Variable selection for structural models

Yutaka Kano
School of Human Sciences, Osaka University, Suita, Osaka 565-0871, Japan
email: kano@hus.osaka-u.ac.jp

Dedicated to the eightieth birthday of Professor C. R. Rao

(Revised on August 17, 2000)

Abstract

Theory of variable selection for structural models that do not have clear dependent variables is developed. Theory is derived within the framework of the curved exponential family of distributions for observed variables. The idea of Rao's score test was taken to construct a test statistic for variable selection, and its statistical properties are examined. In particular, the test statistic is shown to have asymptotic *central* chi-square distribution under a kind of *alternative* hypothesis. This fact will provide an evidence for excellent performance of the score statistic for real data sets.

AMS classification: primary 62H15; secondary 62F05
Keywords: Bias of estimators, factor analysis model, contiguous alternatives, covariance structure model, goodness-of-fit test, likelihood ratio test, noncentrality parameter, Rao's score test.
Abbreviated title: Variable selection for structural models.

---

1. Introduction and motivation

There are numerous researches on variable selection in statistical models including regression models, time series models, and discriminant analysis (e.g., Shibata 1976, Fujikoshi 1985, Miller 1984). Almost all the literature, however, has focused upon the models with dependent variables and discussed selection of independent (or explanatory) variables. In these models, variable selection is made in order to achieve most precise prediction of the dependent variable or smallest misclassification probability, of a future sample.

It is important to develop theory of variable selection for models in multivariate analysis which do not have such clear dependent variables. Examples of such models include a factor analysis model, a model of principal component analysis and some covariance structure models. There is not so much literature on variable selection for such models

(Yanai 1980; Tanaka 1983; Jolliffe 1986; Gorsuch 1988). No one has considered variable selection based on fit measures except for Kano and Ihara (1994) and Kano and Harada (2000).

Consider a simple mean structure model:

$$E[\bar{\boldsymbol{X}}] = \boldsymbol{\eta}(\boldsymbol{u}), \tag{1}$$

where $\bar{\boldsymbol{X}}$ is a sample mean p-vector and $\boldsymbol{\eta}(\boldsymbol{u})$ is a p-dimensional vector-valued function from a domain of R^q $(q < p)$. When the model (1) is not fitted well, a typical approach to the problem is to explore a different mean structure. An alternative potentially useful approach is to identify some variables inconsistent with the model and remove them. The approach is important when variables to be analyzed are not predetermined such as in an analysis of questionnaire data with many items.

To identify inconsistent variables, one can compute fit measures (i.e., goodness-of-fit statistics) for every model obtained by removing possibly inconsistent variables. The procedure is not sensible for nonlinear models with a fairly large number of variables because of heavy computation. A useful way to do it is to use the estimator for $\boldsymbol{u}$ in the (full) model (1) to form fit measures for all models after removing possible inconsistent variables. In this paper, we shall take Rao's score test approach to construct such a statistic. This will be described in Section 3.

One may ask a natural question on the score test approach: when the full model includes an inconsistent variable, say X_1, the estimator $\hat{\boldsymbol{u}}$ must be biased, and the bias will affect the score test statistic including the one for the model of $X_2, \ldots, X_p$. Can the score test statistic with the biased estimator $\hat{\boldsymbol{u}}$ correctly evaluate a fit of the model of $X_2, \ldots, X_p$? The answer is "yes." It will be shown in Section 4 that small deviation from the model does result in bias of $\hat{\boldsymbol{u}}$, but does not affect the score test statistic asymptotically.

Kano and Ihara (1994) and Kano and Harada (2000) have proposed variable selection procedure for exploratory factor analysis. This paper extends their results to a general curved exponential family of distributions. Furthermore, we establish a theorem stated in the paragraph above to provide theoretical foundation with variable selection procedure developed by Kano and his collaborators.

In Section 5 we discuss a forward selection procedure. Two empirical examples are presented in Section 6. We end with remarks.

2. Curved exponential family

Let $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(n)}$ be independently and identically distributed p-dimensional random vectors and assume that their distributions belong to the exponential family with a natural parameter vector $\boldsymbol{\theta}$:

$$\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(n)} \overset{i.i.d.}{\sim} \exp(\boldsymbol{\theta}'\boldsymbol{x} - \psi(\boldsymbol{\theta}))\mu(d\boldsymbol{x}). \tag{2}$$

Here $\mu(\cdot)$ is a carrier measure independent of $\boldsymbol{\theta}$ and $\exp(-\psi(\boldsymbol{\theta}))$ is a normalizing constant (see e.g., Amari, 1985). The parameter space of $\boldsymbol{\theta}$ is given by

$$\Omega_\theta = \left\{ \boldsymbol{\theta} \in R^p \;\Big|\; \int \exp(\boldsymbol{\theta}'\boldsymbol{x})\mu(d\boldsymbol{x}) < \infty \right\}.$$

We suppress the superscript α to write $\boldsymbol{X}$ simply, when no confusion is created. Assume that $\psi(\boldsymbol{\theta})$ is twice continuously differentiable on Ω_θ, and then

$$E[\boldsymbol{X}] = \frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{\psi}(\boldsymbol{\theta}) \ (= \boldsymbol{\eta}, \ \text{say}) \tag{3}$$

$$\mathrm{Var}(\boldsymbol{X}) = \frac{\partial^2 \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \Psi(\boldsymbol{\theta}). \tag{4}$$

See e.g., Lehmann (1986). The parameter space Ω of $\boldsymbol{\eta}$ is then given as

$$\Omega = \boldsymbol{\psi}(\Omega_\theta) \ (\subset R^p).$$

Suppose that $\Psi(\boldsymbol{\theta})$ is positive definite on Ω_θ.

Consider a model with mean structure $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{u})$, where $\boldsymbol{u}$ is a q-vector with $q < p$ and $\boldsymbol{\eta}(\boldsymbol{u})$ is a differentiable p-dimensional vector-valued function from a domain $D_u (\subset R^q)$. Now we are interested in the goodness-of-fit test of the structural model:

$$T_0 \ \ldots \quad H_0 : E[\boldsymbol{X}] = \boldsymbol{\eta}(\boldsymbol{u}) \quad \text{versus} \quad A_0 : E[\boldsymbol{X}] \in \Omega. \tag{5}$$

In this paper, we employ the likelihood ratio criterion to construct a statistic to test a hypothesis as in (5), and call -2 times logarithm of the likelihood ratio statistic an LRT (statistic) briefly. We denote by T_0 the LRT statistic for testing (5) based on the sample in (2). The actual form of the T_0 will be given in (9).

