ASYMPTOTIC CHI-SQUARE TESTS FOR A LARGE CLASS OF FACTOR ANALYSIS MODELS

YASUO AMEMIYA AND T. W. ANDERSON

TECHNICAL REPORT NO. 13
NOVEMBER 1985

PREPARED UNDER
NATIONAL SCIENCE FOUNDATION GRANT DMS 82-19748

ECONOMETRIC WORKSHOP
FOURTH FLOOR, ENCINA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ASYMPTOTIC CHI–SQUARE TESTS FOR A LARGE CLASS OF FACTOR ANALYSIS MODELS

Yasuo Amemiya* and T. W. Anderson
Stanford University

TECHNICAL REPORT NO. 13
NOVEMBER 1985

PREPARED UNDER
NATIONAL SCIENCE FOUNDATION GRANT DMS 82-19748

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

* On leave from Iowa State University
Abstract.

Two types of $\chi^2$ goodness-of-fit tests derived under the normal assumption have been used extensively in factor analysis. Asymptotic behavior of the test statistics is investigated here under weak assumptions on the factor vector and the error vector. It is shown that the limiting $\chi^2$ result holds if the factor vector is either fixed or random with finite second moments and if the components of the error vector are independent and have finite second moments.


Key words and phrases: Factor analysis, goodness-of-fit test, asymptotic robustness.
1. Introduction.

The factor analysis model assumes that a set of \( p \times 1 \) observations \( \zeta_t \) satisfies

\[
\zeta_t = \mu_{\zeta} + \Lambda_f \zeta_t + \epsilon_t, \quad t = 1, 2, \ldots, N,
\]

where \( \zeta_t \) is a \( k \times 1 \) unobservable factor vector, \( \epsilon_t \) is a \( p \times 1 \) unobservable error vector, \( \mu_{\zeta} \) is a \( p \times 1 \) unknown vector, and \( \Lambda \) is a \( p \times k \) unknown matrix of rank \( k \) consisting of factor loadings. The \( \epsilon_t \)'s are assumed to be independently and identically distributed with zero mean vector and a diagonal covariance matrix \( \Sigma_{\epsilon} = \text{diag}(\gamma) \), where \( \gamma \) is the \( p \times 1 \) vector listing the variances of the \( p \) components of \( \epsilon_t \). The \( f_t \)'s are assumed to be either fixed constants or independently and identically distributed with mean \( \mu_f \) and positive definite covariance matrix \( \Sigma_{ff} \). If the \( f_t \)'s are random, the \( f_t \)'s and \( \epsilon_t \)'s are assumed to be independent. Following common practice in factor analysis, we concentrate on inferences based on the sample covariance matrix

\[
m = \frac{1}{N-1} \sum_{i=1}^{N} (\zeta_t - \bar{\zeta})(\zeta_t - \bar{\zeta})^\prime,
\]

where

\[
\bar{\zeta} = \frac{1}{N} \sum_{t=1}^{N} \zeta_t.
\]

If both \( f_t \) and \( \epsilon_t \) are normally distributed, then \( nm \sim W_p(\Sigma, n) \), where \( n = N - 1 \) is the number of degrees of freedom, and

\[
\Sigma = \Lambda \Sigma_{ff} \Lambda' + \Sigma_{\epsilon\epsilon}.
\]

Two commonly used goodness-of-fit test statistics for model (1.1) are

\[
G_1 = n \left[ \log |\hat{\Sigma}| - \log |m| + \text{tr} \left( m^{-1} \hat{\Sigma} \right) - p \right],
\]

\[
G_2 = \frac{n}{2} \text{tr} \left[ (m - \hat{\Sigma})^{-1} \right] ^2,
\]

1
where \( \hat{\Sigma} \) is the maximum Wishart likelihood estimator of \( \Sigma \) obtained by minimizing

\[
\ell(\hat{\Sigma}) = \log |\hat{\Sigma}| + \text{tr}\left( \hat{m}\Sigma^{-1} \right)
\]

under the model specification (1.1) with some specified identification conditions. If \( f_t \) and \( \varepsilon_t \) are normally distributed, then \( G_1 \) is the likelihood ratio test statistic for the null hypothesis of the covariance structure (1.2) against the alternative hypothesis of unrestricted \( \hat{\Sigma} \). Standard computer packages for the maximum likelihood factor analysis output the statistic \( \hat{G}_1 \). The statistic \( G_2 \) is the residual sum of squares at the final stage of the iteratively reweighted generalized least squares method for computing the estimator \( \hat{\Sigma} \). A standard nonlinear least squares computer package can be used to compute the statistic \( G_2 \). If \( f_t \) and \( \varepsilon_t \) are normally distributed, then under the model specification each of \( G_1 \) and \( G_2 \) converges to a \( \chi^2 \) random variable with \( q = [(p-k)(p-k+1)/2-p] \) degrees of freedom. In this paper we investigate the asymptotic behavior of \( G_1 \) and \( G_2 \) for possibly nonnormal \( f_t \) and \( \varepsilon_t \) and present conditions under which the limiting \( \chi^2 \) result for \( \hat{G}_1 \) and \( \hat{G}_2 \) holds.

