

UCLA

Department of Statistics Papers

Title

Semi-parametric X^2 Testing in Mean Structure and Covariance Structure Analysis Using Projections

Permalink

<https://escholarship.org/uc/item/6n22p3nt>

Authors

Jennrich, Robert
Satorra, Albert

Publication Date

2011-02-28

Peer reviewed

Semi-parametric χ^2 testing in mean structure
and covariance structure analysis using
projections

Robert Jennrich

University of California, Los Angeles

and

Albert Satorra*

Universitat Pompeu Fabra and Barcelona GSE

May 13, 2011

Mean structures form a basis for mean, covariance, and other forms of moment structure analysis including structural equation modeling. It is shown how to analyze mean structures using projections. These are used to derive a simple general goodness of fit test statistic that is asymptotically chi-squared and robust to departures from normality. Projections are also used to derive two goodness of fit test statistics for mean structures that are substructures of a more general mean structure. One of these uses the difference of two goodness of fit test statistics, one for the general structure and one for the substructure. It is shown how to use the mean structure results for covariance structure analysis. Best generalized least squares, or ADF estimates are not required. Any asymptotically normal estimates may be use. The primary methods used for testing mean and covariance structures are orthogonal complement methods. A basic difficulty with using these is identified. Specific examples show how the general results may be applied to generalized nonlinear regression and to autoregression with measurement errors. Simulation studies investigate the type one errors and power of the test statistics involved. An appendix contains a review of the basic asymptotic and projection methods used. It also gives conditions that lead to the commonly made assumption that the asymptotic covariance matrix of a vectorized form of a sample covariance matrix is positive definite and that this is a very mild assumption.

Key words: Projections, statistical software, goodness of fit testing, Browne's goodness of fit test, restricted models, difference tests, generalized least squares, nonsingularity of Γ , matrix orthogonal complements, multivariate nonlinear regression, autoregression, Monte Carlo,

* Research of the second author is supported by grant SEJ2006-13537 from the Spanish Ministry of Science and Technology.

Introduction

Moment structure analysis is a popular form of multivariate analysis. Applications include mean structure analysis, confirmatory factor analysis, covariance structure analysis, mean and covariance structure analysis, and structural equation modeling. Rather than deriving test statistics for these and other moment structure analyses, we consider these as special cases of mean structure analysis.

There is a good deal of software for moment structure analysis based on normal sampling. In practice, however, one is seldom sampling from a normal distribution. One may proceed under the hope that the departures from normality faced are mild enough to produce satisfactory analyses. Chau, Bentler, and Satorra (1991) and others have shown, however, that for some problems and some departures from normality, estimates and goodness of fit test statistics can have distributions that differ markedly from distributions derived under the assumption of normal sampling. This has motivated the development of semi-parametric estimation and testing methods and motivates the approach employed here.

A basic problem is given a sequence of asymptotically normally distributed random vectors t_n with asymptotic mean μ_0 find a test statistic for a mean structure $\mu(\theta)$, that is a statistic that may be used to test the goodness of fit hypothesis $\mu(\theta) = \mu_0$ for some θ . It is shown how to derive such a test statistic. The statistic is asymptotically chi-squared under the goodness of fit hypothesis.

It is also shown how to derive two goodness of fit test statistics for a mean structure that is a substructure of a more general structure. One of these uses the difference of two goodness of fit test statistics, one for the general structure and one for the substructure.

These statistics are derived using projections. Projections are used because they can be expressed as continuous functions of basic parameters. Alternative approaches are based on using orthogonal complements or Moore-Penrose inverses. Neither of these are in general continuous functions of basic parameters. One can work around this problem, but not without some difficulty and in the past not always successfully. This difficulty is discussed in the context of Browne's (1984) fundamental goodness of fit test for covariance structure analysis.

The mean structure results are used to derive results for covariance structure analysis. In covariance structure analysis $t_n = s_n$ where s_n is a vectored form of the sample covariance matrix for a sample of size n . Let σ_0 be the asymptotic mean of s_n . Using mean structure results we derive goodness of fit tests for covariance structures $\sigma(\theta)$ and for substructures of $\sigma(\theta)$.

The starting point for our covariance structure analysis and that of others is the asymptotic result $\sqrt{n}(s_n - \sigma_0) \xrightarrow{D} N(0, \Gamma)$. It is common practice to assume that Γ is positive definite. We prove that when the distribution sampled has a nonsingular component, no matter how small, Γ is positive definite and that it is probably positive definite even if the distribution sampled is singular. This is comforting because the positive definiteness of Γ it is a frequently made assumption.

Specific examples show how our general results may be applied to multivariate nonlinear regression and autoregression with measurement errors. Simulation studies investigate the type one errors and power of the test statistics involved.

An appendix contains a review of the basic asymptotic and projection methods used. Generalized projections are used, that is projections in metrics that may differ from the standard euclidian metric. The basic result used is $\|Pz\|_{\Sigma^{-1}}^2 \sim \chi_q^2$ where $z \sim N_p(0, \Sigma)$ and P is a projection in the metric of Σ^{-1} of rank q .

Nowadays there is a good deal of literature and software that has implemented semi-parametric methods for mean and covariance structure analysis. With regard to software there is EQS (Bentler, 2008), LISREL (Jöreskog & Sörbom 1994) and MPlus (Muthén & Muthén, 2007).

This work began with a goodness of fit test for covariance structure analysis first introduced by Browne (1984). This has been discussed and extended to more general moment structure analysis in Satorra (1989), Yuan and Bentler (1997), Satorra and Nuedecker (2003), Boomsma and Hoogland (2001) and Curran, West and Finch (1996). Chou, Bentler and Satorra (1991) have investigated a number of these methods using Monte Carlo methods.

The above references often use somewhat informal arguments that leave something to be desired in the form of mathematical rigor. For example Browne's proof that his basic goodness of fit statistic has an asymptotic χ^2 distribution fails because he assumes without proof that his orthogonal complement function Δ_c exists and is continuous. It turns out, however, that

the computing formula he uses does have an asymptotic χ^2 distribution. This follows from our Theorem 4.

Two methods are used to compute Browne's basic statistic in the software identified above. At present neither has been shown to have an asymptotic distribution. We show that in fact they do using Theorem 1 and Lemma 6.

The difference of two Browne type goodness of fit statistics for testing a restriction of a more general model is often used, but it has not been shown that such a difference has an asymptotic χ^2 distribution. This is proved in Theorem 6. This difference method is easy to carry out by using any of the software identified above for unrestricted testing twice. Taking the difference of the goodness of fit statistics produced gives the required test statistic. This is a timely result, because such differences are widely used in practice but until now it has not been proved this difference has an asymptotic χ^2 distribution.

Mean structure analysis

Let t_1, \dots, t_n be a sequence of random vectors such that

$$\sqrt{n}(t_n - \mu_0) \xrightarrow{D} N_p(0, \Sigma)$$

The sequence t_n is said to be asymptotically normally distributed with asymptotic mean μ_0 and asymptotic covariance matrix Σ .

We do not assume the t_n are sample means because in applications they may not be. They may for example be higher order sample moment vectors.

A function $\mu(\theta)$ of a parameter vector θ with values in R^p will be called a mean structure. We are interested in testing the hypothesis that $\mu_0 = \mu(\theta)$ for some θ . This is called the goodness of fit hypothesis for the mean structure $\mu(\theta)$.

A mean structure $\nu(\beta)$ is called a substructure of $\mu(\theta)$ if there is a function $g(\beta)$ such that $\nu(\beta) = \mu(g(\beta))$. Or more compactly such that $\nu = \mu \circ g$. We are interested in testing the goodness of fit of $\nu(\beta)$ given that of $\mu(\theta)$. Tests of this form can be considerably more powerful than tests that don't use fitting superstructures.

Goodness of fit testing

If f is a function that is differentiable at x , then $\dot{f}(x)$ will denote the Jacobian of f at x .

Theorem 1: If

1. $\sqrt{n}(t_n - \mu_0) \xrightarrow{\mathcal{D}} N_p(0, \Sigma)$ and $\Sigma_n \xrightarrow{p} \Sigma$.
2. Σ is positive definite.
3. μ is a continuously differentiable map from an open subset Θ of R^q into R^p , $\mu(\theta_0) = \mu_0$ for some $\theta_0 \in \Theta$, and $\dot{\mu}(\theta_0)$ has full column rank.
4. $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution.

Then

$$G(\mu) = n\|(I - P_n)(t_n - \mu(\hat{\theta}_n))\|_{\Sigma_n^{-1}}^2 \xrightarrow{\mathcal{D}} \chi_{p-q}^2$$

where $P_n = U_n(U_n'\Sigma_n^{-1}U_n)^{-1}U_n'\Sigma_n^{-1}$ and $U_n = \dot{\mu}(\hat{\theta}_n)$.

Proof: Let $P = U(U'\Sigma^{-1}U)^{-1}U'\Sigma^{-1}$ where $U = \dot{\mu}(\theta_0)$ and note that $P_n \xrightarrow{p} P$. Note that $(I - P)\dot{\mu}(\theta_0) = (I - P)U = U - U = 0$. Using this and the delta method

$$\sqrt{n}(I - P_n)(\mu(\hat{\theta}_n) - \mu(\theta_0)) \stackrel{a}{=} (I - P)\dot{\mu}(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0) = 0$$

Using this and the Slutsky theorem

$$\begin{aligned} & \sqrt{n}(I - P_n)(t_n - \mu(\hat{\theta}_n)) \\ &= \sqrt{n}(I - P_n)(t_n - \mu_0) - \sqrt{n}(I - P_n)(\mu(\hat{\theta}_n) - \mu(\theta_0)) \\ &\stackrel{a}{=} (I - P_n)\sqrt{n}(t_n - \mu_0) \xrightarrow{D} (I - P)z \end{aligned}$$

where $z \sim N(0, \Sigma)$. Using the Slutsky theorem again

$$G(\mu) = n\|(I - P_n)(t_n - \mu(\hat{\theta}_n))\|_{\Sigma_n^{-1}}^2 \xrightarrow{D} \|(I - P)z\|_{\Sigma^{-1}}^2$$

Because $I - P$ is a projection in the metric of Σ^{-1} of rank $p - q$ it follows from Lemma A2 of the Appendix that $\|(I - P)z\|_{\Sigma^{-1}}^2 \sim \chi_{p-q}^2$ and hence that $G(\mu) \xrightarrow{D} \chi_{p-q}^2$. ■

Remark: The assumption that $\mu_0 = \mu(\theta_0)$ for some $\theta_0 \in \Theta$ is a goodness of fit assumption for μ . Theorem 1 may be used to test this assumption.

Remark: Any estimator that satisfies assumption 4 may be used. In particular any generalized least squares estimator may be used. This is shown in Lemma A3 of the Appendix.

Remark: The t_n in assumption 1 could be a variety of things. It for example might be the mean \bar{y}_n of a sample of size n . It might be a vector of sample proportions. It might be a vector of sample variances and covariances. It might be a vector of arbitrary sample moments.

