The Annals of Statistics
2002, Vol. 30, No. 4, 1069–1080

MODELING THROUGH GROUP INVARIANCE: AN INTERESTING EXAMPLE WITH POTENTIAL APPLICATIONS

BY HENG LI[1]

University of Rochester

A particular linear group symmetry model, called the dyadic symmetry model, is studied in some detail. Statistical procedures analogous to (multivariate) analysis of variance are introduced. This model may be suitable for various kinds of data collected on pairs of sampling units. Examples include (complete) diallel cross experiments in genetics and social relations analysis in psychology, for which ad hoc methods of analysis have been developed independently in those disciplines.

Our approach is based entirely on formal data structure following the principle of group symmetry, and hence its applicability is not restricted to any specific substantive areas. This paper illustrates the benefits that can be derived from the exploration of mathematical meanings in the development of statistical methods.

1. Introduction. This paper illustrates through an example the potential wide applicability of the group symmetry (GS) models and linear group symmetry (LGS) models formulated in Andersson and Madsen (1998) [henceforth, AM (1998)]. The example concerns the prototypical layout of a kind of data structure which is not only mathematically interesting in its own right, but also relevant in diverse substantive fields. In this section we describe the model that will be dealt with in this paper, both in its purely mathematical form using the notation and terminology in AM (1998), and in a language more vernacular to statistics. Since the model belongs to the class of LGS models, it is most natural to specify it by locating it within that class; and this calls for a recapitulation of the essentials of AM's (1998) conceptualization of LGS models, which will be carried out below. The following paragraph may be viewed as excerpted from AM (1998), to which the reader is referred for more information.

Let I be a finite set, $\mathbb{R}^I$ be the vector space of all families $x = \{x_h \mid h \in I\}$ of real numbers indexed by I. Let $M(I)$, $GL(I)$, $O(I)$ and $P(I)$ denote the set of all $I \times I$ matrices, the group of all nonsingular $I \times I$ matrices, the group of all orthogonal $I \times I$ matrices, and the cone of all positive definite $I \times I$ matrices, respectively. Let G be a finite group and $\rho : G \to O(I)$ be an orthogonal group representation of G on $\mathbb{R}^I$. Let $P_G(I)$ denote the set of all positive definite $I \times I$ matrices that are G-invariant, that is, $P_G(I) = \{\Sigma \mid \Sigma \in P(I),\ \rho(\pi)\Sigma\rho(\pi)' = \Sigma\ \forall \pi \in G\}$.

Received January 2001; revised November 2001.

[1]Supported in part by NIH Grants M01 RR00044 and P01 DE13539.

AMS 2000 subject classifications. 62H05, 62J10, 62A01, 62E15, 20C99, 62P15, 62P10.

Key words and phrases. Diallel cross, dyadic symmetry model, exchangeability, group symmetry model, linear group symmetry model, patterned covariance, social relations model.

Let $M_G(I)$ denote the set of all $I \times I$ matrices that commute with $\rho(G)$, that is, $M_G(I) = \{A \mid A \in M(I),\ \rho(\pi)A = A\rho(\pi)\ \forall \pi \in G\}$. Let $L \subseteq \mathbb{R}^I$ be an $M_G(I)$-subspace, that is, a subspace satisfying $M_G(I)L = L$.

The set of normal distributions

$$\{N(\xi, \Sigma) \mid (\xi, \Sigma) \in L \times P_G(I)\} \tag{1.1}$$

on $\mathbb{R}^I$ is called a *linear group symmetry* (LGS) model on $\mathbb{R}^I$ determined by L, G and ρ.