Statistical inferences for factor analysis based on the likelihood theory have been developed by Lawley (1940, 1941, 1943, 1953, 1967, 1976), Rao (1955), Anderson and Rubin (1956), Jöreskog (1967), Jennrich and Thayer (1973), Amemiya, Fuller, and Pantula (1985), and Anderson and Amemiya (1985). Most of the results are well described in Lawley and Maxwell (1971) and Anderson (1984a, 1984b). In the literature the derivation of the limiting \( \chi^2 \) result for \( G_1 \) is usually based on the standard likelihood theory under the normality assumption on \( f_t \) and \( \varepsilon_t \). Bartlett (1950) suggested a certain modification of \( G_1 \) to improve the \( \chi^2 \) approximation.

The generalized least squares method was first applied to the factor analysis model by Jöreskog and Goldberger (1972). The method was later applied to more general covariance structure models. See, e.g., Browne (1974, 1982), Jöreskog (1981), and Shapiro (1983). The limiting \( \chi^2 \) result for \( G_2 \) is usually based on the asymptotic equivalence between \( G_1 \) and \( G_2 \) under the normal assumption on observations. See, e.g., Browne (1974).

For a class of functional and structural relationship models containing the factor analysis model as a special case, Amemiya (1985) discussed asymptotic properties of the statistics \( G_1 \)
and $G_2$. In particular, Amemiya (1985) showed that the limiting $\chi^2$ result for $G_1$ and $G_2$ if the factor vectors are either fixed or random with finite second moments and if the fourth order cumulants of the error vector are zero. In this paper we consider asymptotic behavior of $G_1$ and $G_2$ for the factor analysis model. Because of the factor analysis structure the limiting $\chi^2$ result for $G_1$ and $G_2$ holds under a weaker set of assumptions than that of Amemiya (1985). A practical interpretation of Amemiya's (1985) result is that the limiting $\chi^2$ result holds under weak assumptions on the factor vector if the errors are normally distributed. In this paper we shall show that such a $\chi^2$ result holds for the factor analysis model under weak assumptions on the error vector as well as on the factor vector. In particular we shall derive the limiting $\chi^2$ result for $f_t$ either fixed or random with any distribution having finite second moments and for $\varepsilon_t$ with any distribution having finite second moments under the assumption that the $p$ components of $\varepsilon_t$ are independent, not just uncorrelated.

Under the normality assumption on $f_t$ and $\varepsilon_t$ the parameters $\Lambda, \Sigma_{ff}, \nu, \mu_Z$, and $\mu_f$ in model (1.1) are not identified. It is convenient in our theoretical development to specify the model by an identified set of parameters. One such parameterization is to let $\Lambda = (\beta, I_k)$ and $\mu_Z = (\beta_0, 0)$, where $\beta$ is $k \times r$, $\beta_0$ is $r \times 1$, $r = p - k$, and $I_k$ is the $k \times k$ identity matrix. With this parameterization the model (1.1) can be written as

$$Z_t = (\beta_0, 0)' + (\beta, I_k)'f_t + \varepsilon_t, \quad t = 1, 2, \ldots, N.$$  

Under the model (1.5) with random $f_t$ the matrix $\Sigma$ in (1.2) is

$$\Sigma = (\beta, I_k)'\Sigma_{ff}(\beta, I_k) + \Sigma_{\varepsilon\varepsilon}.$$  

In general, if model (1.1) holds with a specified value of $k$, then the model can be parameterized in the form (1.5) by reordering $p$ components of $Z_t$ and transforming $f_t$. Changing the order of $p$ components of $Z_t$ and transforming $f_t$ do not alter the statistics $G_1$ and $G_2$ in (1.3). Hence, without loss of generality, we use the convenient parameterization (1.5) throughout our investigation of the statistics $G_1$ and $G_2$. For a detailed discussion of the identification, see Anderson and Rubin (1956).

The vec and vech operators and their properties are used extensively throughout this paper. We present the definitions and results that are useful in our development. For a $p \times q$ matrix $\mathbf{A} = (a_1, a_2, \ldots, a_q)$, define the $pq \times 1$ vec $\mathbf{A} = (a_1', a_2', \ldots, a_q')'$. For a $p \times p$ symmetric matrix $\mathbf{A}$, let vech $\mathbf{A}$ be the $\frac{1}{2}p(p+1) \times 1$ vector listing the elements of the columns of $\mathbf{A}$ that are on or below the diagonal starting with the first column. Let $\Phi_p$ be the $p^2 \times \frac{1}{2}p(p+1)$ matrix such that for a $p \times p$ symmetric $\mathbf{A}$, vec $\mathbf{A} = \Phi_p$ vech $\mathbf{A}$. Define $\psi_p = (\Phi_p' \Phi_p)^{-1} \Phi_p'$. Then, for a $p \times p$ symmetric $\mathbf{A}$, vech $\mathbf{A} = \psi_p$ vec $\mathbf{A}$. Note that the elements of $\Phi_p$ are zero or one, that each of the $p$ columns of $\Phi_p$ corresponding to the off–diagonal elements has two nonzero elements, and that the columns of $\Phi_p$ are orthogonal. Let $L_p$ be the $p^2 \times p$ matrix consisting of the $p$ columns of $\Phi_p$ corresponding to the diagonal elements. The matrix $L_p$ can be obtained by columnwise Kronecker products of $I_p$ and $I_p$, i.e., the $i$–th column of $L_p$ is zero except for the $[p(i-1)+i]$–th element, which is one. Let $K_p$ be the $p^2 \times \frac{1}{2}p(p-1)$ matrix consisting of the columns of $\Phi_p$ corresponding to the off–diagonal elements, i.e., the columns of $\Phi_p$ that are not the columns of $L_p$. Then, for a $p \times p$ symmetric $\mathbf{A}$,