Consistency of the goodness of fit test based on $G(\mu)$

The goodness of fit hypothesis in Theorem 1 is the assumption that $\mu(\theta_0) = \mu_0$ for some $\theta_0 \in \Theta$. What happens when this hypothesis doesn't hold? One might hope that $G(\mu) \xrightarrow{p} \infty$ as $n \rightarrow \infty$. That is that the test based on $G(\mu)$ is a consistent test. This is very nearly true.

Corollary 1: If the assumption $\mu(\theta_0) = \mu_0$ in Theorem 1 is replaced by the assumption $(I - P)(\mu_0 - \mu(\theta_0)) \neq 0$ where $P = U(U'\Sigma^{-1}U)^{-1}U'\Sigma^{-1}$ and $U = \dot{\mu}(\theta_0)$, then $G(\mu) \xrightarrow{p} \infty$ as $n \rightarrow \infty$.

Proof: Since $P_n \xrightarrow{p} P$ and $\Sigma_n \xrightarrow{p} \Sigma$

$$n^{-1}G(\mu) = \|(I - P_n)(t_n - \mu(\hat{\theta}_n))\|_{\Sigma_n^{-1}}^2 \xrightarrow{p} \|(I - P)(\mu_0 - \mu(\theta_0))\|_{\Sigma^{-1}}^2 \neq 0$$

Thus $G(\mu) \xrightarrow{p} \infty$. ■

Remark: If $\mu_0 - \mu(\theta_0)$ is not zero it seems unlikely that it is precisely in the column space of $\dot{\mu}(\theta_0)$. If it is not $(I - P)(\mu_0 - \mu(\theta_0)) \neq 0$ and this implies the test is consistent. Another way to say this is that if μ_0 is chosen at random the test will be consistent with probability one.

Testing the goodness of fit of a substructure

Let μ be a mean structure and ν be a substructure of μ . More precisely let $\nu = \mu \circ g$ for some function g . One often wants to test the goodness of fit of ν given the goodness of fit of μ . Tests of this form can be considerably more powerful than tests of the form discussed in the previous section.

Theorem 2: If

1. $\sqrt{n}(t_n - \mu_0) \xrightarrow{\mathcal{D}} N_p(0, \Sigma)$ and $\Sigma_n \xrightarrow{p} \Sigma$.
2. Σ is positive definite.
3. μ is a continuously differentiable map from an open subset Θ of R^q into R^p , $\mu_0 = \mu(\theta_0)$ for some $\theta_0 \in \Theta$, and $\dot{\mu}(\theta_0)$ has full column rank.
4. g is a continuously differentiable map from an open subset B of R^k into R^q , $\theta_0 = g(\beta_0)$ for some $\beta_0 \in B$, and $\dot{g}(\beta_0)$ has full column rank.
5. $\nu = \mu \circ g$
6. $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges in distribution.

Then

$$G(\nu|\mu) = n\|(P_n - \tilde{P}_n)(t_n - \nu(\hat{\beta}_n))\|_{\Sigma_n^{-1}}^2 \xrightarrow{\mathcal{D}} \chi_{q-k}^2$$

where $P_n = U_n(U_n'\Sigma_n^{-1}U_n)^{-1}U_n'\Sigma_n^{-1}$, $U_n = \dot{\mu}(\hat{\theta}_n)$, $\hat{\theta}_n = g(\hat{\beta}_n)$, $\tilde{P}_n = X_n(X_n'\Sigma_n^{-1}X_n)^{-1}X_n'\Sigma_n^{-1}$, and $X_n = \dot{\nu}(\hat{\beta}_n)$.

Proof: Let $P = U(U'\Sigma^{-1}U)^{-1}U'\Sigma^{-1}$ where $U = \dot{\mu}(\theta_0)$ and $\theta_0 = g(\beta_0)$. Let $\tilde{P} = X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$ where $X = \dot{\nu}(\beta_0)$. Note that $P_n \xrightarrow{p} P$ and $\tilde{P}_n \xrightarrow{p} \tilde{P}$. Note also that $P\dot{\mu}(\theta_0) = \dot{\mu}(\theta_0)$ and $\tilde{P}\dot{\nu}(\beta_0) = \dot{\nu}(\beta_0)$. Since $\nu = \mu \circ g$

$$\dot{\nu}(\beta_0) = \dot{\mu}(g(\beta_0))\dot{g}(\beta_0) = \dot{\mu}(\theta_0)\dot{g}(\beta_0)$$

Thus $P\dot{\nu}(\beta_0) = \dot{\mu}(\theta_0)\dot{g}(\beta_0) = \dot{\nu}(\beta_0)$ and $(P - \tilde{P})\dot{\nu}(\beta_0) = 0$. By the delta method

$$\sqrt{n}(P_n - \tilde{P}_n)(\nu(\beta_n) - \nu(\beta_0)) \stackrel{a}{=} (P - \tilde{P})\dot{\nu}(\beta_0)\sqrt{n}(\beta_n - \beta_0) = 0$$

Using this and the Slutsky theorem

$$\begin{aligned} \sqrt{n}(P_n - \tilde{P}_n)(t_n - \nu(\hat{\beta}_n)) &= \sqrt{n}(P_n - \tilde{P}_n)(t_n - \mu_0) \\ &\quad - \sqrt{n}(P_n - \tilde{P}_n)(\nu(\hat{\beta}_n) - \nu(\beta_0)) \\ &\stackrel{a}{=} (P_n - \tilde{P}_n)\sqrt{n}(t_n - \mu_0) \xrightarrow{D} (P - \tilde{P})z \end{aligned}$$

where $z \sim N(0, \Sigma)$. By the Slutsky theorem again

$$G(\nu|\mu) = n\|(P_n - \tilde{P}_n)(t_n - \nu(\hat{\beta}_n))\|_{\Sigma_n^{-1}}^2 \xrightarrow{D} \|(P - \tilde{P})z\|_{\Sigma^{-1}}^2$$

Note that $P - \tilde{P}$ is a projection in the metric of Σ^{-1} of rank $q - k$. By Lemma A2, $\|(P - \tilde{P})z\|_{\Sigma^{-1}}^2 \sim \chi_{q-k}^2$ and $G(\nu|\mu) \xrightarrow{D} \chi_{q-k}^2$. ■

Remark: Note that $\nu(\beta_0) = \mu(\theta_0) = \mu_0$. Thus $G(\nu|\mu)$ may be viewed as a statistic for testing the goodness of fit of ν given that of μ .

Remark: Since $\hat{\theta}_n = g(\hat{\beta}_n)$ there is no need to compute a separate $\hat{\theta}_n$. It is only the structure with the smaller number of parameters that needs to be fitted. The statistic $G(\nu|\mu)$ may be viewed as the semi-parametric version of the Rao (1965, p350) statistic.

Theorem 3: If

1. $\sqrt{n}(t_n - \mu_0) \xrightarrow{\mathcal{D}} N_p(0, \Sigma)$ and $\Sigma_n \xrightarrow{p} \Sigma$
2. Σ is positive definite.
3. μ is a continuously differentiable map from an open subset Θ of R^q into R^p , $\mu_0 = \mu(\theta_0)$ for some $\theta_0 \in \Theta$, and $\dot{\mu}(\theta_0)$ has full column rank.
4. g is a continuously differentiable map from an open subset B of R^k into R^q , $\theta_0 = g(\beta_0)$ for some $\beta_0 \in B$, and $\dot{g}(\beta_0)$ has full column rank.
5. $\nu = \mu \circ g$.
6. $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution.
7. $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges in distribution.

Then

$$D(\nu|\mu) = n\|(I - \tilde{P}_n)(t_n - \nu(\hat{\beta}_n))\|_{\Sigma_n^{-1}}^2 - n\|(I - P_n)(t_n - \mu(\hat{\theta}_n))\|_{\Sigma_n^{-1}}^2 \xrightarrow{\mathcal{D}} \chi_{q-k}^2$$

where $P_n = U_n(U_n'\Sigma_n^{-1}U_n)^{-1}U_n'\Sigma_n^{-1}$, $U_n = \dot{\mu}(\hat{\theta}_n)$, $\tilde{P}_n = X_n(X_n'\Sigma_n^{-1}X_n)^{-1}X_n'\Sigma_n^{-1}$, and $X_n = \dot{\nu}(\hat{\beta}_n)$.

Proof: Let $P = U(U'\Sigma^{-1}U)^{-1}U'\Sigma^{-1}$ where $U = \dot{\mu}(\theta_0)$ and $\tilde{P} = X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$ where $X = \dot{\nu}(\beta_0)$. Note that $P_n \xrightarrow{p} P$ and $\tilde{P}_n \xrightarrow{p} \tilde{P}$. By the delta method

$$\sqrt{n}(I - P_n)(\mu(\hat{\theta}_n) - \mu(\theta_0)) \stackrel{a}{=} (I - P)\dot{\mu}(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0) = 0$$

and

$$\begin{aligned} \sqrt{n}(I - P_n)(t_n - \mu(\hat{\theta}_n)) &= \sqrt{n}(I - P_n)(t_n - \mu_0) - \sqrt{n}(I - P_n)(\mu(\hat{\theta}_n) - \mu_0) \\ &\stackrel{a}{=} (I - P)\sqrt{n}(t_n - \mu_0) \end{aligned}$$

Using Lemma A1 of the appendix

$$n\|(I - P_n)(t_n - \mu(\hat{\theta}_n))\|_{\Sigma_n^{-1}}^2 \stackrel{a}{=} n\|(I - P)(t_n - \mu_0)\|_{\Sigma^{-1}}^2$$

Similarly

$$n\|(I - \tilde{P}_n)(t_n - \nu(\hat{\beta}_n))\|_{\Sigma_n^{-1}}^2 \stackrel{a}{=} \|(I - \tilde{P})(t_n - \mu_0)\|_{\Sigma^{-1}}^2$$

Using these asymptotic equalities and equation (1) in the Appendix

$$\begin{aligned} D(\nu|\mu) &\stackrel{a}{=} n\|(I - \tilde{P})(t_n - \mu_0)\|_{\Sigma_n^{-1}}^2 - n\|(I - P)(t_n - \mu_0)\|_{\Sigma_n^{-1}}^2 \\ &= n\|(P - \tilde{P})(t_n - \mu_0)\|_{\Sigma^{-1}}^2 \end{aligned}$$

Using the Slutsky theorem $D(\nu|\mu) \xrightarrow{\mathcal{D}} \|(P - \tilde{P})z\|_{\Sigma^{-1}}^2$ where $z \sim N(0, \Sigma)$. Since $P - \tilde{P}$ is a projection of rank $q - k$ in the metric of Σ^{-1} , it follows from Lemma A2 of the Appendix that $\|(P - \tilde{P})z\|_{\Sigma^{-1}}^2 \sim \chi_{q-k}^2$ and hence $D(\nu|\mu) \xrightarrow{\mathcal{D}} \chi_{q-k}^2$ ■

Remark: It follows from assumptions 4 and 5 that $\nu(\beta_0) = \mu(\theta_0) = \mu_0$. Thus $D(\nu|\mu)$ may be viewed as a statistic for testing the goodness of fit of ν given that of μ .

Remark: This Theorem requires estimates $\hat{\theta}_n$ and $\hat{\beta}_n$ for θ_0 and β_0 . Theorem 2 only required an estimate for β_0 .