Since $\Psi(\boldsymbol{\theta})$ is nonsingular, the inverse function theorem ensures existence of the inverse function $\boldsymbol{\psi}^{-1}(\cdot)$. Then use of $\boldsymbol{\psi}(\boldsymbol{\theta}) = \boldsymbol{\eta}(\boldsymbol{u})$ from (3) shows that

$$\boldsymbol{\theta} = \boldsymbol{\psi}^{-1}(\boldsymbol{\eta}(\boldsymbol{u})) \ (= \boldsymbol{\theta}(\boldsymbol{u}), \ \text{say}).$$

Thus, the population distribution is of *curved* exponential type (Amari, 1985).

Put $\Xi(\boldsymbol{u}) = \dfrac{\partial \boldsymbol{\eta}(\boldsymbol{u})}{\partial \boldsymbol{u}'}$, and we then have

$$\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{u}'} = \Psi^{-1}\left(\boldsymbol{\theta}(\boldsymbol{u})\right)\Xi(\boldsymbol{u}) \ (= \Psi^{-1}(\boldsymbol{u})\Xi(\boldsymbol{u}), \text{ say}). \tag{6}$$

Since the log likelihood based on the data $\boldsymbol{X}^{(\alpha)}$'s is written as $\ell(\boldsymbol{u}) = n\Big(\boldsymbol{\theta}(\boldsymbol{u})'\bar{\boldsymbol{X}} - \psi(\boldsymbol{\theta}(\boldsymbol{u}))\Big)$ with $\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{\alpha=1}^{n}\boldsymbol{X}^{(\alpha)}$, we obtain the score function and Fisher information of the structural parameter vector $\boldsymbol{u}$ as

$$\begin{aligned} \text{score function} &= n\Xi(\boldsymbol{u})'\Psi^{-1}(\boldsymbol{u})\left(\bar{\boldsymbol{X}} - \boldsymbol{\psi}(\boldsymbol{\theta}(\boldsymbol{u}))\right) & (7)\\ \text{Fisher information} &= n\Xi(\boldsymbol{u})'\Psi^{-1}(\boldsymbol{u})\Xi(\boldsymbol{u}) \ (= I(\boldsymbol{u}), \text{ say}). & (8)\end{aligned}$$

The MLE $\hat{\boldsymbol{\theta}}$ under A_0 in (5) is $\hat{\boldsymbol{\theta}} = \boldsymbol{\psi}^{-1}(\bar{\boldsymbol{X}})$. The MLE $\hat{\boldsymbol{u}}$ under H_0 in (5) is defined as a solution to the maximization problem: $\max_{\boldsymbol{u}\in D_u}\ell(\boldsymbol{u})$. Consistency of $\hat{\boldsymbol{u}}$ is guaranteed if the strong identifiability condition holds, i.e., let $\boldsymbol{u}_0$ be the true parameter, and for any $\epsilon > 0$ there is $\delta > 0$ such that if $||\boldsymbol{\eta}(\boldsymbol{u}) - \boldsymbol{\eta}(\boldsymbol{u}_0)|| < \delta$ then $||\boldsymbol{u} - \boldsymbol{u}_0|| < \epsilon$ (see Rao 1973, Section 5e.2; Kano 1986). Assuming that $\boldsymbol{u}_0$ is an interior point of the domain D_u, then the MLE meets $\frac{\partial}{\partial \boldsymbol{u}}\ell(\boldsymbol{u}) = \boldsymbol{0}$ with probability going to one.

The LRT for testing (5) is equivalent to the following T_0, and T_0 has the expansion in (10) if $\sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\eta}) = O_p(1)$:

$$\begin{aligned} T_0 &= 2n\left(\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}(\hat{\boldsymbol{u}})\}'\bar{\boldsymbol{X}} - \{\psi(\hat{\boldsymbol{\theta}}) - \psi(\boldsymbol{\theta}(\hat{\boldsymbol{u}})\}\right) & (9)\\ &= n(\bar{\boldsymbol{X}} - \boldsymbol{\eta})'\left(\Psi^{-1} - \Psi^{-1}\Xi(\Xi'\Psi^{-1}\Xi)^{-1}\Xi'\Psi^{-1}\right)(\bar{\boldsymbol{X}} - \boldsymbol{\eta}) + o_p(1) & (10)\\ &\xrightarrow{L} \chi^2_{p-q} \quad \text{as } n \to \infty. \end{aligned}$$

The convergence to the central chi-square distribution χ^2_{p-q} holds when H_0 is true. For simplicity we occasionally write Ψ and Ξ for $\Psi(\boldsymbol{u})$ and $\Xi(\boldsymbol{u})$, respectively.

It should be noted that the inverse of $\Xi'\Psi^{-1}\Xi$ in (10) is replaced with a generalized inverse matrix $(\Xi'\Psi^{-1}\Xi)^{-}$ if Ξ is not of full column rank. For a generalized inverse, see Rao and Mitra (1971) or Rao (1973).

3. Derivation of test statistic for backward elimination

Partition $\boldsymbol{X} = [\underbrace{\boldsymbol{X}_1{}'}_{p_1}, \underbrace{\boldsymbol{X}_2{}'}_{p_2}]'$. Partition $E[\boldsymbol{X}] = \boldsymbol{\eta}(\boldsymbol{u}) = [\boldsymbol{\eta}_1(\boldsymbol{u})', \boldsymbol{\eta}_2(\boldsymbol{u})']'$ and $\bar{\boldsymbol{X}} = [\bar{\boldsymbol{X}}_1', \bar{\boldsymbol{X}}_2']'$ correspondingly.[1] We shall examine if $\boldsymbol{X}_1$ is inconsistent with the model considered in H_0 in (5). Let $\Omega_2(\subset R^{p_2})$ be the parameter space of $\boldsymbol{\eta}_2$ (without structure).

[1]There may be some elements of $\boldsymbol{u}$ that are unrelated to $\boldsymbol{\eta}_2(\boldsymbol{u})$. The forthcoming $\Xi_2 = \dfrac{\partial \boldsymbol{\eta}_2(\boldsymbol{u})}{\partial \boldsymbol{u}'}$ may not be of full column rank.

In addition to (5) consider the following testing problem:

$$T_2 \ \ldots \quad H_2 : E[\boldsymbol{X}_2] = \boldsymbol{\eta}_2(\boldsymbol{u}) \quad \text{versus} \quad A_2 : E[\boldsymbol{X}_2] \in \Omega_2. \tag{11}$$

The LRT statistic for the hypotheses in (11) is labeled as T_2.

The variables $\boldsymbol{X}_1$ may be regarded as inconsistent if T_0 indicates rejection and T_2 shows acceptance. When the dimension p of an observed vector $\boldsymbol{X}$ is large, there are an extremely large number of combinations of possibly inconsistent variables, and estimation of many structural models in multivariate analysis require extremely heavy computation. In addition, occurrence of improper solutions annoys researchers (e.g., Kano 1998). Thus it will be almost intractable to calculate test statistics T_2 for all models with deleting possible inconsistent variables.