$$\text{vec } \mathbf{A} = L_p \gamma_A + K_p \delta_A,$$

where $\gamma_A$ is the $p \times 1$ vector listing the diagonal elements of $\mathbf{A}$ and $\delta_A$ is the $\frac{1}{2}p(p-1) \times 1$ vector listing the elements of $\mathbf{A}$ that are below the main diagonal. Note that $L_p'K_p = 0$, $L_p'L_p = I_p$, and $K_p'K_p = 2I_p$, where $p_0 = \frac{1}{2}p(p-1)$. Thus, $\gamma_A = L_p$ vec $\mathbf{A}$ and $\delta_A = \frac{1}{2}K_p'$ vec $\mathbf{A}$. If $\mathbf{A}$ is a $p \times p$ diagonal matrix, then vec $\mathbf{A} = L_p \gamma_A$. The $p^2 \times p^2$ matrix $\Upsilon_p = \Phi_p \psi_p = \Phi_p (\Phi_p' \Phi_p)^{-1} \Phi_p'$ is the orthogonal projection operator onto the column space of $\Phi_p$ and satisfies

$$\psi_p \Upsilon_p = \psi_p, \quad \Upsilon_p \Phi_p = \Phi_p,$$

$$\Upsilon_p L_p = L_p, \quad \Upsilon_p K_p = K_p,$$

and for a $p \times q$ matrix $\mathbf{A}$,

$$\Upsilon_p (\mathbf{A} \otimes \mathbf{A}) = (\mathbf{A} \otimes \mathbf{A}) \Upsilon_q,$$
where $\otimes$ denotes the Kronecker product. For detailed discussions of vec and vech operators see, e.g., Browne (1974) and Henderson and Searle (1979).

3. Notations and Assumptions.

Using the parameterization (1.5) we define some notations and present a set of regularity conditions. Let $\gamma_\beta = \text{vec} \beta, \gamma_f = \text{vec} \Sigma_f \gamma_f$, and $\gamma_\epsilon = L_p^{\prime} \text{vec} \Sigma_\epsilon \gamma_\epsilon$, where $L_p$ is defined in Section 2. Define $\gamma = (\gamma_\beta, \gamma_f, \gamma_\epsilon)^\prime$ and

$$
\Sigma(\gamma) = (\beta, I_k)^\prime \Sigma_f (\beta, I_k) + \Sigma_\epsilon.
$$

Let $\Gamma$ be the parameter space for $\gamma$, and let $\hat{\gamma} = (\hat{\gamma}_\beta, \hat{\gamma}_f, \hat{\gamma}_\epsilon)^\prime$ be the maximum Wishart likelihood estimator of $\gamma$ obtained by minimizing $\ell(\Sigma)$ in (1.4) evaluated at $\Sigma = \Sigma(\gamma)$ over $\Gamma$. Note that $\ell(\Sigma)$ is not associated with the likelihood unless both $f_\sim$ and $\epsilon_\sim$ are normally distributed. If the $f_\sim$'s are fixed, then $\Sigma_f$ is not a parameter of the population distribution and $\hat{\gamma}_f$ estimates vec of a $k \times k$ matrix

$$
m_{ff}(n) = \frac{1}{n} \sum_{t=1}^{N} (f_\sim - \bar{f})(f_\sim - \bar{f})',
$$

where $n = N - 1,$

$$
\bar{f} = \frac{1}{N} \sum_{t=1}^{N} f_\sim.
$$

Let $\gamma_f(n) = \text{vech} m_{ff}(n).$ It will be useful to consider the sum of squares and cross products matrix of $\epsilon_\sim.$ Let

$$
m_{\epsilon\epsilon}(n) = \frac{1}{n} \sum_{t=1}^{N} (\epsilon_\sim - \bar{\epsilon})(\epsilon_\sim - \bar{\epsilon})',
$$

where

$$
\bar{\epsilon} = \frac{1}{N} \sum_{t=1}^{N} \epsilon_\sim.
$$
It will also be useful to introduce notations for the diagonal part and the off–diagonal part of $\sim m_{ee}$. Let
\[
\gamma_{e}(n) = \sim L' \sim_p \vec{m}_{ee}(n), \\
\delta_{e}(n) = \frac{1}{2} \sim K' \sim_p \vec{m}_{ee}(n),
\]
where $\sim L_p$ and $\sim K_p$ are defined in Section 2. Thus, $\gamma_{e}(n)$ is the $p \times 1$ vector consisting of the diagonal elements of $\sim m_{ee}(n)$, $(1/n) \sum_{i=1}^{N} (\epsilon_{ii} - \bar{\epsilon}_{i})^2$, $i = 1, 2, \ldots, p$, and $\delta_{e}(n)$ is the $\frac{1}{2}p(p - 1) \times 1$ vector consisting of the off–diagonal elements of $\sim m_{ee}(n)$, $(1/n) \sum_{i=1}^{N} (\epsilon_{ii} - \bar{\epsilon}_{i})(\epsilon_{jj} - \bar{\epsilon}_{j})$, $i > j = 1, 2, \ldots, p$, where $\epsilon_{ii}$ and $\bar{\epsilon}_{i}$ are the $i$–th elements of $\sim \epsilon$ and $\bar{\epsilon}$, respectively. By (2.1),
\[
\vec{m}_{ee}(n) = \sim L_p \gamma_{e}(n) + \sim K_p \delta_{e}(n).
\]

Throughout our discussion we assume the following regularity conditions.