Remark: The proof of this theorem is somewhat more difficult than those of Theorems 1 and 2.

Remark: The first term of $D(\nu|\mu)$ is a statistic for testing the goodness of fit of ν without assuming that of μ and the second term is a statistic for testing the goodness of fit of μ . Since there is fairly available software for computing these goodness of fit statistics $D(\nu|\mu)$ can be computed by using the software twice and subtracting. This is a practical advantage because at present there does not seem to be readily available software for computing the statistic $G(\nu|\mu)$ in Theorem 2. It is possible, however, that Theorem 2 will motivate the creation of such software in part to avoid the need to estimate θ_0 and in part to provide an alternate method of testing.

Remark: A test based on this theorem may be viewed as a semi-parametric version of the likelihood ratio test.

Best generalized least squares estimators

In mean structure analysis an estimator $\hat{\theta}_n$ is called a generalized least squares (GLS) estimator if $\hat{\theta}_n$ minimizes

$$Q(\theta) = \|t_n - \mu(\theta)\|_{W_n}^2$$

where $W_n \xrightarrow{p} W$ and W is positive definite. When $W_n = I$ a GLS estimator is called an ordinary least squares (OLS) estimator.

A GLS estimator is called a best generalized least squares estimator (BGLS) if $W_n = \Sigma_n^{-1}$. There has been some reluctance to use BGLS estimators, but their use can simplify formulas. Assume $\hat{\theta}_n$ in Theorem 1 is a BGLS estimator. Then

$$\dot{\mu}(\hat{\theta}_n)' \Sigma_n^{-1} (t_n - \mu(\hat{\theta}_n)) = 0$$

It follows that

$$P_n(t_n - \mu(\hat{\theta}_n)) = 0$$

where P_n is defined as in Theorem 1. Thus

$$(I - P_n)(t_n - \mu(\hat{\theta}_n)) = t_n - \mu(\hat{\theta}_n)$$

and

$$G(\mu) = n \|t_n - \mu(\hat{\theta}_n)\|_{\Sigma_n^{-1}}^2$$

which is a bit simpler than the more general form for $G(\mu)$ given in Theorem 1. Similarly if $\hat{\beta}_n$ in Theorem 2 is a BGLS estimator

$$G(\nu|\mu) = n \|P_n(t_n - \nu(\hat{\beta}_n))\|_{\Sigma_n^{-1}}^2$$

and if the estimators $\hat{\theta}_n$ and $\hat{\beta}_n$ in Theorem 3 are BGLS estimators then

$$D(\nu|\mu) = n\|t_n - \nu(\hat{\theta}_n)\|_{\Sigma_n^{-1}}^2 - n\|t_n - \mu(\hat{\theta}_n)\|_{\Sigma_n^{-1}}^2$$

One can also show that

$$T_n = n\|\mu(\hat{\theta}_n) - \nu(\hat{\beta}_n)\|_{\Sigma_n^{-1}}^2 \xrightarrow{\mathcal{D}} \chi_{q-k}^2$$

Proof: Let $P = U(U'\Gamma^{-1}U)^{-1}U'\Gamma^{-1}$ where $U = \dot{\mu}(\theta_0)$ and note that

$$\sqrt{n}(I - P_n)(\mu(\hat{\theta}_n) - \mu(\theta_0)) \stackrel{a}{=} (I - P)\dot{\mu}(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0) = 0$$

and

$$\begin{aligned} \sqrt{n}(t_n - \mu(\hat{\theta}_n)) &= \sqrt{n}(I - P_n)(t_n - \mu(\hat{\theta}_n)) \\ &= \sqrt{n}(I - P_n)(t_n - \mu_0) - \sqrt{n}(I - P_n)(\mu(\hat{\theta}_n) - \mu_0) \\ &\stackrel{a}{=} \sqrt{n}(I - P_n)(t_n - \mu_0) \end{aligned}$$

Similarly

$$\sqrt{n}(t_n - \nu(\hat{\beta}_n)) \stackrel{a}{=} \sqrt{n}(I - \tilde{P}_n)(t_n - \mu_0)$$

Subtracting

$$\sqrt{n}(\mu(\hat{\theta}_n) - \nu(\hat{\beta}_n)) \stackrel{a}{=} (P_n - \tilde{P}_n)\sqrt{n}(t_n - \mu_0)$$

Let $z \sim N(0, \Sigma)$. Then

$$\begin{aligned} T_n &= n\|\mu(\hat{\theta}_n) - \nu(\hat{\beta}_n)\|_{\Sigma_n^{-1}}^2 \stackrel{a}{=} \|(P_n - \tilde{P}_n)\sqrt{n}(t_n - \mu_0)\|_{\Sigma_n^{-1}}^2 \\ &\xrightarrow{\mathcal{D}} \|(P - \tilde{P})z\|_{\Sigma^{-1}}^2 \sim \chi_{q-k}^2 \blacksquare \end{aligned}$$

Covariance structure analysis

There are many forms of moment structure analysis to which our previous work might be applied. The most commonly used is covariance structure analysis. This will be considered here.

Given a covariance matrix Σ let $\text{vech}(\Sigma)$ be a listing of the diagonal and lower diagonal elements of Σ as a column vector. The elements are read column-wise. For example

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \longrightarrow \begin{pmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{31} \\ \sigma_{22} \\ \sigma_{32} \\ \sigma_{33} \end{pmatrix}$$

Let x_1, \dots, x_n be a sample from a distribution with covariance matrix Σ_0 . Let $\Sigma(\theta)$ be a covariance structure for Σ_0 . While this is the natural way to formulate a covariance structure it will be convenient to let $\sigma(\theta) = \text{vech}(\Sigma(\theta))$ and $\sigma_0 = \text{vech}(\Sigma_0)$ and view $\sigma(\theta)$ as a covariance structure for σ_0 .

We are interested in testing the goodness of fit hypothesis $\sigma_0 = \sigma(\theta)$ for some θ . We are also interested in testing $\sigma_0 = \tau(\beta)$ for some β when τ is a substructure of σ .

For these purposes we will derive covariance structure analysis analogs for the mean structure analysis Theorems 1, 2, and 3.

We begin with a lemma that shows that under appropriate assumptions $\sqrt{n}(s_n - \sigma_0) \xrightarrow{\mathcal{D}} N(0, \Gamma)$. This suggests that mean structure methods with $t_n = s_n$ and $\mu_0 = \sigma_0$ might be used for covariance structure analysis.

Lemma 3: If

1. x_1, \dots, x_n is a sample from an m dimensional distribution distribution with finite fourth moments and covariance matrix Σ_0 .
2. $S_n = \frac{1}{n} \sum (x_i - \bar{x}_n)(x_i - \bar{x}_n)'$.
3. $s_n = \text{vech}(S_n)$ and $\sigma_0 = \text{vech}(\Sigma_0)$
4. $t_i = \text{vech}((x_i - \bar{x})(x_i - \bar{x})')$

Then

$$\sqrt{n}(s_n - \sigma_0) \xrightarrow{\mathcal{D}} N(0, \Gamma)$$

for some Γ .

Moreover

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^n (t_i - \bar{t}_n)(t_i - \bar{t}_n)' \xrightarrow{P} \Gamma$$

Proof: We may assume without loss of generality that the x_i have mean zero. By the Slutsky theorem $\sqrt{n}\bar{x}\bar{x}' \xrightarrow{p} 0$. Thus

$$\sqrt{n}S_n = \frac{1}{\sqrt{n}} \sum x_i x_i' - \sqrt{n}\bar{x}\bar{x}' \stackrel{a}{=} \frac{1}{\sqrt{n}} \sum x_i x_i'$$

And

$$\sqrt{n}(s_n - \sigma_0) \stackrel{a}{=} \frac{1}{\sqrt{n}} \sum (\text{vech}(x_i x_i') - \sigma_0)$$

Since $\sigma_0 = E(\text{vech}(x_i x_i'))$, by the central limit theorem

$$\sqrt{n}(s_n - \sigma_0) \xrightarrow{\mathcal{D}} N(0, \Gamma)$$

where Γ is the common covariance matrix for the $\text{vech}(x_i x_i')$. This is the first assertion.

Using Jennrich (2008, p585) t_i is the i -th infinitesimal jackknife pseudo value for s_n and

$$\Gamma_n = \frac{1}{n} \sum (t_i - \bar{t}_n)(t_i - \bar{t}_n)' \xrightarrow{p} \Gamma$$

which is the second assertion. ■

Remark: The second assertion of this lemma is often used. The only proof we know of for this is that in the reference given.

The following is a corollary of Theorem 1.

Theorem 4: If

1. x_1, \dots, x_n is a sample from an m dimensional distribution distribution with finite fourth moments and covariance matrix Σ_0 .
2. $S_n = \frac{1}{n} \sum (x_i - \bar{x}_n)(x_i - \bar{x}_n)'$.
3. $s_n = \text{vech}(S_n)$ and $\sigma_0 = \text{vech}(\Sigma_0)$.
4. $\sqrt{n}(s_n - \sigma_0) \xrightarrow{\mathcal{D}} N_p(0, \Gamma)$ where $p = m(m+1)/2$ and $\Gamma_n \xrightarrow{p} \Gamma$.
5. Γ is positive definite.
6. σ is a continuously differentiable map from an open subset Θ of R^q into R^p , $\sigma_0 = \sigma(\theta_0)$ for some $\theta_0 \in \Theta$ and $\dot{\sigma}(\theta_0)$ has full column rank.
7. $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution.

Then

$$G(\sigma) = n \|(I - P_n)(s_n - \sigma(\hat{\theta}_n))\|_{\Gamma_n^{-1}}^2 \xrightarrow{\mathcal{D}} \chi_{p-q}^2$$

where $P_n = U_n(U_n' \Gamma_n^{-1} U_n)^{-1} U_n' \Gamma_n^{-1}$ and $U_n = \dot{\sigma}(\hat{\theta}_n)$.

Proof: The assumptions of Theorem 1 are satisfied when $p = m(m+1)/2$, $t_n = s_n$, $\mu_0 = \sigma_0$, $\Sigma = \Gamma$, $\Sigma_n = \Gamma_n$, and $\mu = \sigma$.

It follows from Theorem 1 that

$$G(\sigma) = n\|(I - P_n)(s_n - \sigma(\hat{\theta}_n))\|_{\Gamma_n^{-1}}^2 \xrightarrow{\mathcal{D}} \chi_{p-q}^2$$

where $P_n = U_n(U_n'\Gamma_n^{-1}U_n)^{-1}U_n'\Gamma_n^{-1}$ and $U_n = \dot{\sigma}(\hat{\theta}_n)$. ■

Remark: The assumption $\sigma_0 = \sigma(\theta_0)$ for some $\theta_0 \in \Theta$ is a goodness of fit hypothesis for the covariance structure σ . Theorem 3 can be used to test this hypothesis.

Remark: The assumptions of Lemma 3 are satisfied. Lemma 3 implies assumption 4 and also provides a simple choice for Γ_n .

Remark: The assumption that Γ is positive definite is very mild. This is shown in the Appendix. In-particular if the distribution from which the x_i are sampled has a continuous component no matter how small, then Γ is positive definite. For example if the distribution has a density or is a mixture of a distribution that has a density and a discrete distribution, then Γ is positive definite.