The purpose of this section is to construct test statistics (asymptotically) equivalent to T_2 not by calculating T_2 exactly but as simple functions of the MLE $\hat{\boldsymbol{u}}$ under H_0. For this, Rao's score test approach is useful (Rao 1947).

One can not directly apply the likelihood ratio criterion to test the hypothesis H_0 against H_2 because the random vectors in these hypotheses are different. We introduce a new model that may play a role of interface between the two models. Let $\underline{\boldsymbol{u}} = [\eta_1 \ldots, \eta_{p_1}, \boldsymbol{u}']'$. Consider the following structure:

$$\underline{\boldsymbol{\eta}}(\underline{\boldsymbol{u}}) = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_{p_1} \\ \boldsymbol{\eta}_2(\boldsymbol{u}) \end{bmatrix}, \tag{12}$$

where $\eta_1, \ldots, \eta_{p_1}$ are free parameters to be estimated. The structure (12) implies that $E[\boldsymbol{X}_1]$ does not have structure. The derivatives of the structure is expressible in the form:

$$\frac{\partial \underline{\boldsymbol{\eta}}(\underline{\boldsymbol{u}})}{\partial \underline{\boldsymbol{u}}'} = \begin{bmatrix} I_{p_1} & O \\ O & \Xi_2(\boldsymbol{u}) \end{bmatrix} \left(= \Xi(\underline{\boldsymbol{u}}), \ \text{say} \right). \tag{13}$$

Consider the testing problems as follows:

$$T_{2'} \ \ldots \qquad H_{2'} : E[\boldsymbol{X}] = \underline{\boldsymbol{\eta}}(\underline{\boldsymbol{u}}) \quad \text{versus} \quad A_0 : E[\boldsymbol{X}] \in R^p \tag{14}$$

$$T_{02'} \ \ldots \qquad H_0 : E[\boldsymbol{X}] = \boldsymbol{\eta}(\boldsymbol{u}) \quad \text{versus} \quad H_{2'} : E[\boldsymbol{X}] = \underline{\boldsymbol{\eta}}(\underline{\boldsymbol{u}}) \tag{15}$$

Here $T_{2'}$ and $T_{02'}$ denote the LRT statistics for (14) and (15), respectively.

We have defined the four LRT statistics, which have the following relations:

$$T_{02'} = T_0 - T_{2'} \tag{16}$$

$$T_2 = T_{2'} + o_p(1) \tag{17}$$

The equality (16) is obvious. The asymptotic equivalence in (17) will be expected since both T_2 and $T_{2'}$ test the structure $E[\boldsymbol{X}_2] = \boldsymbol{\eta}_2(\boldsymbol{u})$. A mathematical proof of (17) will be given in the Appendix. It follows from (16) and (17) that

$$T_2 = T_{2'} + o_p(1) = T_0 - T_{02'} + o_p(1).$$

Thus, it is seen that if a test statistic for (15) is formed as Rao's score test instead of $T_{02'}$, we can construct a test statistic for (11) which is a (simple) function of $\hat{\boldsymbol{u}}$. The score test is sometimes called a Lagrange multiplier test (Aitchison and Silvey 1958). For a comprehensive review, see Buse (1982).

Calculate the following quantities under $H_{2'} : E[\boldsymbol{X}] = \underline{\boldsymbol{\eta}}(\underline{\boldsymbol{u}})$:

$$\begin{aligned}
\boldsymbol{\theta} &= \boldsymbol{\psi}^{-1}(\underline{\boldsymbol{\eta}}(\underline{\boldsymbol{u}})) \; \left(= \underline{\boldsymbol{\theta}}(\underline{\boldsymbol{u}}), \quad \text{say} \right) \\
\frac{\partial \underline{\boldsymbol{\theta}}(\underline{\boldsymbol{u}})}{\partial \underline{\boldsymbol{u}}'} &= \Psi^{-1}(\underline{\boldsymbol{u}}) \frac{\underline{\boldsymbol{\eta}}(\underline{\boldsymbol{u}})}{\partial \underline{\boldsymbol{u}}'} = \Psi^{-1}(\underline{\boldsymbol{u}}) \begin{bmatrix} I_{p_1} & O \\ O & \Xi_2(\boldsymbol{u}) \end{bmatrix} = \Psi^{-1}(\underline{\boldsymbol{u}}) \underline{\Xi}(\underline{\boldsymbol{u}}),
\end{aligned}$$

where $\Xi_2(\boldsymbol{u}) = \dfrac{\partial \boldsymbol{\eta}_2(\boldsymbol{u})}{\partial \boldsymbol{u}'}$. The score function and Fisher information under the model (12) are given as

$$\begin{aligned}
\text{score function} &= n \underline{\Xi}(\underline{\boldsymbol{u}})' \Psi^{-1}(\underline{\boldsymbol{u}}) \left(\bar{\boldsymbol{X}} - \boldsymbol{\psi}(\underline{\boldsymbol{\theta}}(\underline{\boldsymbol{u}})) \right) \\
\text{Fisher information} &= n \underline{\Xi}(\underline{\boldsymbol{u}})' \Psi^{-1}(\underline{\boldsymbol{u}}) \underline{\Xi}(\underline{\boldsymbol{u}}) \; (= I(\underline{\boldsymbol{u}}), \; \text{say}).
\end{aligned}$$

In Rao's score statistic, we estimate $\underline{\boldsymbol{u}}$ by using estimators under H_0, that is,

$$\underline{\hat{\boldsymbol{u}}} = \begin{bmatrix} \widehat{\eta_1} \\ \vdots \\ \eta_{p_1} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\eta}_1(\hat{\boldsymbol{u}}) \\ \hat{\boldsymbol{u}} \end{bmatrix}.$$

Notice that $\underline{\boldsymbol{\eta}}(\underline{\hat{\boldsymbol{u}}}) = \boldsymbol{\eta}(\hat{\boldsymbol{u}})$ and we have

$$\begin{aligned}
\underline{\boldsymbol{\theta}}(\underline{\hat{\boldsymbol{u}}}) &= \boldsymbol{\psi}^{-1}(\underline{\boldsymbol{\eta}}(\underline{\hat{\boldsymbol{u}}})) = \boldsymbol{\psi}^{-1}(\boldsymbol{\eta}(\hat{\boldsymbol{u}})) \\
\boldsymbol{\psi}(\underline{\boldsymbol{\theta}}(\underline{\hat{\boldsymbol{u}}})) &= \boldsymbol{\eta}(\hat{\boldsymbol{u}}).
\end{aligned}$$