Assumption (A).

(i) The model (1.5) holds for $N = p + 1, p + 2, \ldots$.

(ii) The $\sim \epsilon_t$'s satisfy
\[
\lim_{n \to \infty} \sim m_{ee}(n) = \Sigma_{ee}.
\]

(iii) The $\sim \epsilon_t$'s are independent and are independent of the $f_t$'s, and are identically distributed with mean zero and covariance matrix $\Sigma_{ee} = \text{diag}\{\sim \gamma_{e}\}$.

(iv) The true (limiting) value of $\sim \gamma$,
\[
\sim \gamma_0 = (\sim \gamma'_{\beta}, \text{vech} \sim \Sigma_{ff})', \sim \gamma'_{e},
\]

is an interior point of $\sim \Gamma$, where $\sim \Sigma_{ff}$ is defined in (ii).

(v) For any $\xi > 0$ there exists an $\eta_{\xi} > 0$ such that any $\sim \gamma$ in $\sim \Gamma$ with $|\sim \gamma - \sim \gamma_0| > \xi$ satisfies $|\nu_i - 1| > \eta_{\xi}$ for some $i = 1, 2, \ldots, p$, where $\sim \gamma_0$ is defined in (iv), and the $\nu_i$'s are the $p$
roots of

\[ |\Sigma(\gamma) - \nu \Sigma(\gamma_0)| = 0. \]

\( (vi) \, \delta_\iota(n) = 0_p \left( \frac{1}{\sqrt{n}} \right) . \)

If the \( f_i \)'s are fixed, then the probability limit in assumption (ii) reduces to the nonstochastic limit. For random \( f_i \) a sufficient, though not necessary, condition for assumption (ii) is that the \( f_i \)'s are independently and identically distributed with covariance matrix \( \Sigma_{ff} \). The \( f_i \)'s with some elements fixed and other elements random can satisfy assumption (ii). Assumption (iii) does not specify any distributional form of \( \varepsilon_t \) except for the existence of the second moments.

The minimization of \( \ell(\Sigma(\gamma)) \) is over \( \Sigma \) on which \( \Sigma_{ff} \) and \( \Sigma_{ee} \) are nonnegative definite. Thus, the minimum of \( \ell(\Sigma(\gamma)) \) may be attained at a value with singular \( \Sigma_{ff} \) (less than \( k \) factors) or with some zero error variances (Heywood case). Under assumption (iv) the true values of \( \Sigma_{ff} \) and \( \Sigma_{ee} \) are positive definite. Assumption (v) is an identification assumption that guarantees the consistency of the maximum Wishart likelihood estimator \( \hat{\gamma} \). See, e.g., Amemiya, Fuller, and Pantula (1985). Assumption (vi) is an order condition on \( \delta_\iota(n) \), the off-diagonal part of \( m_{ee}(n) \). Because \( \varepsilon_{it}, i = 1, 2, \ldots, p \), are either uncorrelated or independent, assumption (vi) is not considered to be very restrictive. There is no assumption on a rate at which the diagonal part \( \gamma_\iota(n) \) converges to \( \gamma_\iota \). We shall discuss practical implications of the assumptions in the next section.

4. Results.

Under assumption (A) in Section 3 we investigate asymptotic behavior of the goodness-of-fit test statistics \( G_1 \) and \( G_2 \). Amemiya (1985) treated the same problem for a more general class of models under the assumption that the fourth-order moments of \( \varepsilon_t \) have a certain structure. Amemiya's (1985) results are not applicable here because under assumption (A) \( \varepsilon_t \) may not have moments of order higher than two. The key result here will be that because of the factor analysis structure the asymptotic behavior of \( G_1 \) and \( G_2 \) are free of the diagonal part \( \gamma_\iota(n) \) of \( m_{ee}(n) \). To obtain this result we shall use an expansion of the maximum Wishart likelihood estimator \( \hat{\gamma} \) around \( \gamma(n) = [\gamma_B'_{\hat{\gamma}}, \gamma_f'(n), \gamma_e'(n)] \) rather than around the true value \( \gamma_0 \). The following two lemmas treat such an expansion.
Lemma 1. Under assumption (A),

\[
\hat{\gamma} - \gamma(n) = (F' \Omega_{\gamma}^{-1} F)^{-1} F' \Omega_{\gamma}^{-1} \vech\{m - \Sigma[\gamma(n)]\} + o_p\left(\frac{1}{\sqrt{n}}\right),
\]

where

\[
F = \frac{\partial \vech \Sigma(\gamma_0)}{\partial \gamma'},
\]

\[
\Omega = \psi_p'[\Sigma(\gamma_0) \otimes \Sigma(\gamma_0)] \psi_p' .
\]

Proof. The result that \(\hat{\gamma} - \gamma(n) = O_p\left(\frac{1}{\sqrt{n}}\right)\) is a special case of Theorem 2 of Anderson and Amemiya (1985), which is a slight modification of Theorem 12.3 of Anderson and Rubin (1956, p. 149). The expression for the leading terms follows from the proof of Theorem 2 in Anderson and Amemiya (1984).