Remark: It is shown in the appendix that if $\hat{\theta}_n$ is a generalized least squares estimate of θ_0 , then $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution.

The following is a corollary of Theorem 2.

Theorem 5: If

1. x_1, \dots, x_n is a sample from an m dimensional distribution distribution with finite fourth moments and covariance matrix Σ_0 .
2. $S_n = \frac{1}{n} \sum (x_i - \bar{x}_n)(x_i - \bar{x}_n)'$.
3. $s_n = \text{vech}(S_n)$ and $\sigma_0 = \text{vech}(\Sigma_0)$.
4. $\sqrt{n}(s_n - \sigma_0) \xrightarrow{\mathcal{D}} N_p(0, \Gamma)$ where $p = m(m+1)/2$ and $\Gamma_n \xrightarrow{p} \Gamma$.
5. Γ is positive definite.
6. σ is a continuously differentiable map from an open subset Θ of R^q into R^p , $\sigma_0 = \sigma(\theta_0)$ for some $\theta_0 \in \Theta$, and $\dot{\sigma}(\theta_0)$ has full column rank.
7. g is a continuously differentiable map from an open subset B of R^k into R^q , $\theta_0 = g(\beta_0)$ for some $\beta_0 \in B$, and $\dot{g}(\beta_0)$ has full column rank.
8. $\tau = \sigma \circ g$.
9. $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges in distribution.

Then

$$G(\tau|\sigma) = n \|(P_n - \tilde{P}_n)(s_n - \tau(\hat{\beta}_n))\|_{\Gamma_n^{-1}}^2 \xrightarrow{\mathcal{D}} \chi_{p-q}^2$$

where $P_n = U_n(U_n' \Gamma_n^{-1} U_n)^{-1} U_n' \Gamma_n^{-1}$, $U_n = \dot{\sigma}(\hat{\theta}_n)$, $\hat{\theta}_n = g(\hat{\beta}_n)$,
 $\tilde{P}_n = X_n(X_n' \Gamma_n X_n)^{-1} X_n' \Gamma_n$, and $X_n = \dot{\tau}(\hat{\beta}_n)$.

Proof: The assumptions of Theorem 2 are satisfied when $p = m(m+1)/2$, $t_n = s_n$, $\mu_0 = \sigma_0$, $\Sigma = \Gamma$, $\Sigma_n = \Gamma_n$, $\mu = \sigma$, and $\nu = \tau$.

It follows from Theorem 2 that

$$G(\tau|\sigma) = n\|(P_n - \tilde{P}_n)(s_n - \tau(\hat{\beta}_n))\|_{\Gamma_n^{-1}}^2 \xrightarrow{\mathcal{D}} \chi_{p-q}^2$$

where $P_n = U_n(U_n'\Gamma_n^{-1}U_n)^{-1}U_n'\Gamma_n^{-1}$, $U_n = \dot{\sigma}(\hat{\theta}_n)$, $\hat{\theta}_n = g(\hat{\beta}_n)$, $\tilde{P} = X_n(X_n'\Gamma_n X_n)^{-1}X_n'\Gamma_n$, and $X_n = \dot{\tau}(\hat{\beta}_n)$. ■

Remark: It follows from assumptions 6, 7, and 8 that $\tau(\beta_0) = \sigma(\theta_0) = \sigma_0$. Thus $D(\tau|\sigma)$ may be viewed as a statistic for testing the goodness of fit of τ given that of σ .

Remark: The assumptions of Lemma 3 are satisfied. Lemma 3 implies assumption 4 and also provides a simple choice for Γ_n .

Remark: As noted above the assumption that Γ is positive definite is very mild.

Remark: It is shown in the Appendix that if $\hat{\beta}_n$ is a GLS estimator of β_0 , then $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges in distribution and assumption 9 is satisfied.

The following is a corollary of Theorem 3.

Theorem 6: If

1. x_1, \dots, x_n is a sample from an m dimensional distribution with finite fourth moments and covariance matrix Σ_0 .
2. $S_n = \frac{1}{n} \sum (x_i - \bar{x}_n)(x_i - \bar{x}_n)'$.
3. $s_n = \text{vech}(S_n)$ and $\sigma_0 = \text{vech}(\Sigma_0)$.
4. $\sqrt{n}(s_n - \sigma_0) \xrightarrow{\mathcal{D}} N_p(0, \Gamma)$ where $p = m(m+1)/2$ and $\Gamma_n \xrightarrow{p} \Gamma$.
5. Γ is positive definite.
6. σ is a continuously differentiable map from an open subset Θ of R^q into R^p , $\sigma_0 = \sigma(\theta_0)$ for some $\theta_0 \in \Theta$, and $\dot{\sigma}(\theta_0)$ has full column rank.
7. g is a continuously differentiable map from an open subset \mathcal{B} of R^k into R^q , $\theta_0 = g(\beta_0)$ for some $\beta_0 \in \mathcal{B}$, and $\dot{g}(\beta_0)$ has full column rank.
8. $\tau = \sigma \circ g$.
9. $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution.
10. $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges in distribution.

Then

$$D(\tau|\sigma) = n\|(I - \tilde{P}_n)(s_n - \tau(\hat{\beta}_n))\|_{\Gamma_n^{-1}}^2 - n\|(I - P_n)(s_n - \sigma(\hat{\theta}_n))\|_{\Gamma_n^{-1}}^2 \xrightarrow{\mathcal{D}} \chi_{q-k}^2$$

where $P_n = U_n(U_n' \Sigma^{-1} U_n)^{-1} U_n' \Sigma_n^{-1}$, $U_n = \dot{\sigma}(\hat{\theta}_n)$, $\tilde{P}_n = X_n(X_n' \Sigma_n^{-1} X_n)^{-1} X_n' \Sigma_n^{-1}$, and $X_n = \dot{\tau}(\hat{\beta}_n)$.

Proof: The assumptions of Theorem 3 are satisfied when $p = m(m+1)/2$, $t_n = s_n$, $\mu_0 = \sigma_0$, $\Sigma = \Gamma$, $\Sigma_n = \Gamma_n$, $\mu = \sigma$, $\nu = \tau$ and $\mu = \sigma$.

It follows from Theorem 3 that

$$D(\tau|\sigma) = n\|(I - \tilde{P}_n)(s_n - \tau(\hat{\beta}_n))\|_{\Gamma_n^{-1}}^2 - n\|(I - P_n)(s_n - \sigma(\hat{\theta}_n))\|_{\Gamma_n^{-1}}^2 \xrightarrow{D} \chi_{q-k}^2$$

where $P_n = U_n(U_n'\Sigma^{-1}U_n)^{-1}U_n'\Sigma_n^{-1}$, $U_n = \dot{\sigma}(\hat{\theta}_n)$, $\tilde{P}_n = X_n(X_n'\Sigma_n^{-1}X_n)^{-1}X_n'\Sigma_n^{-1}$, and $X_n = \dot{\tau}(\hat{\beta}_n)$. ■

Remark: It follows from assumptions 6, 7, and 8 that $\tau(\beta_0) = \sigma(\theta_0) = \sigma_0$. Thus $D(\tau|\sigma)$ may be viewed as a statistic for testing the goodness of fit of τ given that of σ .

Remark: This Theorem requires estimators $\hat{\theta}_n$ and $\hat{\beta}_n$ for θ_0 and β_0 . Theorem 5 only required an estimator for β_0 .

Remark: The test statistic $D(\tau|\sigma)$ can be written in the form $D(\tau|\sigma) = G(\tau) - G(\sigma)$ that is as the difference of two basic goodness of fit statistics. Since there is fairly readily available software for computing these basic goodness of fit statistics, $D(\tau|\sigma)$ can be computed by using the software twice and subtracting.

Remark: A test based on this theorem may be viewed as a semi-parametric version of the likelihood ratio test.

Alternative expressions for $G(\mu)$ and $G(\sigma)$

Using orthogonal complements the goodness of fit test statistics $G(\mu)$ and $G(\sigma)$ can be expressed in an interesting alternative form.

If matrices X and Y have p rows, then they are matrix orthogonal complements if $X'Y = 0$ and the columns of (X, Y) are a basis for R^p .

Lemma 4: If U is a $p \times q$ matrix with full column rank, P projects onto the column space of U in the metric of Σ^{-1} , and H is a matrix orthogonal complement of U , then

$$(I - P)' \Sigma^{-1} (I - P) = H(H' \Sigma H)^{-1} H'$$

Proof: Let V be a matrix orthogonal complement of $\Sigma^{-1}U$. Then the column space of V is the orthogonal complement of the column space of U in the metric of Σ^{-1} and $Q = I - P$ is a projection onto the column space of V in the metric of Σ^{-1} .

Since both H and $\Sigma^{-1}V$ are matrix orthogonal complements of U their columns are linearly independent and span the same space. Thus $\Sigma^{-1}V = HM$ for some nonsingular matrix M . Using the fact that $Q = V(V' \Sigma^{-1} V)^{-1} V' \Sigma^{-1}$

$$\begin{aligned} (I - P)' \Sigma^{-1} (I - P) &= Q' \Sigma^{-1} Q = \Sigma^{-1} Q Q = \Sigma^{-1} Q \\ &= \Sigma^{-1} V (V' \Sigma^{-1} V)^{-1} V' \Sigma^{-1} = HM (M' H' \Sigma HM)^{-1} M' H' \\ &= H (H' \Sigma H)^{-1} H' \blacksquare \end{aligned}$$

The following is a corollary of Lemma 4.

Lemma 5: If H_n is a matrix orthogonal complement of the U_n in Theorem 1, then

$$G(\mu) = n(t_n - \mu(\hat{\theta}_n))' H_n (H_n' \Sigma_n H_n)^{-1} H_n' (t_n - \mu(\hat{\theta}_n))$$

Proof: Using Lemma 4

$$\begin{aligned} G(\mu) &= n \|(I - P_n)(t_n - \mu(\hat{\theta}_n))\|_{\Sigma_n^{-1}}^2 \\ &= (t_n - \mu(\hat{\theta}_n))' (I - P_{U_n})' \Sigma_n^{-1} (I - P_{U_n})(t_n - \mu(\hat{\theta}_n)) \\ &= (t_n - \mu(\hat{\theta}_n))' H_n (H_n' \Sigma_n H_n)^{-1} H_n' (t_n - \mu(\hat{\theta}_n)) \blacksquare \end{aligned}$$

In the context of covariance structure analysis a similar argument gives

Lemma 6: If H_n is a matrix orthogonal complement of the U_n of Theorem 4

$$G(\sigma) = n(t_n - \sigma(\hat{\theta}_n))' H_n (H_n' \Sigma_n H_n)^{-1} H_n' (t_n - \sigma(\hat{\theta}_n)) \blacksquare$$

Remark: Lemmas 5 and 6 provide alternate computing formulas for $G(\mu)$ and $G(\sigma)$. LISREL uses the formula in Lemma 6 to compute there goodness of fit statistic for covariance structure analysis. With this exception no statistical software we know of uses the formulas in Lemmas 5 and 6 or the projection formulas in Theorems 1 and 4. The next section will identify what is used.

