Metrika (1991) 38:345-355

Invariance of Covariance Structures Under Groups of Transformations

By M.W. Browne[1] and A. Shapiro[2]

Summary: The invariance of covariance structures under Lie groups of transformations is discussed. Implications for minimum discrepancy estimates of parameters are considered.

1 Introduction

Arguments based on the invariance of a parametric family of distributions under groups of transformations have been employed in statistical inference for a long time (e.g. Lehman 1959; Eaton 1972; Muirhead 1982). The present paper will study the implications of additional assumptions concerning the differential structure of transformation groups. Groups with differential structure have become important in many branches of modern mathematics and are known under the name *Lie groups* after the Norwegian mathematician Sophus Lie, 1842-1899.

We shall consider the invariance under Lie groups of transformations of symmetric matrix valued functions employed as models for covariance matrices. It will be shown that this invariance has certain implications for minimum discrepancy estimates of model parameters. The results obtained will extend and unify results due to Dijkstra (1990) who separately treated two types of transformation. We shall also mention potential applications of the invariance principle to studies of asymptotic robustness in the analysis of covariance structures.

[1] M.W. Browne, Department of Psychology, 142 Townshend Hall, 1885 Neil Avenue, Columbus OH 43210-1222, USA.
[2] A. Shapiro, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta GA 30332-0205, USA.

0026-1335/91/6/345-355 $2.50

2 Fitting a Covariance Structure

Suppose that the vector variate x has a distribution with a $p \times p$ covariance matrix Σ. A covariance structure is a symmetric matrix valued function $\Sigma(\theta)$ which relates a parameter vector θ from a subset $\mathscr{P}_\theta$ of $\mathbb{R}^q$ to Σ.

The discrepancy between a model $\Sigma(\theta)$ and a sample covariance matrix S is measured by means of a discrepancy function $F(S,\Sigma)$. This is a nonnegative twice continuously differentiable real valued function of two positive definite matrix variables S and Σ such that $F(S,\Sigma) = 0$ if and only if $S = \Sigma$. A covariance structure $\Sigma(\theta)$ is fitted to a sample covariance matrix S by minimising $F(S,\Sigma(\theta))$ over $\mathscr{P}_\theta$ to obtain the parameter estimate $\hat{\theta}$ and fitted structure $\hat{\Sigma} = \Sigma(\hat{\theta})$.

Some examples of discrepancy functions follow.

(a) The normal theory maximum likelihood discrepancy function:

$$F_a(S,\Sigma) = \log|\Sigma| - \log|S| + \operatorname{tr}[S\Sigma^{-1}] - p \ . \tag{1}$$

(b) A specific type of generalised least squares discrepancy function with a weight matrix independent of Σ:

$$F_b(S,\Sigma) = \tfrac{1}{2}\operatorname{tr}[V^{-1}(S-\Sigma)]^2 \ . \tag{2}$$

Possible choices for the $p \times p$ positive definite weight matrix V, are $V = S$ yielding normal theory generalised least squares estimates and $V = I_p$ yielding ordinary least squares estimates.

(c) A generalised least squares discrepancy function with a weight matrix that is a function of Σ:

$$F_c(S,\Sigma) = \tfrac{1}{2}\operatorname{tr}[\Sigma^{-1}(S-\Sigma)]^2 \ . \tag{3}$$

Here the matrix V in (2) is replaced by the function $\Sigma = \Sigma(\theta)$.

(d) A general form of generalised least squares discrepancy function:

$$F_d(S,\Sigma) = (s-\sigma)' W^{-1}(s-\sigma) \tag{4}$$

where s and σ are vectors with $p^* = \frac{1}{2}p(p+1)$ elements formed from the distinct elements of S and Σ respectively, and W is a $p^* \times p^*$ positive definite matrix. Generalised least squares discrepancy functions of this sort are used when normality assumptions for x are inappropriate. The discrepancy function

$F_b(S,\Sigma)$ in (2) is a special case of $F_d(S,\Sigma)$ that has received considerable attention (e.g. Jöreskog and Goldberger 1972; Browne 1974; Dijkstra 1990).

