PROPERTIES OF TESTS FOR DETERMINING THE NUMBER OF FACTORS IN THE MULTIVARIATE STRUCTURAL AND FUNCTIONAL RELATIONSHIPS

YASUO AMEMIYA

TECHNICAL REPORT NO. 16
MAY 1986

PREPARED UNDER
NATIONAL SCIENCE FOUNDATION GRANT DMS 82-19748

ECONOMETRIC WORKSHOP
FOURTH FLOOR, ENCINA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
PROPERTIES OF TESTS FOR DETERMINING THE NUMBER
OF FACTORS IN THE MULTIVARIATE STRUCTURAL
AND FUNCTIONAL RELATIONSHIPS

Yasuo Amemiya
Iowa State University and Stanford University

TECHNICAL REPORT NO. 16
May 1986

PREPARED UNDER
NATIONAL SCIENCE FOUNDATION GRANT DMS 82-19748

Econometric Workshop
Fourth Floor, Encina Hall
Stanford University
Stanford, California
Abstract.

In the multivariate structural and functional relationship models, the most commonly used test statistics for choosing the number of factors are the likelihood ratio and the residual sum of squares goodness-of-fit statistics. The models with the error covariance matrix known and known up to a multiple are considered. When the number of fitted factors is less than the true number, the asymptotic properties of the two statistics are derived for both the fixed alternative and the local alternative cases. The asymptotic behavior of the test statistics is studied also under the assumption that the number of fitted factors is greater than the true number. A Monte Carlo result supporting the asymptotic theory is presented. Some recommendations concerning the use of the two tests in determining the number of factors are given.


Key words and phrases: Tests of rank, asymptotic comparison of tests, errors in variables.
1. Introduction.

The multivariate structural and functional relationship models have been used widely in social and behavioral sciences. One of practically important statistical problems associated with the models is the determination of the number of factors (equivalently, the number of relationships, or the rank of the systematic part). For many types of structural relationship models, certain goodness-of-fit tests derived under the normal assumption have been used extensively in choosing the number of factors. The goodness-of-fit tests are designed to examine the fit of a model against a class of alternatives that includes, but is not restricted to, the models with an incorrectly specified number of factors. Because the goodness-of-fit tests have a relatively simple null distribution (limiting $\chi^2$) and are widely available in many computer packages, the use of the goodness-of-fit tests for selecting the number of factors is very popular in the application of various structural relationship models including the factor analysis model. In this paper we shall investigate the powers of the two goodness-of-fit tests in detecting incorrectly specified number of factors for two particular versions of functional and structural relationships.

Suppose that a set of $p \times 1$ observable vectors $Z_t$ satisfies

\begin{align}
Z_t = \mu_Z + \Lambda f_t + \varepsilon_t, \quad t = 1, 2, \ldots, N,
\end{align}

where the $\mu_Z$ is a $p \times 1$ fixed vector, the $f_t$'s are fixed or random $k \times 1$ unobservable vectors, the $\varepsilon_t$'s are $p \times 1$ unobservable random vectors with mean zero and covariance matrix $\Sigma_{ee}$, and $\Lambda$ is a $p \times k$ unknown matrix of rank $k$. The systematic part of the $p \times 1$ observation $Z_t$ is explained by the $k \times 1$ factor vector $f_t$. The model (1.1) is equivalent to assume

\begin{align}
Z_t = z_t + \varepsilon_t, \\
B'z_t = \beta_0, \quad t = 1, 2, \ldots, N,
\end{align}

where the $z_t$'s are fixed or random $p \times 1$ unobservable vectors, $\beta_0$ is an $r \times 1$ unknown vector, $B$ is a $p \times r$ unknown matrix of rank $r$, and $p = k + r$. The $p \times 1$ systematic part $z_t$ satisfies $r$ linear relationships. The model with fixed $f_t$ or $z_t$ is called the functional relationship. The
model with random \( f_t \) or \( z_t \) is called the structural relationship. For the structural case the \( \xi_t \)'s are assumed to be independent of the \( f_t \)'s or the \( z_t \)'s. We call the model the normal structural model, if both the factor and error are normally distributed. Several versions of the model exist depending on different types of structure assumed about \( \Sigma_{\xi\xi} \). In this paper we consider two cases: (a) \( \Sigma_{\xi\xi} \) known and (b) \( \Sigma_{\xi\xi} = \sigma^2 \Sigma_0 \) and \( \Sigma_0 \) known. The case (a) arises in practice when the error covariance matrix \( \Sigma_{\xi\xi} \) is estimated based on a large number of degrees of freedom either by replications or from past data. The case (a) plays an important role in theoretical development of many other models, because estimation of the systematic part of the model with any error covariance structure can be formulated as estimation in case (a) with the maximum likelihood estimator of \( \Sigma_{\xi\xi} \) treated to be known. See, e.g., Amemiya (1985, 1986a). The case (b) with \( \Sigma_0 = I_p \), the \( p \times p \) identity matrix, is the classical errors–in–variables model which has been discussed extensively in the literature. The structural model with case (b) is equivalent to the principal component analysis model where unimportant components have a common variance. We shall assume that in both cases (a) and (b) \( \Sigma_{\xi\xi} \) is positive definite. We do not discuss the cases with singular \( \Sigma_{\xi\xi} \) to keep our development in a reasonable simplicity. For many normal structural models with various structures on \( \Sigma_{\xi\xi} \), two popular goodness–of–fit test statistics are the likelihood ratio statistic \( L \) and the residual sum of squares statistic \( R \). Let the sample covariance matrix be denoted by

\[
m = \frac{1}{n} \sum_{i=1}^{N} (\tilde{z}_t - \bar{Z})(\tilde{z}_t - \bar{Z})',
\]

where \( \bar{Z} = (1/N) \sum_{i=1}^{N} \tilde{z}_t \), and \( n = N - 1 \) is the number of degrees of freedom. If in model

(1.2) \( \tilde{z}_t \sim N(\mu_x, \Sigma_{xx}) \) and \( \tilde{\xi}_t \sim N(0, \Sigma_{\xi\xi}) \), then \( m \sim W_p(\Sigma, n) \), where \( n = N - 1 \),

(1.3)

\[
\Sigma = \Sigma_{xx} + \Sigma_{\xi\xi},
\]

\[
B'\Sigma_{xx}B = 0.
\]

The statistic \( L \) is \(-2 \log \lambda \), where \( \lambda \) is the likelihood ratio criterion based on \( m \) for the model specification (1.3) against the alternative that \( \Sigma \) is an unrestricted positive definite matrix. The statistic \( R \) is the residual sum of squares obtained from the final stage of the iteratively
rewighted least squares estimation based on \( m \). We shall consider the use of \( L \) and \( R \) for the functional model and structural model where the distribution of \( f \) is not necessarily normal.

Assume that the model with \( k_0 \) factors is fitted. For case (a) where \( \Sigma_{ee} \) is a known positive definite matrix, the two statistics are

\[
L_a(k_0) = n \sum_{i=k_0+1}^{p} (-\log \hat{\lambda}_i + \hat{\lambda}_i - 1),
\]

\[
R_a(k_0) = \frac{n}{2} \sum_{i=k_0+1}^{p} (\hat{\lambda}_i - 1)^2,
\]

where \( \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p \) are the roots of

\[
| \tilde{m} - \lambda \tilde{\Sigma_{ee}} | = 0,
\]

\( k_0 = \min\{k_0, r\} \), and \( r \) is the number of \( \hat{\lambda}_i > 1 \). For case (b) where \( \tilde{\Sigma_{ee}} = \sigma^2 \tilde{\Sigma_0} \) and \( \tilde{\Sigma_0} \) is known, the two statistics are

\[
L_b(k_0) = n \left[ r_0 \log \hat{\sigma}^2(k_0) - \sum_{i=k_0+1}^{p} \log \hat{\gamma}_i \right],
\]

\[
R_b(k_0) = \frac{n}{2} \sum_{i=k_0+1}^{p} \left[ \frac{\hat{\gamma}_i}{\hat{\sigma}^2(k_0)} - 1 \right]^2,
\]

where \( \hat{\gamma}_1 \geq \hat{\gamma}_2 \geq \ldots \geq \hat{\gamma}_p \) are the roots of \( | \tilde{m} - \gamma \tilde{\Sigma_0} | = 0, r_0 = p - k_0 \), and

\[
\hat{\sigma}^2(k_0) = \frac{1}{r_0} \sum_{i=k_0+1}^{p} \hat{\gamma}_i.
\]

See, e.g., Lawley (1956), Browne (1974), Anderson (1984), and Amemiya (1985). The residual sum of squares statistics \( R_a \) and \( R_b \) have not been used widely, because for cases (a) and (b) the computation of the maximum likelihood estimator does not require an iterative procedure.