Using those expressions above, we obtain Rao's score statistic for (15) as follows:

$$\begin{aligned}
T_{02'RAO} &= (\widehat{\text{score}})' (\text{Fisher } \widehat{\text{information}})^{-} (\widehat{\text{score}}) \\
&= n \left(\bar{\boldsymbol{X}} - \boldsymbol{\eta}(\hat{\boldsymbol{u}}) \right)' \Psi^{-1}(\underline{\hat{\boldsymbol{u}}}) \underline{\Xi}(\underline{\hat{\boldsymbol{u}}}) \left(\underline{\Xi}(\underline{\hat{\boldsymbol{u}}})' \Psi^{-1}(\underline{\hat{\boldsymbol{u}}}) \underline{\Xi}(\underline{\hat{\boldsymbol{u}}}) \right)^{-} \\
&\qquad \cdot \underline{\Xi}(\underline{\hat{\boldsymbol{u}}})' \Psi^{-1}(\underline{\hat{\boldsymbol{u}}}) \left(\bar{\boldsymbol{X}} - \boldsymbol{\eta}(\hat{\boldsymbol{u}}) \right). \qquad (18)
\end{aligned}$$

Here we denote by A^- a generalized inverse of a matrix A.

The next formula is useful when $\Xi = \begin{bmatrix} I & O \\ O & \Xi_2 \end{bmatrix}$ (cf. Kano and Ihara 1994 formula (A4)):

$$\Psi^{-1}\Xi\left(\Xi'\Psi^{-1}\Xi\right)^{-}\Xi'\Psi^{-1}$$
$$= \Psi^{-1} - \begin{bmatrix} O & O \\ O & \Psi_{22}^{-1} - \Psi_{22}^{-1}\Xi_2\left(\Xi_2'\Psi_{22}^{-1}\Xi_2\right)^{-}\Xi_2'\Psi_{22}^{-1} \end{bmatrix}, \tag{19}$$

where Ψ_{22} is the $(2,2)$ block of Ψ. Application of the formula to (18) leads to

$$T_{02'RAO} = n\left(\bar{\boldsymbol{X}} - \boldsymbol{\eta}(\hat{\boldsymbol{u}})\right)'\Psi^{-1}(\underline{\hat{\boldsymbol{u}}})\left(\bar{\boldsymbol{X}} - \boldsymbol{\eta}(\hat{\boldsymbol{u}})\right) \tag{20}$$
$$-n\left(\bar{\boldsymbol{X}}_2 - \boldsymbol{\eta}_2(\hat{\boldsymbol{u}})\right)'\left\{\Psi_{22}^{-1}(\underline{\hat{\boldsymbol{u}}}) - \Psi_{22}^{-1}(\underline{\hat{\boldsymbol{u}}})\Xi_2(\underline{\hat{\boldsymbol{u}}})\left(\Xi_2(\underline{\hat{\boldsymbol{u}}})'\Psi_{22}^{-1}(\underline{\hat{\boldsymbol{u}}})\Xi_2(\underline{\hat{\boldsymbol{u}}})\right)^{-}\right.$$
$$\left.\cdot\Xi_2(\underline{\hat{\boldsymbol{u}}})'\Psi_{22}^{-1}(\underline{\hat{\boldsymbol{u}}})\right\}\left(\bar{\boldsymbol{X}}_2 - \boldsymbol{\eta}_2(\hat{\boldsymbol{u}})\right). \tag{21}$$

Under the assumption that $\sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\eta}) = O_p(1)$ we have

$$\sqrt{n}\left(\bar{\boldsymbol{X}} - \boldsymbol{\eta}(\hat{\boldsymbol{u}})\right) = (I_p - \Xi(\Xi'\Psi^{-1}\Xi)^{-1}\Xi'\Psi^{-1})\sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\eta}) + o_p(1). \tag{22}$$

We can verify the asymptotic equivalence between (20) and T_0 by substitution of the expression (22) into (20). Thus, the LRT statistic T_2 is asymptotically equivalent to (21).

In sum, we have

$$T_2 = T_0 - T_{02'} + o_p(1) = T_0 - T_{02'RAO} + o_p(1) = T_{2RAO} + o_p(1),$$

where

$$T_{2RAO} = n\left(\bar{\boldsymbol{X}}_2 - \boldsymbol{\eta}_2(\hat{\boldsymbol{u}})\right)'\left\{\Psi_{22}^{-1}(\underline{\hat{\boldsymbol{u}}}) - \Psi_{22}^{-1}(\underline{\hat{\boldsymbol{u}}})\Xi_2(\underline{\hat{\boldsymbol{u}}})\left(\Xi_2(\underline{\hat{\boldsymbol{u}}})'\Psi_{22}^{-1}(\underline{\hat{\boldsymbol{u}}})\Xi_2(\underline{\hat{\boldsymbol{u}}})\right)^{-}\right.$$
$$\left.\cdot\Xi_2(\underline{\hat{\boldsymbol{u}}})'\Psi_{22}^{-1}(\underline{\hat{\boldsymbol{u}}})\right\}\left(\bar{\boldsymbol{X}}_2 - \boldsymbol{\eta}_2(\hat{\boldsymbol{u}})\right). \tag{23}$$

The test statistic T_{2RAO} is basically a function of $\hat{\boldsymbol{u}}$ and one can use it to test (11).

We shall study statistical properties of the test statistic (23) in the next section.

4. Distribution of the test statistic: Null and nonnull cases

In this section, we shall focus upon how the test statistic (23) behaves when $\boldsymbol{X}_1$ is inconsistent. While the inconsistent $\boldsymbol{X}_1$ does not influence upon the exact LRT statistic T_2, it does affect the statistic (23) since the estimator $\hat{\boldsymbol{u}}$ used in (23) depends on $\boldsymbol{X}_1$.

To study the behavior we introduce a contiguous hypothesis or population drift (see e.g., LeCam 1960; Stroud 1972; Browne and Shapiro 1989):

$$H_{2''}: E[\boldsymbol{X}] = \boldsymbol{\eta}(\boldsymbol{u}) + \frac{1}{\sqrt{n}} \begin{bmatrix} \boldsymbol{d}_1 \\ \boldsymbol{0} \end{bmatrix} \ \left(= \boldsymbol{\eta}(\boldsymbol{u}) + \frac{1}{\sqrt{n}}\boldsymbol{d}, \text{ say} \right). \tag{24}$$

The hypothesis reduces to H_0 in (5) when $\boldsymbol{d} = \boldsymbol{0}$. Under the $H_{2''}$, we have

$$\begin{aligned} \sqrt{n}(\bar{\boldsymbol{X}}_2 - \boldsymbol{\eta}_2) &\xrightarrow{L} N_{p_2}(\boldsymbol{0}, \Psi_{22}) \\ \sqrt{n}(\hat{\boldsymbol{u}} - \boldsymbol{u}) &= (\Xi'\Psi^{-1}\Xi)^{-1}\Xi\Psi^{-1}\sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\eta}) + o_p(1) \\ &\xrightarrow{L} N_q\left((\Xi'\Psi^{-1}\Xi)^{-1}\Xi\Psi^{-1}\boldsymbol{d}, (\Xi'\Psi^{-1}\Xi)^{-1}\right). \end{aligned} \tag{25}$$

As expected, the (asymptotic) bias $\frac{1}{\sqrt{n}}(\Xi'\Psi^{-1}\Xi)^{-1}\Xi\Psi^{-1}\boldsymbol{d}$ is introduced to the estimator $\hat{\boldsymbol{u}}$. To what extent does the bias influence on the test statistics (23)? It should be noted from (25) that no bias is introduced to $\bar{\boldsymbol{X}}_2$.