3 Invariance of Covariance Structures

Consider a multiplicative group $\mathscr{G}$ of nonsingular $p \times p$ matrices. That is, if $A \in \mathscr{G}$ and $B \in \mathscr{G}$ then $AB^{-1} \in \mathscr{G}$. In particular, this implies that the identity matrix I_p is in $\mathscr{G}$. Suppose that $\mathscr{G}$ possesses the structure of a differentiable manifold near I_p. This means that there exists a neighbourhood $\mathscr{N}$ of zero in $\mathbb{R}^t$, a neighbourhood $\mathscr{M}$ of I_p in the linear space of $p \times p$ matrices, and a continuously differentiable mapping φ of $\mathscr{N}$ onto $\mathscr{G} \cap \mathscr{M}$ such that $\varphi(0) = I_p$ and the Jacobian matrix of φ at zero is of full rank t.

Thus $\mathscr{G}$ constitutes a Lie group with matrix multiplication as the group operation. Associated with $\mathscr{G}$ is a corresponding Lie group $\mathscr{G}^*$ of transformations defined on the set of symmetric positive definite matrices by $\Sigma \to A\Sigma A'$, where A' stands for the transpose of $A \in \mathscr{G}$. We shall examine covariance structures that are invariant under $\mathscr{G}^*$. Notice that the covariance matrix $A\Sigma A'$ is obtained from Σ by the linear transformation $x \to Ax$ of the vector variate x.

Definition: A covariance structure $\Sigma(\theta)$ is said to be invariant under the group $\mathscr{G}^*$ if for every $\theta \in \mathscr{P}_\theta$ and $A \in \mathscr{G}$ there exists a $\theta^* \in \mathscr{P}_\theta$ such that

$$\Sigma(\theta^*) = A\Sigma(\theta)A' \ . \tag{5}$$

□

This means that for any $A \in \mathscr{G}$ the set

$$\mathscr{P}_\sigma = \{\Omega : \Omega = \Sigma(\theta),\ \theta \in \mathscr{P}_\theta\} \tag{6}$$

of positive definite matrices corresponding to the given model, remains invariant under the transformation $\Sigma \to A\Sigma A'$.

Remark: Suppose that A and θ in (5) determine θ^* uniquely. Then to every $A \in \mathscr{G}$ corresponds the transformation $\theta \to \bar{A}(\theta) = \theta^*$ of $\mathscr{P}_\theta$ into itself. It can be shown that these transformations are one-to-one and onto and form a group $\bar{\mathscr{G}}$ of transformations of P_θ, the so-called induced group, such that the mapping $\mathscr{G} \to \bar{\mathscr{G}}$ is a homomorphism (e.g. Muirhead, 1982, p. 202).

Given an algebraic expression for $\Sigma(\theta)$ and a group $\mathscr{G}$ it is usually straightforward to verify that the corresponding group $\mathscr{G}^*$ of transformations leaves $\Sigma(\theta)$ invariant. We now give some examples of invariance which appear naturally in the analysis of covariance structures.

Invariance under a Constant Scaling Factor

Consider the group

$$\mathscr{G}_1 = \{A : A = \tau I_p,\ \tau \neq 0\}$$

which is homomorphic to the multiplicative group of nonzero real numbers. Most covariance structures used in practice are at least invariant under $\mathscr{G}_1^*$. Implications of this type of invariance are pointed out by Swain (1975), Browne (1982), Shapiro and Browne (1987) and Dijkstra (1990) amongst others. This is probably the simplest nontrivial invariance useful in applications. An example of a structure that is invariant under $\mathscr{G}_1^*$, but that is not invariant under the richer groups of transformations considered subsequently, is the intraclass correlation model

$$\Sigma = 1\varphi 1' + \psi I_p \ , \tag{7}$$

with $\mathscr{P}_\theta = \{\theta = (\varphi, \psi) : \varphi > 0,\ \psi > 0\}$. Here 1 denotes the $p \times 1$ vector of ones.

Invariance under Changes of Scale

Covariance structures that are not destroyed by scale changes of the elements of x, are invariant under transformations associated with

$$\mathscr{G}_2 = \{A : A = \mathrm{Diag}(\tau),\ \tau_i \neq 0,\ i = 1, \dots, p\} \ ,$$

where Diag (τ) denotes the diagonal matrix with diagonal elements given by the components of the vector $\tau = (\tau_1, \dots, \tau_p)$.