We shall show that \( R_a \) and \( R_b \) are strong competitors to \( L_a \) and \( L_b \), respectively. By studying these two special cases, we hope to obtain some insight into the general comparison of the statistics \( L \) and \( R \) for the model with any error covariance structure.

The multivariate structural and functional relationships were discussed thoroughly by Anderson (1984). The normal structural model with known \( \Sigma_{ee} \) was studied by Lawley (1953).
and Theobald (1975). Tintner (1945) and Theobald (1975) discussed the functional model with known $\Sigma_{ee}$. Gleser (1981) treated the functional model with $\Sigma_{ee} = \sigma^2 I_p$. The normal structural model with $\Sigma_{ee} = \sigma^2 I_p$ and its relationship to the principal component analysis was discussed in Anderson (1963). The literature on other types of the multivariate structural and functional relationships includes Anderson (1951), Anderson and Rubin (1956), Lawley and Maxwell (1971), Jöreskog (1981), and Amemiya and Fuller (1984). Bartlett (1954) proposed the use of the statistics $L_a$ and $L_b$. Lawley (1956) suggested certain modifications to $L_a$ and $L_b$ to improve the $\chi^2$ approximation for the normal structural model. The residual sum of squares statistics are discussed in, e.g., Browne (1974, 1984) and Bentler (1983). Amemiya (1985) discussed both the likelihood ratio and the residual sum of squares statistics for a large class of functional and structural relationships with general error covariance structure. Amemiya and Anderson (1985) studied the two test statistics for the factor analysis models. All existing works on the two test statistics have concentrated on the null distribution. In this paper we consider the nonnull distribution and power comparison.

Section 2 presents a canonical reduction of the problem that will be useful in studying properties of the test statistics. We study the tests under the model where the number of fitted factors is smaller than the true number in Section 3 (fixed alternative) and in Section 4 (local alternative). Section 5 deals with the case where the number of fitted factors is larger than the true number. A numerical result is given in Section 6. Some recommendations concerning the use of the two tests in selecting the number of factors are given in Section 7. The two test statistics considered here are functions of the characteristic roots. The limiting distribution results for the characteristic roots are presented in the appendix.

2. Canonical Reduction.

To discuss distributional properties of the statistics $L_a, R_a, L_b,$ and $R_b$, it is convenient to transform the model into a canonical form. For case (b), using the relationship $\tilde{\Sigma}_{ee} = \sigma^2 \tilde{\Sigma}_0$
where \( o^2 \) is an unknown true value, we can write

\[
L_0(k_0) = n \left[ r_0 \log \bar{\lambda}(k_0) - \sum_{i=k_0+1}^{p} \log \hat{\lambda}_i \right]
\]

\[
R_0(k_0) = \frac{n}{2} \sum_{i=k_0+1}^{p} \left( \frac{\hat{\lambda}_i}{\bar{\lambda}(k_0)} - 1 \right)^2,
\]

where \( \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p \) are the roots of (1.5), \( r_0 = p - k_0 \) and

\[
\bar{\lambda}(k_0) = \frac{1}{r_0} \sum_{i=k_0+1}^{p} \hat{\lambda}_i.
\]

Thus, the two types of statistics are functions of the roots of (1.5) in both cases (a) and (b). The roots of (1.5) are invariant with respect to a transformation \( \tilde{Z}_t = T'Z_t \), i.e.,

\[
m^* = T'm, \quad \Sigma^* = T'\Sigma_T T,
\]

where \( T \) is a nonsingular matrix. We shall choose a particular matrix \( T \) so that the problem is simplified. In the models (1.1) and (1.2), the matrices \( \Lambda \) and \( B \) are not identified. See, e.g., Anderson (1984). To discuss distributional properties of the test statistics, we assume that the model (1.1) holds and is identified by some restrictions on \( \Lambda \), and treat \( \Lambda \) and \( B \) as particular fixed true values satisfying \( B'\Lambda = 0 \). For the structural case, let \( \Sigma_{ff} \) be the covariance matrix of \( f_t \). For the functional case, the corresponding matrix is the sum of squares and cross product matrix

\[
m_{ff} = \frac{1}{n} \sum_{i=1}^{N} (f_t - \bar{f})(f_t - \bar{f})',
\]

where \( \bar{f} = (1/N) \sum_{t=1}^{N} f_t \). We assume that \( \Sigma_{ff} \) and \( m_{ff}(n \geq k) \) are positive definite. Then,

\[
\mathcal{E}\{m\} = \Lambda m_{ff} \Lambda' + \Sigma_{cco}, \quad \text{functional case},
\]

\[
= \Lambda \Sigma_{ff} \Lambda' + \Sigma_{cco}, \quad \text{structural case}.
\]

We choose a matrix \( T \) such that \( T'\Sigma_{cco} T = I_p \) and \( T'\mathcal{E}\{m\} T \) is diagonal. Let

\[
G = (\Lambda' \Sigma^{-1}_{cco} \Lambda)^{1/2} m_{ff} (\Lambda' \Sigma^{-1}_{cco} \Lambda)^{1/2}, \quad \text{functional case},
\]

\[
= (\Lambda' \Sigma^{-1}_{cco} \Lambda)^{1/2} \Sigma_{ff} (\Lambda' \Sigma^{-1}_{cco} \Lambda)^{1/2}, \quad \text{structural case},
\]

(2.1)
where for a symmetric positive definite $A$, $A^{\frac{1}{2}}$ is the symmetric positive square root matrix satisfying $A^{\frac{1}{2}}A^{\frac{1}{2}} = A$. Also let $Q$ be a $k \times k$ orthogonal matrix and $D_\eta = \text{diag}\{\eta_1, \eta_2, \ldots, \eta_k\}$ be a $k \times k$ diagonal matrix such that $\eta_1 \geq \eta_2 \geq \ldots \geq \eta_k > 0$, and $G = QD_\eta Q'$. Let

\begin{equation}
\lambda_i = \eta_i + 1, \quad i = 1, 2, \ldots, k,
\end{equation}

\begin{equation}
= 1, \quad i = k + 1, \ldots, p,
\end{equation}

\begin{equation}
D_\lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_p\}.
\end{equation}

The model structure with $k$ factors and $r$ relationships is characterized by the fact that $k$ $\lambda_i$'s are greater than one and $r$ $\lambda_i$'s are one. Define $T = (T_1, T_2)$, where

\begin{equation}
T_1 = \Sigma_\epsilon^{-1} \Lambda (\Lambda' \Sigma_\epsilon^{-1} \Lambda)^{-\frac{1}{2}} Q,
\end{equation}

\begin{equation}
T_2 = B (B' \Sigma_\epsilon B)^{-\frac{1}{2}}.
\end{equation}

Then, $T_1' \Sigma_\epsilon T_1 = I_p$, and $T_1' \xi \{m\} T_2 = D_\lambda$. We note that for the functional case $T_1$ and $\lambda_i, i = 1, 2, \ldots, k$, depend on $n$ because $m_{ff}$ depends on $n$. Using this $T$, the roots $\hat{\lambda}_i$'s of (1.5) are the characteristic roots of

\begin{equation}
S = T'mT,
\end{equation}

and estimators of the $\lambda_i$'s.