Let us partition $\Xi = \begin{bmatrix} \Xi_1 \\ \Xi_2 \end{bmatrix}$, and we then have

$$\begin{aligned} \sqrt{n}\left(\bar{\boldsymbol{X}}_2 - \boldsymbol{\eta}_2(\hat{\boldsymbol{u}})\right) &= \sqrt{n}(\bar{\boldsymbol{X}}_2 - \boldsymbol{\eta}_2) - \sqrt{n}(\boldsymbol{\eta}_2(\hat{\boldsymbol{u}}) - \boldsymbol{\eta}_2) \\ &= \sqrt{n}(\bar{\boldsymbol{X}}_2 - \boldsymbol{\eta}_2) - \Xi_2(\Xi'\Psi^{-1}\Xi)^{-1}\Xi'\Psi^{-1}\sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\eta}) + o_p(1). \end{aligned} \tag{26}$$

The bias appears in the second term in the RHS of (26). Substituting (26) into (23), the second term of (26) drops out, and finally we obtain

$$T_{2RAO} = n\left(\bar{\boldsymbol{X}}_2 - \boldsymbol{\eta}_2\right)'\left\{\Psi_{22}^{-1} - \Psi_{22}^{-1}\Xi_2\left(\Xi_2'\Psi_{22}^{-1}\Xi_2\right)^{-}\Xi_2'\Psi_{22}^{-1}\right\}\left(\bar{\boldsymbol{X}}_2 - \boldsymbol{\eta}_2\right) + o_p(1), \tag{27}$$

which is free from the bias.

Thus, convergence to the *central* chi-square distribution under $H_{2''}$ is established in view of (25) and (27). The degrees of freedom are $p_2 - q_2$, where $q_2 = \text{rank}(\Xi_2(\boldsymbol{u}))$.

On the other hand, we can show that if $\boldsymbol{X}_2$ is inconsistent, T_{2RAO} is asymptotically distributed according to the NONcentral chi-square. In fact, if

$$E[\boldsymbol{X}_2] = \boldsymbol{\eta}(\boldsymbol{u}) + \frac{1}{\sqrt{n}}\boldsymbol{d}_2,$$

then the noncentrality parameter is given as

$$\boldsymbol{d}_2'\left\{\Psi_{22}^{-1} - \Psi_{22}^{-1}\Xi_2\left(\Xi_2'\Psi_{22}^{-1}\Xi_2\right)^{-}\Xi_2'\Psi_{22}^{-1}\right\}\boldsymbol{d}_2.$$

It has been shown that T_{2RAO} asymptotically has the central chi-square distribution for a correctly specified structure for $\boldsymbol{X}_2$ (no inconsistent variable is contained in $\boldsymbol{X}_2$)

and the noncentral chi-square distribution for a misspecified structure for $\boldsymbol{X}_2$ (inconsistent variables are contained in $\boldsymbol{X}_2$), whether or not the structure for $\boldsymbol{X}_1$ is correctly specified. In other words, the statistic T_{2RAO} does work for the testing problem (11).

5. Test statistic for forward selection

Partition $\boldsymbol{X} = [\underbrace{\boldsymbol{X}_1{}'}_{p_1}, \underbrace{\boldsymbol{X}_2{}'}_{p_2}]'$, and partition $E[\boldsymbol{X}] = \boldsymbol{\eta}(\boldsymbol{u}) = [\boldsymbol{\eta}_1(\boldsymbol{u})', \boldsymbol{\eta}_2(\boldsymbol{u})']'$ correspondingly. Using an estimator for $\boldsymbol{u}$ based on the data $\boldsymbol{X}_2$, we shall construct a test statistic to test the hypothesis (5), that is, to examine whether $\boldsymbol{X}_1$ should be added to $\boldsymbol{X}_2$.

Assume that the structural model for $\boldsymbol{X}_2$ is true, that is, $E[\boldsymbol{X}_2] = \boldsymbol{\eta}_2(\boldsymbol{u})$. When the assumption is false, one should first take backward elimination procedure to identify inconsistent variables, rather than adding a new variable.

As noted in the footnote in Section 3, there may be parameters that are not actively related with $\boldsymbol{\eta}_2(\boldsymbol{u})$. We partition $\boldsymbol{u}$ as

$$\begin{aligned} \boldsymbol{u} &= [\boldsymbol{u}_1', \boldsymbol{u}_2']' \\ \boldsymbol{\eta}_2(\boldsymbol{u}) &= \boldsymbol{\eta}_2(\boldsymbol{u}_2) \\ \Xi_2 &= \frac{\partial \boldsymbol{\eta}_2(\boldsymbol{u})}{\partial \boldsymbol{u}_2'} \quad \text{full column rank.} \end{aligned}$$

That is, $\boldsymbol{u}_1$ is a parameter vector that is related only with $\boldsymbol{X}_1$.

Assume that we have an estimator $\hat{\boldsymbol{u}}_2$ at hand and that it is possible to construct an estimator $\tilde{\boldsymbol{u}}_1$ for $\boldsymbol{u}_1$ in some way. The purpose of this section is to form a test statistic for testing (5) based on $[\tilde{\boldsymbol{u}}_1', \hat{\boldsymbol{u}}_2']'$.

Let $\tilde{\boldsymbol{u}} = [\tilde{\boldsymbol{u}}_1', \hat{\boldsymbol{u}}_2']'$. Define

$$\begin{aligned} T &= n\left(\bar{\boldsymbol{X}} - \boldsymbol{\eta}(\tilde{\boldsymbol{u}})\right)' \left\{\Psi^{-1}(\tilde{\boldsymbol{u}}) - \Psi^{-1}(\tilde{\boldsymbol{u}})\Xi(\tilde{\boldsymbol{u}})\Big(\Xi(\tilde{\boldsymbol{u}})'\Psi^{-1}(\tilde{\boldsymbol{u}})\Xi(\tilde{\boldsymbol{u}})\Big)^{-1} \right. \\ &\qquad \left. \cdot \Xi(\tilde{\boldsymbol{u}})'\Psi^{-1}(\tilde{\boldsymbol{u}})\right\} \left(\bar{\boldsymbol{X}} - \boldsymbol{\eta}(\tilde{\boldsymbol{u}})\right). \end{aligned} \tag{28}$$

The test statistic is an extension of the goodness-of-fit statistic developed by Browne (1982, 1984) in a covariance structure model.