An example of a model that is invariant under $\mathscr{G}_2^*$ is the factor analysis model,

$$\Sigma = \Lambda\Lambda' + \mathrm{Diag}(\psi) \tag{8}$$

where Λ is a $p\times m$ matrix of factor loadings and

$$\mathscr{P}_\theta = \{\theta = (\Lambda, \psi) : \psi_i > 0,\ i = 1, \ldots, p\} .$$

Invariance under Block Diagonal Transformations

When x consists of k subsets of variates, one may employ block diagonal transformation matrices of the form

$$\mathscr{G}_3 = \{A : A = \text{Diag}\,(A_{11}, \ldots, A_{kk}),\ \det A_{ii} \neq 0,\ i = 1, \ldots, k\}$$

where A_{ii} is a $p_i\times p_i$ matrix, $p = p_1 + \ldots + p_k$ and Diag $(A_{11}, \ldots, A_{kk})$ denotes the $p\times p$ block diagonal matrix with diagonal blocks $A_{11}, \ldots, A_{kk}$. An example of a model that is invariant under $\mathscr{G}_3^*$ is the multiple battery factor analysis model (Tucker 1958),

$$\Sigma = \Lambda\Lambda' + \text{Diag}\,(\Psi_{11}, \ldots, \Psi_{kk}) \tag{9}$$

where the matrices Ψ_{ii}, $i = 1, \ldots, k$ are required to be positive definite.

Kronecker Product Transformation Matrices

Swain (1975) proposed a Kronecker product covariance structure for situations where each element of x stands for a measurement taken under the combination of one of p_1 conditions of one type and one of p_2 conditions of another type with $p = p_1 \times p_2$. The structure is

$$\Sigma = \Sigma_1 \otimes \Sigma_2 , \tag{10}$$

where Σ_1 and Σ_2 are positive definite matrices of order $p_1\times p_1$ and $p_2\times p_2$ respectively, and $\otimes$ denotes the Kronecker product of matrices (see e.g. Muirhead 1982, p. 73, for the definition and basic properties of the Kronecker product).

This covariance structure is invariant under transformations associated with the group $\mathscr{G}_4$ of $p\times p$ transformation matrices $A = A_1 \otimes A_2$ where A_i is a $p_i\times p_i$ nonsingular matrix, $i = 1, 2$.

Orthogonal Transformation Matrices

The factor analysis model with homogeneous unique variances has the covariance structure

$$\Sigma = \Lambda\Lambda' + \psi I_p \tag{11}$$

where ψ is a nonnegative scalar. This structure is invariant under transformations associated with the group $\mathscr{G}_5$ of nonsingular matrices with the property

$$AA' = \tau^2 I_p$$

where $\tau \neq 0$. Clearly

$$\mathscr{G}_5 = \mathscr{G}_1 \times O(p)$$

where $O(p)$ is the group of $p\times p$ orthogonal matrices.

4 Properties of Fitted Structures Under Transformation Invariance

In this section we study implications of transformation invariance for the minimum discrepancy estimators $\hat{\theta}$ and the corresponding fitted structure $\hat{\Sigma} = \Sigma(\hat{\theta})$. First we derive a general result and then we consider applications to particular models and discrepancy functions.

We start by noting that with every Lie group $\mathscr{G}$ is associated a linear space $\mathscr{T}$ tangent to $\mathscr{G}$ at I_p. Usually this tangent space is not difficult to find and we shall show subsequently how it can be calculated for every example mentioned in the previous section.

Suppose that a covariance structure $\Sigma(\theta)$ is invariant under the group of transformations $\mathscr{G}^*$ corresponding to a Lie group $\mathscr{G}$, and let $\Sigma = \Sigma(\theta)$, $\theta \in \mathscr{P}_\theta$, be a point in the set $\mathscr{P}_\sigma$ defined in (6). Consider the set

$$\mathscr{O}(\Sigma) = \{\Omega : \Omega = A\Sigma A',\ A \in \mathscr{G}\}\ ,$$

which is called the *orbit* of Σ under $\mathscr{G}^*$. The transformation invariance of $\Sigma(\theta)$ means that $\mathscr{O}(\Sigma) \subset \mathscr{P}_\sigma$. A differential structure on $\mathscr{O}(\Sigma)$ is yielded by the differential structure of $\mathscr{G}$ at I_p. More specifically, it follows that $\mathscr{O}(\Sigma)$ is a differentiable manifold and its tangent space at Σ can easily be calculated as follows. Consider the matrix $I_p + dA$, $dA \in \mathscr{T}$. Then

$$(I_p + dA)\Sigma(I_p + dA)' = \Sigma + (dA)\Sigma + \Sigma(dA)' + \text{higher order terms} \ .$$