Since $B' \Lambda = 0$, the corresponding transformation of the observation $Z_t$ can be written as

\begin{equation}
Z^*_t = T'_t Z_t = \mu^*_t + (f^*_t)^{\sim} + (u_t)^{\sim} (v_t)^{\sim},
\end{equation}

where

\begin{equation}
\mu^*_t = T'_t \mu_z, \quad f^*_t = Q' (\Lambda' \Sigma_\epsilon^{-1} \Lambda)^{\frac{1}{2}} f_t,
\end{equation}

\begin{equation}
u_t = T'_t \xi t, \quad \upsilon_t = T'_2 \xi t.
\end{equation}

The $f^*_t$ is the normalized factor vector satisfying

\begin{equation}
\frac{1}{n} \sum_{t=1}^{N} (f^*_t - \bar{f}^*) (f^*_t - \bar{f}^*)' = D_\eta
\end{equation}
for the functional case, and \( \text{Var}\{f_i\} = D_\eta \) for the structural case. The \( \sim u_i \) and \( \sim v_i \) are the normalized error vectors satisfying \( \text{Var}\{u_i', v_i'\} = I_p \). The systematic part is concentrated in the first \( k \) components of the \( p \times 1 \) vector \( Z_i^* \), and the remaining part \( v_i \) is the \( r \times 1 \) residual vector. Under weak assumptions the roots \( \lambda_i \)'s of (1.5) are consistent for the \( \lambda_i \)'s in (2.2). Thus, \( \tilde{\lambda}_i, i = k + 1, k + 2, \ldots, p, \) are estimators of one. Both \( L_a(k_0) \) and \( R_a(k_0) \) in (1.4) have the form of the sum of \( p - k_0 \) terms where each term measures a distance between \( \tilde{\lambda}_i \) and one.

For case (b), the \( \hat{\lambda}_i \)'s are unobservable, and the observed roots \( \hat{\gamma}_i = \sigma^2 \tilde{\lambda}_i, i = k + 1, k + 2, \ldots, p, \) are estimators of \( \sigma^2 \). The statistics \( L_b(k_0) \) and \( R_b(k_0) \) compare \( \hat{\gamma}_i, i = k_0 + 1, k_0 + 2, \ldots, p, \) to their average \( \delta^2(k_0) \).

3. Fixed Alternatives.

We assume the model with \( k \) factors, and study the behavior of \( L_a(k_0), R_a(k_0), L_b(k_0), \) and \( R_b(k_0) \) under the assumption that \( k_0 < k \), i.e., that the number of fitted factors is less than the true number of factors, or equivalently, the number of fitted relationships is greater than the true number of relationships. In this section the true number of factors is assumed to stay fixed as \( n \to \infty \). We shall consider a local alternative case in Section 4 where the true number of factors approaches to a given number as \( n \to \infty \).

First we present assumptions that will be used later. For both the structural and functional cases, we assume

(i) The model (1.1) holds with some fixed \( \tilde{A} \) of rank \( k \), and \( \tilde{B} \) of rank \( r = p - k \) satisfying \( \tilde{B}' \tilde{A} = 0 \).

(ii) The \( \tilde{\varepsilon}_i \)'s are independently and identically distributed \( N(0, \Sigma_{\varepsilon}) \) random vectors.

For the structural model, we assume

(S1) For each \( n \geq p \), the \( \tilde{f}_i \)'s, \( t = 1, 2, \ldots, n \), are independent of the \( \tilde{\varepsilon}_i \)'s, and are independently and identically distributed with mean \( \tilde{\mu}_f \), positive definite covariance matrix \( \Sigma_{ff} \), and finite fourth moments.
(S2) The characteristic roots \( \eta_i, i = 1, 2, \ldots, k, \) of \( G \) defined in (2.1) have a general multiplicity structure

\[
D_\gamma = \text{diag}\{\eta_1, \eta_2, \ldots, \eta_k\} = \text{block diag}\{\gamma_1 I_{q_1}, \gamma_2 I_{q_2}, \ldots, \gamma_s I_{q_s}\},
\]

where \( \gamma_1 > \gamma_2 > \ldots > \gamma_s > 0 \) and \( \sum_{j=1}^{s} q_j = k. \)

Under (S1) and (S2), \( f_i \) can be any nonnormal random vector with nonsingular covariance matrix and finite fourth moments. Note that if (S1) holds then (S2) is simply a notation rather than a restriction. Under (S2) \( D_\gamma \) defined in (2.2) is

\[
D_\gamma = \text{block diag}\{(\gamma_1 + 1) I_{q_1}, \ldots, (\gamma_s + 1) I_{q_s}, I_r\}.
\]

For the functional case, we recall that \( G, \eta_i, \) and \( D_\gamma \) depend on \( n \). To distinguish from the structural case, and to emphasize the dependency on \( n \), we write \( G(n), \eta_i(n), \) and \( D_\gamma(n) \). For the functional case, we assume

(F1) For each \( n \geq p, \) the \( f_i \)'s, \( t = 1, 2, \ldots, n, \) are fixed constants.

(F2) The characteristic roots \( \eta_i(n), i = 1, 2, \ldots, k, \) of \( G(n) \) defined in (2.1) have a general multiplicity structure

\[
D_\eta(n) = \text{diag}\{\eta_1(n), \ldots, \eta_k(n)\} = \text{block diag}\{\gamma_1(n) I_{q_1}, \ldots, \gamma_s(n) I_{q_s}\},
\]

where

\[
\gamma_j(n) \rightarrow \gamma_j, \quad j = 1, 2, \ldots, s, \quad \text{as } n \rightarrow \infty,
\]

\[
\gamma_1 > \gamma_2 > \ldots > \gamma_s > 0.
\]

Note that under (F2) the multiplicity of \( \eta_i(n) \) is the same for all \( n \). The limit condition in (F2) holds if for some positive definite \( \bar{m}_{ff} \),

\[
\bar{m}_{ff} \rightarrow \bar{m}_{ff}, \quad \text{as } n \rightarrow \infty.
\]

Under (F2),

\[
D_\gamma(n) = \text{block diag}\{[\gamma_1(n) + 1] I_{q_1}, \ldots, [\gamma_s(n) + 1] I_{q_s}, I_r\}.
\]
We first note that the distributions of the statistics under the null model with \( k = k_0 \) have been studied in the literature. If assumptions (i), (ii), and either (S1) and (S2) or (F1) and (F2) hold, then as \( n \to \infty \)

\[
\begin{align*}
L_a(k) &\overset{L}{\to} \chi^2_r, \\
R_a(k) &\overset{L}{\to} \chi^2_r, \\
L_b(k) &\overset{L}{\to} \chi^2_{r-1}, \\
R_b(k) &\overset{L}{\to} \chi^2_{r-1},
\end{align*}
\]

where \( r = \frac{1}{2}r(r+1) \). In fact, for the structural case, (3.1) holds if \( f \) has only finite second moments. See Amemiya (1985). The goodness-of-fit test procedures reject the model (a) if \( L_a(k_0) \) or \( R_a(k_0) \) is larger than \( \chi^2_{r,\alpha} \), the upper \( \alpha \) quantile, and reject the model (b) if \( L_b(k_0) \) or \( R_b(k_0) \) is larger than \( \chi^2_{r-1,\alpha} \). Such procedures have an asymptotic significance level \( \alpha \). We consider the powers of the test when \( k_0 = k - q, 0 < q < k \), i.e., when \( q \) factors are ignored in the model fitting. We assume that \( q \) is any integer satisfying \( 0 < q < k \) and may be greater than \( q \) in assumptions (S2) and (F2). The following theorem gives the first order approximations to \( L_a(k_0), R_a(k_0), L_b(k_0), \) and \( R_b(k_0) \) for both the structural model and the functional model.

**Theorem 1.** Let assumptions (i) and (ii) hold. Assume that \( k_0 = k - q, 0 < q < k \). If (S1) and (S2) hold, then

\[
\begin{align*}
L_a(k_0) &= nC_{La} + O_p\left(n^{\frac{1}{2}}\right), \quad R_a(k_0) = nC_{Ra} + O_p\left(n^{\frac{1}{2}}\right), \\
L_b(k_0) &= nC_{Lb} + O_p\left(n^{\frac{1}{2}}\right), \quad R_b(k_0) = nC_{Rb} + O_p\left(n^{\frac{1}{2}}\right),
\end{align*}
\]

and if (F1) and (F2) hold, then

\[
\begin{align*}
L_a(k_0) &= nC_{La} + o_p(n), \quad R_a(k_0) = nC_{Ra} + o_p(n), \\
L_b(k_0) &= nC_{Lb} + o_p(n), \quad R_b(k_0) = nC_{Rb} + o_p(n),
\end{align*}
\]
where

\[ C_{La} = \sum_{i=k_0+1}^{k} (-\log(1 + \eta_i) + \eta_i), \]
\[ C_{Ra} = \frac{1}{2} \sum_{i=k_0+1}^{k} \eta_i^2, \]
\[ C_{Lb} = r_0 \log \left( 1 + \frac{1}{r_0} \sum_{i=k_0+1}^{k} \eta_i \right) - \sum_{i=k_0+1}^{k} \log(1 + \eta_i), \]
\[ C_{Rb} = \frac{1}{2} r_0 (r_0 - 1) \left( r_0 + \sum_{i=k_0+1}^{k} \eta_i \right)^{-2} \left( \sum_{i=k_0+1}^{k} \eta_i \right)^{2}, \]

and \( r_0 = p - k_0. \)

