= -\operatorname{tr}\{I_{lk}\Sigma^{-1}\sum_{i=1}^{N}(\boldsymbol{s}_i - B\boldsymbol{y}_i)\boldsymbol{y}_i'\} \\ &= -\{\Sigma^{-1}\sum_{i=1}^{N}(\boldsymbol{s}_i - B\boldsymbol{y}_i)\boldsymbol{y}_i'\}_{kl},\end{aligned} \tag{6}$$

where I_{lk} denotes a matrix of an appropriate size, only the (l, k)-th element of which is one, others zero.

The estimate of Σ follows from

$$\frac{\partial f}{\partial \sigma_{kl}} = \frac{1}{2}(2-\delta_{kl})[N\sigma^{kl} - \{\Sigma^{-1}\sum_{i=1}^{N}(\boldsymbol{s}_i - \boldsymbol{\mu}_i)(\boldsymbol{s}_i - \boldsymbol{\mu}_i)'\Sigma^{-1}\}_{kl}], \tag{7}$$

where σ^{kl} denotes the (k, l)-th element of Σ^{-1} and δ_{kl} is the Kronecker delta.

Except for the case, $\boldsymbol{\mu}_i = B\boldsymbol{y}_i$, $\hat{\boldsymbol{\theta}}_\mu$ cannot, in general, be obtained algebraically. For these cases we use an iterative method in the following way. Starting with an initial value of $\boldsymbol{\theta}_\mu$, we replace $\boldsymbol{\theta}_\mu$ by an updated value. For Σ, we use

$$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{s}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{s}_i - \hat{\boldsymbol{\mu}}_i)', \tag{8}$$

which is obtained from (7).

We repeat the cycle until convergence is attained. The estimates of the variance-covariance matrix of parameters are obtained by

$$E\left(\frac{\partial^2 f}{\partial \boldsymbol{\theta}_\mu \partial \boldsymbol{\theta}_\mu'}\right)=\sum_{i=1}^{N}\frac{\partial \boldsymbol{\mu}'_i}{\partial \boldsymbol{\theta}_\mu}\Sigma^{-1}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\theta}_\mu'},$$
$$E\left(\frac{\partial^2 f}{\partial \sigma_{kl}\partial \sigma_{uv}}\right)=\frac{N}{8}(2-\delta_{kl})(2-\delta_{uv})\times \mathrm{tr}\{\Sigma^{-1}(I_{kl}+I_{lk})\Sigma^{-1}(I_{uv}+I_{vu})\}$$
$$=\frac{N}{4}(2-\delta_{kl})(2-\delta_{uv})(\sigma^{ku}\sigma^{lv}+\sigma^{kv}\sigma^{ul}). \tag{9}$$

The multivariate regression model represented by (5) is the model in which each dependent variable is predicted by the same set of independent variables. But, when the sets of independent variables are different from dependent variable to variable, the result of the univariate regression analysis is not in general equivalent to that of the multivariate one. The elements of B, corresponding to the independent variables which are not used for the prediction of each dependent variable, are zero (i.e. μ_{ki} (the k-th element of $\boldsymbol{\mu}_i$)$=\beta_{k1}+\beta_{k2}y_{1i}+\cdots+\beta_{k,\ qk+1}y_{qk,\ i}$, when using the first q_k y_{li}'s). But, (6) holds for non-zero β_{kl}. Thus, when Σ is given, $\hat{B}$ can be obtained algebraically. Conversely, when B is given, $\hat{\Sigma}$ is obtained by (8). We update B and Σ alternately until none of the absolute value of the elements of $\partial f/\partial \boldsymbol{\theta}\sigma$ is greater than some small value. In this case the number of independent parameters in B is $\sum_{k=1}^{p} q_k+p$, where q_k is the order of the polynomial for the k-th variable.

3.2 Constant correlation matrix model

Let D_i be the diagonal matrix consisting of the standard deriations of observed variables for its diagonal elements when y_i is given. Then

$$D_i=\begin{bmatrix} \sigma_{1i} & & O \\ & \sigma_{2i} \ddots & \\ O & & \sigma_{pi} \end{bmatrix}, \tag{10}$$
$$\Sigma_i=D_iR_iD_i, \tag{11}$$

where R_i is the correlation matrix of p variables given y_i. Now we consider the model, $R_i=R$ $(i=1, \cdots, N)$. This is the model in which the scales of the variables vary with y, but the correlations are constant.

The model is

$$\Sigma_i=D_iRD_i, \tag{12}$$

and

$$\frac{\partial f}{\partial \boldsymbol{\theta}_\mu}=-\sum_{i=1}^{N}\frac{\partial \boldsymbol{\mu}'_i}{\partial \boldsymbol{\theta}_\mu}D_i^{-1}R^{-1}D_i^{-1}(\boldsymbol{s}_i-\boldsymbol{\mu}_i). \tag{13}$$

Let $\boldsymbol{\theta}_s'=(\boldsymbol{\theta}\sigma_1', \cdots, \boldsymbol{\theta}\sigma_k', \cdots, \boldsymbol{\theta}\sigma_p')$ and $\boldsymbol{\theta}\sigma_k'=(\theta\sigma_{k1}, \cdots, \theta\sigma_{kl}, \cdots, \theta\sigma_{k,\ sk})$, where s_k is the number of the independent parameters in σ_{ki}. Assume that $\theta\sigma_k$ is related only to the k-th variable. Then the partial derivatives with respect to $\theta\sigma_{kl}$ and r_{kl}, the (k, l)-th element of R, $(k\neq l)$, are

$$\frac{\partial f}{\partial \theta\sigma_{kl}}=\sum_{i=1}^{N}\frac{1}{\sigma_{ki}}\frac{\partial \sigma_{ki}}{\partial \theta\sigma_{kl}}-\sum_{i=1}^{N}\frac{1}{\sigma_{ki}^2}\{R^{-1}D_i^{-1}(\boldsymbol{s}_i-\boldsymbol{\mu}_i)(\boldsymbol{s}_i-\boldsymbol{\mu}_i)'\}_{kk}\frac{\partial \sigma_{ki}}{\theta\sigma_{kl}}, \tag{14a}$$
$$\frac{\partial f}{\partial r_{kj}}=Nr^{kj}-\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{s}_i-\boldsymbol{\mu}_i)'D_i^{-1}R^{-1}(I_{kj}+I_{jk})R^{-1}D_i^{-1}(\boldsymbol{s}_i-\boldsymbol{\mu}_i)$$

$$= Nr^{kj} - \{R^{-1}\sum_{i=1}^{N} D_i^{-1}(\boldsymbol{s}_i - \boldsymbol{\mu}_i)(\boldsymbol{s}_i - \boldsymbol{\mu}_i)' D_i^{-1} R^{-1}\}_{kj}. \tag{14b}$$

Setting (14b) equal to zero,

$$\hat{R} = \frac{1}{N}\sum_{i=1}^{N} \hat{D}_i^{-1}(\boldsymbol{s}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{s}_i - \hat{\boldsymbol{\mu}}_i)' \hat{D}_i^{-1}. \tag{15}$$

Thus, given D_i and $\boldsymbol{\mu}_i$, ($i=1, \cdots, N$), we obtain $\hat{R}$ algebraically. Next we have the information matrix as follows.