Consequently the tangent space to $\mathscr{O}(\Sigma)$ is given by

$$\{Z : Z = B\Sigma + \Sigma B',\ B \in \mathscr{T}\} \ . \tag{12}$$

By definition, the fitted matrix $\hat{\Sigma}$ is a minimiser of the function $F(S, \cdot)$ over the set $\mathscr{P}_\sigma$. Consider the orbit $\mathscr{O}(\hat{\Sigma})$ of $\hat{\Sigma}$ under $\mathscr{G}^*$. Since $\mathscr{O}(\hat{\Sigma}) \subset \mathscr{P}_\sigma$ and $\hat{\Sigma} \in \mathscr{O}(\hat{\Sigma})$, it follows that $\hat{\Sigma}$ is also a minimiser of $F(S, \cdot)$. By the standard optimality conditions this implies that the gradient of $F(S, \cdot)$ at $\hat{\Sigma}$ is orthogonal to the tangent space of $\mathscr{O}(\hat{\Sigma})$ at $\hat{\Sigma}$. We now formulate these conditions in a computationally convenient manner. Consider the $p \times p$ matrix

$$Q = \partial F(S, \Sigma)/\partial \Sigma$$

of partial derivatives of the discrepancy function with respect to the elements of Σ. Assume, that for any square matrix Σ, $F(S, \Sigma) = F(S, \Sigma')$, so that the matrix Q is symmetric. Define the scalar product

$$\langle A, B \rangle = \operatorname{tr}[AB']$$

on the linear space of $p \times p$ matrices. The optimality conditions discussed earlier then mean that $\operatorname{tr}[QZ'] = 0$ for any matrix Z from the tangent space of $\mathscr{O}(\hat{\Sigma})$ at $\hat{\Sigma}$. Since Q is symmetric, use of (12) yields the following result.

Proposition 1: Suppose that the structure $\Sigma(\theta)$ is invariant under $\mathscr{G}^*$. Let $\mathscr{T}$ be the tangent space to $\mathscr{G}$ at I_p and let

$$\hat{\Omega} = [\partial F(S, \hat{\Sigma})/\partial \Sigma]\hat{\Sigma} \ . \tag{13}$$

Then

$$\operatorname{tr}[\hat{\Omega} B'] = 0 \tag{14}$$

for all $B \in \mathscr{T}$. □

We shall call $\hat{\Omega}$ in (13) the *reflector* matrix because it reflects transformation invariance properties of $\Sigma(\theta)$. Specific formulae for the reflector matrix corresponding to the discrepancy functions (1), (2), (3), (4) considered earlier are:

$$\hat{\Omega}_a = \hat{\Sigma}^{-1}(\hat{\Sigma} - S) \ ,$$

$$\hat{\Omega}_b = V^{-1}(\hat{\Sigma} - S) V^{-1}\hat{\Sigma} \ ,$$

$$\hat{\Omega}_c = \hat{\Sigma}^{-1} S \hat{\Sigma}^{-1}(\hat{\Sigma} - S) \ ,$$

and

$$\hat{\Omega}_d = Q\hat{\Sigma}$$

in which the elements of the symmetric matrix Q are related to the elements of the vector $W^{-1}(\hat{\sigma} - s)$ by

$$[Q]_{ij} = \tfrac{1}{2}(2 - \delta_{ij})[W^{-1}(\hat{\sigma} - s)]_g \ ,$$

where δ_{ij} is the Kronecker delta and

$$g = \begin{vmatrix} \tfrac{1}{2}j(j-1)+1 \ , & i \le j \\ \tfrac{1}{2}i(i-1)+j \ , & i > j \ , \end{vmatrix}$$

when the elements of the matrix $(\hat{\Sigma} - S)$ forming the vector $(\hat{\sigma} - s)$ are chosen in the order 11, 12, 22, 13, 23, 33, ...