**Proof.** If (S1) and (S2) hold, then by Lemma 2 in the appendix, \( \sqrt{n}(\hat{\lambda}_i - \lambda_i), i = 1, 2, \ldots, p, \) converges in distribution to a random variable, where

\[ \lambda_i = 1 + \eta_i, \quad i = 1, 2, \ldots, k, \]
\[ \lambda_i = 1, \quad i = k + 1, k + 2, \ldots, p. \]

If (F1) and (F2) hold, then by Lemma 2 in the appendix, \( \sqrt{n}(\hat{\lambda}_i - \lambda_i(n)), i = 1, 2, \ldots, p, \) converges in distribution to a random variable, where

\[ \lambda_i(n) = 1 + \eta_i(n), \quad i = 1, 2, \ldots, k, \]
\[ \lambda_i(n) = 1, \quad i = k + 1, k + 2, \ldots, p. \]

By (F2), \( \eta_i(n) \to \eta_i, i = 1, 2, \ldots, k. \) Thus, under (F1) and (F2), \( \hat{\lambda}_i - \lambda_i = o_p(1), i = 1, 2, \ldots, p. \)

Note that under either (S1) and (S2) or (F1) and (F2), as \( n \to \infty, \)

\[ P\{k_0^* = k_0\} = P\{\hat{\lambda}_{k_0} > 1\} \to 1. \]

Hence, if we substitute \( \lambda_i \) and \( k_0 \) in place of \( \hat{\lambda}_i \) and \( k_0^* \), then the error is \( O_p(n^{-\frac{1}{2}}) \) under (S1) and (S2), and is \( o_p(1) \) under (F1) and (F2). Therefore, the results follow by the substitution.

\[ \blacksquare \]

It is easy to see that if \( \eta_i > 0 \) for some \( i = k_0 + 1, \ldots, k, \) then \( C_{La} > 0, C_{Ra} > 0, C_{Lb} > 0 \) and \( C_{Rb} > 0. \) Hence, by Theorem 1, the four statistics based on the \( k_0 \) factor model tend to
∞ at the rate \( n \) for both the structural and functional models with \( k = k_0 + q \) factors. Thus, the following corollary is immediate. Note that \( L_a(k_0) \) and \( R_a(k_0) \) are compared to \( \chi^2_{\tau_0, \alpha} \), and \( L_b(k_0) \) and \( R_b(k_0) \) are compared to \( \chi^2_{\tau_0 - 1, \alpha} \), where \( \tau_0 = \frac{1}{2} r_0(r_0 + 1), r_0 = p - k_0. \)

**Corollary 1.** Let assumptions (i), (ii), and either (S1) and (S2) or (F1) and (F2) hold. Assume that \( k_0 = k - q, 0 < q < k \). Then, as \( n \to \infty \)

\[
\begin{align*}
P\{L_a(k_0) > \chi^2_{\tau_0, \alpha}\} &\to 1, \\
P\{R_a(k_0) > \chi^2_{\tau_0, \alpha}\} &\to 1, \\
P\{L_b(k_0) > \chi^2_{\tau_0 - 1, \alpha}\} &\to 1, \\
P\{R_b(k_0) > \chi^2_{\tau_0 - 1, \alpha}\} &\to 1.
\end{align*}
\]

**Proof.** For example, \( L_a(k_0)/n \to C_{L_a} > 0 \). Hence, for large \( n \)

\[
P\{L_a(k_0)/n > \chi^2_{\tau, \alpha}/n\} > P\{L_a(k_0)/n > C_{L_a}/2\},
\]

and the right hand side approaches one. Other results follow by the same argument. To show \( C_{L_b} > 0 \), we note that \( \log(1 + \lambda) \) is concave and use Jensen’s inequality for a random variable taking values \( \eta_i, i = k_0 + 1, \ldots, k \), and 0 \( r \) times with probability \( (1/r_0) \) each.

Therefore, for both cases (a) and (b), and for both the structural and functional models, the goodness-of-fit tests based on the likelihood ratio and the residual sum of squares have the usual nice property of power tending to one as \( n \to \infty \) when a less than enough number of factors are fitted. As we shall see in Section 5, neither tests have this property when a more than enough number of factors are fitted.

To compare the likelihood ratio test and the residual sum of squares test, we define the asymptotic p-value (asymptotic observed significance level) of a test.

**Definition.** Suppose a test statistic \( T \) satisfies that under the null hypothesis \( H_0, T \to \chi^2 \) as \( n \to \infty \). Suppose also that an asymptotic size \( \alpha \) test rejects \( H_0 \) if \( T > t_\alpha \) where \( P\{T_0 > t_\alpha\} = \alpha \). Then, the asymptotic p-value (asymptotic observed significance level) of this test given an observed value \( T = t \) is

\[
p(t) = P\{T_0 > t\}.
\]
The use of the asymptotic p-value is the same as that of the exact p-value. Hence, a smaller asymptotic p-value indicates a stronger evidence against the null hypothesis. Since \( p(t) \) is a function of \( t \), we can consider a random variable \( p(T) \). The next corollary compares the asymptotic p-values of the likelihood ratio test \( L_a \) and the residual sum of squares test \( R_a \) for case (a).

**Corollary 2.** Let assumptions (i), (ii), and either (S1) and (S2) or (F1) and (F2) hold. Assume that \( k_0 = k - q, 0 < q < k \). Then, as \( n \to \infty \)

\[
P \{ p[R_a(k_0)] < p[L_a(k_0)] \} \to 1.
\]

**Proof.** We note that two tests based on \( R_a(k_0) \) and \( L_a(k_0) \) have the common cut-off point \( \chi^2_{\alpha} \), and that \( R_a(k_0) \) and \( L_a(k_0) \) have the common asymptotic distribution \( \chi^2_{\alpha} \) under the null hypothesis of true number of factors \( k_0 \). Hence, it suffices to show that as \( n \to \infty \),

\[
P \left\{ \frac{1}{n} R_a(k_0) > \frac{1}{n} L_a(k_0) \right\} \to 1.
\]

Because by Theorem 1,

\[
\frac{1}{n} R_a(k_0) \xrightarrow{p} C_{Ra}, \quad \frac{1}{n} L_a(k_0) \xrightarrow{p} C_{La}
\]

it suffices to show that \( C_{Ra} > C_{La} \). Since for \( \eta > 0 \),

\[
\frac{1}{2} \eta^2 > -\log(1 + \eta) + \eta,
\]

it follows that \( C_{Ra} > C_{La} \). 

By Corollary 2, for case (a), under both the structural and functional models, we expect that for large \( n \) the residual sum of squares test tends to provide a stronger evidence against the model with \( k_0 \) factors when the true model has more than \( k_0 \) factors. In this sense we consider the residual sum of squares test to be more powerful than the likelihood ratio test for case (a).
4. Local Alternatives.

In the previous section we obtained some evidence showing that for case (a) the residual sum of squares test may be more powerful than the likelihood ratio test for large \( n \). Because the powers of the both types of tests against a fixed alternative tend to one as \( n \to \infty \), the comparison was difficult, and no useful procedure for obtaining approximate powers of the tests against a specific alternative was given. A standard tool for investigating powers in such a situation is to consider a local alternative under which the model approaches to the null model as \( n \to \infty \). It is known that if certain regularity conditions are satisfied, then the distribution of the likelihood ratio test approaches to a central \( \chi^2_d \) distribution under the null and to a noncentral \( \chi^2_d \) distribution under a local alternative, where the number of degrees of freedom \( d \) is common. Such a standard result does not apply directly to our problems, because the \( f_i \)'s are fixed for the functional model and have an unspecified distribution for the structural model. Even for the normal structural model, it does not seem to be trivial to apply the standard result to our testing problem. This is because it is not obvious to formulate our problem of testing the rank of the systematic part in the form of testing a hypothesis that a given subset of parameters are equal to specified values. We attempt to provide a direct and meaningful approach to formulating a local alternative for our testing problem.