Note that

\[
m = (\beta, I_k)' m_{ff}(n)(\beta, I_k) + (\beta, I_k)' m_{fe}(n) + m_{ef}(n)(\beta, I_k) + m_{ee}(n),
\]

\[
\Sigma[\gamma(n)] = (\beta, I_k)' m_{ff}(n)(\beta, I_k) + \text{diag}\{\gamma_\epsilon(n)\},
\]

where

\[
m_{fe}(n) = m_{ef}(n) = \frac{1}{n} \sum_{i=1}^{N} (f_i - \bar{f}) e_i .
\]

Thus, the quantity \(m - \Sigma[\gamma(n)]\) is a function of \(m_{fe}(n)\) and \(\delta_\epsilon(n)\), the off-diagonal part of \(m_{ee}(n)\), and is free of \(m_{ff}(n)\) and \(\gamma_\epsilon(n)\). Under assumption (A) both \(m_{fe}\) and \(\delta_\epsilon(n)\) are \(O_p\left(\frac{1}{\sqrt{n}}\right)\).

To derive expansions of \(G_1\) and \(G_2\), we need an explicit expression of the leading term in the expansion of \(\hat{\gamma}_\epsilon - \gamma_\epsilon(n)\).
**Lemma 2.** Under assumption (A),

\[
\tilde{\gamma}_e - \gamma_e(n) = (L_p^T H L_p)^{-1} L_p^T H K_p \delta_e(n) + o_p \left( \frac{1}{\sqrt{n}} \right)
\]

where

\[
H = (I_r, -\beta')^T \Sigma_{uv}^{-1}(I_r, -\beta') \otimes (I_r, -\beta')^T \Sigma_{uv}^{-1}(I_r, -\beta'),
\]

\[
\Sigma_{uv} = (I_r, -\beta') \Sigma_{ee}(I_r, -\beta)',
\]

\[r = p - k,
\]

and \(L_p\) and \(K_p\) are defined in Section 2.

The proof of Lemma 2 is given in the Appendix. We see from Lemma 2 that \(\tilde{\gamma}_e - \gamma_e(n)\) depends asymptotically only on the off–diagonal part \(\delta_e(n)\).

The next lemma gives alternative expressions for the statistics \(G_1\) and \(G_2\) defined in (1.3).

**Lemma 3.** Under assumption (A), with probability approaching one as \(n \to \infty\),

\[
G_1 = n \{- \log |I_r + \hat{\Lambda}| + \text{tr}(\hat{\Lambda})\},
\]

\[
G_2 = \frac{n}{2} \text{tr}(\hat{\Lambda}^2),
\]

where

\[
\hat{\Lambda} = \hat{\Sigma}_{uu}^{-1/2} \hat{m}_{uu} \hat{\Sigma}_{uu}^{-1/2} - I_r,
\]

\[
\hat{\Sigma}_{uu} = (I_r, -\hat{\beta}') \text{diag}\{\hat{\gamma}_e\}(I_r, -\hat{\beta}')',
\]

\[
\hat{m}_{uu} = (I_r, -\hat{\beta}') \text{diag}\{\hat{\gamma}_e\}(I_r, -\hat{\beta}')',
\]

\[
\text{vec} \hat{\beta} = \hat{\gamma}_e.
\]

**Proof.** This result is a special case of Lemma 2 of Amemiya (1985). The proof is based on the fact that \(\hat{\gamma}_e\) and \(\hat{\gamma}_f\) are the maximum Wishart likelihood estimators for the model with the error covariance matrix \(\text{diag}\{\hat{\gamma}_e\}\) treated as known.
Thus, both $G_1$ and $G_2$ are functions of $\tilde{\Delta}$. The following lemma gives an expansion of $\tilde{\Delta}$.

**Lemma 4.** Under assumption (A),

$$\text{vec} \tilde{\Delta} = \sum_{\sim} (I - L_p (L_p L_p)^{-1} L_p H K_p \delta_e(n) + o_p \left( \frac{1}{\sqrt{n}} \right),$$

where

$$R = \sum_{\sim}^{1/2} (I_r, -\beta^t) \otimes \sum_{\sim}^{1/2} (I_r, -\beta^t).$$

The proof of the Lemma 4 is given in the Appendix. Combining Lemmas 3 and 4 we obtain the following expansion of the statistics $G_1$ and $G_2$.

**Theorem 1.** Under assumption (A),

$$G_i = n \delta_e(n) P \delta_e(n) + o_p(1), \quad i = 1, 2,$$

where

$$P = \frac{1}{2} K_p (I - H L_p (L_p L_p)^{-1} L_p H K_p).$$

**Proof.** The first derivative vector and the second derivative matrix of the expression for $G_1$ given in Lemma 3 with respect to vech $\tilde{\Delta}$ evaluated at $\tilde{\Delta} = 0$ are, respectively, the zero vector and $n \Phi_e^t \Phi_e$. Thus because $\tilde{\Delta} = o_p \left( \frac{1}{\sqrt{n}} \right)$ by Lemma 4,

$$G_1 = \frac{n}{2} (\text{vec} \tilde{\Delta})^t \Phi_e^t \Phi_e \text{vec} \tilde{\Delta} + o_p \left( \frac{1}{\sqrt{n}} \right) = \frac{n}{2} (\text{vec} \tilde{\Delta})^t \text{vec} \tilde{\Delta} + o_p \left( \frac{1}{\sqrt{n}} \right) = G_2 + o_p \left( \frac{1}{\sqrt{n}} \right).$$

The result follows by substituting the expansion of vec $\tilde{\Delta}$ given in Lemma 4 and using $R^t R = H$.