Browne's goodness of fit statistic

The formula for $G(\sigma)$ given by Lemma 6, is similar to the goodness of fit test statistic (2.20a) of Browne (1984) which has the form

$$B = n(s_n - \sigma(\hat{\theta}_n))' \hat{\Delta}_c (\hat{\Delta}_c' \hat{Y} \hat{\Delta}_c)^{-1} \hat{\Delta}_c' (s_n - \sigma(\hat{\theta}_n))$$

where $\hat{\Delta}_c = \Delta_c(\hat{\theta}_n)$, Δ_c is a function such that $\Delta_c(\theta)$ is a matrix orthogonal complement of $\Delta(\theta) = \dot{\sigma}(\theta)$, and \hat{Y} is a consistent estimator of the asymptotic covariance matrix of s_n .

Browne fails to prove that B is asymptotically χ_{p-q}^2 because he assumes without proof that Δ_c exists and is continuous. He also provides no way to evaluate B in the form displayed above. He claims, however, that in the notation of our Theorem 4 it can be expressed in the computable form

$$B = n(s_n - \sigma(\hat{\theta}_n))' (\Gamma_n^{-1} - \Gamma_n^{-1} P_n) (s_n - \sigma(\hat{\theta}_n))$$

This is his formula (2.20b). It is the formula used by EQS to evaluate Browne's statistic. Note that

$$B = n(s_n - \sigma(\hat{\theta}_n))' \Gamma_n^{-1} (I - P_n) (s_n - \sigma(\hat{\theta}_n)) = n \|(I - P_n)(s_n - \sigma(\hat{\theta}_n))\|_{\Sigma_n^{-1}}^2 = G(\mu)$$

This means that as implemented by EQS Browne's statistic is equal to $G(\sigma)$.

If the $\hat{\Delta}_c$ in the first expression for B were replaced by H_n from the previous section and \hat{Y} were replaced by Γ_n , the resulting

$$B = n(s_n - \sigma(\hat{\theta}_n))' H_n (H_n' \Gamma_n H_n)^{-1} H_n' (s_n - \sigma(\hat{\theta}_n)) = G(\sigma)$$

This is the formula used by LISREL to compute Browne's statistic and it too is equal to $G(\sigma)$.

Since what is computed by EQS and LISREL is equal to $G(\sigma)$ it follows from Theorem 4 that these statistics have an asymptotic χ^2_{p-q} distribution. This is something that has not been proved previously and is an important result because EQS and LISREL are used extensively for goodness of fit testing in covariance structure analysis. We don't know what Mplus and other statistical software use to compute Browne's statistic, but it seems likely that they use one of the two formulas given above.

Best generalized least squares estimators

In covariance structure analysis an estimator $\hat{\theta}_n$ is called a BGLS estimator if $\hat{\theta}_n$ minimizes

$$Q(\theta) = \|s_n - \sigma(\theta)\|_{\Gamma_n^{-1}}^2$$

If $\hat{\theta}_n$ and $\hat{\beta}_n$ in Theorems 4, 5, and 6 are BGLS estimators

$$\begin{aligned} G(\sigma) &= n\|s_n - \sigma(\hat{\theta}_n)\|_{\Sigma_n^{-1}}^2 \\ G(\tau|\sigma) &= n\|P_n(s_n - \sigma(\hat{\theta}_n))\|_{\Gamma_n^{-1}}^2 \\ D(\tau|\sigma) &= n\|s_n - \tau(\hat{\theta}_n)\|_{\Gamma_n^{-1}}^2 - n\|s_n - \sigma(\hat{\theta}_n)\|_{\Gamma_n^{-1}}^2 \end{aligned}$$

Moreover

$$T_n = \|\sigma(\hat{\theta}) - \tau(\hat{\beta})\|_{\Gamma_n^{-1}}^2 \xrightarrow{\mathcal{D}} \chi_{q-k}^2$$

Because covariance structure analysis is a form of mean structure analysis these results follow from those in the section on best generalized least squares estimators for mean structure analysis.

Remark: While the use of BGLS estimators simplifies the formulas for $G(\sigma)$, $G(\tau|\sigma)$, and $D(\tau|\sigma)$ from a practical point BGLS estimators have proved to be quite difficult to use with modest sample sizes and moderate sized models. The main problem is that programs used to compute them fail to converge and often give improper solutions. These problems are the main reason we have developed our results in a way that does not require the use of BGLS estimators.

First and second moment structure analysis

Let x be a random vector. Let

$$y = \begin{pmatrix} x \\ \text{vech}(xx') \end{pmatrix}$$

a vector containing the first and second powers of the components of x . Let $\xi_0 = E(y)$. And let $\xi(\theta)$ be a structure or model for ξ_0 . We are interested in testing the goodness of fit of $\xi(\theta)$.

Let μ_0 and Σ_0 be the mean vector and covariance matrix for x . Then

$$\xi_0 = \begin{pmatrix} \mu_0 \\ \text{vech}(\Sigma_0 + \mu_0\mu_0') \end{pmatrix}$$

For applications to mean and covariance structure analysis one defines $\xi(\theta)$ in terms of structures $\mu(\theta)$ and $\Sigma(\theta)$ for μ_0 and Σ_0 as

$$\xi(\theta) = \begin{pmatrix} \mu(\theta) \\ \text{vech}(\Sigma(\theta) + \mu(\theta)\mu(\theta)') \end{pmatrix}$$

Let x_1, \dots, x_n be a sample from the distribution of x and

$$y_i = \begin{pmatrix} x_i \\ \text{vech}(x_i x_i') \end{pmatrix}$$

Then

$$\sqrt{n}(\bar{y}_n - \xi_0) \xrightarrow{\mathcal{D}} N(0, \Gamma)$$

where Γ is the covariance matrix of y . One can use mean structure analysis methods to construct tests for the goodness of fit of $\xi(\theta)$ and substructures of $\xi(\theta)$. In a paper that is already too long we will not consider the details of this.

Corrected test statistics

It has been known for some time that in covariance structure analysis Browne's (1984, Proposition 4) goodness of fit test statistic can seriously over reject when the goodness of fit hypothesis holds. Yuan and Bentler (1997) show this problem can be ameliorated by using a corrected version of Browne's statistic B of the form

$$\text{CB}(\sigma) = \frac{B(\sigma)}{1 + B(\sigma)/n}$$

We will use this for our $G(\sigma)$ statistic. That is

$$\text{CG}(\sigma) = \frac{G(\sigma)}{1 + G(\sigma)/n}$$

And for the difference statistic $D(\tau|\sigma)$ use

$$\text{CD}(\tau|\sigma) = \text{CG}(\tau) - \text{CG}(\sigma)$$

where $\text{CG}(\tau) = G(\tau)/(1 + G(\tau)/n)$.

Examples

Multivariate nonlinear regression

Let $f(\theta)$ be a p -component continuously differentiable function of θ for all θ in a compact subset of Θ of R^q and assume $f(\theta)$ is one to one. For $i = 1, \dots, n$ let

$$y_i = f(\theta_0) + \epsilon_i$$

where θ_0 is an interior point of Θ and $\epsilon_1, \dots, \epsilon_n$ is a sample from a distribution with mean zero and positive definite covariance matrix Σ . This is a multivariate nonlinear regression model.

From the central limit theorem

$$\sqrt{n}(\bar{y}_n - f(\theta_0)) \xrightarrow{\mathcal{D}} N_p(0, \Sigma)$$

From the law of large numbers

$$S_n = \frac{1}{n} \sum (y_i - \bar{y}_n)(y_i - \bar{y}_n)' \xrightarrow{p} \Sigma$$

For $\theta \in \Theta$ let

$$Q_n(\theta) = (\bar{y}_n - f(\theta))' S_n^{-1} (\bar{y}_n - f(\theta))$$

and let $\hat{\theta}_n$ be any minimizer of $Q_n(\theta)$. Assume $\dot{f}(\theta_0)$ has full column rank. The assumptions of Lemma A3 are satisfied when $t_n = \bar{y}_n$ and $W_n = S_n^{-1}$. It follows from Lemma A3 that $\hat{\theta}_n$ is an asymptotically normal estimator of θ_0 .

The assumptions of Theorem 1 are satisfied when $t_n = \bar{y}_n$, $\Sigma_n = S_n^{-1}$ and $\mu = f$. It follows from Theorem 1 that

$$G(f) = n \|(I - P_n)(\bar{y} - f(\hat{\theta}_n))\|_{S_n^{-1}} \xrightarrow{\mathcal{D}} \chi_{p-q}^2$$

where $P_n = U_n(U_n'S_n^{-1}U_n)^{-1}U_n'S_n^{-1}$ and $U_n = \dot{f}(\hat{\theta}_n)$.

We will use simulation to investigate the performance the statistic $G(f)$.

Remark: Here $\hat{\theta}_n$ is a BGLS estimator and the formula for $G(f)$ can be simplified. We will not do this because we also want to consider the use of ordinary least squares estimators and for these the simplified formula is not asymptotically χ_{p-q}^2 .

Simulation study

Let

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

This is a one-way analysis of variance design matrix. Let the i -th component of $f(\theta)$ be $\exp(x_i\theta)$ where x_i is the i -th row of X . Thus $f(\theta)$ is an exponential model. Clearly $f(\theta)$ and $\dot{f}(\theta)$ are continuous functions of θ for all θ and it is easy to show $\dot{f}(\theta)$ has full column rank for all θ .

Let Σ be the randomized block covariance matrix.

$$\Sigma = \begin{pmatrix} 1 & .5 & 0 & 0 \\ .5 & 1 & 0 & 0 \\ 0 & 0 & 1 & .5 \\ 0 & 0 & .5 & 1 \end{pmatrix}$$

Let z_{ij} , $i = 1, \dots, n$ and $j = 1, \dots, 4$ be a sample from the affine transformation $(\chi_3^2 - 3)/\sqrt{6}$ of the chi-squared distribution with 3 degrees of freedom. This distribution is decidedly non-normal. It is quite skewed and fairly long tailed on the right. The z_{ij} , however, have mean zero, variance one, and are independent. For $i = 1, \dots, n$ let

$$e_i = \Sigma^{1/2} \begin{pmatrix} z_{i1} \\ z_{i2} \\ z_{i3} \\ z_{i4} \end{pmatrix}$$

These are a sample from a distribution with mean zero and covariance matrix Σ . For $i = 1, \dots, n$ let $y_i = f(\theta_0) + e_i$ where $\theta_0 = (1, 2)'$. From the discussion above

$$G(f) \xrightarrow{\mathcal{D}} \chi_{p-q}^2 = \chi_2^2$$

In applications one would reject the hypothesis that the y_i have the mean structure $f(\theta)$ if

$$G(f) > cv$$

where cv is the 5% upper quantile of the chi-squared distribution with two degrees of freedom. The nominal type-one error error for this test is 5%. We will use simulation to estimate the actual type-one error for this test.

To this end for various values of n , $N = 1000$ independent samples y_1, \dots, y_n were generated and the percentage of these whose $G(f)$ values exceeded cv were computed. These percentages are estimates of the type-one errors of the tests based on $G(f)$. The results are displayed in the first column of Table 1.