We consider a sequence of models indexed by \( n \) under which each model has \( k \) factors and the number of factors tends to \( k_0 = k - q \) as \( n \to \infty \). To present the local alternative model assumptions, we use the transformed factor vector \( \tilde{f}_t^* \) defined in (2.4). Let

\[
\tilde{f}_t^* = (\tilde{f}_{t1}^*, \tilde{f}_{t2}^*)' = (\tilde{f}_{t1}^*, n^{-\frac{1}{2}} \tilde{f}_{t2}^*)',
\]

where \( \tilde{f}_{t2}^* \) and \( \tilde{f}_{t2}^0 \) are \( q \times 1 \), and \( 0 < q < k \). For the structural and functional models, respectively, we replace assumptions (S2) and (F2) by

(S2-a) The \( \tilde{f}_{t2}^0 \) has finite fourth moments, and the characteristic roots \( \eta_i, i = 1, 2, \ldots, k \), of \( G \) have a structure

\[
\text{diag}\{\eta_1, \ldots, \eta_k\} = \text{block diag}\left\{\gamma_1, \ldots, \gamma_{s-1}, \gamma_{s-1} - \frac{\delta_1}{\sqrt{n}}, \frac{\delta_2}{\sqrt{n}}, \ldots, \frac{\delta_q}{\sqrt{n}}\right\},
\]

where \( \gamma_1 > \gamma_2 > \ldots > \gamma_{s-1} > 0, \delta_1 \geq \delta_2 \geq \ldots \geq \delta_q > 0 \), and \( \sum_{j=1}^{q-1} q_j = k - q \).
(F2-a) The characteristic roots $\eta_i(n), i = 1, 2, \ldots, k$ of $G(n)$ have a structure

$$\text{diag}\{\eta_1(n), \ldots, \eta_k(n)\} = \text{block diag} \left\{ \gamma_1(n)_{\sim \delta_1/n}, \ldots, \gamma_{s-1}(n)_{\sim \delta_{s-1}/n}, \frac{\delta_1}{\sqrt{n}}, \frac{\delta_2}{\sqrt{n}}, \ldots, \frac{\delta_q}{\sqrt{n}} \right\},$$

where $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_q > 0, \sum_{j=1}^{s-1} q_j = k - q,$

$$\gamma_j(n) \to \gamma_j, \quad j = 1, 2, \ldots, s - 1, \quad \text{as } n \to \infty,$$

$$\gamma_1 > \gamma_2 > \ldots > \gamma_{s-1} > 0.$$

Under assumptions (i) and either (S2-a) or (F2-a), the model has $k$ factors for every finite $n,$ and has $k_0 = k - q$ factors in the limit. Under (S1) and (S2-a), the uncorrelated $q$ components of $f_{,t2}^*$ have variances $\delta_i/\sqrt{n}, i = 1, 2, \ldots, q.$ Under (F1) and (F2-a), the orthogonal $q$ components of $f_{,t2}^*, t = 1, 2, \ldots, N,$ have sums of squares $\delta_i/\sqrt{n}, i = 1, 2, \ldots, q.$ We consider asymptotic distributions of $L_a(k_0), R_a(k_0), L_b(k_0),$ and $R_a(k_0),$ where $k_0 = k - q.$ Thus, $q$ less than the true number of factors are fitted for every finite $n.$ The statistics $L_a(k_0)$ and $R_a(k_0)$ are compared to $\chi^2_{r_0,\alpha}$ and the statistics $L_b(k_0)$ and $R_b(k_0)$ are compared to $\chi^2_{r_0-1,\alpha},$ where $r_0 = \frac{1}{2}r_0(r_0 + 1), r_0 = p - k_0 = p - k + q.$

**Theorem 2.** Let assumptions (i), (ii), and either (S1) and (S2-a) or (F1) and (F2-a) hold. Assume that $k_0 = k - q, 0 < q < k.$ Then, as $n \to \infty,$

$$L_a(k_0) \xrightarrow{L} \chi^2_{r_0}(\phi_a),$$

$$R_a(k_0) \xrightarrow{L} \chi^2_{r_0}(\phi_a),$$

$$L_b(k_0) \xrightarrow{L} \chi^2_{r_0-1}(\phi_b),$$

$$R_b(k_0) \xrightarrow{L} \chi^2_{r_0-1}(\phi_b),$$

where $\chi^2_d(\phi)$ denotes the noncentral chi-squared distribution with degrees of freedom $d$ and
noncentrality parameter $\phi, r_0 = \frac{1}{2} r_0 (r_0 + 1), r_0 = p - k_0,$ and
\[
\phi_a = \frac{1}{2} \sum_{i=1}^{q} \delta_i^2,
\]
\[
\phi_b = \frac{1}{2} \left[ \sum_{i=1}^{q} \delta_i^2 - \frac{1}{r_0} \left( \sum_{i=1}^{q} \delta_i \right)^2 \right].
\]

Proof. We use the vec and vech operators. For a $p \times p$ symmetric $A = (a_1, a_2, \ldots, a_p)$, let $\text{vec } \tilde{A} = (a_1', a_2', \ldots, a_p')', p^2 \times 1$, and let $\text{vech } \tilde{A}$ be the $\frac{1}{2} p(p + 1) \times 1$ vector listing the elements of $\tilde{A}$ that are on or below the diagonal starting with the first column. Let $\tilde{\Phi}_p$ be the $p^2 \times \frac{1}{2} p(p+1)$ matrix such that $\text{vec } \tilde{A} = \tilde{\Phi}_p \text{ vech } \tilde{A}$, and let $\tilde{\psi}_p = (\tilde{\Phi}_p' \tilde{\Phi}_p)^{-1} \tilde{\Phi}_p^{-1}$. Then, vech $\tilde{A} = \tilde{\psi}_p \text{ vec } \tilde{A}$. For detailed discussions of vec and vech operators, see, e.g., Browne (1974) and Henderson and Searle (1979).

By Lemma 1 in the appendix, under either (S1) and (S2-a) or (F1) and (F2-a), as $n \to \infty$, the joint distribution of $\sqrt{n} (\tilde{\lambda}_i - 1), i = k - q + 1, \ldots, p$, converges to the joint distribution of the ordered characteristic roots of $\tilde{\Omega}$, where
\[
\text{vech } V \sim N \left[ \text{vech } \begin{pmatrix} D & 0 \\ 0 & \tilde{\Omega} \end{pmatrix}, \tilde{\Omega} \right],
\]
\[
\tilde{\Omega} = 2\psi_r \psi'_r,
\]
\[
D = \text{diag}\{\delta_1, \delta_2, \ldots, \delta_q\}.
\]

Since $\tilde{\lambda}_{k-q} P^{1} + \gamma_{s-1} > 1$,
\[
P\{k_0^* = k - q\} \to 1.
\]

Thus, under either the structural model or the functional model, by Taylor expansion,
\[
L_0(k_0) = F_0 + O_p \left( n^{-\frac{1}{2}} \right),
\]
\[
R_0(k_0) = F_0 + O_p \left( n^{-\frac{1}{2}} \right),
\]
\[
L_0(k_0) = F_0 + O_p \left( n^{-\frac{1}{2}} \right),
\]
\[
R_0(k_0) = F_0 + O_p \left( n^{-\frac{1}{2}} \right),
\]

15
where
\[
F_a = \frac{n}{2} \sum_{i=h^{-}q+1}^{p} (\hat{\lambda}_i - 1)^2, \\
F_b = \frac{n}{2} \sum_{i=h^{-}q+1}^{p} (\hat{\lambda}_i - 1)^2 - \frac{nr_0}{2} [\lambda(k - q) - 1]^2.
\]

As \( n \to \infty \),

\[
F_a \overset{L}{\to} \frac{1}{2} \text{tr } V^2 = (\text{vech } V)' P_a (\text{vech } V),
\]

(4.1)

\[
F_b \overset{L}{\to} \frac{1}{2} \text{tr } V^2 - \frac{1}{2r_0} (\text{tr } V)^2 = (\text{vech } V)' P_b (\text{vech } V),
\]

(4.2)

where
\[
P_a = \frac{1}{2} \tilde{\phi}'_r \tilde{\phi}_r, \\
P_b = \frac{1}{2} \tilde{\phi}'_r \left[ \frac{f_2}{r_0^2} - \frac{1}{r_0^2} \tilde{\ell}'_r \tilde{\ell}_r \right] \tilde{\phi}_r,
\]

and \( \tilde{\ell}_r \) is the \( r_0^2 \times 1 \) vector such that \( \tilde{\ell}'_r \text{vec } V = \text{tr } V \). Because
\[
(\tilde{\phi}'_r \tilde{\phi}_r)^{-1} = \psi_r \psi'_r, \\
\tilde{\ell}'_r \tilde{\phi}_r \psi_r = \tilde{\ell}'_r,
\]

it follows that \( P_a \Omega \) is an idempotent of rank \( r_0 \) and \( P_b \Omega \) is idempotent of rank \( r_0 - 1 \). Hence, \( F_a \) and \( F_b \) converge in distribution to noncentral \( \chi^2 \) distributions with \( r_0 \) and \( r_0 - 1 \) degrees of freedom, respectively. The noncentrality parameters are obtained by substituting
\[
\epsilon \{ V \} = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}
\]

into the right hand sides of (4.1) and (4.2).