$$E\left(\frac{\partial^2 f}{\partial\boldsymbol{\theta}_\mu \partial\boldsymbol{\theta}_\mu'}\right) = \sum_{i=1}^{N} \frac{\partial\boldsymbol{\mu}_i'}{\partial\boldsymbol{\theta}_\mu} D_i^{-1} R^{-1} D_i^{-1} \frac{\partial\boldsymbol{\mu}_i}{\partial\boldsymbol{\theta}_\mu'}. \tag{16a}$$

$$\begin{aligned}
E\left(\frac{\partial^2 f}{\partial\theta\sigma_{kl}\partial\theta\sigma_{uv}}\right) &= \frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}\{D_i^{-1}R^{-1}D_i^{-1}(I_{kk}RD_i + D_iRI_{kk})D_i^{-1}R^{-1}D_i^{-1} \\
&\qquad (I_{uu}RD_i + D_iRI_{uu})\}\frac{\partial\sigma_{ki}}{\partial\theta_{kl}}\frac{\partial\sigma_{ui}}{\partial\theta\sigma_{uv}} \\
&= \sum_{i=1}^{N} \mathrm{tr}(D_i^{-1}I_{kk}D_i^{-1}\mathrm{I}_{uu} + RI_{kk}D_i^{-1}R^{-1}D_i^{-1}I_{uu})\frac{\partial\sigma_{ki}}{\partial\theta\sigma_{kl}}\frac{\partial\sigma_{ui}}{\partial\theta\sigma_{uv}} \\
&= \sum_{i=1}^{N} \frac{\delta_{ku} + r^{ku}r_{ku}}{\sigma_{ki}\sigma_{ui}}\frac{\partial\sigma_{ki}}{\partial\theta\sigma_{kl}}\frac{\partial\sigma_{ui}}{\partial\theta_{uv}}.
\end{aligned} \tag{16b}$$

$$\begin{aligned}
E\left(\frac{\partial^2 f}{\partial\theta\sigma_{kl}\partial r_{mn}}\right) &= \frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}\{D_i^{-1}R^{-1}D_i^{-1}(I_{kk}RD_i + D_iRI_{kk})D_i^{-1}R^{-1}D_i^{-1} \\
&\qquad (D_iI_{mn}D_i + D_iI_{nm}D_i)\}\frac{\partial\sigma_{ki}}{\partial\theta\sigma_{kl}} \\
&= \frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}\{(R^{-1}D_i^{-1}I_{kk} + I_{kk}D_i^{-1}R^{-1})(I_{mn} + I_{nm})\}\frac{\partial\sigma_{ki}}{\partial\theta\sigma_{kl}} \\
&= \sum_{i=1}^{N} \mathrm{tr}\{(I_{mn} + I_{nm})(R^{-1}D_i^{-1}I_{kk})\}\frac{\partial\sigma_{ki}}{\partial\theta\sigma_{kl}}.
\end{aligned}$$

$$\begin{aligned}
&(k \neq m,\ k \neq n) \quad E\left(\frac{\partial^2 f}{\partial\theta\sigma_{kl}\partial r_{mn}}\right) = 0, \\
&(k = m) \quad E\left(\frac{\partial^2 f}{\partial\theta\sigma_{ml}\partial r_{mn}}\right) = \sum_{i=1}^{N} \frac{r^{mn}}{\sigma_{mi}} \times \frac{\partial\sigma_{mi}}{\partial\theta\sigma_{ml}}, \\
&(k = n) \quad E\left(\frac{\partial^2 f}{\partial\theta\sigma_{nl}\partial r_{mn}}\right) = \sum_{i=1}^{N} \frac{r^{mn}}{\sigma_{ni}} \times \frac{\partial\sigma_{ni}}{\partial\theta\sigma_{nl}}.
\end{aligned} \tag{16c}$$

$$\begin{aligned}
E\left(\frac{\partial^2 f}{\partial r_{kl}\partial r_{uv}}\right) &= \frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}\{D_i^{-1}R^{-1}D_i^{-1}D_i(I_{kl} + I_{lk})D_iD_i^{-1}R^{-1}D_i^{-1}D_i(I_{uv} + I_{vu})D_i\} \\
&= \frac{N}{2}\mathrm{tr}\{R^{-1}(I_{kl} + I_{lk})R^{-1}(I_{uv} + I_{vu})\} \\
&= N\,\mathrm{tr}\{R^{-1}(I_{kl} + I_{lk})R^{-1}I_{uv}\} \\
&= N(r^{kv}r^{ul} + r^{ku}r^{lv}).
\end{aligned} \tag{16d}$$

The vector $\boldsymbol{\mu}_i$ is specified by each application independently of D_i. When polynomials are chosen for $\boldsymbol{\mu}_i$, the same discussion in the previous section applies here. The vector $\boldsymbol{\theta}_\sigma$ consists of $\boldsymbol{\theta}_s$ and the elements of unique off-diagonal elements of R. Suppose that $\boldsymbol{\mu}_i$ and σ_{ki} are described by polynomials. Let $q_{\mu k}$ and $q_{\sigma k}$ be the orders of the polynomials for μ_{ki} (the k-th element of $\boldsymbol{\mu}_i$) and σ_{ki}, respectively (i.e. $\sigma_{ki} = \theta\sigma_{k1} + \theta\sigma_{k2}y_i + \cdots + \theta_{\sigma_k,\ q\sigma_k+1}\ y_i^{q\sigma k}$). Then the total number of the independent parameters is $\sum_{k=1}^{p}(q_{\mu k} + q_{\sigma_k}) + p(p+3)/2$.

3.3 One-factor model

In the following two sections, we deal with factor analysis models. We assume two types of changes with an external variable in the case of factor analysis. The one is the change of factor loadings and the other is that of the variance-covariance matrix of factor scores.

In this section, we deal with the change of factor loadings in a particular case where the one common-factor structure is assumed. The model is

$$\Sigma_i = \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i' + \Psi_i^2, \tag{17}$$

where $\boldsymbol{\lambda}_i$ is the vector of the factor loadings of a single common factor and Ψ_i is the diagonal matrix consisting of the factor loadings of unique factors, given y_i.

Let λ_{ki} be the k-th element of $\boldsymbol{\lambda}_i$ and $\boldsymbol{\theta}_{\lambda k}$ be the vector of the parameters with respect to the k-th variable. Let $\theta_{\lambda kl}$ be the l-th element of $\boldsymbol{\theta}_{\lambda k}$. Similarly, $\boldsymbol{\theta}_{\psi k}$ and $\theta_{\psi kl}$ are defined. Further, we suppose that each parameter in $\boldsymbol{\lambda}_i$ and Ψ_i corresponds to only a single variable.

The derivatives except those for $\boldsymbol{\theta}_\mu$ are

$$\frac{\partial f}{\partial \theta_{\lambda kl}} = \sum_{i=1}^{N} \{(\Sigma_i^{-1} - \Sigma_i^{-1}(\boldsymbol{s}_i - \boldsymbol{\mu}_i)(\boldsymbol{s}_i - \boldsymbol{\mu}_i)'\Sigma_i^{-1})\boldsymbol{\lambda}_i\}_{k\text{-th row}} \frac{\partial \lambda_{ki}}{\partial \theta_{\lambda kl}} \tag{18a}$$

$$\frac{\partial f}{\partial \theta_{\psi kl}} = \sum_{i=1}^{N} (\Sigma_i^{-1} - \Sigma_i^{-1}(\boldsymbol{s}_i - \boldsymbol{\mu}_i)(\boldsymbol{s}_i - \boldsymbol{\mu}_i)'\Sigma_i^{-1})_{kk} \psi_{ki} \frac{\partial \psi_{ki}}{\partial \theta_{\psi kl}} \tag{18b}$$