Table 1: Type-one error and power estimates for the goodness of fit tests based on $G(f)$ and $CG(f)$ for samples of size n using 1000 replications of the analysis for each sample size. The margin of error for these estimates is .0138

n	Type-one error		Power	
	$G(f)$	$CG(f)$	$G(f)$	$CG(f)$
25	.091	.042	.280	.184
50	.061	.039	.460	.390
100	.048	.038	.736	.707

The first column of Table 1 contains the the type-one error estimates when using $G(f)$. For $n = 25$ its type-one error estimate differs from its nominal value .05 by much more the estimates margin of error .0138. This suggests that samples of 25 are too small for the test using $G(f)$ to be reliable. The other type-one error estimates when using $G(f)$ are within their margin of error to .05.

The second column in Table 1 are the type one error estimates when using $CG(f)$ the corrected version of $G(f)$. These are all within there margin of error of to .05.

The last two columns of Table 1 display the estimated power of the tests based on $G(f)$ and $CG(f)$ when the goodness of fit hypothesis fails. More specifically a small interaction component $\gamma_0 = (.1, -.1, -.1, .1)'$ was added

to $f(\theta_0)$ and the data y_1, \dots, y_n generated using

$$y_i = f(\theta_0) + \gamma_0 + e_i$$

The mean structure $f(\theta)$ no longer satisfies the goodness of fit hypothesis and hopefully the goodness of fit tests based $G(f)$ and $CG(f)$ will reject this hypothesis. Estimates of the power of these tests are given in the last two columns of Table 1. As expected the estimated power increases as the sample size increases and is fairly large when $n = 100$.

As noted above we have used BGLS estimators in the simulations that produced Table 1. We have also run our simulations using ordinary least squares estimators rather than BGLS estimators. To our surprise this had no effect at all on the results in Table 1. The values of the parameter estimates were a bit different, but the values of the test statistics were very similar for the two types of estimators. So similar they had no effect on the simulation estimates.

Autoregression with measurement errors

We consider a covariance structure for an autoregression model with measurement errors. Models of this form are used to analyze panel data in economics and social sciences.

For our example consider the following model equations

$$\begin{aligned}x_t &= \eta_t + u_t, & t = 1, \dots, 5 \\ \eta_t &= \alpha \eta_{(t-1)} + v_t, & t = 2, \dots, 5\end{aligned}$$

where the u_t 's, v_t 's and η_1 are independent samples from densities with zero means and finite fourth moments.

We assume the u_t have a common variance ψ . Let $\phi_1 = \text{var}(\eta_1)$ and for $t = 2, \dots, 5$ let $\phi_t = \text{var}(v_t)$. Let $\theta = (\psi, \alpha, \phi_1, \dots, \phi_5)'$ be a vector containing the parameters of the model. Let Θ be all θ such that ψ and ϕ_1, \dots, ϕ_5 are positive and $|\alpha| < 1$. This is an open subset of R^7 .

This model is a particular case of the state dependence model of Anderson and Hsiao (1982), an autoregressive model with measurement error used to analyze panel or longitudinal data. For a recent application of this model to financial data, see Bou and Satorra (2007). In applications the x_t 's are the only observable variables. The u_t 's are measurement errors, and the v_t are disturbance terms for the autoregression model.

Let $x = (x_1, \dots, x_5)'$, $v_1 = \eta_1$, $v = (v_1, \dots, v_5)'$ and $u = (u_1, \dots, u_5)'$.

Then

$$x = Lv + u$$

Where

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \alpha & 1 & 0 & 0 & 0 \\ \alpha^2 & \alpha & 1 & 0 & 0 \\ \alpha^3 & \alpha^2 & \alpha & 1 & 0 \\ \alpha^4 & \alpha^3 & \alpha^2 & \alpha & 1 \end{pmatrix}$$

Because the components of u are independent and have densities, the distribution of u has a density. Similarly v has a density. Because L is nonsingular and u and v are independent, x has a density.

The covariance matrix for x is

$$\Sigma(\theta) = L\Phi L' + \psi I_5$$

where Φ is a diagonal matrix with diagonal elements ϕ_1, \dots, ϕ_5 . Note that $\Sigma(\theta)$ is continuously differentiable for all $\theta \in \Theta$ and $\sigma(\theta) = \text{vech}(\Sigma(\theta))$ is also.

Let vectors x_1, \dots, x_n be a sample from this autoregressive model with a parameter vector $\theta_0 \in \Theta$. Then the covariance matrix for the population sampled is $\Sigma_0 = \Sigma(\theta_0)$. Let $S_n = \frac{1}{n} \sum (x_i - \bar{x}_n)(x_i - \bar{x}_n)'$, $s_n = \text{vech}(S_n)$, and $\sigma_0 = \text{vech}(\Sigma_0)$.

It follows from Lemma 3 that $\sqrt{n}(s_n - \sigma_0) \xrightarrow{\mathcal{D}} N(0, \Gamma)$ for some Γ and that $\Gamma_n \xrightarrow{p} \Gamma$ where Γ_n is defined in Lemma 3. We will use this Γ_n .

Since x has a density it follows from Lemma A4 that Γ is positive definite.

Note that σ is a continuously differentiable map from Θ into R^p where $p = 15$. Moreover $\sigma_0 = \text{vech}(\Sigma_0) = \sigma(\theta_0)$.

We will use GLS to estimate θ_0 . It follows from Lemma A3 that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution.

Thus the assumptions of Theorem 4 will be satisfied if $\dot{\sigma}(\theta_0)$ has full column rank. If so it follows from Theorem 4 that

$$G(\sigma) = n\|(I - P_n)(s_n - \sigma(\hat{\theta}_n))\|_{\Gamma_n^{-1}}^2 \xrightarrow{D} \chi_{p-q}^2 = \chi_8^2$$

where $P_n = U_n(U_n'\Gamma_n^{-1}U_n)^{-1}U_n'\Gamma_n^{-1}$ and $U_n = \dot{\sigma}(\hat{\theta}_n)$.

In specific applications one needs to show or assume $\dot{\sigma}(\theta_0)$ has full column rank. For the simulations below this was shown by computing $\dot{\sigma}(\theta_0)$ and its singular values.

For the simulations below NT-GLS was used for the estimation. This is a form of GLS estimation. More specifically the estimator $\hat{\theta}_n$ of θ_0 minimizes

$$Q_n(\theta) = (s_n - \sigma(\theta))'W_n(s_n - \sigma(\theta))$$

$W_n = .5D'(S_n^{-1} \otimes S_n^{-1})D$, and D is the duplication matrix for vech and vec. See Magnus and Neudecker (1999). An alternate expression for $Q_n(\theta)$ is

$$Q_n(\theta) = \text{tr}(S_n^{-1}(S_n - \Sigma(\theta)))^2$$

Consider next testing the goodness of fit of a substructure τ of σ where τ is the restriction of σ obtained by assuming ϕ_2, \dots, ϕ_5 have a common value ϕ . We would like to use Theorem 6 to test the goodness of fit of τ given that of σ by showing that under this hypothesis $D(\tau|\sigma)$ is asymptotically χ_3^2 .

The first 5 assumptions of Theorem 6 are satisfied when $m = 5$ and $p = 15$.

Let \mathcal{B} be the set of all $\beta = (\beta_1, \dots, \beta_4)'$ such that $\beta_1, \beta_3,$ and β_4 are positive and $|\beta_2| < 1$. Then \mathcal{B} is an open set. For all $\beta \in \mathcal{B}$ let

$$g(\beta) = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_4, \beta_4, \beta_4)'$$

Then g is a continuously differentiable map from \mathcal{B} into R^7 . Note that $\dot{g}(\theta)$ has full column rank for all θ .

Given a $\beta_0 \in \mathcal{B}$ let $\theta_0 = g(\beta_0)$ and use this θ_0 to generate the data x_1, \dots, x_n . Then g satisfies assumption 7 of Theorem 6 with $k = 4$ and $q = 7$.

Let $\tau = \sigma \circ g$. Then assumption 8 of Theorem 6 is satisfied.

NT-GLS will be used to estimate θ_0 and β_0 . Then by Lemm 3A of the appendix assumptions 9 and 10 of Theorem 6 are satisfied,

Thus the assumptions of Theorem 6 are satisfied and it follows from Theorem 6 that

$$D(\tau|\sigma) \xrightarrow{D} \chi_3^2$$

where $D(\tau|\sigma)$ is the statistic defined in Theorem 6.

Results of the Monte Carlo simulation study

The data were generated using $\psi = .2$, $\alpha = .6$, and $\phi_t = .2$ for $t = 1, \dots, 5$.

This means

$$\theta_0 = (.2, .6, 1, .2, .2, .2, .2)$$

A Monte Carlo simulation was used to estimate the type-one errors when using $G(\sigma)$ and its corrected version $CG(\sigma)$. These are asymptotically χ_8^2 distributed. Also considered was the likelihood ratio goodness of fit statistic $LR(\sigma)$ derived under the assumption of normal sampling. Since the sampling is not normal, one does not expect its asymptotic distribution to be χ_8^2 , but this method is often used in the hope that its behavior will not depart too seriously from that expected under normal sampling. $LR(\sigma)$ is -2 times the likelihood ratio statistic for testing $\sigma_0 = \sigma(\theta)$ for some θ . This is computed using the maximum likelihood estimate $\hat{\theta}_n$ of θ_0 . The results are given in Table 2.

Table 2: Rejection rate estimates in percent for nominal 5% tests using $G(\sigma)$, its corrected version $CG(\sigma)$, and the corresponding likelihood ratio statistic $LR(\sigma)$ for samples of size n . The data were generated as described above. The number of Monte-Carlo replications was $N = 1000$. The margin of error for these estimates is 1.37%.

n	$G(\sigma)$	$CG(\sigma)$	$LR(\sigma)$
50	15.8	2.7	12.3
100	9.2	3.7	12.6
500	5.8	4.6	11.1
1000	5.8	4.6	11.1

The type-one error estimate for the test based on $G(\sigma)$ is within a margin of error of 5% for samples of 500 or more. Its corrected version has this property for samples of 100 or more.

The likelihood ratio statistic rejects much too often at all sample sizes. This clearly indicates the need for a semi-parametric method for this problem and for semi-parametric methods more generally.

The simulation was also used to estimate the type-one errors when using the difference statistic $D(\tau|\sigma)$ and its corrected version $CD(\tau|\sigma)$. These are asymptotically χ_3^2 distributed. The corresponding likelihood ratio statistic for testing the goodness of fit of τ given that of σ was also considered. The results are given in Table 3.

Table 3: Rejection rate estimates in percent for nominal 5% tests using the difference statistic $D(\tau|\sigma)$, its corrected version $CD(\tau|\sigma)$, and the corresponding likelihood ratio statistic $LR(\tau|\sigma)$ for samples of size n . The data were generated as described above. The number of Monte-Carlo replications was $N = 1000$. The margin of error for these estimates is 1.37%.

n	$D(\tau \sigma)$	$CD(\tau \sigma)$	$LR(\tau \sigma)$
50	28.6	8.6	23.1
100	14.7	6.2	21.1
500	6.7	5.7	23.0
1000	6.5	6.2	22.1

The type-one error estimate for the test based on the difference statistic $D(\tau|\sigma)$ seems to require a sample of over 1000 to be within a margin of error of 5%. For samples of 500 or more, however, its type-one error doesn't seem to be a great deal over 5%, perhaps an acceptable value for model building purposes at least. The type one error estimates for the test based on the corrected statistic $CD(\tau|\sigma)$ are within a margin of error of 5% for sample sizes of 100 or more. Clearly correction helps. The test based on the likelihood ratio statistic $LR(\tau|\sigma)$ rejects much too often.