We see from Theorem 1 that the leading terms in the expansions of $G_1$ and $G_2$ are common and depend only on the off–diagonal part $\delta_e(n)$ of $\tilde{m}_{ee}(n)$. For the functional and structural
relationship models with general error covariance structure, Amemiya (1985) gave expansions of $G_1$ and $G_2$ whose common leading term is free of the factor vector but depends on all elements of $[m_{ee}(n) - \Sigma_{ee}]$. Because of the factor analysis structure here the expansions in Theorem 1 are not only derived under weaker assumptions than those of Amemiya (1985) but also have the common leading term that is free of the diagonal elements of $m_{ee}(n)$. The common limiting distribution of $G_1$ and $G_2$ can be obtained by assuming the limiting normality of $\eta_{ee}(n)$, the off-diagonal part of $m_{ee}(n)$. Under the factor analysis structure, the limiting normality of $\eta_{ee}(n)$ holds under relatively weak assumptions that are satisfied by many practical situations. We present two theorems, each of which gives a sufficient condition for $G_1$ and $G_2$ to have the common limiting $\chi^2$ distribution with degrees of freedom $q = \frac{1}{2}(p - k)(p - k + 1) - p$.

**Theorem 2.** Let assumptions (i)-(v) hold. Assume

\[(vi-a) \quad \sqrt{n} \eta_{ee} \overset{L}{\rightarrow} N(0, V),\]

where

$$V = \frac{1}{2} \eta_{ee}'(\Sigma_{ee} \otimes \Sigma_{ee}) \eta_{ee}.$$

Then,

$$G_i \overset{L}{\rightarrow} \chi^2_q, \quad i = 1, 2,$$

where $q = \frac{1}{2}(p - k)(p - k + 1) - p$.

**Proof.** If the $\eta_i$'s are normally distributed, then by Theorem 3 of Amemiya (1985)

$$G_i \overset{L}{\rightarrow} \chi^2_q \quad i = 1, 2.$$

By Theorem 1, under assumption (vi-a) the common limiting distribution of $G_1$ and $G_2$ depends on the distribution of $\eta_i$ only through the limiting covariance matrix $V$. But, the $V$ specified in assumption (vi-a) is the limiting covariance matrix of $\sqrt{n} \eta_{ee}(n)$ if the $\eta_i$'s are normally distributed. Hence, the result follows.
The covariance matrix $\mathcal{V}$ of the limiting distributions of $\sqrt{n} \varepsilon(n)$ given in assumption (vi-a) is a diagonal matrix with relatively simple diagonal elements. Let $\delta_{ij}(n), i > j$, denote a typical element of $\varepsilon(n)$, i.e., let

$$
\delta_{ij}(n) = \frac{1}{n} \sum_{i=1}^{N} (\varepsilon_{it} - \bar{\varepsilon}_l)(\varepsilon_{jt} - \bar{\varepsilon}_j).
$$

Also let $\sigma_{eeit}$ be the $i$-th diagonal element of $\Sigma_{ee}$. Then, the forms of $\mathcal{V}$ in assumption (vi-a) assumes that the limiting variance of $\sqrt{n} \delta_{ij}(n)$ is $\sigma_{eeit}\sigma_{eer}\delta_{jj}$ and that the limiting covariance between $\sqrt{n} \delta_{ij}(n)$ and $\sqrt{n} \delta_{kk}(n), (i, j) \neq (k, l)$, is zero. If the $\varepsilon_t$'s are normally distributed, then assumption (vi-a) is satisfied. However, because the factor analysis model assumes the diagonality of $\Sigma_{ee}$ and because no restriction is placed on the diagonal elements of $\Sigma_{ee}(n)$, assumption (vi-a) is not very restrictive and is satisfied in many practical situations without a higher order moment assumption on $\varepsilon_t$. The next theorem gives another sufficient condition for the limiting $\chi^2_q$ distribution of $G_1$ and $G_2$ which has an important practical meaning.

**Theorem 3.** Let assumptions (i)-(v) hold. Assume

(vi-b) The $p$ components $\varepsilon_{it}, i = 1, 2, \ldots, p$, of $\varepsilon_t$ are independent.

Then,

$$
G_i \xrightarrow{L} \chi^2_q, \quad i = 1, 2.
$$

**Proof.** Under assumption (vi-b) the $\frac{1}{2}p(p - 1) \times 1$ vectors $\varepsilon_t$ consisting of $\varepsilon_{it}\varepsilon_{jt}, i > j$, are independently and identically distributed with mean zero and covariance matrix $\mathcal{V}$ given in assumption (vi-a). Hence, assumption (vi-a) holds. Thus, the result follows from Theorem 2.

Theorem 3 has shown that the goodness-of-fit test statistics $G_1$ and $G_2$ have the limiting $\chi^2_q$ distribution under the model specification for a large class of $f_t$ and $\varepsilon_t$ provided that the $p$ elements of $\varepsilon_t$ are not just uncorrelated but independent. Besides the independence assumption there is no assumption on the $\varepsilon_{it}$'s concerning moments of order greater than two.
The elements of the factor vector \( \tilde{f}_t \) can be either fixed or random with finite second moments. There is no assumption on the distributional form for \( \tilde{f}_t \) nor \( \varepsilon_t \). Thus, the characteristics of the distribution of the observation \( \tilde{Z}_t \) may be quite different from those of the normal distribution. For example, the goodness-of-fit test using \( G_1 \) of \( G_2 \) and \( \chi^2 \) cut-off points is asymptotically valid for a discrete factor analysis model, where each component of \( \tilde{f}_t \) takes integer values, \( \varepsilon_{it} \)'s are independent, and each \( \varepsilon_{it} \) takes values \(-1, 0, \) and \( 1 \).