Under the assumption that $\sqrt{n}(\tilde{\boldsymbol{u}} - \boldsymbol{u}) = O_p(1)$, we can make the expansion as

$$\sqrt{n}\left(\bar{\boldsymbol{X}} - \boldsymbol{\eta}(\tilde{\boldsymbol{u}})\right) = \sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\eta}) - \Xi\sqrt{n}(\tilde{\boldsymbol{u}} - \boldsymbol{u}) + o_p(1), \tag{29}$$

so that the T is expressible in the form:

$$T = n(\bar{\boldsymbol{X}} - \boldsymbol{\eta})'\left(\Psi^{-1} - \Psi^{-1}\Xi(\Xi'\Psi^{-1}\Xi)^{-1}\Xi'\Psi^{-1}\right)(\bar{\boldsymbol{X}} - \boldsymbol{\eta}) + o_p(1).$$

It follows from the expression that T asymptotically has the central chi-square distribution under H_0: $E[\boldsymbol{X}] = \boldsymbol{\eta}(\boldsymbol{u})$ and the noncentral chi-square distribution under the $H_{2''}$ in (24) with the noncentrality parameter given as

$$\boldsymbol{d}' \left(\Psi^{-1} - \Psi^{-1}\Xi(\Xi'\Psi^{-1}\Xi)^{-1}\Xi'\Psi^{-1} \right) \boldsymbol{d}.$$

One question may arise why one needs to add external variables irrespective of the fact that $\boldsymbol{X}_2$ is correctly specified (no inconsistent variable is contained in $\boldsymbol{X}_2$). It is because a correctly specified larger model will make better or more efficient inference on $\boldsymbol{u}_2$ even though nuisance parameters are introduced. For details, see Kano, Bentler and Mooijaart (1993).

6. Examples

A typical example of curved exponential family of distributions for observed variables is a covariance structure model under multivariate normality assumption (e.g., Jöreskog 1970; Bentler 1986; Bollen 1989). Mean and covariance structure models under multivariate normal assumption also belong to this family (e.g., Browne and Arminger 1995). For an observed random p-vector $\boldsymbol{X}$, the model can be characterized as

$$\begin{cases} E[\boldsymbol{X}] &= \boldsymbol{\mu}(\boldsymbol{u}) \\ Var(\boldsymbol{X}) &= \Sigma(\boldsymbol{u}) \end{cases}. \tag{30}$$

Under multivariate normality assumption on $\boldsymbol{X}$, the probability density function is written as

$$\exp\left(\boldsymbol{\theta}_1'\boldsymbol{X} + \boldsymbol{\theta}_2'v(\boldsymbol{X}\boldsymbol{X}') - \psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \right),$$

where $\exp\{-\psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\}$ is defined as a normalizing constant. We have used a *vec*-operator $v(\cdot)$. Notation on *vec*-operators is described. The notation $vec(A)$ for a matrix A denotes a vector formed by stacking all column vectors of A in order, and the $v(A)$ for a square matrix A represents a vector consisting of the elements of lower triangular part, including diagonals, of A. We denote by D_p the operator such that $vec(A) = D_p v(A)$ for any symmetric matrix A of order p.

The mean and covariance structure model in (30) introduces a structure as

$$\begin{aligned} \boldsymbol{\theta}_1 &= \Sigma^{-1}(\boldsymbol{u})\boldsymbol{\mu}(\boldsymbol{u}) \\ \boldsymbol{\theta}_2 &= -D_p' vec(\Sigma^{-1}(\boldsymbol{u}))/2. \end{aligned}$$

In this section, we shall take an exploratory factor analysis model to study how close the statistic T_{2RAO} is to T_2, the exact likelihood ratio test statistic, and to illustrate usefulness of the variable selection procedure proposed. The factor analysis model with k factors is defined as follows: an observed random p-vector $\boldsymbol{X} = [X_1, \ldots, X_p]'$ is expressed as

$$\boldsymbol{X} = \boldsymbol{\mu} + \Lambda \boldsymbol{f} + \boldsymbol{e},$$

where $\boldsymbol{\mu}$ is a general mean vector, $\boldsymbol{f}$ is a random k-vector of common factors, $\boldsymbol{e}$ is a random p-vector of unique factors, and Λ is a factor loading matrix. Assume that $\mathrm{Var}(\boldsymbol{f}) = I_k$, $\mathrm{Cov}(\boldsymbol{f}, \boldsymbol{e}) = O$ and $\mathrm{Var}(\boldsymbol{e}) = \Psi$, a $p \times p$ diagonal matrix called a unique variance matrix. Under these assumptions, the model introduces a covariance structure:

$$\mathrm{Var}(\boldsymbol{X}) = \Lambda\Lambda' + \Psi. \tag{31}$$

The communality of the i-th variable X_i is defined as proportion of the squared length of the i-th row vector of Λ to the variance of X_i. Maximum likelihood factor analysis was developed by Lawley (1941) and is equivalent to Rao's canonical factor analysis (Rao 1955). For details on factor analysis, see e.g., Lawley and Maxwell (1971).

Variable selection in factor analysis has been made based on the magnitude of communalities (see e.g., Gorsuch 1988). Variables with a small value of communality are suggested to be dropped. From the statistical point of view, however, fit of the model considered or discrepancy between sample and the model should be of primary importance.

The first example is a data set of customer satisfaction analyzed in Churchill and Surprenant (1982). They studied customer satisfaction of virtual purchasers ($n = 180$) of flowers (chrysanthemum) with five questions X_1 to X_5 in a questionnaire. The sample correlation matrix is reproduced in Table 1. One dimensional latent factor, F_1, of customer satisfaction was introduced behind the observed questions $X_1, \ldots, X_5$, that is, one-factor analysis model is assumed for the observed questions. They discussed causal mechanism of the customer satisfaction with structural equation modeling (covariance structure analysis). They suggested deletion of X_5 in the one-factor model since the reliability of X_5 is row (page 498).

The reliability corresponds to the communality in factor analysis, which is reported in the last column of Table 2. As pointed out by Churchill and Surprenant, the communality estimate for X_5 is .420, which is rather lower than those of the other variables.