To consider power we changed the θ_0 for data generation to

$$\theta_0 = (.2, .6, 1, .2, .175, .225, .250)$$

With this choice $\tau(\beta_0) \neq \sigma(\theta_0)$, the goodness of fit hypothesis for τ given σ is not satisfied, and we expect to see tests of this hypothesis rejected. The following table gives power estimates in percent for the goodness of fit of τ using tests based on $G(\tau)$ and $D(\tau|\sigma)$ and their corrected versions.

Table 4: Power estimates in percent for nominal 5% tests using $G(\tau)$ and $D(\tau|\sigma)$ and their corrected versions $CG(\tau)$ and $CD(\tau|\sigma)$ for samples of size n . The data were generated as described above. The number of Monte-Carlo replications was $N = 1000$.

n	$G(\tau)$	$CG(\tau)$	$D(\tau \sigma)$	$CD(\tau \sigma)$
50	30.9	4.7	31.3	9.2
100	18.9	7.4	20.6	10.5
500	16.3	14.4	28.3	26.0
1000	28.5	26.5	48.1	46.7

As expected power increases with n . The powers for the tests based on $D(\tau|\sigma)$ and $CD(\tau|\sigma)$ can be considerable larger than those for the tests based on $G(\tau)$ and $CG(\tau)$. This is what motivated the development of the difference tests.

Discussion

We have discussed semi-parametric χ^2 methods for testing goodness of fit hypotheses in moment structure analysis. These methods are of interest because they do not require sampling from a normal distribution and because the distributions sampled in practice are not normal.

Projection methods are used to derive new results and prove “known” results. The strategy is to derive results for mean structure analysis and extend them to other forms of moment structure analysis.

A basic theoretical result is that if $y \sim N(0, \Sigma)$ and P is a projection in the metric of Σ^{-1} of rank q , then $\|Py\|_{\Sigma^{-1}}^2 \xrightarrow{\mathcal{D}} \chi_q^2$. This result motivated our use of generalized projections. This was also motivated by the fact that the required projections can be expressed as continuous functions of model parameters. This is not true for approaches based on orthogonal complements and Moore-Penrose inverses.

For mean structure analysis we began with an asymptotically normal sequence of statistics t_n such that $\sqrt{n}(t_n - \mu_0) \xrightarrow{\mathcal{D}} N_p(0, \Sigma)$ with Σ positive definite.

Given an assumed mean structure μ for μ_0 we have shown how to test its goodness of fit using the statistic $G(\mu)$ which under natural assumptions has an asymptotic χ^2 distribution.

We have shown that when the goodness of fit hypothesis fails, $G(\mu) \xrightarrow{p} \infty$ as $n \rightarrow \infty$ except in some very special cases. Thus in general tests based on $G(\mu)$ are consistent. We conjecture this is also true for the other goodness

of fit test statistics we have derived.

We have shown how to test the goodness of fit a sub-structure ν of μ . Two test statistics for this were derived $G(\nu|\mu)$ and $D(\nu|\mu)$ and it was shown that if μ_0 is in the range of ν , under some fairly natural additional assumptions these test statistics are asymptotically χ^2 . These statistics may be used to test the goodness of fit of ν given that of μ .

The test statistic $D(\nu|\mu)$ is the difference of two goodness of fit statistics. This corresponds to a well known result when using normal sampling that has been conjectured to hold in the semi-parametric case as well. We have shown that it does.

We have shown how mean structure analysis can be extended to covariance structure analysis. The latter can be viewed as an application of mean structure analysis in which x_1, \dots, x_n is a sample from the population of interest, S_n is its sample covariance matrix, $s_n = \text{vech}(S_n)$, and $t_n = s_n$. Assuming the population sampled has finite fourth moments it follows that $\sqrt{n}(s_n - \sigma_0) \xrightarrow{D} N(0, \Gamma)$ for some σ_0 and Γ , but it is not necessarily true that Γ is positive definite. We have shown, however, that this assumption is very mild.

Our covariance structure analysis statistic $G(\sigma)$ is closely related to Browne's (1984) goodness of fit statistic B . The important difference is that we have proved $G(\sigma)$ is asymptotically χ^2 , but Browne has failed to do this for B because he has assumed, without proof, that his function Δ_c exists and is continuous.

It is sometimes difficult to deal with full column rank assumptions such as

the assumption that $\dot{\mu}(\theta_0)$ has full column rank. In the nonlinear regression example this was easy because there it was easy to see that $\dot{\mu}(\theta)$ has full column rank for all θ . In the auto regression example we showed $\dot{\sigma}(\theta_0)$ had full column rank by computing its singular values which can be done because in the simulation θ_0 is known. In general, however, this cannot be done. In our examples the Gauss-Newton algorithm was used to compute $\hat{\theta}_n$. For this algorithm to converge $\dot{\sigma}(\hat{\theta}_n)$ must have full column rank. If not the program will stop with an error message. Successful convergence indicates that $\dot{\sigma}(\hat{\theta}_n)$ has full column rank. This does not prove $\dot{\sigma}(\theta_0)$ has full column rank, but it does motivate assuming this. Covariance structures $\sigma(\theta)$ for which $\dot{\sigma}(\theta_0)$ fails to have full column rank are often called over parameterized because they can be replaced by a covariance structure with fewer parameters that has the same range and and has a full column rank Jacobian at θ_0 . Over parameterized covariance structures are sometimes encountered in early stages of an analysis.

Other assumptions that must be dealt with is that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ and $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converge in distribution. We have shown that these assumptions are satisfied when θ_0 and β_0 are estimated using GLS.

We conclude with applications to multivariate nonlinear regression and auto-regression with measurement errors. The former is a mean structure analysis problem and the latter is a covariance structure analysis problem. For covariance structure analysis satisfactory performance can require large samples, but smaller samples can give satisfactory performance by using the simple correction procedure introduced by Yuan and Bentler. It is probably

the case that in general semi-parametric covariance structure analysis will require large or at least moderately large samples. However, these days large samples are often encountered. One of the authors is working with an application that has a sample size of 30,000.

Appendix

Delta method

The following is the form of the delta method used in this paper.

If $\sqrt{n}(t_n - \mu_0)$ converges in distribution and f is a function that is differentiable at μ_0 , then

$$\sqrt{n}(f(t_n) - f(\mu_0)) \stackrel{a}{\approx} \dot{f}(\mu_0)\sqrt{n}(t_n - \mu_0)$$

Slutzky method

The following is a general form of the Slutzky method for convergence in distribution.

If $x_n \xrightarrow{p} x$, $y_n \xrightarrow{\mathcal{D}} y$, and f is a continuous mapping, then

$$f(x_n, y_n) \xrightarrow{\mathcal{D}} f(x, y)$$

We use this, immediate corollaries such as

$$x_n + y_n \xrightarrow{\mathcal{D}} x + y \quad , \quad y_n x_n \xrightarrow{\mathcal{D}} yx \quad , \quad x_n y_n \stackrel{a}{\approx} x y_n$$

and

Lemma A1: If $x_n \stackrel{a}{\approx} y_n \xrightarrow{\mathcal{D}} y$ and g is continuous, then $g(x_n) \stackrel{a}{\approx} g(y_n)$.

Proof: Let $f(u, v) = g(u + v) - g(v)$. Then f is continuous and

$$\begin{aligned} g(x_n) - g(y_n) &= g(x_n - y_n + y_n) - g(y_n) = f(x_n - y_n, y_n) \\ &\xrightarrow{\mathcal{D}} f(0, y) = g(y) - g(y) = 0 \blacksquare \end{aligned}$$

Projections

A discussion of projections may be found in books and courses on linear algebra. See for example Halmos (1958). Properties we use are summarized here.

Let W be a $p \times p$ positive definite matrix. For any $x \in R^p$ let

$$\|x\|_W = (x'Wx)^{1/2}$$

This is the norm of x in the metric of W . Let \mathcal{X} be any q dimensional subspace of R^p , let $y \in R^p$, and let \hat{y} be the vector in \mathcal{X} that is closest to y in the metric of W . That is $x = \hat{y}$ minimizes $\|y - x\|_W$ over all $x \in \mathcal{X}$. The vector \hat{y} is the projection of y on \mathcal{X} . Moreover \hat{y} is a linear function of y . The $p \times p$ matrix P for this linear transformation is called a projection matrix because Py is the projection of y onto \mathcal{X} in the metric of W . We need a computing formula for P . To this end let the columns of X be a basis for \mathcal{X} and

$$\beta = \hat{\beta} \min \|y - X\beta\|_W$$

over all $\beta \in R^q$. Using GLS

$$\hat{\beta} = (X'WX)^{-1}X'Wy$$

and

$$\hat{y} = X\hat{\beta} = X(X'WX)^{-1}X'Wy$$

is the projection of y onto \mathcal{X} in the metric of W . It follows that

$$P = X(X'WX)^{-1}X'W$$

If P projects onto \mathcal{X} in the metric of W and \tilde{P} projects onto a subspace \mathcal{Y} of \mathcal{X} in the metric of W , then $P - \tilde{P}$ is a projection in the metric of W that projects onto $\mathcal{X} \cap \mathcal{Y}^\perp$ where \mathcal{Y}^\perp is the orthogonal complement of \mathcal{Y} in the metric of W . Moreover $P - \tilde{P}$ has rank $q - k$ where q is the dimension of \mathcal{X} and k is the dimension of \mathcal{Y} . Also

$$\|(P - \tilde{P})x\|_W^2 = \|Px\|_W^2 - \|\tilde{P}x\|_W^2$$

In particular $I_p - P$ is a projection in the metric of W that projects onto the orthogonal complement of \mathcal{X} in the metric of W and has rank $p - q$. Also

$$\|(I - \tilde{P})x\|_W^2 = \|(I - P)x\|_W^2 - \|\tilde{P}x\|_W^2 \quad (1)$$

Vectors v_1, \dots, v_q are an ortho-normal basis in the metric of W for a vector space \mathcal{X} if v_1, \dots, v_q are a basis for \mathcal{X} , $v_i'Wv_i = 1$ for all i , and $v_i'Wv_j = 0$ for all $i \neq j$. If $x \in \mathcal{X}$

$$\|x\|_W^2 = \sum (v_i'Wx)^2$$

Lemma A2: If

1. If $y \sim N_p(0, \Sigma)$ and Σ is positive definite.
2. P is a projection in the metric of Σ^{-1} onto a q dimensional subspace of R^p .

