Statistical Inference of Moment/Covariance Structures

Alexander Shapiro

Georgia Institute of Technology
School of Industrial and Systems Engineering

Conference Honoring the Scientific Contributions of Michael W. Browne
Let $\xi = (\xi_1, ..., \xi_m)'$ be a vector variable representing a parameter vector of some statistical population. For example, in the analysis of covariance structures, $\xi$ will represent the elements of a $p \times p$ covariance matrix $\Sigma$, i.e., $\xi = \text{vec}(\Sigma)$. We assume that $\xi$ varies in a set $\Xi \subset \mathbb{R}^m$ representing a saturated model for the population vector $\xi$. For example, in the analysis of covariance structures, $\Xi$ corresponds to the set of positive definite matrices.

A model for $\xi$ is a subset $\Xi_0$ of $\Xi$. Parametric model

$$\Xi_0 = \{\xi \in \Xi : \xi = g(\theta), \ \theta \in \Theta\},$$

where $\Theta \subset \mathbb{R}^k$ is the parameter space and $g : \mathbb{R}^k \to \mathbb{R}^m$. 
Following Browne (1982), we say that a real valued function $F(s, \xi)$, of two vector variables $s, \xi \in \Xi$ is a discrepancy function if it satisfies the following conditions:

(i) $F(s, \xi) \geq 0$ for all $s, \xi \in \Xi$,

(ii) $F(s, \xi) = 0$ if and only if $s = \xi$,

(iii) $F(s, \xi)$ is twice continuously differentiable jointly in $s$ and $\xi$.

In the analysis of covariance structures often used normal distribution ML discrepancy function is

$$F_{ML}(S, \Sigma) = \ln |\Sigma| - \ln |S| + \text{Trace}(S\Sigma^{-1}) - p.$$
Let \( \hat{\xi} \) be a sample estimate of the population value \( \xi_0 \in \Xi \), based on a sample of size \( N \). In the analysis of covariance structures \( \hat{\xi} \) usually is the sample covariance matrix \( \hat{\xi} = \text{vec}(S) \). Note that the model holds if \( \xi_0 \in \Xi_0 \), i.e., if there exists \( \theta_0 \in \Theta \) such that \( \xi_0 = g(\theta_0) \).

The MDF estimator \( \hat{\theta} \) is given by an optimal solution of the optimization problem

\[
\min_{\theta \in \Theta} F(\hat{\xi}, g(\theta)).
\]

The optimal value \( \hat{F} \) of this problem can be used for testing the model. Note that \( \hat{F} \) is also the optimal value of problem

\[
\min_{\xi \in \Xi_0} F(\hat{\xi}, \xi).
\]
We assume that $N^{1/2}(\hat{\xi} - \xi_0)$ converges in distribution to normal $\mathcal{N}(0, \Gamma)$. This usually is ensured by the Central Limit Theorem. For example, in the analysis of covariance structures, assuming that the population distribution is normal, $N^{1/2}(s - \sigma_0)$ converges in distribution to $\mathcal{N}(0, \Gamma_N)$ with

$$\Gamma_N = 2M_p(\Sigma_0 \otimes \Sigma_0)M_p,$$

where $s = \text{vec}(S)$ and $M_p$ is a certain $p^2 \times p^2$ symmetric idempotent matrix of rank $p(p+1)/2$ (Browne, 1974).
Consider function

$$\phi(s) = \min_{\xi \in \Xi_0} F(s, \xi).$$

Note that $\hat{F} = \phi(\hat{\xi})$. Let $\xi^*$ be an optimal solution of the above problem for $s = \xi_0$, i.e., $\xi^*$ is a minimizer of $F(\xi_0, \xi)$ over $\xi \in \Xi_0$, and

$$F_0 = \min_{\xi \in \Xi_0} F(\xi_0, \xi) = F(\xi_0, \xi^*).$$

Let us make the following observations.
For every $s \in \Xi$ it holds that $\phi(s) \geq 0$ and $\phi(s) = 0$ iff $s \in \Xi_0$. In particular $F_0 \geq 0$ and $F_0 = 0$ iff $\xi_0 \in \Xi_0$. 
Suppose that the minimizer $\xi^*$ is unique. Then $\phi(s)$ is differentiable at $s = \xi_0$ and

$$\frac{\partial \phi(\xi_0)}{\partial s} = \frac{\partial F(\xi_0, \xi^*)}{\partial s}$$

(this follows by Danskin Theorem). That is\(^1\)

$$\hat{F} = F_0 + \gamma'(\hat{\xi} - \xi_0) + o(\|\hat{\xi} - \xi_0\|),$$

where $\gamma = \frac{\partial F(s, \xi^*)}{\partial s} \bigg|_{s=\xi_0}$. It follows that $N^{1/2}(\hat{F} - F_0)$ converges in distribution to normal $\mathcal{N}(0, \gamma' \Gamma \gamma)$.