Now we shall examine goodness-of-fit of the factor analysis model. The LRT statistic T_0 for goodness of fit shows 11.033, whose p-value is 0.0507. The model is barely accepted at $\alpha = .05$ level. The column of T_2 in Table 2 shows results of the LRT for the five models obtained by deleting one variable in order from the five-variable model. The Churchill and Surprenant model, removing X_5, receives a poorer fit ($T_2 = 6.585$, p-value=0.0372). It is seen from the table that the Churchill and Surprenant model is worst, and the model removing X_2 or X_4 receives a quite good fit. Thus, variable selection with the magnitude of communalities does not lead to a well-fitted model in this example.

Results of the score statistic developed in this paper for the data set are shown in the column of T_{2RAO} in the table. Those values are enough close to the exact LRT T_2. No misleading will occur concerning the variable selection if T_{2RAO} is used for T_2. Computational burden of T_{2RAO} is much lighter than that of T_2.

The next example is a fairly large data set, Twenty-four Psychological Variables ($p = 24$, n=145, $k = 4$) presented in Harman (1976, page 124). While the four-factor solution has been accepted as an interpretable model (e.g., Jöreskog 1978; Akaike 1987), the chi-square test rejects the model [$\chi^2_{186} = 227.140$, P-value=0.0213].

Table 3 shows results of T_{2RAO}, T_2 and their differences along with communality estimates. Those statistics are listed in the order of the magnitude of T_2. It is seen that the order of the magnitude of T_{2RAO} is almost identical with that of T_2, and the differences of those statistics are very small and negligible. Thus, the T_{2RAO} performs very nicely.

We have five statistically accepted models. Which model should be chosen among the five may depend on interpretability.

Communality estimates are shown in the last column of the table along with the order of the magnitude of the communalities from the lowest in the parenthesis. Again, the traditional magnitude-of-communality rule does not work to find a well-fitted model.

The example was analyzed also by Kano and Harada (2000). The comparison between the score test and the LRT was not made there.

The Web-based application, namely SEFA, was developed for variable selection in exploratory factor analysis. The reader who wants to use the SEFA can access at http://koko15.hus.osaka-u.ac.jp/~harada/factor/stepwise/

6. Final remarks

In this paper we have discussed theory of variable selection in a general structural model in multivariate analysis. Rao's score test approach to construct a goodness-of-fit test statistic for all models obtained by removing every set of possibly inconsistent variables successfully reduces computation, and resultant statistics seem enough accurate.

The accuracy comes from the theoretical foundation, which is established in this paper, that when inconsistent variables $\boldsymbol{X}_1$ are included, Rao's score test statistic for the model removing $\boldsymbol{X}_1$ can be approximated by the *central* chi-square distribution even though the $\boldsymbol{X}_1$ cause bias to the MLE for the structural parameter vector, which is used to form the score test statistic.

A natural question may arise: what if a model and a (original) model removing some variables from the original model are both acceptable? Which model should be adopted? One answer is that the original larger model should be chosen because a larger model will make better statistical inference in general, as noted in Section 4 (Kano, Bentler and Mooijaart 1993).

An alternative approach would be to apply information criteria such as AIC (e.g., Akaike 1987). Note, however, that one can not compare between a large model and a small model directly, that is, AIC can not say anything about

$$H_0 : E[\boldsymbol{X}] = \boldsymbol{\eta}(\boldsymbol{u}) \quad \text{versus} \quad H_2 : E[\boldsymbol{X}_2] = \boldsymbol{\eta}_2(\boldsymbol{u}).$$

Instead, we can propose users to use AIC to compare

$$H_0 : E[\boldsymbol{X}] = \boldsymbol{\eta}(\boldsymbol{u}) \quad \text{versus} \quad H_{2'} : E[\boldsymbol{X}] = \underline{\boldsymbol{\eta}}(\underline{\boldsymbol{u}}),$$

where $\underline{\boldsymbol{\eta}}(\underline{\boldsymbol{u}})$ was given in (12). For getting AIC, one needs to calculate $T_{2'}$ for every model. If T_{2RAO} is used for $T_{2'}$, computation will be reduced greatly.

Notice that

$$T_0 - T_{2'} = T_0 - T_2 + o_p(1) = T_0 - T_{2RAO} + o_p(1).$$

If the difference $T_0 - T_{2RAO}$ is larger than the twice of the difference in degrees of freedom, AIC indicates preference of the smaller model in $H_{2'}$ or H_2. Thus, AIC provides information as to whether a variable should be removed.

Acknowledgment The author expresses his special thanks to Professor Michael Browne of Ohio State University. His comments to the paper of Kano and Harada (2000) concerning the (alternative) distribution of a test statistic when a variable inconsistent with a model exists enhanced the author's motivation for carrying out this research. He also thanks Professor Yasunori Fujikoshi of Hiroshima University for affording him an excellent opportunity to contribute to this special number.

Appendix

We shall provide a concise proof of the asymptotic equivalence in (17) here. As in (10), we have

$$T_{2'} = n(\bar{\boldsymbol{X}} - \boldsymbol{\eta})' \left(\Psi^{-1} - \Psi^{-1}\Xi(\Xi'\Psi^{-1}\Xi)^{-}\Xi'\Psi^{-1}\right)(\bar{\boldsymbol{X}} - \boldsymbol{\eta}) + o_p(1).$$

Recall that Ξ is given as in (13), and apply the identity in (19). We then have the asymptotic equivalence.

References

Aitchison, J., Silvey, S. D., 1958. Maximum likelihood estimation of parameters subject to restrains. Ann. Math. Statist., 29, 813-828.

Akaike, H., 1987. Factor analysis and AIC. Psychometrika, 52, 317-332.

Amari, S. 1985. Differential-Geometrical Methods in Statistics. Lecture Notes in Statistics, 28, Springer-Verlag, Berlin.

Bentler, P. M., 1986. Structural modeling and Psychometrika: An historical perspective on growth and achievements. Psychometrika, 51, 35-51.

Bollen, K. A., 1989. Structural Equations with Latent Variables. Wiley Interscience, New York.

Browne, M. W., 1982. Covariance structures. In Hawkins, D. M. (Ed.), Topics in Applied Multivariate Analysis. Cambridge University Press, Cambridge, pp. 72-141.

Browne, M. W., 1984. Asymptotically distribution-free methods for the analysis of covariance structures. Br. Jour. Math. Statist. Psychol., 37, 62-83.

Browne, M. W., Arminger, G., 1995. Specification and estimation of mean- and covariance-structure model. In Arminger, G., Clogg, C. C., Sobel, M. E. (Eds.), Handbook of Statistical Modeling for the Social and Behavioral Sciences, Plenum Press, New York.

Browne, M. W., Shapiro, A., 1989. Robustness of normal theory methods in the analysis of linear latent variates models. Br. Jour. Math. Statist. Psychol., 41, 193-208.