Theorem 2 has shown that under our very general conditions on \( f_t \), the standard noncentral \( \chi^2 \) result holds for both the likelihood ratio and the residual sum of squares tests. The results of Theorem 2 can be used to obtain approximate powers of the tests for a specific model with more factors than fitted. However, the limiting distributions of the two types of
test statistics are common under a local alternative. Hence, combining the results of this section and the previous section, we conclude that for both structural and functional models and for both cases (a) and (b) the two goodness-of-fit tests based on the likelihood ratio and the residual sum of squares are effective tests in detecting the model with more factors than fitted and have similar powers against such alternatives with the residual sum of squares test being slightly more powerful for case (a). Up to the order of approximation we have used, there is no difference in powers between the structural model and the functional model with common $\eta_i$’s, i.e., common $C(m)$. That is, whether $f_i$ is fixed or random with an unspecified distribution, the asymptotic powers of the two types of tests depend only on $C(m)$. By Theorem 2, for all four statistics, the approximate powers increase if the $\delta_i$’s increase. Thus, if $k - q$ factors are fitted to the model with $k$ factors, then the powers are large when the $q$ smallest variances of the $k$ normalized factors are large compared to the error variances. In terms of the $\lambda_i$’s in (2.2), the powers are large if $\lambda_i, i = k - q + 1, \ldots, k$, are much larger than one. Comparing the two cases (a) and (b) where the case (a) assumes more information about the error covariance structure, we see that the limiting distribution under the local alternative for the case (a) has not only a larger degrees of freedom but also a larger noncentrality parameter than for the case (b).

5. Alternative of a Smaller Rank.

In this section we discuss the behavior of the test statistics when we fit more factors than needed. We assume the true model with $k$ factors, and consider $L_a(k_0), R_a(k_0), L_b(k_0)$, and $R_a(k_0)$ where $k_0 = k + \ell, 0 < \ell < r$. Under either case (a) or (b) the model with $k + \ell$ factors has a larger number of parameters than the model with $k$ factors. Thus, we may expect the four tests to have small powers, if any, against the model with a smaller number of factors. For the normal structural model defined in (1.3) it can be shown that $L_a(k_0)$ and $L_b(k_0)$ are the likelihood ratio test statistics for either $H_1$: the rank of $\Sigma_{xz}$ is $k_0$ or $H_2$: the rank of $\Sigma_{xz}$ is at most $k_0$, against the alternative of unrestricted $\Sigma$. The $R_a(k_0)$ and $R_b(k_0)$ can be used to test either $H_1$ or $H_2$. By considering properties of the tests when the true rank is less than $k_0$, we hope to obtain some useful information on the use of the four goodness-of-fit tests as tests
for choosing a proper number of factors. As before, we consider the problem for the functional model and the structural model with the distribution of the factor vector unspecified.

**Theorem 3.** Let assumption (i), (ii), and either (S1) and (S2) or (F1) and (F2) hold. Assume that \( k_0 = k + \ell, 0 < \ell < r \). Then, as \( n \to \infty \),

\[
L_a(k_0) \overset{L}{\to} G_a, \quad R_a(k_0) \overset{L}{\to} G_a,
\]

\[
L_b(k_0) \overset{L}{\to} G_b, \quad R_b(k_0) \overset{L}{\to} G_b,
\]

where

\[
G_a = \frac{1}{2} \left[ \sum_{i=1}^{\ell} \ell_i^2 I_{\{\ell_i < 0\}} + \sum_{i=\ell+1}^{r} \ell_i^2 \right],
\]

\[
G_b = \frac{1}{2} \left[ \sum_{i=\ell+1}^{r} \ell_i^2 - \frac{1}{r-\ell} \left( \sum_{i=\ell+1}^{r} \ell_i \right)^2 \right],
\]

\( I_{\{\}} \) is the indicator function, \( \ell_1 \geq \ell_2 \geq \ldots \geq \ell_r \) are the characteristic roots of \( \tilde{U} \), and

\[
\text{vech } \tilde{U} \sim N(0, 2\psi_r \psi_r').
\]

**Proof.** By Lemma 2 in the appendix, as \( n \to \infty \),

\[
P(k_0^* \geq k) = P(\hat{\lambda}_{k+1} > 1) \to 1,
\]

and \( \sqrt{n}(\hat{\lambda}_i - 1), i = k + 1, k + 2, \ldots, p \), converges in law to \( \ell_1 \geq \ell_2 \geq \ldots \geq \ell_r \). Thus, under either (S1) and (S2) or (F1) and (F2),

\[(5.1) \quad L_a(k_0) = n \left[ \sum_{i=k+1}^{k+\ell} (- \log \hat{\lambda}_i + \hat{\lambda}_i - 1) I_{\{\hat{\lambda}_i < 1\}} + \sum_{i=k+\ell}^{p} (- \log \hat{\lambda}_i + \hat{\lambda}_i - 1) \right] \]

\[
= \frac{n}{2} \left[ \sum_{i=k+1}^{k+\ell} (\hat{\lambda}_i - 1)^2 I_{\{\sqrt{n}(\hat{\lambda}_i - 1) < 0\}} + \sum_{i=k+\ell}^{p} (\hat{\lambda}_i - 1)^2 \right] + O_p \left( n^{-\frac{1}{2}} \right)
\]

\[
\overset{L}{\to} G_a = \frac{1}{2} \left[ \sum_{i=1}^{\ell} \ell_i^2 I_{\{\ell_i < 0\}} + \sum_{i=\ell+1}^{r} \ell_i^2 \right],
\]

where we used that the indicator function is bounded and that the expression in (5.1) is a continuous function of \( \sqrt{n}(\hat{\lambda}_i - 1), i = k + 1, \ldots, p \), except for a set with zero probability under
the distribution of \( \ell_i, i = 1, 2, \ldots, r \). See, e.g., Chernoff (1956, Theorem 2). The results for \( R_a(k_0), L_b(k_0) \), and \( R_a(k_0) \) follow by the same argument.

Based on the results of Theorem 3, we can make some observations concerning the use of the four tests. When a larger number of factors than the true number are fitted, the powers of the four tests do not increase with \( n \). To study the distributions of \( G_a \) and \( G_b \), we note that

\[
P(\ell_i \geq 0) > 0,
\]

\[
\frac{1}{2} \sum_{i=1}^{r} \ell_i^2 \sim \chi^2_r,
\]

\[
\frac{1}{2} \left[ \sum_{i=1}^{r} \ell_i^2 \sim \frac{1}{r} \left( \sum_{i=1}^{r} \ell_i \right)^2 \right] \sim \chi^2_{r-1},
\]

where \( \bar{r} = \frac{1}{2} r (r + 1) \). Thus, \( G_a \) is stochastically smaller than \( \chi^2_r \), and \( G_b \) is stochastically smaller than \( \chi^2_{r-1} \). The asymptotic null distributions of \( L_a(k_0) \) and \( R_a(k_0) \) are \( \chi^2_{\bar{r}_0} \), and the asymptotic null distributions of \( L_a(k_0) \) and \( R_b(k_0) \) are \( \chi^2_{\bar{r}_0-1} \), where

\[
\bar{r}_0 = \frac{1}{2} r_0 (r_0 + 1)
\]

\[
= \frac{1}{2} (r - \ell) (r - \ell + 1) < \bar{r}.
\]

Thus, it is unclear whether or not \( G_a \) and \( G_b \) are stochastically larger than the null distributions \( \chi^2_{\bar{r}_0} \) and \( \chi^2_{\bar{r}_0-1} \), respectively. But, we can conclude that the powers of the four tests when too many factors are fitted are small. Hence, the four statistics may not be very useful to test \( H_1 \): there exist exactly \( k \) factors, because the four tests may accept \( H_1 \) often when there are indeed less than \( k \) factors. On the other hand, in testing \( H_2 \): there are at most \( k \) factors, the four tests may not have a correct asymptotic significance level, because the tests may have a power larger than the asymptotic significance level when the true number of factors is strictly less than \( k \). As we shall recommend in Section 7, a meaningful use of these four tests in choosing the number of factors seems to be a successive application of the tests starting from a small value of \( k_0 \) and increasing the value of \( k_0 \).