A common interpretation of the factor analysis model (1.1) is that all the inter-dependency among the \( p \) components of \( \tilde{Z}_t \) is explained by a \( k \times 1 \) factor vector \( \tilde{f}_t \). From this point of view, assumption (vi–b), the independence of the \( \varepsilon_{it} \)'s is a condition included in the factor analysis model specification. If we consider the factor analysis model (1.1) as a model for the inter-dependency with a corresponding second moment structure, then the conditions of Theorem 3 are essentially the assumptions of the factor analysis model. Hence, the \( \chi^2 \) goodness-of-fit test based on \( G_1 \) or \( G_2 \) can be used for practically any factor analysis model.
Appendix

Proof of Lemma 2. The leading term in the expansion of \( \hat{\gamma} - \gamma(n) \) given in Lemma 1 has the form of the generalized least squares estimator. We shall apply a linear transformation to the rows of \( F \) and vech\{\( m - \Sigma(\gamma(n)) \)} and the corresponding rows and columns of \( \Omega \) so that \( \hat{\gamma}_e - \gamma_e(n) \) is isolated and the transformed \( \tilde{\Omega} \) is diagonal. We first note that \( F \) has the form

\[
\tilde{F} = (F_\beta, F_f, F_e),
\]

where

\[
F_\beta = \frac{\partial \text{vech} \Sigma(\gamma_0)}{\partial \gamma_\beta} = \frac{2\psi_p[(I_r, 0_{rk})' \otimes (\beta, I_k)' \Sigma_{xx}]}{\tilde{\gamma_\beta}},
\]

\[
F_f = \frac{\partial \text{vech} \Sigma(\gamma_0)}{\partial \gamma_f} = \psi_p[(\tilde{\beta}, I_k)' \otimes (\tilde{\beta}, I_k)'] \Phi_k,
\]

\[
F_e = \frac{\partial \text{vech} \Sigma(\gamma_0)}{\partial \gamma_e} = \psi_p L_p.
\]

Define the transformation matrix

\[
Q = J \psi_p(T' \otimes T') \Phi_p,
\]

where

\[
T = (T_1, T_2),
\]

\[
T_1 = (I_r, -\beta')',
\]

\[
r = p - k,
\]

\[
T_2 = \Sigma^{-1}_{ee}(\beta, I_k)' \Sigma_{pp}
\]

\[
\Sigma_{pp} = [(\beta, I_k) \Sigma^{-1}_{ee}(\beta, I_k)']^{-1},
\]

and \( J \) is the \( \frac{1}{2}p(p+1) \times \frac{1}{2}p(p+1) \) permutation matrix such that for a \( p \times p \) symmetric \( A_1 \), a \( p \times s \) \( A_2 \), and a \( p \times u \) \( A_3 \),
\[ Q \text{vech } A_1 = \left( \text{vech } T'_1 A_1 T_1 \right)' \], \( \left( \text{vech } T'_2 A_2 T_2 \right)' \),
\[ Q \psi_p (A_2 \otimes A_3) = J \psi_p (T'_2 A_2 \otimes T'_1 A_3) \]
\[ = \left\{ [\psi_r (T'_1 A_2 \otimes T'_1 A_3)]', C_2, [\psi_k (T'_2 A_2 \otimes T'_2 A_3)]' \right\}', \]
where
\[ C_2 = \frac{1}{2}(T'_1 A_2 \otimes T'_2 A_3) + \frac{1}{2} \left\{ [T'_2 A_2 \otimes (T'_1 A_3)_1]', \ldots, [T'_2 A_2 \otimes (T'_1 A_3)_r]' \right\}', \]
and \( (T'_i A_3)_i \) is the \( i \)-th row of \( T'_i A_3 \). Using (2.2), (2.3), and
\[ T'_1 (\beta, I_k)' = 0, \]
\[ T'_2 (\beta, I_k)' = I_k, \]
\[ T'_1 \Sigma_{xx} T_2 = 0, \]
we obtain
\[ Q.F = \begin{pmatrix}
0 & 0 & \psi_r (T'_1 \otimes T'_1) L_p \\
I_r \otimes \Sigma_{xx} & 0 & \left( T'_1 \otimes T'_2 \right) L_p \\
\Sigma_0 \otimes \Sigma_{xx} & I & \psi_k (T'_2 \otimes T'_2) L_p
\end{pmatrix}, \]
\[ Q_{\sim} Q'_{\sim} = \text{block diag} \left\{ \psi_r (\Sigma_{uu} \otimes \Sigma_{uu}) \psi_r, \frac{1}{2} \Sigma_{uu} \otimes \Sigma_{\eta \eta}, \psi_k (\Sigma_{\eta \eta} \otimes \Sigma_{\eta \eta}) \psi_k \right\}, \]
where
\[ \Sigma_0 = \Sigma_{pp} (\beta, I_k) \Sigma_{xx} (I_r, 0)', \]
\[ \Sigma_{uu} = T'_1 \Sigma_{xx} T_1, \]
\[ \Sigma_{\eta \eta} = \Sigma_{pp} + \Sigma_{xx}. \]
Also,
\[ Q \text{vech } \{ m - \Sigma [\gamma (n)] \} = (w'_1, w'_2, w'_3), \]
where