$$\begin{aligned} E\left(\frac{\partial^2 f}{\partial \theta_{\lambda kl} \partial \theta_{\lambda uv}}\right) &= \frac{1}{2} \sum_{i=1}^{N} \operatorname{tr}\{\Sigma_i^{-1}(\boldsymbol{e}_k \boldsymbol{\lambda}_i' + \boldsymbol{\lambda}_i \boldsymbol{e}_k')\Sigma_i^{-1}(\boldsymbol{e}_u \boldsymbol{\lambda}_i' + \boldsymbol{\lambda}_i \boldsymbol{e}_u')\} \frac{\partial \lambda_{ki}}{\partial \theta_{\lambda kl}} \frac{\partial \lambda_{ui}}{\partial \theta_{\lambda uv}} \\ &= \sum_{i=1}^{N} \operatorname{tr}(\Sigma_i^{-1} \boldsymbol{e}_k \boldsymbol{\lambda}_i' \Sigma_i^{-1} \boldsymbol{e}_u \boldsymbol{\lambda}_i' + \Sigma_i^{-1} \boldsymbol{e}_k \boldsymbol{\lambda}_i' \Sigma_i^{-1} \boldsymbol{\lambda}_i \boldsymbol{e}_u') \frac{\partial \lambda_{ki}}{\partial \theta_{\lambda kl}} \frac{\partial \lambda_{ui}}{\partial \theta_{\lambda uv}} \\ &= \sum_{i=1}^{N} \{(\boldsymbol{\lambda}_i' \Sigma_i^{-1} \boldsymbol{e}_k)(\boldsymbol{\lambda}_i' \Sigma_i^{-1} \boldsymbol{e}_u) + (\Sigma_i^{-1})_{ku} \boldsymbol{\lambda}_i' \Sigma_i^{-1} \boldsymbol{\lambda}_i\} \frac{\partial \lambda_{ki}}{\partial \theta_{\lambda kl}} \frac{\partial \lambda_{ui}}{\partial \theta_{\lambda uv}} \end{aligned} \tag{18c}$$

$$\begin{aligned} E\left(\frac{\partial^2 f}{\partial \theta_{\lambda kl} \partial \theta_{\psi uv}}\right) &= \sum_{i=1}^{N} \operatorname{tr}\{\Sigma_i^{-1}(\boldsymbol{e}_k \boldsymbol{\lambda}_i' + \boldsymbol{\lambda}_i \boldsymbol{e}_k')\Sigma_i^{-1} I_{uu} \psi_{ui}\} \frac{\partial \lambda_{ki}}{\partial \theta_{\lambda kl}} \frac{\partial \psi_{ui}}{\partial \theta_{\psi uv}} \\ &= 2 \times \sum_{i=1}^{N} (\Sigma_i^{-1} \boldsymbol{\lambda}_i)_u (\Sigma_i^{-1})_{ku} \, \psi_{ui} \frac{\partial \lambda_{ki}}{\partial \theta_{\lambda kl}} \frac{\partial \psi_{ui}}{\partial \theta \psi_{uv}}, \end{aligned} \tag{18d}$$

$$E\left(\frac{\partial^2 f}{\partial \theta_{\psi kl} \partial \theta_{\psi uv}}\right) = 2 \times \sum_{i=1}^{N} (\Sigma_i^{-1})_{ku}^2 \, \psi_{ki} \psi_{ui} \frac{\partial \psi_{ki}}{\partial \theta_{\psi kl}} \frac{\partial \psi_{ui}}{\partial \theta_{\psi uv}}, \tag{18e}$$

where $\boldsymbol{e}_k$ is the p-dimensional vector, the k-th element of which is one, others zero.

The discussion on $\boldsymbol{\mu}_i$ is the same as the previous section and is omitted here. The vector $\boldsymbol{\theta}_\sigma$ consists of $\boldsymbol{\theta}_{\lambda 1}, \cdots, \boldsymbol{\theta}_{\lambda p}, \boldsymbol{\theta}_{\psi 1}, \cdots, \boldsymbol{\theta}_{\psi p}$. Suppose the polynomials for $\boldsymbol{\mu}_i$, $\boldsymbol{\lambda}_i$ and Ψ_i, then the total number of the independent parameters for this model sums to $\sum_{k=1}^{p}(q_{\mu k} + q_{\lambda k} + q_{\psi k}) + 3p$, where $q_{\lambda k}$ and $q_{\psi k}$ are defined in the similar way to $q_{\mu k}$ (i.e. $\lambda_{ki} = \theta_{\lambda k1} + \theta_{\lambda k2} y_i + \cdots + \theta_{\lambda k, q_{\lambda k}+1} y_i^{q_{\lambda k}}$, $\psi_{ki} = \theta_{\psi k1} + \theta_{\psi k2} y_i + \cdots + \theta_{\psi k, q_{\psi k}+1} y_i^{q_{\psi k}}$).

3.4 Oblique multiple-factor model

In this section, we deal with a model in which factor loadings remain unchanged, but the variance-covariance matrix of oblique factors changes with an external variable. Factor loadings are the regression coefficients of observed variables onto factors and represent the properties of the measurement of observed variables. Thus, we can assume that factor loadings are constant when the measurement properties are the same in spite of the change of the external variable. But, still in some situations covariances among factors change, since they reflect individual values of latent variables, which may be very likely to change.

The variance-covariance matrix of the observed variables, given y_i is

$$\Sigma_i = \Lambda \Phi_i \Lambda' + \Psi_i \tag{19}$$

where Λ is a $p \times q$ (the number of common factors) matrix of factor loadings and Φ_i and Ψ_i are the variance-covariance matrices of common factors and unique factors given y_i, respectively. This model is an extension of Jöreskog's (1971) model in which Φ_i and Ψ_i have different values from group to group.

Now let T be a $q \times q$ non-singular matrix. If we replace Λ and Φ_i by ΛT and $T^{-1}\Phi_i T'^{-1}$, respectively, (19) is unchanged. That is, the model (19) has no identification. Therefore at least q^2 constraints must be imposed. The situation is the same as in the case of single population. But in the latter case the variances of common factors are often set to one and (q^2-q) constraints are imposed on Λ and Φ. Thus, the off-diagonal elements of Φ are interpreted as correlations. But in the case of (19) we cannot use this method, since Φ_i varies with i. It may be appropriate to set $(1/N)\sum_{i=1}^{N}\mathrm{Diag}(\Phi_i)=I$ or to impose q^2 constraints on Λ.

Now we derive the partial derivatives and the information matrix except those for $\boldsymbol{\theta}_\mu$. Let λ_{kl} be the (k, l)-th element of Λ and, $\theta\varphi_{kjl}$ be the l-th element of the vector, $\boldsymbol{\theta}\varphi_{kj}$, which consists of the parameters with respect to φ_{kj}, the (k, j)-th element of Φ_i. Similarly, $\theta_{\psi kl}$, $\boldsymbol{\theta}_{\psi k}$ and ψ_{ki} are defined. Again we suppose that each parameter in Φ_i and Ψ_i corresponds to a single variable. The final expressions in the following equations are set to conform to the results shown by Jöreskog (1971) in the case of several populations as much as possible.

$$\begin{aligned}\frac{\partial f}{\partial \lambda_{kj}} &= \frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}[\{\Sigma_i^{-1}-(\boldsymbol{s}_i-\boldsymbol{\mu}_i)(\boldsymbol{s}_i-\boldsymbol{\mu}_i)'\Sigma_i^{-1}\}(I_{kj}\Phi_i\Lambda'+\Lambda\Phi_i I_{jk})] \\ &= \sum_{i=1}^{N}[\{\Sigma_i^{-1}-\Sigma_i^{-1}(\boldsymbol{s}_i-\boldsymbol{\mu}_i)(\boldsymbol{s}_i-\boldsymbol{\mu}_i)'\Sigma_i^{-1}\}\Lambda\Phi_i]_{kj}.\end{aligned} \tag{20a}$$

$$\frac{\partial f}{\partial \theta\varphi_{kjl}} = \frac{1}{2}(2-\delta_{kj})\sum_{i=1}^{N}[\Lambda'\{\Sigma_i^{-1}-\Sigma_i^{-1}(\boldsymbol{s}_i-\boldsymbol{\mu}_i)(\boldsymbol{s}_i-\boldsymbol{\mu}_i)'\Sigma_i^{-1}\}\Lambda]_{kj}\frac{\partial \varphi_{kji}}{\partial \theta\varphi_{kjl}}. \tag{20b}$$

$$\frac{\partial f}{\partial \theta_{\psi kl}} = \frac{1}{2}\sum_{i=1}^{N}\{\Sigma_i^{-1}-\Sigma_i^{-1}(\boldsymbol{s}_i-\boldsymbol{\mu}_i)(\boldsymbol{s}_i-\boldsymbol{\mu}_i)'\Sigma_i^{-1}\}_{kk}\frac{\partial \psi_{ki}}{\partial \theta_{\psi kl}}. \tag{20c}$$