In this section we discuss small sample properties of the tests based on $L_a$ and $R_a$ for case (a) using a Monte Carlo experiment. The model (1.1) with $p = 5, k = 2, \text{ and } r = 3$ is considered. Observations were generated using the model

$$Z_t = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} f_{t1} \\ f_{t2} \end{pmatrix} + \varepsilon_t, \quad t = 1, 2, \ldots, N,$$

where $\varepsilon_t \sim N(0, I_5)$. Two sample sizes $N = 29$ and $N = 43$ were used. For the factor vector $f_t$, the following four cases were considered for each of the two sample sizes $N = 29$ and 43.

[F-1] $f_t, t = 1, 2, \ldots, N$, are fixed over replications, and satisfy

$$m_{ ff } = \frac{1}{N-1} \sum_{t=1}^{N} (f_t - \bar{f})(f_t - \bar{f})' = \begin{pmatrix} 3 & 0 \\ 0 & \frac{1}{4} \end{pmatrix}.$$  

[F-2] $f_t, t = 1, 2, \ldots, N$, are fixed over replications, and satisfy

$$m_{ ff } = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}.$$  

[S-1] $f_t \sim N \left[ \begin{pmatrix} 3 \\ 0 \\ 0 \\ 0 \\ \frac{1}{4} \end{pmatrix} \right], t = 1, 2, \ldots, N.$  

[S-2] $f_t \sim N \left[ \begin{pmatrix} 3 \\ 0 \\ 0 \\ 0 \\ 2 \end{pmatrix} \right], t = 1, 2, \ldots, N.$

Note that $\Sigma = \mathcal{E} \{ m \}$ is common within each of two parameterization groups $[1] = \{ [F-1], [S-1] \}$ and $[2] = \{ [F-2], [S-2] \}$. The characteristic roots $\lambda_i$ defined in (2.2) are

$$\lambda_1 = 10, \lambda_2 = 1.5, \lambda_3 = \lambda_4 = \lambda_5 = 1, \quad \text{for group [1],}$$

$$\lambda_1 = 10, \lambda_2 = 5, \lambda_3 = \lambda_4 = \lambda_5 = 1, \quad \text{for group [2].}$$

Because of the difference in the values of $\lambda_2$, distinguishing the true two factor model from an incorrect one factor model based on a sample is more difficult for cases [F1] and [S1] than cases
[F2] and [S2], as suggested by the results of Section 4. We consider only case (a) where Σξ is known to be I₆. Lawley (1956) suggested two types of modifications for Łₐ(k₀) to improve χ² approximation under the null hypothesis that k = k₀. Amemiya (1985) studied Lawley's modifications numerically. He found that the null χ² approximation works much better for one of Lawley's statistic Łₐ(k₀) than for Łₐ(k₀), where

\begin{align}
Łₐ(k₀) &= \frac{Ł_a}{n} Ł_a(k₀), \\
Ł_a &= n - k₀ - \frac{1}{6} \left( 2r₀ + 1 - \frac{2}{r₀ + 1} \right) - \frac{k₀^2}{r₀ + 1}, \\
n &= N - 1.
\end{align}

We present the results only for Łₐ(k₀) and Ɍₐ(k₀) defined in (6.1) and (1.4). Note that up to the order of approximation we have used, the asymptotic properties of Łₐ(k₀) are the same as those of Łₐ(k₀). Hence, all the results on Łₐ(k₀) given in Section 3, 4, and 5 apply to Łₐ(k₀).

The values of the six statistics Łₐ(k₀) and Ɍₐ(k₀), k₀ = 1, 2, 3, were computed over 1000 Monte Carlo samples for each of eight configurations: two sample sizes N = 29 and 43, and four cases [F-1], [S-1], [F-2], and [S-2]. The true number of factors is k = 2. Thus, k₀ = 2 is the null case, and k₀ = 1 and 3 correspond to the cases with one less and one more number of factors than the true model, respectively. For k₀ = 1, 2, 3, and for α = 0.1, 0.05, 0.01, the percentage of times the values of Łₐ(k₀) and Ɍₐ(k₀) exceeded χ²(α) were recorded, where

\begin{align*}
r₀ &= \frac{1}{2}(p - k₀)(p - k₀ + 1),
\end{align*}

and χ²(α) is the upper α quantile of the χ² distribution with d degrees of freedom. The results are given in Table 1. We note that as a measure of sampling variability the approximate standard error of an observed percentage can be obtained based on the binomial distribution. To provide a rough idea on the sampling variability, we present the approximate standard errors for a few observed percentages.

<table>
<thead>
<tr>
<th>Observed percentage</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximate standard error</td>
<td>0.31</td>
<td>0.69</td>
<td>0.95</td>
<td>1.26</td>
<td>1.45</td>
<td>1.55</td>
</tr>
</tbody>
</table>

The first conclusion drawn from the table is about the difference between the functional model and the structural model. For each N = 29 and 43 and each of parameterizations [1] and [2], the observed percentages of Łₐ(k₀) and Ɍₐ(k₀) are very similar between the functional case and the structural case. Although not reported here, we have also considered the case
with one factor normal and another factor normalized $\chi^2_3$. The percentages for such a case were also very similar to those for the functional and normal structural cases. We conclude that as indicated by the asymptotic results in Sections 3, 4, and 5, the null and nonnull properties of $\tilde{L}_a$ and $R_a$ depend largely on the second moment $\mathcal{E}\{m\}$, or equivalently on the roots $\lambda_i$'s, and are not affected seriously by the distributional form of the factors $f_i$.

The case with $k_0 = 1$ is the situation where the number of fitted factors is one less than the true number. As supported by Corollary 1, the powers of both tests increase as $N$ increases. Both tests have a large power of detecting the case [2] with $\lambda_2 = 5$ representing the situation where one ignores the existence of a factor with a large effect compared to the error term. For the case [1] with $\lambda_2 = 1.5$ where a factor has a small variability, the powers are modest, but as supported by Corollary 2, the power of $R_a$ is significantly larger than that of $\tilde{L}_a$.

The case with $k_0 = 2$ is the null case where the asymptotic distribution is $\chi^2_6$ for $\tilde{L}_a$ and $R_a$. For both $N = 29$ and 43, $\tilde{L}_a$ has a nearly exact significance level. The test based on $R_a$ and the $\chi^2$ approximation is very conservative in the sense that the probability of rejecting the null is much smaller than the nominal level for both $N = 29$ and 43.

The case with $k_0 = 3$ is the situation where the number of fitted factors is one more than the true number. As supported by Theorem 3, the powers of $\tilde{L}_a$ and $R_a$ do not increase with $N$, in fact decrease slightly. For all cases $\tilde{L}_a$ has a modest power. If $\tilde{L}_a$ was used to test $H_2$: the number of factors is at most $k$, then the significance level based on the asymptotic $\chi^2$ distribution for the model with exactly $k$ factors would be incorrect, because the probability of rejecting $H_2$ is larger when there are only $k - 1$ factors. For case [1] where one of the normalized factors has a small variance, $\tilde{L}_a$ has a larger power when $k_0 = 3$ than when $k_0 = 1$. It seems possible in practice that the test based on $\tilde{L}_a$ accepts the model with $\tilde{k}$ factors for some $\tilde{k}$ and rejects the model with $\tilde{k} + 1$ factors. The power of $R_a$ is generally very small when more factors than needed are fitted.