\(^1\)Recall that $\hat{F} = \phi(\hat{\xi})$ and $F_0 = \phi(\xi_0)$. 
For example in the analysis of covariance structures, with the ML discrepancy function,

\[ \gamma = \frac{\partial F_{ML}(\mathbf{s}, \sigma^*)}{\partial \mathbf{s}} \bigg|_{\mathbf{s}=\sigma_0} = \text{vec} \left[ (\Sigma^*)^{-1} - \Sigma_0^{-1} \right]. \]

In particular, if the population distribution is normal, then the asymptotic variance is

\[ \gamma' \Gamma \gamma = 2 \text{Trace} \left[ (\Sigma^* - I_p) \Sigma_0 (\Sigma^* - I_p) \right]. \]

In that case this can be derived from Vuong (1989) result for the likelihood ratio test statistic.
For the Generalized Least Squares (GLS) discrepancy function

\[ F_{GLS}(S, \Sigma) = \frac{1}{2} \text{Trace} \left\{ [S - \Sigma]S^{-1} \right\}^2 \]

we have that

\[
\gamma = \left. \frac{\partial F_{GLS}(s, \sigma^*)}{\partial S} \right|_{s=\sigma_0} = \text{vec} \left[ \Sigma_0^{-1} \left( \Sigma^* - \Sigma^* \Sigma_0^{-1} \Sigma^* \right) \Sigma_0^{-1} \right]
\]

and the asymptotic variance, assuming normal distribution of

the population,

\[
\gamma' \Gamma \gamma = 2 \text{Trace} \left[ (\Sigma_0^{-1} \Sigma^* \Sigma_0^{-1} \Sigma^* - \Sigma_0^{-1} \Sigma^*)^2 \right].
\]

Note that here \(\Sigma^*\) is the minimizer corresponding to the GLS

discrepancy function.
Second order approximation of the MDF test statistic

Second order Taylor expansion of $\phi(\hat{\xi})$ at the point $\xi^*$:

$$\hat{F} = (\hat{\xi} - \xi^*)' Q (\hat{\xi} - \xi^*) + o(\|\hat{\xi} - \xi^*\|^2),$$

where $Q = \frac{1}{2} \frac{\partial^2 \phi(\xi^*)}{\partial s \partial s'}$. Hence

$$N\hat{F} \approx \left[ N^{1/2}(\hat{\xi} - \xi^*) \right]' Q \left[ N^{1/2}(\hat{\xi} - \xi^*) \right].$$

Recall that $N^{1/2}(\hat{\xi} - \xi_0)$ converges in distribution to $\mathcal{N}(0, \Gamma)$ and note that

$$N^{1/2}(\hat{\xi} - \xi^*) = N^{1/2}(\hat{\xi} - \xi_0) + N^{1/2}(\xi_0 - \xi^*).$$
If $\hat{\xi}$ is an unbiased estimator of $\xi_0$, then $\mathbb{E}[\hat{\xi} - \xi^*] = \xi_0 - \xi^*$. Denote $\mu = N^{1/2}(\xi_0 - \xi^*)$.

Sequence of local alternatives or the population drift: population value $\xi_{0,N}$ of the parameter vector depends on the sample size $N$ and converges to a point $\xi^* \in \Xi_0$ such that $N^{1/2}(\xi_{0,N} - \xi^*)$ converges to a limit $\mu$.

Then distribution of the MDF test statistic $N\hat{F}$ can be approximated by the distribution of the quadratic form $Z'QZ$, where $Z \sim \mathcal{N}(\mu, \Gamma)$. Note that $NF_0 = N\phi(\xi_0)$ and

$$NF_0 \approx \left[N^{1/2}(\xi_0 - \xi^*)\right]' Q \left[N^{1/2}(\xi_0 - \xi^*)\right] = \mu' Q \mu.$$
Suppose that the set $\Xi_0$ is a smooth manifold near the point $\xi^* \in \Xi_0$. In a sense this holds generically with the tangent space to $\Xi_0$ at $\xi^* = g(\theta^*)$ generated by columns of the $m \times k$ Jacobian matrix $\Delta = \partial g(\theta^*)/\partial \theta$. Then

$$Q = \Phi (\Phi' V^{-1} \Phi)^{-1} \Phi',$$

where $V = \frac{1}{2} \partial^2 F / \partial \xi \partial \xi'$ calculated at point $(\xi^*, \xi^*)$, and $\Phi$ is an $m \times (m - r)$ matrix of full rank $m - r$ such that $\Phi' \Delta = 0$, $r = \text{rank}(\Delta)$.

Then the quadratic form $Z' Q Z$ has (noncentral) chi-square distribution with noncentrality parameter $\delta = \mu' Q \mu$ and $df = m - r$ if

$$V^{-1} = \Gamma + \Delta C \Delta'.$$

Asymptotic robustness theory (Browne and Shapiro, 1988).
Let $\Theta_1$ be a subset of the parameter space $\Theta$ and

$$\Xi_1 = \{\xi \in \Xi : \xi = g(\theta), \theta \in \Theta_1\}.$$ 

Of course, $\Xi_1 \subset \Xi_0$. Let

$$\hat{F}_0 = \min_{\xi \in \Xi_0} F(\hat{\xi}, \xi),$$

$$\hat{F}_1 = \min_{\xi \in \Xi_1} F(\hat{\xi}, \xi).$$

We have the following (Steiger, Shapiro, Browne, 1985).
Suppose that \( \xi_{0,N} \) converges to a point \( \xi^* \in \Xi_1 \) such that \( N^{1/2}(\xi_{0,N} - \xi^*) \rightarrow \mu \) (assumption of the sequence of local alternatives), and that \( \Xi_1 \) is a smooth submanifold of the smooth manifold \( \Xi_0 \) near \( \xi^* \).

Then: (i) \( \hat{N}F_1 \) converges in distribution to noncentral chi-square, (ii) \( \hat{N}F_1 - \hat{N}F_0 \) converges in distribution to noncentral chi-square with noncentrality parameter \( \delta_1 - \delta_0 \) and \( \nu_1 - \nu_0 \) degrees of freedom, (iii) \( \hat{N}F_0 \) and \( \hat{N}F_1 - \hat{N}F_0 \) are asymptotically independent of each other.