With the above definition in place, we can now specify the particular LGS model this paper deals with, by specifying I, G, ρ and L. Let U be a finite set containing u elements. The index set I is specified to be $I = U \times U$. Let the group G be the symmetric group on U. Denote the elements in I by pairs of lower case letters, such as ij, with $i \in U$ and $j \in U$, so that a vector x in $\mathbb{R}^I$ can be expressed as $\{x_{ij} \mid ij \in I\}$, and an $I \times I$ matrix A can be expressed as $\{A_{ij,kl} \mid ij \in I,\ kl \in I\}$. Under this notation, we can specify ρ as

$$\begin{cases} \rho(\pi)_{ij,kl} = 1, & \text{if } i = g(k),\ j = g(l), \\ \rho(\pi)_{ij,kl} = 0, & \text{otherwise.} \end{cases} \tag{1.2}$$

Finally, let e^{diag} be a vector in $\mathbb{R}^I$ with $e^{\text{diag}}_{ij} = 1$ if $i = j$ and $e^{\text{diag}}_{ij} = 0$ if $i \neq j$; e^{off} be a vector in $\mathbb{R}^I$ with $e^{\text{off}}_{ij} = 1$ if $i \neq j$ and $e^{\text{off}}_{ij} = 0$ if $i = j$. The vectors e^{diag} and e^{off} span L. This completes the specification of our model, which will be given the name *dyadic symmetry model*. Henceforth, the symbols I, G, ρ and L will be used as defined in the current paragraph, and u will be assumed to be greater than 3, except in a brief discussion of the situations when $u \leq 3$.

The dyadic symmetry model can be a suitable candidate for modeling the situations in which measurements are made on pairs of units. Let U be a set of u units, y_{ij} denote the measurement made on the pair consisting of distinct units i and j, z_{ii} denote the measurement made on the pair formed by the same unit i, and x denote the u^2-dimensional vector of measurements on all the possible pairs of units. It is conventional in statistics to represent those measurements by arranging them in a square table whose row and column headings consist of labels for the same set of units, $1, \dots, u$, in ascending order. Here we use integers to label units in order to simplify notation. It will become clear later that using different letters to denote the "diagonal" and "off-diagonal" measurements (in their tabular representation) is not only notationally but also conceptually advantageous. Henceforth, the vector of $u(u-1)$ "off-diagonal" measurements is denoted by y and the vector of u "diagonal" measurements is denoted by z, so that x can also be considered as (y, z). Now if we consider the units in U to be exchangeable, then any probability distribution postulated for the observation vector x in $\mathbb{R}^I = \mathbb{R}^{U \times U}$ should be invariant under G in the sense that x and $\rho(\pi)x$ should have the same distribution. The dyadic symmetry model follows from the additional distributional assumption of normality, since it is precisely the collection of all the normal distributions on $\mathbb{R}^I = \mathbb{R}^{U\times U}$ that are invariant under G.

The data structure as described above closely parallels the kind of factorial models discussed in McCullagh (2000). Indeed, it is an immediate extension of the data structure considered in Li's (2000) discussion of McCullagh (2000), by the inclusion of the "diagonal" measurements. Therefore it is natural to expect that results on the dyadic symmetry model documented in this paper would have applications in at least some of the substantive areas mentioned in McCullagh (2000) and Li (2000). The layout in the dyadic symmetry model may remind us of another class of models involving pairs of units, namely those related to the method of paired comparisons [David (1988)]. A fundamental distinction exists, however, in their emphasis: the former is essentially a method of summarizing an aggregate of units, whereas the latter are more concerned with comparing individual units.

Having defined the dyadic symmetry model from both mathematical and statistical perspectives, the objective of the rest of this paper is to study some specific properties of this model. In Section 2 the likelihood function will be expressed as a product of χ^2 and Wishart densities, under a particular parameterization of the covariance structure in the dyadic symmetry model.

A latent variable model will be shown to give rise to a covariance structure that has the same pattern as that of the dyadic symmetry model. Section 3 adapts results in Section 2 to a submodel that has already arisen in practical situations. Arithmetic expressions for the sufficient statistics will be provided in a form resembling the usual sums of squares and cross products found in (multivariate) analysis of variance, to facilitate potential application. Section 4 contains a few summarizing and concluding remarks.