$$\begin{aligned}E\left(\frac{\partial^2 f}{\partial \lambda_{kj}\partial \lambda_{st}}\right) &= \frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}\{\Sigma_i^{-1}(I_{kj}\Phi_i\Lambda'+\Lambda\Phi_i I_{jk})\Sigma_i^{-1}(I_{st}\Phi_i\Lambda'+\Lambda\Phi_i I_{ts})\} \\ &= \sum_{i=1}^{N}\mathrm{tr}(\Sigma_i^{-1}I_{kj}\Phi_i\Lambda'\Sigma_i^{-1}\Lambda\Phi_i I_{ts}+\Sigma_i^{-1}\Lambda\Phi_i I_{jk}\Sigma_i^{-1}\Lambda\Phi_i I_{ts}) \\ &= \sum_{i=1}^{N}\{(\Sigma_i^{-1})_{sk}(\Phi_i\Lambda'\Sigma_i^{-1}\Lambda\Phi_i)_{jt}+(\Sigma_i^{-1}\Lambda\Phi_i)_{sj}(\Sigma_i^{-1}\Lambda\Phi_i)_{kt}\}.\end{aligned} \tag{20d}$$

$$\begin{aligned}E\left(\frac{\partial^2 f}{\partial \lambda_{kj}\partial \theta\varphi_{stl}}\right) &= \frac{1}{4}(2-\delta_{st})\sum_{i=1}^{N}\mathrm{tr}\{\Sigma_i^{-1}(I_{kj}\Phi_i\Lambda'+\Lambda\Phi_i I_{jk})\Sigma_i^{-1}\Lambda(I_{st}+I_{ts})\Lambda'\}\frac{\partial \varphi_{sti}}{\partial \theta\varphi_{stl}} \\ &= \frac{1}{2}(2-\delta_{st})\sum_{i=1}^{N}\mathrm{tr}(\Lambda'\Sigma_i^{-1}I_{kj}\Phi_i\Lambda'\Sigma_i^{-1}\Lambda I_{st}+\Lambda'\Sigma_i^{-1}I_{kj}\Phi_i\Lambda'\Sigma_i^{-1}\Lambda I_{ts})\frac{\partial \varphi_{sti}}{\partial \theta\varphi_{stl}} \\ &= \frac{1}{2}(2-\delta_{st})\sum_{i=1}^{N}\{(\Lambda'\Sigma_i^{-1})_{tk}(\Phi_i\Lambda'\Sigma_i^{-1}\Lambda)_{js}+(\Lambda'\Sigma_i^{-1})_{sk}(\Phi_i\Lambda'\Sigma_i^{-1}\Lambda)_{jt}\}\frac{\partial \varphi_{sti}}{\partial \theta\varphi_{stl}}.\end{aligned}$$

$$(20e)$$

$$E\left(\frac{\partial^2 f}{\partial\lambda_{kj}\partial\theta_{\psi uv}}\right)=\frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}\{\Sigma_i^{-1}(I_{kj}\Phi_i\Lambda'+\Lambda\Phi_i I_{jk})\Sigma_i^{-1}I_{uu}\}\frac{\partial\psi_{ui}}{\partial\theta_{\psi uv}}$$
$$=\sum_{i=1}^{N}\mathrm{tr}(\Sigma_i^{-1}\Lambda\Phi_i I_{jk}\Sigma_i^{-1}I_{uu})\frac{\partial\psi_{ui}}{\partial\theta_{\psi uv}}$$
$$=\sum_{i=1}^{N}(\Sigma_i^{-1}\Lambda\Phi_i)_{uj}(\Sigma_i^{-1})_{ku}\frac{\partial\psi_{ui}}{\partial\theta_{\psi uv}}. \qquad (20f)$$

$$E\left(\frac{\partial^2 f}{\partial\theta\varphi_{kjl}\partial\theta\varphi_{stm}}\right)=\frac{1}{8}(2-\delta_{kj})(2-\delta_{st})\sum_{i=1}^{N}\mathrm{tr}\{\Sigma_i^{-1}\Lambda(I_{kj}+I_{jk})\Lambda'$$
$$\Sigma_i^{-1}\Lambda(I_{st}+I_{ts})\Lambda'\}\frac{\partial\varphi_{kji}}{\partial\theta\varphi_{kjl}}\frac{\partial\varphi_{sti}}{\partial\theta\varphi_{stm}}$$
$$=\frac{1}{4}(2-\delta_{kj})(2-\delta_{st})\sum_{i=1}^{N}\mathrm{tr}(\Lambda'\Sigma_i^{-1}\Lambda I_{kj}\Lambda'\Sigma_i^{-1}\Lambda I_{st}$$
$$+\Lambda'\Sigma_i^{-1}\Lambda I_{kj}\Lambda'\Sigma_i^{-1}\Lambda I_{ts})\frac{\partial\varphi_{kji}}{\partial\theta\varphi_{kjl}}\frac{\partial\varphi_{sti}}{\partial\theta\varphi_{stm}}$$
$$=\frac{1}{4}(2-\delta_{kj})(2-\delta_{st})\sum_{i=1}^{N}\{(\Lambda'\Sigma_i^{-1}\Lambda)_{tk}(\Lambda'\Sigma_i^{-1}\Lambda)_{js}$$
$$+(\Lambda'\Sigma_i^{-1}\Lambda)_{sk}(\Lambda'\Sigma_i^{-1}\Lambda)_{jt}\}\frac{\partial\varphi_{kji}}{\partial\theta\varphi_{kjl}}\frac{\partial\varphi_{sti}}{\partial\theta\varphi_{stm}}. \qquad (20g)$$

$$E\left(\frac{\partial^2 f}{\partial\theta\varphi_{kjl}\partial\theta_{\psi uv}}\right)=\frac{1}{4}(2-\delta_{kj})\sum_{i=1}^{N}\mathrm{tr}\{\Sigma_i^{-1}\Lambda(I_{kj}+I_{jk})\Lambda'\Sigma_i^{-1}\}\frac{\partial\varphi_{kji}}{\partial\theta\varphi_{kjl}}\frac{\partial\psi_{ui}}{\partial\theta_{\psi uv}}$$
$$=\frac{1}{2}(2-\delta_{kj})\sum_{i=1}^{N}(\Lambda'\Sigma_i^{-1})_{ku}(\Lambda'\Sigma_i^{-1})_{ju}\frac{\partial\varphi_{kji}}{\partial\theta_{\varphi kjl}}\frac{\partial\psi_{ui}}{\partial\theta_{\psi uv}} \qquad (20h)$$

$$E\left(\frac{\partial^2 f}{\partial\theta_{\psi kl}\partial\theta_{\psi jm}}\right)=\frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}(\Sigma_i^{-1}\Lambda I_{kk}\Sigma_i^{-1}\Lambda I_{jj})\frac{\partial\psi_{ki}}{\partial\theta_{\psi kl}}\frac{\partial\psi_{ji}}{\partial\theta_{\psi jm}}$$
$$=\frac{1}{2}\sum_{i=1}^{N}(\Sigma_i^{-1})^2_{kj}\frac{\partial\psi_{ki}}{\partial\theta_{\psi kl}}\frac{\partial\psi_{ji}}{\partial\theta_{\psi jm}}. \qquad (20i)$$

The discussion on $\boldsymbol{\mu}_i$ is the same as before and is omitted here. We suppose polynomials for μ_{ki}, φ_{kji} and ψ_{ki} and that q^2 elements of Λ are pre-assigned. Let $q_{\varphi kj}$, and $q_{\psi k}$ be the orders of the polynomials for φ_{kji} and ψ_{ki} (i.e. $\varphi_{kji}=\theta_{\varphi kj,1}+\theta_{\varphi kj,2}y_i+\cdots+\theta_{\varphi kj,q\varphi kj+1}y_i^{q\varphi ki}$, $\psi_{ki}=\theta_{\psi k,1}+\theta_{\psi k,2}y_i+\cdots+\theta_{\psi k,\ q\theta k+1}y_i^{q\psi k}$). The total number of parameters is $\sum_{k=1}^{p}(q_{\mu k}+q_{\psi k})+\sum_{k\leq j}^{q}q_{\varphi kj}+pq-q^2/2+2p+q/2$.