For both $\tilde{L}_a$ and $R_a$, the only case detected by a large power is the situation where the number of fitted factors is smaller than the true number and the effect of all existing factors is large compared to the error variances. When the number of fitted factors is larger than the true number, the two tests have powers that may be larger than the significance level.
Hence, a sensible way to use either $\bar{L}_a$ or $R_a$ in choosing the number of factors is a successive application of the test in which the value of $k_0$ increases from the smallest plausible value. Such a procedure stops at the value of $k_0$ giving a nonsignificant test result. It is interesting to note that in this procedure where the successive hypothesis of the form $H_0: k = k_0$ is $H_A: k > k_0$ are tested, $R_a$ has a smaller type I error and a smaller type II error than $\bar{L}_a$. If $\bar{L}_a$ is used in such a successive procedure, then, instead of stopping at a nonsignificant value of $k_0$, one may proceed to test the model with one more factor, because $\bar{L}_a$ has some power against a model with too many factors.

7. Conclusion.

We have discussed the properties of the likelihood ratio $L$ and residual sum of squares $R$ goodness-of-fit test statistics when they are used to choose the number of factors in the multivariate functional and structural relationships. Two versions of the model, (a) the error covariance matrix known and (b) the error covariance matrix known up to a multiple are considered. The null and nonnull properties of $L$ and $R$ depend largely on the second moment structure of the model, and are approximately free of the distributional form of the factors. When the number of fitted factors is less than the true number, the both types of test have a reasonable power that increase as the sample size increases and as the variances of normalized factors increase relative to the error variances. Under such an alternative, the residual sum of squares test is believed to be more powerful than the likelihood ratio test for case (a).

When the number of fitted factors is larger than the true number, both tests have small powers that do not increase with the sample size. Such a power seems to be larger than the probability of rejecting the model with the correct number of factors. The Monte Carlo results for case (a) support the theoretical asymptotic results. Based on our study, we can make the following recommendations. Using the likelihood ratio test or the residual sum of squares test, a procedure for choosing the number of factors with a reasonable power function would be a successive application of the test in which the number of fitted factors is increased starting with the smallest plausible number for the given problem. One would stop such a sequential procedure when the model with some number of factors is accepted. If the likelihood ratio test
is used for case (a), it is advisable to proceed to fit the model with one more factor than the nonsignificant model in order to make sure the result. For case (a), such a successive procedure using the modified likelihood ratio statistic has a nearly exact type I error at each stage for moderate to large sample sizes. The successive procedure using the residual sum of squares statistic for case (a) has a smaller type I error and a larger power at each stage than the procedure using the likelihood ratio statistic. It is desired to develop some modification to the residual sum of squares statistic that improves the $\chi^2$ approximation to its null distribution.

Appendix

We present two lemmas on the limiting distribution of the characteristic roots in the functional and structural relationships. The lemmas were used in the proofs of Theorem 1, 2, and 3. The standard results on the limiting distribution of the characteristic roots of the Wishart matrix do not apply to our problem, because the factors are either fixed or random with an unspecified distribution. Our study also requires the limiting distribution for the local alternative case where the limiting multiplicity of the population roots is different from the multiplicity in the finite sample. We only present the relevant results without proofs. The results here are special cases of more general results covered in Amemiya (1986b). The proofs of the lemmas are given in Amemiya (1986b). For both the functional and structural models, the local alternative case is covered in Lemma 1, and the fixed multiplicity case is covered in Lemma 2.

**Lemma 1.** Let assumptions (i), (ii), and either (S1) and (S2-a) or (F1) and (F2-a) hold. Let $p_0 = 0$, and for $j = 1, 2, \ldots, s - 1$, let

$$p_j = p_{j-1} + g_j,$$

$$g_j(n) = (\hat{\lambda}_{p_{j-1}+1} - \gamma_j - 1, \hat{\lambda}_{p_{j-1}+2} - \gamma_j - 1, \ldots, \hat{\lambda}_{p_j} - \gamma_j - 1), \quad \text{under (S1) and (S2-a)},$$

$$= [\hat{\lambda}_{p_{j-1}+1} - \gamma_j(n) - 1, \hat{\lambda}_{p_{j-1}+2} - \gamma_j(n) - 1, \ldots, \hat{\lambda}_{p_j} - \gamma_j(n) - 1], \quad \text{under (F1) and (F2-a)}.$$
Also let
\[ \sim h_1(n) = [\sim g_1(n), \sim g_2(n), \ldots, \sim g_{s-1}(n)], \]
\[ \sim h_2(n) = (\hat{\lambda}_{k-q+1} - 1, \hat{\lambda}_{k-q+2} - 1, \ldots, \hat{\lambda}_p - 1), \]
\[ \sim h = [\sim h_1(n), \sim h_2(n)]. \]

Then, as \( n \to \infty, \)
\[ \sim h(n) \overset{L}{\to} \sim h = (\sim h_1, \sim h_2), \]

where \( \sim h_2 \) is the \( 1 \times (p - k + q) \) vector of the characteristic roots of an \( (p - k + q) \times (p - k + q) \) symmetric random matrix \( \sim V \) such that
\[ \sim V = \sim U + \begin{pmatrix} \sim D & 0 \\ 0 & 0 \end{pmatrix}, \]
\[ \sim D = \text{diag}\{\sim \delta_1, \sim \delta_2, \ldots, \sim \delta_q\}, \]
\[ \text{vech } \sim U \sim N(0, 2 \psi_{p-k+q} \sim \psi_{p-k+1}), \]
and \( \sim \psi_{p-k+q} \) is defined in the proof of Theorem 2.

**Lemma 2.** Let assumptions (i), (ii), and either (S1) and (S2) or (F1) and (F2) hold. Then the result of Lemma 1 holds with replacing \( s - 1 \) by \( s \) and setting \( q = 0 \) and \( \sim D = 0. \)

**Acknowledgement**

The author wishes to thank T. W. Anderson for useful comments.
Table 1. Monte Carlo powers (%) for problem (a)
(1000 samples)

<table>
<thead>
<tr>
<th>Number of factors fitted</th>
<th>$k_0 = 1$</th>
<th>$k_0 = 2$</th>
<th>$k_0 = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referred distribution</td>
<td>$\chi^2_{10}$</td>
<td>$\chi^2_5$</td>
<td>$\chi^2_3$</td>
</tr>
<tr>
<td>Upper percentiles</td>
<td>10</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>[F-1] N=29 $\bar{L}_a$</td>
<td>19.9</td>
<td>12.1</td>
<td>3.7</td>
</tr>
<tr>
<td>$R_a$</td>
<td>35.8</td>
<td>26.9</td>
<td>16.2</td>
</tr>
<tr>
<td>[F-1] N=43 $\bar{L}_a$</td>
<td>27.9</td>
<td>17.5</td>
<td>5.9</td>
</tr>
<tr>
<td>$R_a$</td>
<td>41.4</td>
<td>34.3</td>
<td>20.3</td>
</tr>
<tr>
<td>[S-1] N=29 $\bar{L}_a$</td>
<td>19.9</td>
<td>11.5</td>
<td>4.0</td>
</tr>
<tr>
<td>$R_a$</td>
<td>33.1</td>
<td>24.9</td>
<td>15.3</td>
</tr>
<tr>
<td>[S-1] N=43 $\bar{L}_a$</td>
<td>26.2</td>
<td>18.3</td>
<td>6.1</td>
</tr>
<tr>
<td>$R_a$</td>
<td>41.4</td>
<td>32.6</td>
<td>19.6</td>
</tr>
<tr>
<td>[F-2] N=29 $\bar{L}_a$</td>
<td>100.0</td>
<td>99.9</td>
<td>99.9</td>
</tr>
<tr>
<td>$R_a$</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>[F-2] N=43 $\bar{L}_a$</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$R_a$</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>[S-2] N=29 $\bar{L}_a$</td>
<td>99.3</td>
<td>98.9</td>
<td>97.0</td>
</tr>
<tr>
<td>$R_a$</td>
<td>99.9</td>
<td>99.9</td>
<td>99.5</td>
</tr>
<tr>
<td>[S-2] N=43 $\bar{L}_a$</td>
<td>100.0</td>
<td>100.0</td>
<td>99.9</td>
</tr>
<tr>
<td>$R_a$</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>
References