2. The covariance structure.

We can characterize the pattern in the covariance matrices in the dyadic symmetry model by giving a description of $P_G(I)$ as a subset of $\mathbb{R}^{I\times I}$, that is, a subset of the set of all real functions on $I \times I$. From this perspective, $P_G(I)$ is the set of real functions on $I \times I$ which are symmetric, positive definite and constant on each orbit on $I \times I$ under the group action induced by that of G on U. Table 1 lists all the orbits by their representative elements as displayed in the first column. The symbols for the values of covariance on each orbit are listed under the column "Covariance."

A latent variable model giving rise to the covariance pattern in Table 1, and hence justified in the sense of Dawid (1988), is

$$\begin{aligned} y_{ij} &= \xi^{\text{off}} + \mu_y + g_i + g_j + s_{ij} + d_i - d_j + r_{ij}, \\ z_{ii} &= \xi^{\text{diag}} + \mu_z + a_i, \end{aligned} \tag{2.1}$$

where ξ^{diag} and ξ^{off} are constants, and the rest of the terms are random variables with mean 0 subject to the constraints $s_{ij} = s_{ji}$ and $r_{ij} = -r_{ji}$. The g, d and a

TABLE 1
Covariance structure

	Parameters		
Orbit	Latent variable	Covariance	Canonical parameters
ij, ij	$\sigma_{\mu y}^2 + 2\sigma_g^2 + \sigma_s^2 + 2\sigma_d^2 + \sigma_r^2$	σ_y^2	$\frac{\lambda_y-\lambda_s}{u(u-1)} + \frac{(u-2)(\lambda_s+\lambda_r)}{2u} + \frac{\lambda_g+\lambda_d}{u}$
ij, kl	$\sigma_{\mu y}^2 (= \mathrm{var}(\mu_y))$	$\sigma_y^2 \rho_{y0}$	$\frac{\lambda_y-\lambda_s}{u(u-1)} + \frac{2(\lambda_s-\lambda_g)}{u(u-2)}$
ji, ij	$\sigma_{\mu y}^2 + 2\sigma_g^2 + \sigma_s^2 - 2\sigma_d^2 - \sigma_r^2$	$\sigma_y^2 \rho_{y1}$	$\frac{\lambda_y-\lambda_s}{u(u-1)} + \frac{(u-2)(\lambda_s-\lambda_r)}{2u} + \frac{\lambda_g-\lambda_d}{u}$
ij, jk	$\sigma_{\mu y}^2 + \sigma_g^2 - \sigma_d^2$	$\sigma_y^2 \rho_{y2}$	$\frac{\lambda_y-\lambda_s}{u(u-1)} + \frac{(u-4)(\lambda_g-\lambda_s)}{2u(u-2)} + \frac{\lambda_r-\lambda_d}{2u}$
ij, ik	$\sigma_{\mu y}^2 + \sigma_g^2 + 2\sigma_{gd} + \sigma_d^2$	$\sigma_y^2 \rho_{y3}$	$\frac{\lambda_y-\lambda_s}{u(u-1)} + \frac{(u-4)(\lambda_g-\lambda_s)}{2u(u-2)} + \frac{\lambda_d-\lambda_r}{2u} + \frac{\lambda_{gd}}{\sqrt{u(u-2)}}$
ij, kj	$\sigma_{\mu y}^2 + \sigma_g^2 - 2\sigma_{gd} + \sigma_d^2$	$\sigma_y^2 \rho_{y4}$	$\frac{\lambda_y-\lambda_s}{u(u-1)} + \frac{(u-4)(\lambda_g-\lambda_s)}{2u(u-2)} + \frac{\lambda_d-\lambda_r}{2u} - \frac{\lambda_{gd}}{\sqrt{u(u-2)}}$
ii, ii	$\sigma_{\mu z}^2 + \sigma_a^2$	σ_z^2	$\frac{(u-1)\lambda_a+\lambda_z}{u}$
ii, jj	$\sigma_{\mu z}^2 (= \mathrm{var}(\mu_z))$	$\sigma_z^2 \rho_{z0}$	$\frac{\lambda_z-\lambda_a}{u}$
ii, kl	$\sigma_{\mu yz} (= \mathrm{cov}(\mu_z, \mu_y))$	$\sigma_y \sigma_z \rho_{zy0}$	$\frac{\lambda_{zy}}{u\sqrt{u-1}} - \frac{\sqrt{2}\lambda_{ag}}{u\sqrt{u-2}}$
ii, ij	$\sigma_{\mu yz} + \sigma_{ag} + \sigma_{ad}$	$\sigma_y \sigma_z \rho_{zy1}$	$\frac{\sqrt{u-2}\lambda_{ag}}{\sqrt{2}N} + \frac{\lambda_{ad}}{\sqrt{2u}} + \frac{\lambda_{zy}}{u\sqrt{u-1}}$
ii, ji	$\sigma_{\mu yz} + \sigma_{ag} - \sigma_{ad}$	$\sigma_y \sigma_z \rho_{zy2}$	$\frac{\sqrt{u-2}\lambda_{ag}}{\sqrt{2}N} - \frac{\lambda_{ad}}{\sqrt{2u}} + \frac{\lambda_{zy}}{u\sqrt{u-1}}$