15
\[ w_1 = \text{vech} \tilde{T}_1' [m_{\tilde{e}}(n) - \text{diag}(\tilde{\gamma}_e(n))] T_1, \]
\[ w_2 = \text{vec} \tilde{T}_2' \{ m - \Sigma [\tilde{\gamma}(n)] \} T_1, \]
\[ w_3 = \text{vech} \tilde{T}_2' \{ m - \Sigma [\tilde{\gamma}(n)] \} T_2. \]

Because the matrix

\[
\begin{pmatrix}
I_r \otimes \Sigma_{zz} & 0 \\
\Sigma_0 \otimes \Sigma_{zz} & I
\end{pmatrix}
\]

is nonsingular, and because \( Q\tilde{Q}' \) is block diagonal, it follows that the last \( p \) rows of \([F'Q'(Q\tilde{Q}' Q F')^{-1} F'Q'(Q\tilde{Q}' Q F')^{-1}] \) are

\[
\{ [L_p'(T_1 \otimes T_1) \psi_r^{-1} \psi_r (T_1' \otimes T_1') L_p]^{-1} L_p'(T_1 \otimes T_1) \psi_r^{-1} V_{vv}^{-1}, 0, 0 \},
\]

where

\[ V_{vv}^{-1} = \psi_r (\Sigma_{uu} \otimes \Sigma_{vv}) \psi_r^{-1} = \hat{\psi}_r (\Sigma_{uu}^{-1} \otimes \Sigma_{vv}^{-1}) \hat{\psi}_r. \]

Thus, using (2.1), (2.2), and (2.3),

\[
\tilde{\gamma}_e - \gamma(n) = [L_p'(T_1 \Sigma_{uu}^{-1} T_1' \otimes T_1 \Sigma_{uu}^{-1} T_1') L_p]^{-1} L_p'(T_1 \Sigma_{uu}^{-1} \otimes T_1 \Sigma_{uu}^{-1}) \hat{\psi}_r \psi_r^{-1} \\
= [L_p'(T_1 \Sigma_{uu}^{-1} T_1' \otimes T_1 \Sigma_{uu}^{-1} T_1') L_p]^{-1} L_p'(T_1 \Sigma_{uu}^{-1} T_1' \otimes T_1 \Sigma_{uu}^{-1} T_1') K_p \delta_e(n).
\]

**Proof of Lemma 4.** We first note that by Lemma 1

\[
\tilde{\beta} = \beta + O_p \left( \frac{1}{\sqrt{n}} \right), \\
\tilde{\gamma}_e = \gamma_e(n) + O_p \left( \frac{1}{\sqrt{n}} \right) = \gamma_0 + o_p(1).
\]

Thus,
\[ 
\hat{\mu}_{uv} - \hat{\Sigma}_{uv} = (I_r, -\beta') [m - \text{diag}(\hat{\gamma}_e)] (I_r, -\beta')' - C - C' \\
+ (\hat{\beta} - \beta)' (0, I_k) [m - \text{diag}(\hat{\gamma}_e)] (0, I_k)' (\hat{\beta} - \beta) \\
= (I_r, -\beta') [m_{ce} - \text{diag}(\hat{\gamma}_e)] (I_r, -\beta')' + o_p \left( \frac{1}{\sqrt{n}} \right),
\]

where we have used

\[
C = (\hat{\beta} - \beta)' (0, I_k) [m - \text{diag}(\hat{\gamma}_e)] (I_r, -\beta')' \\
= (\hat{\beta} - \beta)' \left\{ m_{fe}(n) (I_r, -\beta')' + (0, I_k) [m_{ce}(n) - \text{diag}(\hat{\gamma}_e)] (I_r, -\beta')' \right\} \\
= O_p \left( \frac{1}{\sqrt{n}} \right).
\]

By Lemma 2,

\[
\text{vec} \left[ \text{diag}(\hat{\gamma}_e) - \text{diag}(\gamma_e(n)) \right] = L_p [\hat{\gamma}_e - \gamma_e(n)] \\
= L_p (L_p H L_p)^{-1} L_p H L_p \delta_e(n) + o_p \left( \frac{1}{\sqrt{n}} \right).
\]

Hence, by (2.1),

\[
\text{vec}(\hat{\mu}_{uv} - \hat{\Sigma}_{uv}) = [(I_r, -\beta') \otimes (I_r, -\beta')] \text{vec} [m_{ce} - \text{diag}(\gamma_e(n)) + \text{diag}(\gamma_e(n) - \hat{\gamma}_e)] \\
+ o_p \left( \frac{1}{\sqrt{n}} \right) \\
= [(I_r, -\beta') \otimes (I_r, -\beta')] [I - L_p (L_p H L_p)^{-1} L_p H] K_p \delta_e(n) + o_p \left( \frac{1}{\sqrt{n}} \right).
\]

The result follows because

\[ \hat{\Sigma}_{uv} = \Sigma_{uv} + o_p(1). \]
TECHNICAL REPORTS
NATIONAL SCIENCE FOUNDATION - GRANT DMS 82-19748
Econometric Workshop