Let us first consider the group $\mathscr{G}_1$. It follows immediately from the definition of $\mathscr{G}_1$ that its tangent space is the one dimensional linear space generated by the identity matrix I_p. Consequently,

Corollary 1.1: If $\Sigma(\theta)$ is invariant under $\mathscr{G}_1^*$, then the sum of the diagonal elements of $\hat{\Omega}$ is zero.

For example, in the case of the discrepancy function F_a and transformations in $\mathscr{G}_1^*$ we obtain that

$$\operatorname{tr}[S\hat{\Sigma}^{-1}] = p \ .$$

This result was pointed out by Browne (1974). Swain (1975) also considered transformations in $\mathscr{G}_1^*$ and obtained related results for a family of discrepancy functions that includes F_a, F_b with $V = S$, and F_c.

The tangent space of the group $\mathcal{G}_2$ is the p-dimensional linear space of diagonal matrices. Therefore,

Corollary 1.2: If $\Sigma(\theta)$ is invariant under $\mathcal{G}_2^*$, then all diagonal elements of $\hat{\Omega}$ are zero.

For example, if we take the discrepancy function F_b then invariance under changes of scale implies that

$$\text{Diag}\,[(V^{-1}\hat{\Sigma})^2] = \text{Diag}\,[V^{-1}SV^{-1}\hat{\Sigma}] \ .$$

This result was obtained by Dijkstra (1990). Dijkstra (1990) also provided special cases of Corollaries 1.1 and 1.2 for F_b, with an arbitrary weight matrix V, and the Swain family of discrepancy functions.

The tangent space associated with $\mathcal{G}_3$ is the linear space of corresponding block diagonal matrices. Consequently,

Corollary 1.3: If $\Sigma(\theta)$ is invariant under $\mathcal{G}_3^*$, then the diagonal blocks of $\hat{\Omega}$ are null.

The tangent space of $\mathcal{G}_4$ is generated by matrices of the form $A_1 \otimes I_{p_2}$ and $I_{p_1} \otimes A_2$ where A_1 and A_2 are arbitrary matrices of order $p_1 \times p_1$ and $p_2 \times p_2$ respectively. Let $\hat{\Omega}$ be partitioned into p_1^2 submatrices of order $p_2 \times p_2$:

$$\hat{\Omega} = \begin{bmatrix} \hat{\Omega}_{11} & \hat{\Omega}_{12} & \cdots \\ \hat{\Omega}_{21} & \hat{\Omega}_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \ .$$

Corollary 1.4: If $\Sigma(\theta)$ is invariant under $\mathcal{G}_4^*$, then $\text{tr}\,[\hat{\Omega}_{ij}] = 0$, $\forall i,j$ *and* $\sum_{i=1}^{p_1} \hat{\Omega}_{ii} = 0$.

The tangent space of $\mathcal{G}_5$ is generated by the tangent space of the group $O(p)$ of orthogonal matrices and the tangent space of $\mathcal{G}_1$. It is well known, and can easily be calculated from the equation $AA' = I_p$, that the tangent space of $O(p)$ at I_p is the linear space of skew-symmetric matrices. Consequently the tangent space of $\mathcal{G}_5^*$ is the direct sum of the space of skew-symmetric matrices and the one dimensional space generated by I_p. Therefore,

Corollary 1.5: If $\Sigma(\theta)$ is invariant under $\mathcal{G}_5^*$, then $\hat{\Omega}$ is symmetric and $\text{tr}\,[\hat{\Omega}] = 0$.

These corollaries will have practical applications. Computation of the reflector matrix $\hat{\Omega}$ in computer programs for the analysis of covariance structures will help in the detection of user errors in specifying a model which does not have transformation invariance properties that are required by the input data. In particular there have been many erroneous analyses where a model that is not invariant under $\mathscr{G}_2^*$ has been fitted to a correlation matrix. Cudeck (1989, p. 317) has given a long list of published articles where this error has been made. Verification by computer programs that diagonal elements of the reflector matrix are null whenever a correlation matrix has been input would help to prevent further occurrences of the error. Any nonnull diagonal elements of the reflector matrix imply that the analysis of a correlation matrix as if it were a covariance matrix is incorrect. On the other hand, if all diagonal elements of the reflector matrix are null one cannot conclude that the analysis of the correlation matrix instead of the covariance matrix is appropriate. For example, if the model fits the correlation matrix perfectly, the reflector matrix will be null and consequently have null diagonal elements, even if the model is not invariant under changes of scale.