3.5 Description of change

Although in the earlier sections σ_k, λ_k and the elements of Φ and Ψ were the functions of y, actual expressions of these functions were not given up to here except for polynomials. We have to specify appropriate functions carefully considering characteristics of the application situation. However, for descriptive purposes and to confirm roughly the tendency of change, polynomial functions may be used. In the analysis of covariance structures in a single population, Ogasawara (1979) and McDonald (1980) used polynomial functions for expressing the relationships between factor loadings.

4. Examples

4.1 Data

The data which will be analyzed in the following sections, is the same as that of

Ogasawara (1986). The ability tests A, B and C (tentative names) consist of four, six and two subtests, respectively, which are administered with time limits. The task, the time limit and the number of items for each subtest are shown in Table 1. The first halves of Tests A, B and C are supposed to measure mainly perceptual speed. The second halves of them measure mainly abilities of reasoning and spatial orientation, and are relatively difficult compared to the first halves.

Subjects were Japanese railway employees aged 20 to 54. Tests A and C were administered to the same subject group and Test B to part of this group and another group. The numbers of subjects of Tests A and C, and Test B were 1495 and 1493, respectively. These data do not show noticeable group differences except for age.

The scores of the tests are the numbers right, except that the score of Subtest B1 is the number right minus the number wrong.

4.2 Application to the data of Tests A and C

All the foregoing models were fitted to the data of Tests A and C. The vectors and matrices $\boldsymbol{\mu}$, σ, $\boldsymbol{\lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ were supposed to be expressed by polynomial functions of age, y. We used the results of the separate analysis for each variable (Ogasawara, 1986) to specify the orders of the polynomial functions in $\boldsymbol{\mu}$. He showed that the orders of polynomial functions in the elements of $\boldsymbol{\mu}$, which minimized the AIC, were 2, 3, 1, 1, 2, and 2 for Subtests A1, A2, A3, A4, C1 and C2, respectively. (In the original paper, the optimal order of the polynomial for A4 was 3, but should be corrected to be 1.) These polynomial functions were used for fitting the foregoing models. As a result, it was shown that none of the absolute t-value (estimate/estimate of standard error) of the coefficient for the highest term in the polynomial function in each element of $\boldsymbol{\mu}$ was less than 2, indicating adequacy of the model of $\boldsymbol{\mu}$.

With respect to σ the orders of the polynomials of the minimum AIC in Ogasawara (1986) were 0, 1, 2, 2, 0 and 1 in the order of the subtests, where "0" meant that the standard

Table 1
Contents of subtests

Test	Subtest	Task	Time (min)	Number of items
A	A1	Letter search	2	45
	A2	Finding figures	2.5	42
	A3	Symmetric figure I	3	32
	A4	Rearranging figures	3.5	35
C	C1	Digit symbol	1.5	80
	C2	Symmetric figure II	2.5	32
B	B1	Figure comparison	2	66
	B2	Figure matching	2	54
	B3	Figure classification	2	45
	B4	Part and whole	2	34
	B5	Construction of square	2	30
	B6	Surface development	3	35

Table 2
Results of fitting the constant-Σ model and the constant-R models to the data of Tests A and C

Model No.	Subtest :	A1	A2	A3	A4	C1	C2	AIC (††)
1	Σ : constant							50391.35 (38)
2	R : constant							
	Order of pylynomials in σ :	0	1	2	2	0	1	50342.97
	$t^{\dagger}$:	57.56	2.42	−1.13	−3.23	58.35	−3.95	(44)
3	R : constant							
	Order of polynomials in σ :	0	1	1	2	0	1	50342.29
	$t^{\dagger}$:	57.56	2.41	−5.30	−3.20	58.35	−3.98	(43)
	$\hat{R}=$	1.00						
		0.36	1.00					
		0.25	0.35	1.00				
		0.19	0.25	0.34	1.00			
		0.31	0.30	0.28	0.29	1.00		
		0.20	0.33	0.52	0.42	0.37	1.00	

† t-value of the estimated coefficient for the highest term in each polynomial
†† Values in parentheses under AIC's indicate the numbers of independent parameters.

Table 3
Results of fitting the one-factor models to the data of Tests A and C

Model No.	Subtest :	A1	A2	A3	A4	C1	C2	AIC (††)
4	Order of polynomials in λ' :	0	1	2	2	0	1	
	$t^{\dagger}$:	14.36	−0.09	−1.99	−0.25	18.77	−3.37	50464.82
	Order of polynomials in Diag (Ψ) :	0	1	2	2	0	1	(41)
	$t^{\dagger}$:	51.25	2.19	0.11	−3.27	48.26	−1.78	
5	Order of polynomials in λ' :	0	0	2	1	0	1	
	$t^{\dagger}$:	14.31	19.15	−2.00	−0.58	18.78	−4.64	50460.30
	Order of polynomials in Diag (Ψ) :	0	1	1	2	0	0	(37)
	$t^{\dagger}$:	51.28	2.04	−1.98	−3.50	48.27	36.13	
6	Order of polynomials in λ' :	0	0	2	0	0	1	
	$t^{\dagger}$:	14.31	19.14	−2.04	19.67	18.77	−4.59	50458.60
	Order of polynomials in Diag (Ψ) :	0	1	1	2	0	0	(36)
	$t^{\dagger}$:	51.30	2.09	−1.97	−3.56	48.30	36.12	

†,†† as before

deviation of the subtest score was constant.

We fitted twelve models named Models 1-12, the results of which are shown in Tables 2-6.

Table 2 shows the results of fitting the constant variance-covariance matrix model and the constant correlation matrix models (hereafter referred to as "constant-Σ model" and

Table 4
Results of fitting the two-factor models to the data of Tests A and C

Model No.	Subtest :	A1	A2	A3	A4	C1	C2	AIC (†††)
7	$\Lambda'=$	$3^{\dagger\dagger}$	1.34	$0^{\dagger\dagger}$	0.004	2.90	−1.00	
	(t-value)		(4.34)		(0.02)	(3.96)	(−2.32)	50396.46
		$0^{\dagger\dagger}$	1.05	$3^{\dagger\dagger}$	1.98	4.26	4.14	(34)
			(5.97)		(12.48)	(8.52)	(10.04)	
	Order of polynomials in :	0	0	0				
	φ_{11}, φ_{21} and φ_{22}, $t^{\dagger}$:	5.51	9.47	11.69				
	Order of polynomials in :	0	0	0	0	0	0	
	Diag (Ψ), $t^{\dagger}$:	6.37	18.21	19.48	24.00	23.30	6.58	
8	$\hat{\Lambda}'=$	$3^{\dagger\dagger}$	1.35	$0^{\dagger\dagger}$	$0^{\dagger\dagger}$	3.06	−0.94	
	(t-value)		(4.70)			(4.31)	(−2.53)	50355.19
		$0^{\dagger\dagger}$	1.05	$3^{\dagger\dagger}$	2.01	4.22	4.12	(42)
			(6.26)		(16.68)	(8.49)	(10.85)	
	Order of polynomials in :	1	1	1				
	φ_{11}, φ_{21} and φ_{22}, $t^{\dagger}$:	0.72	0.28	−3.76				
	Order of polynomials in :	1	1	1	1	1	1	
	Diag (Ψ), $t^{\dagger}$:	−1.60	2.64	−3.17	−2.97	−1.14	−0.38	
9	Order of polynomials in :	0	0	2				50347.04
	φ_{11}, φ_{21} and φ_{22}, $t^{\dagger}$:	6.06	10.28	−0.52				(41)
	Order of polynomials in :	0	2	2	2	0	0	
	Diag (Ψ), $t^{\dagger}$:	6.95	−0.80	−0.19	−3.04	23.33	7.92	
10	Order of polynomials in :	0	0	1				
	φ_{11}, φ_{21} and φ_{22}, $t^{\dagger}$:	6.11	10.33	−4.42				50343.81
	Order of polynomials in :	0	1	1	3	0	0	(39)
	Diag (Ψ), $t^{\dagger}$:	7.17	2.63	−3.33	0.27	23.29	7.76	
11	$\hat{\Lambda}'=$	$3^{\dagger\dagger}$	1.33	$0^{\dagger\dagger}$	$0^{\dagger\dagger}$	3.00	−0.97	50341.88
	(t-value)		(4.84)			(4.35)	(−2.78)	(38)
		$0^{\dagger\dagger}$	1.06	$3^{\dagger\dagger}$	2.00	4.22	4.14	
			(6.57)		(16.72)	(8.69)	(11.74)	
	Order of polynomials in :	0	0	1				
	φ_{11}, φ_{21} and φ_{22}, $t^{\dagger}$:	6.11	10.33	−4.43				
	Order of polynomials in :	0	1	1	2	0	0	
	Diag (Ψ), $t^{\dagger}$:	7.16	2.63	−3.33	−3.06	23.29	7.78	
12	$\hat{\Lambda}'=$	$3^{\dagger\dagger}$	1.13	$0^{\dagger\dagger}$	$0^{\dagger\dagger}$	2.34	−0.66	50426.61
	Grouping (t-value)		(4.25)			(3.81)	(−2.52)	(112)
	Method	$0^{\dagger\dagger}$	1.14	$3^{\dagger\dagger}$	1.98	4.53	3.81	
			(7.20)		(17.02)	(9.85)	(13.30)	