Buse, A., 1982. The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. Amer. Statist., 36, 153-157.

Churchill, G. A., Surprenant, C., 1982. An investigation into the determinations of customer satisfaction. Jour. Marketing Res., 19, 491-504.

Fujikoshi, Y., 1985. Selection of variables in two-group discriminant analysis by error rate and Akaike's information criteria. Jour. Multivariate Anal., 17, 27-37.

Gorsuch, R. L., 1988. Exploratory factor analysis. In Nesselroade, J. R., Cattell, R. B. (Eds.), Handbook of Multivariate Experimental Psychology (2nd ed.), Plenum Press, New York and London, pp. 231-258.

Harman, H. H., 1976. Modern Factor Analysis (3rd ed.), The University of Chicago Press, Plenum Press.

Jolliffe, I. T., 1986. Principal Component Analysis. Springer-Verlag, New York.

Jöreskog, K. G., 1970. A general method for analysis of covariance structures. Biometrika, 57, 239-251.

Jöreskog, K. G., 1978. Structural analysis of covariance and correlation matrices. Psychometrika, 43, 443-477.

Kano, Y., 1986. Conditions on consistency of estimators in covariance structure model. Jour. Japan Statist. Soc., 16, 75-80.

Kano, Y., 1998. Improper solutions in exploratory factor analysis: Causes and treatments. In Rizzi, A., Vichi, M., Bock, H. (Eds.), Advances in Data Sciences and Classification, Springer, Berlin, pp. 375-382.

Kano, Y., Bentler, P. M., Mooijaart, A., 1993. Additional information and precision of estimators in multivariate analysis. In Matusita, K., Puri, M. L., Hayakawa, T. (Eds.), Statistical Sciences and Data Analysis: Proceedings of the Third Pacific Area Statistical Conference. VSP International Science Publisher, Zeist, The Netherlands, pp. 187-196.

Kano, Y., Harada, A., 2000. Stepwise variable selection in factor analysis. Psychometrika, 65, 7-22.

Kano, Y., Ihara, M., 1994. Identification of inconsistent variates in factor analysis. Psychometrika, 59, 5-20.

Lawley, D. N., 1941. Further investigations in factor estimation. Proc. Roy. Soc. Edinb., A 61, 176-185.

Lawley, D. N., Maxwell, A. E., 1971. Factor Analysis as a Statistical Method (2nd ed.). London, Butterworths.

LeCam, L., 1960. Locally asymptotically normal families of distributions. Univ. California Press Statist., 3, 37-88.

Lehmann, E. L., 1986. Testing Statistical Hypotheses (2nd ed.). John Wiley and Sons, New York.

Miller, A. J., 1984. Selection of subsets of regression variables. Jour. Roy. Statist. Soc., A 147, 389-425.

Rao, C. R., 1947. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. Proc. Cambridge Philos. Soc., 44, 50-57.

Rao, C. R., 1955. Estimation and tests of significance in factor analysis. Psychometrika, 20, 93-111.

Rao, C. R., 1973. Linear Statistical Inference and its Application (2nd ed.), Wiley, New York.

Rao, C. R., Mitra, S. K., 1971. Generalized Inverse of Matrices and its Applications. John Wiley and Sons, New York.

Shibata, R., 1976. Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika, 63, 117-126.

Stroud, Y. W., 1972. Fixed alternatives and Wald's formulation of the noncentral asymptotic behaviour of the likelihood ratio statistic. Ann. Math. Statist., 43, 447-454.

Tanaka, Y., 1983. Some criteria for variable selection in factor analysis. Behaviormetrika, 13, 31-45.

Yanai, H., 1980. A proposition of generalized method for forward selection of variables. Behaviormetrika, 7, 95-107.

TABLE 1

Sample correlation matrix of customer satisfaction data (n=180), reproduced from Churchill and Surprenant (1982)

X1	1				
X2	.76	1			
X3	.69	.71	1		
X4	.74	.86	.75	1	
X5	.57	.57	.57	.59	1

TABLE 2

Comparison of T_{2RAO} with T_2 in customer satisfaction data

$\chi^2_5(.05) = 5.991$, $\chi^2_5(.05) = 11.070$

variable to remove	df	T_{2RAO}	T_2	difference	communality (MLE for 5 variables)
X1	2	4.062	4.136	-0.074	0.676
X2	2	0.192	0.408	-0.216	0.839
X3	2	3.990	4.020	-0.030	0.647
X4	2	0.946	1.304	-0.358	0.860
X5	2	6.604	6.585	0.019	0.420
all variables included	5	11.033	11.033	0.000	

T_{2RAO} and T_2 denote the statistic developed here and the usual likelihood ratio test statistic, respectively. The difference means that $T_{2RAO} - T_2$.

TABLE 3

Comparison of T_{2RAO} with T_2 in 24 Psychological Data

$\chi^2_{186}(.05) = 218.820$, $\chi^2_{167}(.05) = 198.154$

variable to remove	df	T_{2RAO}	T_2	difference	communality(its order) (MLE for 24 variables)
X11	167	190.126	190.014	0.112	0.449 (12)
X3	167	193.219	193.222	-0.003	0.358 (6)
X5	167	195.376	195.252	0.124	0.648 (20)
X19	167	197.481	197.423	0.058	0.239 (2)
X13	167	197.726	197.548	0.178	0.511 (16)
X23	167	199.521	199.489	0.032	0.503 (15)
X9	167	201.029	200.717	0.312	0.744 (23)
X17	167	202.537	202.428	0.109	0.389 (7)
X8	167	203.195	203.157	0.038	0.516 (17)
X4	167	203.322	203.274	0.048	0.349 (5)
X7	167	204.121	203.767	0.354	0.718 (22)
X6	167	204.475	204.353	0.122	0.689 (21)
X21	167	204.474	204.438	0.036	0.416 (11)
X18	167	205.314	205.102	0.212	0.409 (10)
X14	167	205.391	205.177	0.214	0.340 (4)
X2	167	206.812	206.801	0.011	0.219 (1)
X12	167	207.305	207.215	0.090	0.567 (19)
X20	167	207.247	207.221	0.026	0.409 (9)
X16	167	208.798	208.568	0.230	0.452 (13)
X22	167	208.927	208.897	0.030	0.399 (8)
X10	167	211.196	209.215	1.981	0.760 (24)
X24	167	209.542	209.405	0.137	0.501 (14)
X1	167	209.703	209.661	0.042	0.561 (18)
X15	167	214.625	214.595	0.030	0.306 (3)
all variables included	186	227.140	227.140	0.000	

T_{2RAO} and T_2 denote the statistic developed here and the usual likelihood ratio test statistic, respectively. The difference means that $T_{2RAO} - T_2$.