terms having the same subscript are allowed to be correlated with each other, and so are μ_z and μ_y. When a member of $P_G(I)$ is induced by model (2.1), the parameters in the former can be expressed in terms of those in the latter. The column "Latent variable" in Table 1 contains expressions of the covariance parameters in terms of the parameters in (2.1). In the next section we will explain how the latent variable model (2.1) for the "off-diagonal" measurements y_{ij} $(i \neq j)$ is essentially the same as Model (b) in Cockerham and Weir (1977). For the moment we only need to note that we inherit notation from Cockerham and Weir (1977) for all the parameters that are common to their Model (b) and our model (2.1), except for $\mathrm{Cov}(g, d)$, which we denote by σ_{gd}. The rest of the parameters as they appear in the column "Latent variable" in Table 1 are annotated parenthetically. Note that not all members of $P_G(I)$ can be induced by model (2.1), due to the constraints the latter introduces by requiring that the covariance matrices of the latent variables be nonnegative definite.

The set of covariance matrices as parameterized by the covariance and the latent variable parameters in Table 1 is not straightforward to deal with directly in the context of statistical analysis and inference. However, it can be written into a form in which it is, using what we will call canonical parameters. The fourth column of Table 1 defines a set of canonical parameters by expressing covariance parameters as linear functions of those parameters, each of which is denoted by the letter λ with a subscript. Since canonical parameters are linearly related to the covariance parameters, the covariance matrices Σ in the dyadic symmetry model can be expressed in terms of the set of canonical parameters as follows:

$$\begin{aligned}\Sigma = {} & \lambda_{uz}E_{uz} + \lambda_{uy}E_{uy} + \lambda_{zy}\Delta_{zy} \\ & + \lambda_a E_a + \lambda_s E_s + \lambda_r E_r + \lambda_g E_g + \lambda_d E_d \\ & + \lambda_{gd}\Delta_{gd} + \lambda_{ad}\Delta_{ad} + \lambda_{ag}\Delta_{ag},\end{aligned} \tag{2.2}$$

where the E's and Δ's are known symmetric matrices numerically determined by the expressions in the fourth column of Table 1. For example, the (ji, ij)th element of E_s can be obtained from Table 1 as follows: go to the row whose first column is ji, ij and in the same row find the coefficient of λ_s in the fourth column, which is $\frac{u-2}{2u} - \frac{1}{u(u-1)}$. A specification of the E and Δ matrices in terms of irreducible G-subspaces and bijective G-linear mappings will be given in the next paragraph. For any Σ in the form of (2.2), the inverse of Σ has the following expression:

$$\begin{aligned}\Sigma^{-1} = {} & \eta_{uz}E_{uz} + \eta_{uy}E_{uy} + \eta_{zy}\Delta_{zy} + \lambda_s^{-1}E_s + \lambda_r^{-1}E_r \\ & + \eta_a E_a + \eta_g E_g + \eta_d E_d + \eta_{gd}\Delta_{gd} + \eta_{ad}\Delta_{ad} + \eta_{ag}\Delta_{ag},\end{aligned} \tag{2.3}$$

where

$$\Lambda_{gda} = \begin{pmatrix} \lambda_g & \lambda_{gd} & \lambda_{ag} \\ \lambda_{gd} & \lambda_d & \lambda_{ad} \\ \lambda_{ag} & \lambda_{ad} & \lambda_a \end{pmatrix} = \begin{pmatrix} \eta_g & \eta_{gd} & \eta_{ag} \\ \eta_{gd} & \eta_d & \eta_{ad} \\ \eta_{ag} & \eta_{ad} & \eta_a \end{pmatrix}^{-1},$$

$$\Lambda_{zy} = \begin{pmatrix} \lambda_{uy} & \lambda_{zy} \\ \lambda_{zy} & \lambda_{uz} \end{pmatrix} = \begin{pmatrix} \eta_{uy} & \eta_{zy} \\ \eta_{zy} & \eta_{uz} \end{pmatrix}^{-1}.$$

Under the above notation, the determinant of Σ can be written as

$$|\Sigma| = \lambda_s^{u(u-3)/2}\lambda_r^{(u-1)(u-2)/2}|\Lambda_{zy}||\Lambda_{gda}|^{u-1}. \tag{2.4}$$

The matrix Σ in (2.2) is positive definite if and only if $\lambda_s > 0$, $\lambda_r > 0$, Λ_{zy} is positive definite, and Λ_{gda} is positive definite. The above expressions are a direct extension of Li (2000), and are directly verifiable. However, an interpretation of those matrices in terms of group representation theory would make things more transparent.

A central part in the definition of any GS or LGS model is the mapping $\rho : G \to GL(\mathbb{R}^I)$, which is a group representation of G on $\mathbb{R}^I$. For the dyadic symmetry model this mapping is given in (1.2). Under this particular representation, $\mathbb{R}^I$ as a G-space can be decomposed into the direct sum of a set of seven mutually orthogonal irreducible G-subspaces. The E matrices in (2.2) are the orthogonal projections onto the set of seven mutually orthogonal irreducible G-subspaces to be specified below. Subspace I, the range of E_{uz}, is the one-dimensional space spanned by $\{x : x_{ii} = 1\ \forall i,\ x_{ij} = 0\ \forall i \neq j\}$; subspace II, the range of E_{uy}, is the one-dimensional space spanned by $\{x : x_{ii} = 0\ \forall i,\ x_{ij} = 1\ \forall i \neq j\}$; subspace III, the range of E_a, is the $(u-1)$-dimensional space $\{x : \sum_{i=1}^{u} x_{ii} = 0,\ x_{ij} = 0\ \forall i \neq j\}$; subspace IV, the range of E_s, is the $u(u-3)/2$-dimensional space

$\{x : x_{ii} = 0,\ x_{ij} = x_{ji},\ \sum_{k \neq i} x_{ik} = 0,\ \forall i, j\}$; subspace V, the range of E_r, is the $(u-1)(u-2)/2$-dimensional space $\{x : x_{ij} = -x_{ji}\ \forall i, j,\ \sum_{j \neq i} x_{ij} = 0\ \forall i\}$; subspace VI, the range of E_g, is the $(u-1)$-dimensional space $\{x : x_{ii} = 0\ \forall i,\ x_{ij} = g_i + g_j,\ \sum_{i=1}^{u} g_i = 0\}$; subspace VII, the range of E_d, is the $(u-1)$-dimensional space $\{x : x_{ij} = d_i - d_j,\ \sum_{i=1}^{u} d_i = 0\}$.