†,††† as before
†† pre-assigned values

Table 5
Estimated values at various ages (Model 11)

Age	$\hat{\Psi}$						$\hat{\varphi}_{11}$	$\hat{\varphi}_{21}$	$\hat{\varphi}_{22}$	$\hat{r}$†
	A1	A2	A3	A4	C1	C2				
22	7.64	5.70	11.77	8.18	78.63	5.30	0.76	0.43	1.03	0.48
27	7.64	6.11	11.03	9.06	78.63	5.30	0.76	0.43	0.95	0.50
32	7.64	6.51	10.30	9.38	78.63	5.30	0.76	0.43	0.86	0.52
37	7.64	6.92	9.56	9.14	78.63	5.30	0.76	0.43	0.78	0.55
42	7.64	7.32	8.82	8.34	78.63	5.30	0.76	0.43	0.69	0.59
47	7.64	7.73	8.09	6.97	78.63	5.30	0.76	0.43	0.61	0.63
52	7.64	8.13	7.35	5.05	78.63	5.30	0.76	0.43	0.52	0.67

† correlation between two factors

Table 6
Estimated values of the parameters in the grouping method (Model 12)

Age group	$\hat{\Psi}$						$\hat{\varphi}_{11}$	$\hat{\varphi}_{21}$	$\hat{\varphi}_{22}$	$\hat{r}$†
	A1	A2	A3	A4	C1	C2				
-24	5.16	5.22	10.51	8.61	84.00	7.23	1.06	0.30	0.75	0.33
25-29	8.05	6.82	10.59	8.06	74.10	6.93	0.75	0.42	1.18	0.45
30-34	9.28	7.15	10.75	7.92	94.57	6.25	0.73	0.44	0.98	0.52
35-39	5.61	6.42	9.44	10.41	78.37	5.30	1.05	0.56	0.77	0.63
40-44	6.98	8.20	9.39	7.99	85.90	4.54	0.72	0.45	0.83	0.58
45-49	4.77	7.59	6.78	6.50	66.76	5.57	1.06	0.33	0.55	0.44
50-	4.32	7.68	8.17	5.22	70.60	6.29	0.99	0.44	0.58	0.59

†as before

"constant-R model"). The AIC of the constant-Σ model is much larger than that of any constant-R model, showing inadequacy of the constant-Σ model. In Model 2 (the constant-R model), the absolute t-value of the coefficient for the highest term in the polynomial function (hereafter referred to as "t-value") for Subtest A3 is small. Hence, replacing the order by 1, we have Model 3, the AIC of which has decreased slightly from Model 2. Table 2 shows the estimate of R in Model 3, suggesting the two groups of Subtests (A1, A2, C1 and A3, A4, C2). We stopped fitting other constant-R models, since all the t-values in Model 3 were large enough.

Table 3 shows the results of fitting the one-factor models. Initially, the orders of polynomials for λ and Diag (Ψ) were set to those for σ which had been supposed to be appropriate by Ogasawara (1986). Reducing the orders of polynomials which have small absolute t-values, we have obtained the minimum AIC model (Model 6) in the one-factor models. But the AIC was much larger than that of the constant-Σ model or any constant-R model.

Table 4 shows the results of fitting the two-factor models. Four elements of Λ were set to be 3 or 0. That is, the scores of Subtests A1 and A3 were made to be the reference variables representing Factor 1 and Factor 2, respectively.

In Model 7, all the elements of Φ and Diag (Ψ) were set to be constant. This model is different from the ordinary factor analysis model in that the means of observed variables change with an external variable y. All the t-values in Λ were large except that of Factor 1 for Subtest A4. In Model 8, the loading in this position was set to be zero. In addition, in this model, all the orders of polynomials in Φ and Diag (Ψ) were set to be one in order to check the linear tendency of change.

The AIC of Model 8 decreased considerably from Model 7. But the t-values with respect to φ_{11}, φ_{21} and the elements of Diag (Ψ) for Subtests A1, C1 and C2, were still small. Hence, the next model, Model 9, was fitted to the data setting the orders in these positions equal to zero and remaining orders to two to investigate appropriate orders. Although the AIC again decreased, the t-values of φ_{22} and the elements of Diag (Ψ) for Subtests A2 and A3 were small, indicating that only the order of the polynomial of Diag (Ψ) for Subtest A4 should have been increased. Hence, restoring the orders of the polynomials with small t-values and increasing the order of Diag (Ψ) for Subtest A4, Model 10 was fitted to the data. The result shows that the order of the polynomial which was increased, should have been as before.

Model 11 was the final model, where all the t-values were large enough to stop fitting other models. The AIC of this model is the smallest in all the models fitted, but not so different from that of Model 3(the constant-R model). From the pattern of $\hat{\Lambda}$ in the results of fitting Model 11(Table 4), we see that the first halves and the second halves of Tests A and C represent different abilities in spite of the situation where the means and the variances change with age.

For illustrative purposes we show some of the estimates of the means and the factor variance-covariance matrices in Model 11 as follows:

$\hat{\mu}_1=8.44+0.623y-0.0106y^2$,

$\hat{\mu}_2=-2.79+1.74y-0.0482y^2+0.000388y^3$,

....

$\hat{\varphi}_{22}=1.41-0.0170y$,

$\hat{\psi}_1=7.64$,

$\hat{\psi}_2=3.92+0.0810y$,

....

Table 5 shows the estimated values of the variances and the covariances of factors fitted by Model 11 at several ages. In the elements of $\hat{\Phi}$, $\hat{\varphi}_{22}$ decreases with age, others constant. Among the diagonal elements of $\hat{\Psi}$, the one for Subtest A2 increases, the one for Subtest A3 decreases and the one for Subtest A4 increases for a while and decreases with age, respectively. As a whole the variances of both the common and the unique factors corresponding to the second halves of Tests A and C tend to decrease with age. On the other hand , those corresponding to the first halves of the tests seem to be relatively constant.

The correlation between the two common factors calculated from $\hat{\Phi}$ is also included in Table 5, showing the increasing tendency with age reflecting the fact that only the element, $\hat{\varphi}_{22}$in $\hat{\Phi}$ decreases with age, others constant.