There is another interesting consequence of the invariance principle. Consider the covariance structure $\Sigma(\theta)$ in the vector form $\sigma(\theta) = \text{vec}\{\Sigma(\theta)\}$, where vec (Σ) stands for the $p^2 \times 1$ vector formed from the elements of Σ stacked columnwise. We say that a point θ_0 is a *regular* point of the model $\Sigma(\theta)$ if: (i) θ_0 is an interior point of $\mathscr{P}_\theta$, (ii) $\sigma(\theta)$ is continuously differentiable in a neighbourhood of θ_0, (iii) the $p^2 \times q$ Jacobian matrix $\partial\sigma(\theta)/\partial\theta'$ has constant rank in a neighbourhood of θ_0.

If follows from the regularity of θ_0 that the set $\mathscr{P}_\sigma$, considered in the vector form, is a differentiable manifold near $\sigma_0 = \sigma(\theta_0)$ and the tangent space to $\mathscr{P}_\theta$ at σ_0 is given by the column space of the Jacobian matrix $\Delta_0 = \partial\sigma(\theta_0)/\partial\theta'$. Since the orbit $\mathscr{O}(\Sigma_0)$ is in $\mathscr{P}_\sigma$ we have that the tangent space of $\mathscr{O}(\Sigma_0)$ is contained in the tangent space of $\mathscr{P}_\sigma$. Together with formula (12) this implies the following result.

Proposition 2: Let θ_0 be a regular point of $\Sigma(\theta)$ and suppose that $\Sigma(\theta)$ is invariant under $\mathscr{G}^*$. Then for every $B \in \mathscr{T}$ there exists a vector ζ such that

$$\text{vec}\,(B\Sigma_0 + \Sigma_0 B') = \Delta_0 \zeta \ . \qquad \square$$

The special case of Proposition 2 for $\mathscr{G}_1^*$ was found useful in studies of asymptotic robustness in the analysis of covariance structures (Shapiro 1986; Shapiro and Browne 1987). Also, Kano, Berkane and Bentler (1990, proof of Theorem 2) have made use of the special case for $\mathscr{G}_2^*$ to show that two discrepancy functions yield estimators which have the same asymptotic properties.

Acknowledgement: The authors thank a referee for helpful comments.

References

Browne MW (1974) Generalised least squares estimators in the analysis of covariance structures. South African Statistical J 8:1–24

Browne MW (1982) Covariance structures. In: Hawkins DM (ed) Topics in Applied Multivariate Analysis. Cambridge University Press, Cambridge UK, pp 72–141

Cudeck R (1989) Analysis of correlation matrices using covariance structure models. Psychological Bulletin 105:317–327

Dijkstra TK (1991) Some properties of estimated scale invariant covariance structures. Psychometrika 55:327–336

Eaton ML (1972) Multivariate Statistical Analysis. Institute of Mathematical Statistics, University of Copenhagen

Jöreskog KG, Goldberger A (1972) Factor analysis by generalised least squares. Psychometrika 37:243–260

Kano Y, Berkane M, Bentler PM (1990) Covariance structure analysis with heterogeneous kurtosis parameters. Biometrika 77 (in press)

Lehman EL (1959) Testing Statistical Hypotheses. Wiley, New York

Muirhead RJ (1982) Aspects of Multivariate Statistical Theory. Wiley, New York

Shapiro A (1986) Asymptotic theory of overparameterized structural models. J Amer Statist Assoc 81:142–149

Shapiro A, Browne MW (1987) Analysis of covariance structures under elliptical distributions. J Amer Statist Assoc 82:1092–1097

Swain AJ (1975) Analysis of Parametric Structures for Variance Matrices. Unpublished PhD Thesis, University of Adelaide

Tucker LR (1958) An inter-battery method of factor analysis. Psychometrika 23:111–136

Received 23 March 1990
Revised version 8 October 1990