Then

$$\|Py\|_{\Sigma^{-1}}^2 \sim \chi_q^2$$

Proof: Let v_1, \dots, v_q be a basis for \mathcal{X} that is ortho-normal in the metric of Σ^{-1} and $V = (v_1, \dots, v_q)$. Then $P = VV'\Sigma^{-1}$ and

$$\|Py\|_{\Sigma^{-1}}^2 = y'\Sigma^{-1}VV'\Sigma^{-1}VV'\Sigma^{-1}y = y'\Sigma^{-1}VV'\Sigma^{-1}y = \sum_{i=1}^q (v_i'\Sigma^{-1}y)^2$$

Let $z_i = v_i'\Sigma^{-1}y$. Then the z_i are normally distributed, have variance 1, and $\text{cov}(z_i, z_j) = 0$ when $i \neq j$. It follows that

$$\|Py\|_{\Sigma^{-1}}^2 = \sum_{i=1}^q z_i^2 \sim \chi_q^2 \blacksquare$$

Asymptotic normality of generalized least squares estimators

Generalized least squares estimators are frequently used for estimating parameter vectors. They are the estimators used in our examples. The following lemma gives conditions under which generalized least squares estimators are asymptotically normal.

Lemma A3: If

1. $\sqrt{n}(t_n - f_0) \rightarrow N(0, \Sigma)$.
2. f is a continuously differentiable one to one mapping from a compact subset Θ of R^q into R^p and $\dot{f}(\theta)$ has full column rank for all $\theta \in \Theta$.
3. θ_0 is an interior point of Θ and $f(\theta_0) = f_0$.
4. $Q_n(\theta) = (t_n - f(\theta))'W_n(t_n - f(\theta))$, $W_n \xrightarrow{p} W$, and W is positive definite.
5. $\hat{\theta}_n$ minimizes $Q_n(\theta)$.

Then $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normally distributed.

Proof: We will show first that $\hat{\theta}_n \xrightarrow{p} \theta_0$. Note that

$$\|f(\theta_n) - f_0\| = \|t_n - f_0 - (t_n - f(\theta_n))\| \leq \|t_n - f_0\| + \|t_n - f(\hat{\theta}_n)\| \leq 2\|t_n - f_0\| \xrightarrow{p} 0$$

Now **Assume** $\|\hat{\theta}_n - \theta_0\| \not\xrightarrow{p} 0$. Then for some $\epsilon > 0$, $\delta > 0$, and N

$$P(\|\hat{\theta}_n - \theta_0\| > \epsilon) > \delta$$

for all $n > N$. By the mean value theorem for some $\bar{\theta}_n$ on the line between θ_0 and θ_n .

$$f(\theta_n) - f(\theta_0) = \dot{f}(\bar{\theta}_n)(\hat{\theta}_n - \theta_0)$$

Let $\lambda_{\min}(\theta)$ be the minimum singular value of $\dot{f}(\theta)$ and γ be the minimum of λ_{\min} over Θ . Because $\dot{f}(\theta)$ has full column rank for all $\theta \in \Theta$, $\lambda_{\min}(\theta) > 0$ for all $\theta \in \Theta$. Because λ_{\min} is continuous and Θ is compact, $\gamma = \min_{\theta \in \Theta} \lambda_{\min}(\theta) > 0$. Note that

$$\|f(\hat{\theta}_n) - f(\theta_0)\| = \|\dot{f}(\bar{\theta}_n)(\hat{\theta}_n - \theta_0)\| \geq \lambda_{\min}(\bar{\theta}_n)\|\hat{\theta}_n - \theta_0\| \geq \gamma\|\hat{\theta}_n - \theta_0\| \geq \gamma\epsilon$$

with probability δ for $n > N$. This implies $\|f(\hat{\theta}_n) - f(\theta_0)\| \not\xrightarrow{p} 0$ which **contradicts** the the assumption above. Thus $\|\hat{\theta}_n - \theta_0\| \xrightarrow{p} 0$ and $\hat{\theta}_n \xrightarrow{p} \theta_0$.

Because $\hat{\theta}_n \xrightarrow{p} \theta_0$ and θ_0 is an interior point of Θ

$$\sqrt{n} \dot{f}(\hat{\theta}_n)' W_n (t_n - f(\hat{\theta}_n)) \stackrel{a}{=} 0$$

and

$$\sqrt{n} \dot{f}(\hat{\theta}_n)' W_n (t_n - f_0 - (f(\hat{\theta}_n) - f_0)) \stackrel{a}{=} 0$$

Since $f(\hat{\theta}_n) - f(\theta_0) = \dot{f}(\bar{\theta}_n)(\hat{\theta}_n - \theta_0)$

$$\sqrt{n} \dot{f}(\hat{\theta}_n)' W_n \dot{f}(\bar{\theta}_n) \sqrt{n}(\hat{\theta}_n - \theta_0) \stackrel{a}{=} \sqrt{n} \dot{f}(\hat{\theta}_n)' W_n \sqrt{n}(t_n - f_0)$$

Multiply both sides by $(\dot{f}(\theta_0)' W \dot{f}(\theta_0))^{-1}$. The resulting left side is asymptotically equal to $\sqrt{n}(\hat{\theta}_n - \theta_0)$. Thus

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \stackrel{a}{=} (\dot{f}(\theta_0)' W \dot{f}(\theta_0))^{-1} \dot{f}(\theta_0)' W \sqrt{n}(t_n - f_0)$$

Since by assumption 1, $\sqrt{n}(t_n - f_0)$ is asymptotically normal $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal. ■

The non-singularity of Γ

If x is a p -component random vector, then $\text{cov}(x)$ is singular if and only if x lies in a proper linear sub-manifold of \mathcal{R}^p with probability one. This is often used to motivate the assumption that Σ is non-singular. We seek a similar motivation for the assumption that Γ is nonsingular. A distribution is said to be singular if all of its mass is on a set of Lebesgue measure zero.

Lemma A4: If

1. x_1, \dots, x_n is a sample from a m dimensional distribution \mathcal{D} with finite fourth moments and covariance matrix Σ_0 .
2. $S_n = \frac{1}{n} \sum (x_i - \bar{x}_n)(x_i - \bar{x}_n)'$.
3. $s_n = \text{vech}(S_n)$ and $\sigma_0 = \text{vech}(\Sigma_0)$.
4. $\sqrt{n}(s_n - \sigma_0) \xrightarrow{\mathcal{D}} N(0, \Gamma)$ and Γ is singular.

Then \mathcal{D} is singular.

Proof: We may assume without loss of generality that \mathcal{D} has mean zero.

Using the Slutsky theorem $\sqrt{n}\text{vech}(\bar{x}_n\bar{x}'_n) \xrightarrow{p} 0$ and

$$\sqrt{n}(s_n - \sigma_0) = \frac{1}{\sqrt{n}} \sum (\text{vech}(x_i x'_i) - \sigma_0) - \sqrt{n}\text{vech}(\bar{x}_n\bar{x}'_n) \stackrel{a}{=} \frac{1}{\sqrt{n}} \sum (\text{vech}(x_i x'_i) - \sigma_0)$$

Note that

$$\frac{1}{\sqrt{n}} \sum (\text{vech}(x_i x'_i) - \sigma_0) \stackrel{a}{=} \sqrt{n}(s_n - \sigma_0) \xrightarrow{\mathcal{D}} N(0, \Gamma)$$

Note that $\sigma_0 = E\text{vech}(xx')$ where x is a sample from \mathcal{D} . It follows from the central limit theorem that Γ is the covariance matrix for $\text{vech}(xx')$.

Since Γ is singular there is an $\ell \neq 0$ such that $\ell'\Gamma\ell = 0$. Thus

$$\text{cov}(\ell'\text{vech}(xx')) = \ell'\Gamma\ell = 0$$

and

$$Q(x) = \ell'\text{vech}(xx') = 0$$

with probability one. Let Z be the zeros of Q . Viewed as a function of the first component of x , $Q(x)$ is a polynomial equation of degree at most two. Thus given the last $p - 1$ components of x there are at most two values of the first component for which $Q(x) = 0$.

Let v be any vector of length $p - 1$ and let Z_v be all z in Z whose last $p - 1$ components are equal to v . Then Z_v has at most two points and hence has Lebesgue measure zero for each value of v . From Fubini's theorem (e.g. Theorem A of Halmos, 1950, p. 147) Z has Lebesgue measure zero. Since x is in Z with probability one, the distribution \mathcal{D} is singular. ■

Remark: We have shown that if the matrix Γ is singular, then the distribution \mathcal{D} is singular. Thus if \mathcal{D} is nonsingular, Γ is nonsingular.

Remark: If \mathcal{D} has a density or is a mixture of a distribution with a density and a discrete distribution it is nonsingular and hence Γ is nonsingular.

Remark: It should be pointed out that while the non-singularity of \mathcal{D} is a sufficient condition for the non-singularity of Γ , it is not a necessary condition.

Remark: Can Γ be singular when the covariance matrix Σ_0 of \mathcal{D} is nonsingular? We believe there are examples where this happens, but to our knowledge no such example exists in the literature and we believe there are few applications of interest for which Γ is singular when Σ_0 is nonsingular. Hence the non-singularity of Γ is a very mild assumption when Σ_0 is nonsingular.

References

- Anderson T.W and Hsiao C. (1982). Estimation of dynamic models using panel data. *Journal of Econometrics* 18: 47-82.
- Bentler, P. M. (2008). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.
- Boomsma, A. and Hoogland, J.J. (2001). The robustness of lisrel modeling revisited, in *Structural Equation Modeling: Present and Future*, R. Cudeck, S. Du Toit and D. Srbom. (eds.), SSI Scientific Software International, Lincolnwood, IL.
- Bou J.C. and Satorra A. (2007). The persistence of abnormal returns at industry and firm levels: evidence from Spain. *Strategic Management Journal* 28(7): 707-722.
- Browne, M.B. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*. **37**, 62-83.
- Chou, C.C., Bentler, P.M. and Satorra, A. (1991). Scaled Test Statistics and Robust Standard Errors for Nonnormal Data in Covariance Structure Analysis: A Monte Carlo Study. *British Journal of Mathematical and Statistical Psychology*. **44**, 347-357.
- Curran, P. J., West, S. G., and Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16-29.
- Halmos, P.R. (1950). *Measure Theory*. Princeton New Jersey: D. Van Nos-

- trand, Inc.
- Halmos, P.R. (1958). *Finite dimensional vector spaces, Second edition*. Princeton New Jersey: D. Van Nostrand, Inc.
- Jöreskog, K. & Sörbom, D. (1994). *ISREL 8 Users Reference Guide. Program and Applications*. Scientific Software, Mooresville, IN.
- Magnus, J. R. and Neudecker, H. 1999. *Matrix differential calculus with applications in statistics and econometrics, Revised edition*. New York: John Wiley.
- Muthén, L.K., & Muthén, B.O. (2007). *Mplus users guide (5th ed.)*. Los Angeles, CA: Muthn & Muthn.
- Rao, C.R. (1965). *Linear statistical inference and its applications*. New York: Wiley.
- Satorra, A., (1989). Alternative test criteria in covariance structure analysis: a unified approach. *Psychometrika* 54, 131151.
- Satorra, A. and Neudecker, N. (2003). A matrix equality useful in goodness-of-fit testing of structural equation models. *Journal of Statistical Planning and Inference* 114 63-80.
- Yuan, K.-H. and Bentler, P.M. (1997). Mean and covariance structure analysis: Theoretical and practical improvements. *Journal of the American Statistical Association*, **92**, 767-774.