The results of fitting the model by the grouping method are presented in Table 4. The age groups are defined by the equal length of age interval (five years), that is ~24, 25~29, ⋯, 50~. The model is identical with Jöreskog's (1971) one, which consists of a common Λ to the age groups, and $\boldsymbol{\mu}_i$, $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Psi}_i$ peculiar to the i-th group. The pattern of $\hat{\Lambda}$ in Model 12 is similar to that of Model 11. However, the AIC is much larger than that of Model 11. Table 6 presents the $\hat{\boldsymbol{\Phi}}_i$' s, $\hat{\boldsymbol{\Psi}}_i$'s and the correlations between the two common factors, in which we see the tendencies which vary with age. But, the results also reflect the sampling variations. Thus, Model 11 has advantage over Model 12 not only in view of AIC but also taking into account the tendency of development and decay of intelligence with age.

4.3 Application to the data of Test B

As for the data of Test B, the separate analysis of each observed variable by Ogasawasa (1986) showed that the appropriate orders of polynomial functions for $\boldsymbol{\mu}$ were 2, 2, 3, 3, 3 and 2 in the order of the subtests. And the appropriate orders for σ were 0, 1, 1, 1, 0, and 1.

We fitted thirteen models named Models 1-13, the results of which are shown in Tables 7-11.

Table 7 shows the results of fitting the constant-Σ model (Model 1) and the constant-R models to the data. In view of the AIC, Model 1 obviously lacks the goodness-of-fit to the data, favoring the constant-R models. But in Model 2 the t-value of σ for Subtest B3 is not large enough. Reducing the order of the polynomial in this position leads to Model 3. But the comparison between the two models is difficult, considering both the t-values and the AIC's.

Table 8 shows the results of fitting the one-factor models to the data. The large AIC's

Table 7
Results of fitting the constant-Σ model and the constant-R models to the data of Test B

Model No.	Subtest :	B1	B2	B3	B4	B5	B6	AIC (††)
1	Σ : constant							51433.51
								(42)
2	R : constant							
	Order of polynomials in σ :	0	1	1	1	0	1	
	t† :	58.43	−2.22	−1.73	−3.52	62.99	−5.42	
	$\hat{R}$ =	1.00						51397.64
		0.43	1.00					(46)
		0.38	0.47	1.00				
		0.29	0.49	0.38	1.00			
		0.29	0.47	0.37	0.57	1.00		
		0.20	0.30	0.28	0.41	0.45	1.00	
3	R : constant							
	Order of polynomials in σ :	0	1	0	1	0	1	51398.45
	t† :	58.42	−2.10	59.85	−3.47	62.99	−5.40	(45)

†,†† as before

Table 8
Results of fitting the one-factor models to the data of Test B

Model No.	Subtest :	B1	B2	B3	B4	B5	B6	AIC (**)
4	Order of polynomials in λ' :	0	1	1	1	0	1	
	t^* :	17.94	−2.44	−3.22	−1.35	29.09	−2.51	51555.77
	Order of polynomials in Diag (Ψ) :	0	1	1	1	0	1	(41)
	t^* :	50.77	−0.48	0.49	−3.11	40.43	−4.20	
5	Order of polynomials in λ' :	0	1	1	0	0	1	
	t^* :	17.95	−2.50	−3.08	29.00	29.03	−2.33	51552.11
	Order of polynomials in Diag (Ψ) :	0	0	0	1	0	1	(38)
	t^* :	50.77	42.19	48.05	−3.71	40.55	−4.21	

*,** as before

indicate that an additional factor is needed.

Table 9 shows the results of fitting the two-factor models. For identification of model, the scores of Subtests B1 and B4 were set to be the reference variables, representing Factor 1 and Factor 2, respectively.

In the results of fitting Model 6, the t-value of the loading of Factor 1 for Subtest B5 is small. Thus, the next model, where the order of the polynomial was replaced by zero (constant), was fitted. Since the revised model, Model 7, did not show the improvement of the AIC and the just identified model would be useful to grasp the tendency of Φ and Ψ independently of additional restrictions on Λ, only the four elements of Λ were set to be fixed. The models, where the orders of polynomials for Φ and Diag (Ψ) were one (Model 8), two (Model 9) and again one (Model 10, 11, 12), were fitted successively.

Models 10, 11 and 12 show the similar values of AIC less than those of other two-factor models. Among the elements of $\hat{\Phi}$ in Models 10 and 11, $\hat{\varphi}_{21}$ and $\hat{\varphi}_{22}$ decrease with age. Although the t-values of φ_{21} and φ_{22} in Models 10 and 11 are not large enough to conclude the change occurred with age, the two of the subtests in the second half of Test B show the significant decrease of Diag (Ψ) with age, which is the similar tendency in the case of Tests A and C.

In contrast to the case of Tests A and C the correlation between the two common factors decreases with age in Model 11 (shown in Table 10) or is constant in Model 12.

We present some of the estimates of the means and the factor variance-covariance matrices in Model 11 as follows:

....

$\hat{\mu}_2 = 27.2 + 0.192y - 0.00566y^2$,

$\hat{\mu}_3 = -1.64 + 2.45y - 0.0702y^2 + 0.000579y^3$,

....

$\hat{\varphi}_{21} = 0.997 - 0.00452y$,

$\hat{\varphi}_{22} = 0.993$,

....

$\hat{\Phi}_5 = 6.02$,

$\hat{\Phi}_6 = 32.0 - 0.386y$.

Table 9
Results of fitting the two-factor models to the data of Test B

Model No.	Subtest :	B1	B2	B3	B4	B5	B6	AIC ($\dagger\dagger\dagger$)
	$\hat{\Lambda}'=$	$3^{\dagger\dagger}$	2.07	2.10	$0^{\dagger\dagger}$	−0.35	−0.60	
	(t-value)		(6.64)	(6.86)		(−1.49)	(−2.26)	51433.55
		$0^{\dagger\dagger}$	1.39	1.01	$3^{\dagger\dagger}$	3.42	3.51	(38)
			(4.36)	(3.04)		(10.70)	(10.68)	
6	Order of polynomials in :	0	0	0				
	φ_{11}, φ_{21} and φ_{22}, $t^{\dagger}$:	8.89	11.99	14.03				
	Order of polynomials in :	0	0	0	0	0	0	
	Diag (Ψ), $t^{\dagger}$:	17.18	14.35	19.85	17.27	12.19	21.57	
	$\hat{\Lambda}'=$	$3^{\dagger\dagger}$	2.08	2.13	$0^{\dagger\dagger}$	$0^{\dagger\dagger}$	−0.40	
	(t-value)		(7.07)	(7.23)			(−1.79)	51434.12
		$0^{\dagger\dagger}$	1.35	0.97	$3^{\dagger\dagger}$	3.00	3.27	(37)
			(4.54)	(3.09)		(23.88)	(11.53)	
7	Order of polynomials in :	0	0	0				
	φ_{11}, φ_{21} and φ_{22}, $t^{\dagger}$:	8.81	11.88	14.84				
	Order of polynomials in :	0	0	0	0	0	0	
	Diag (Ψ), $t^{\dagger}$:	16.68	14.63	19.78	17.76	17.26	21.65	
	Order of polynomials in :	1	1	1				
8	φ_{11}, φ_{21} and φ_{22}, $t^{\dagger}$:	−1.08	−2.22	−1.75				51407.04
	Order of polynomials in :	1	1	1	1	1	1	(47)
	Diag (Ψ), $t^{\dagger}$:	−1.38	−1.11	−0.73	−3.34	1.54	−4.87	
	Order of polynomials in :	0	2	2				51412.60
9	φ_{11}, φ_{21} and φ_{22}, $t^{\dagger}$:	8.94	−0.18	0.42				(46)
	Order of polynomials in :	0	0	0	2	0	2	
	Diag (Ψ), $t^{\dagger}$:	17.33	14.72	19.73	−0.95	12.21	−0.07	
	Order of polynomials in :	0	1	1				51405.99
10	φ_{11}, φ_{21} and φ_{22}, $t^{\dagger}$:	8.94	−1.80	−1.24				(42)
	Order of polynomials in :	0	0	0	1	0	1	
	Diag (Ψ), $t^{\dagger}$:	17.34	14.71	19.72	−3.24	12.35	−4.84	
	$\hat{\Lambda}'=$	$3^{\dagger\dagger}$	2.07	2.15	$0^{\dagger\dagger}$	−0.40	−0.62	51405.48
	(t-value)		(6.75)	(6.94)		(−1.71)	(−2.42)	(41)
		$0^{\dagger\dagger}$	1.39	0.97	$3^{\dagger\dagger}$	3.51	3.49	
			(4.40)	(2.87)		(10.92)	(10.90)	
11	Order of polynomials in :	0	1	0				
	φ_{11}, φ_{21} and φ_{22}, $t^{\dagger}$:	8.96	−1.33	14.14				
	Order of polynomials in :	0	0	0	1	0	1	
	Diag (Ψ), $t^{\dagger}$:	17.67	14.53	19.56	−3.34	11.98	−4.88	

Table 9 (continued)

Model No.	Subtest:	B1	B2	B3	B4	B5	B6	AIC (†††)
12	$\hat{\Lambda}' =$	3††	2.08	2.09	0††	−0.38	−0.64	51405.30
	(t-value)		(6.64)	(6.85)		(−1.61)	(−2.45)	(40)
		0††	1.39	1.03	3††	3.48	3.52	
			(4.32)	(3.08)		(10.82)	(10.75)	
	Order of polynomials in :	0	0	0				
	φ_{11}, φ_{21} and φ_{22}, $t^\dagger$:	8.91	12.05	14.13				
	Order of polynomials in :	0	0	0	1	0	1	
	Diag (Ψ), $t^\dagger$:	17.29	14.25	19.92	−3.25	12.10	−4.79	
13	Grouping method $\hat{\Lambda}' =$	3††	2.12	2.05	0††	−0.39	−0.76	51397.15
	(t-value)		(6.99)	(7.12)		(−1.71)	(−2.87)	(113)
		0††	1.33	1.04	3††	3.47	3.67	
			(4.19)	(3.21)		(11.26)	(10.93)	

†,††,††† as before

Table 10
Estimated values at various ages (Model 11)

Age	$\hat{\Psi}$ B1	B2	B3	B4	B5	B6	$\hat{\varphi}_{11}$	$\hat{\varphi}_{21}$	$\hat{\varphi}_{22}$	$\hat{\gamma}^\dagger$
22	25.32	9.98	16.87	8.96	6.02	23.54	1.54	0.90	0.93	0.75
27	25.32	9.98	16.87	8.35	6.02	21.61	1.54	0.87	0.93	0.73
32	25.32	9.98	16.87	7.74	6.02	19.68	1.54	0.85	0.93	0.71
37	25.32	9.98	16.87	7.14	6.02	17.75	1.54	0.83	0.93	0.69
42	25.32	9.98	16.87	6.53	6.02	15.82	1.54	0.81	0.93	0.67
47	25.32	9.98	16.87	5.92	6.02	13.89	1.54	0.78	0.93	0.65
52	25.32	9.98	16.87	5.31	6.02	11.96	1.54	0.76	0.93	0.64

† as before

Table 11
Estimated values of the parameters in the grouping method (Model 13)

Age group	$\hat{\Psi}$ B1	B2	B3	B4	B5	B6	$\hat{\varphi}_{11}$	$\hat{\varphi}_{21}$	$\hat{\varphi}_{22}$	$\hat{\gamma}^\dagger$
-24	27.83	10.60	16.11	8.72	5.90	23.38	1.37	0.89	0.96	0.78
25-29	27.00	10.21	18.26	7.91	5.49	20.93	1.90	0.96	1.04	0.69
30-34	29.00	11.55	17.02	7.23	5.02	20.79	1.55	1.07	1.09	0.82
35-39	20.58	7.71	17.65	8.29	6.26	15.95	1.63	0.90	0.97	0.71
40-44	21.99	10.51	17.08	7.14	6.31	16.36	1.45	0.69	0.75	0.67
45-49	26.22	9.34	19.59	5.43	6.45	13.61	1.03	0.63	0.88	0.65
50-	24.24	7.09	10.60	4.45	8.47	11.51	2.11	0.88	0.80	0.68

† as before

The results of Model 13 (the grouping method) show that $\hat{\Lambda}$ is similar to those of the previous models. In view of the AIC, Medel 13 seems to have advantage. However, the grouping method does not consider the tendencies of the changes of the subtest scores with age. Further, since the original data is not grouped and the grouping method depends on how to group the non-grouped data, we do not accept Model 13.

In the data of Tests A and C, the optimal model was the oblique two-factor model (Model 11), but the AIC of which was not so different from that of the constant-R model. In the case of Test B the optimal model was the constant-R model. Thus, we conclude that the correlations between the subtests are relatively constant in spite of the change of Φ and Ψ. In addition, we obtained the tendency of continuous changes of the variances and the covariances of the factors, which had not been possible until our model was applied.

5. Some concluding remarks

The models we have proposed in this paper deal with the moments which change continuously with an external variable. These models are not covered by present computer programs such as LISREL and COSAN. Up to now, these problems have been dealt with by grouping data accordng to some criteria. However, considering the loss of information by grouping and the difference of results made by the difference of grouping methods, we feel that the proposed model should have some advantages over the grouping methods.

But, the proposed model requires much CPU time, because the calculation of the gradient vector, the expected value of the Hessian matrix and the inverse of Σ_i is very time-consuming. One of the reasons for this is that Σ_i^{-1} must be computed for each i, ($i=1, \cdots, N$)) at each iteration. The CPU time can be reduced considerably by computing only the several Σ_i^{-1}'s corresponding to typical i's at each stage of iteration. After the iteration converges, we can obtain the exact value of the information matrix.

In the previous examples, age was the only external variable. But, in general, multiple external variables may be considered such as age, years of education and income. But, this extension should not be very difficult, and, in fact, the algorithm necessary for the calculation of polynomials can be used for the multiple external variables, as was described in 3.1. In addition, the combination of discreate external variables, e.g. sex, and continuous variables may be possible.

Acknowledgement

The author is indebted to Dr. Kazuo Shigemasu for his kind and helpful comments.

References

Akaike, H. (1976). What is An Information Criterion?: *Suuri-kagaku,* **14**(3), 5-11, (in Japanese).

Akaike, H. (1987). Factor analysis and AIC; *Psychometrika,* **52**, 317-332.

Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations; *Psychometrika,* **36**, 409-426.

Jöreskog, K.G. and Sörbom, D. (1981). *LISREL User's guide*, Chicago, National Educational Services.

Lee, S.Y. and Jennrich, R.I. (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis; *Psychometrika*, **44**, 99-113.

McDonald, R.P. (1980). A simple comprehensive model for the analysis of covariance structures: Some remarks on applications; *British Journal of Mathematical and Statistical Psychology*, **33**, 161-183.

McGaw, B. and Jöreskog, K.G. (1971). Factorial invariance of ability measures in groups differing in intelligence and socio-economic status; *British Journal of Mathematical and Statistical Psychology*, **24**, 154-168.

Mislevy, R.J. (1987). Exploiting auxiliary information about examinees in the estimation of item parameters; *Applied Psychological Measurement*, 11, 81-91.

Ogasawara, H. (1979). A method of estimating the parameters in the polynomial model of the simplex structure by least squares; *Railway Labor Science*, **33**, 75-91, (in Japanese).

Ogasawara, H, (1986). Maximum likelihood estimation of criteria for age adjustment in intelligence test; *Railway Labor Science*, **40**, 37-53, (in Japanese).

Ogasawara, H. (1988). Normalization of test scores by the structured log-normal distribution; *Railway Technical Research Institute Reports*, **2**(3), 55-65, (in Japanese).

Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups; *British Journal of Mathematical and Statistical Psychology*, **27**, 229-239.

Sörbom, D. (1981). Structural equation models with structured means; In Jöreskog, K.G. and Wold, H. (eds.): *Systems under indirect observation: Causality, structure and prediction*, Amsterdam, North-Holland, 183-195.

(Received April, 1988. Revised January, 1989)