Among the above set of irreducible G-subspaces, I and II are equivalent, and III, VI, VII are also equivalent. Between those equivalent G-subspaces there are bijective G-linear mappings [AM (1998)]. The Δ matrices in (2.2) play the dual roles of both such bijective G-linear mappings and their inverses. For example, Δ_{gd} as a linear operator maps subspace VI (the range of E_g) onto subspace VII (the range of E_d), subspace VII onto subspace VI, all the other vectors to 0, and commutes with $\rho(\pi)$ for all $\pi \in G$; additionally, it is also an isometry (preserves inner product), and satisfies the identities $E_g \Delta_{gd} = \Delta_{gd} E_d$, $E_d \Delta_{gd} = \Delta_{gd} E_g$, $\Delta_{gd}^2 = E_g + E_d$ and $\Delta_{gd}^3 = \Delta_{gd}$.

The rest of the Δ matrices can be described in the same fashion in an obvious way, with the subscripts indicating the subspaces involved. By definition all the Δ matrices are G-invariant, and it is easily seen that all the E matrices are also G-invariant. Those relations among E and Δ matrices, together with their properties, lead directly to (2.3) and (2.4).

The G-space $\mathbb{R}^I$ can be decomposed into seven mutually orthogonal irreducible G-subspaces as described in the previous paragraph only when $u > 3$, which is the reason for the declaration of $u > 3$ made earlier in this paper. For $u \leq 3$ the G-space $\mathbb{R}^I$ does not have as many irreducible components. When $u = 3$, subspace IV is eliminated, leaving us with six mutually orthogonal irreducible G-subspaces. When $u = 2$, subspaces IV, V, VI, VII collapse into a single one-dimensional irreducible G-subspace, which is equivalent to subspace III, resulting in four mutually orthogonal irreducible G-subspaces. Of course, $\mathbb{R}^I$ is one-dimensional when $u = 1$, and therefore no decomposition is possible. Given the above comments, the general results obtained for $u > 3$ are easily adapted to handle the cases when $u \leq 3$.

Using (2.2), (2.3) and (2.4), the likelihood function for the realization $x = (y, z)$ of a random vector following the dyadic symmetry model can be expressed as

$$
\begin{aligned}
&l(\lambda_s, \lambda_r, \Lambda_{gda}, \Lambda_{zy} \mid y, z) \\
&\quad \propto \lambda_r^{-(u-1)(u-2)/4} \lambda_s^{-u(u-3)/4} |\Lambda_{gda}|^{-(u-1)/2} |\Lambda_{zy}|^{-1/2} \\
&\quad \times \exp\left[\frac{u}{2}\Big((z_{\cdot\cdot} - \xi^{\text{diag}}), \sqrt{(u-1)}(y_{\cdot\cdot} - \xi^{\text{off}})\Big) \Lambda_{zy}^{-1} \begin{pmatrix} (z_{\cdot\cdot} - \xi^{\text{diag}}) \\ \sqrt{(u-1)}(y_{\cdot\cdot} - \xi^{\text{off}}) \end{pmatrix}\right] \\
&\quad \times \exp\left(-\frac{\operatorname{tr}(E_r y y')}{2\lambda_r}\right) \exp\left(-\frac{\operatorname{tr}(E_s y y')}{2\lambda_s}\right) \exp\left(-\frac{1}{2} \operatorname{tr} \Lambda_{gda}^{-1} S\right),
\end{aligned} \tag{2.5}
$$

where

$$S = \begin{pmatrix} \operatorname{tr}(E_g yy') & \frac{1}{2}\operatorname{tr}(\Delta_{gd} yy') & \frac{1}{2}\operatorname{tr}(\Delta_{ga} y\mathbf{y}') \\ \frac{1}{2}\operatorname{tr}(\Delta_{gd} yy') & \operatorname{tr}(E_d yy') & \frac{1}{2}\operatorname{tr}(\Delta_{ad} y\mathbf{y}') \\ \frac{1}{2}\operatorname{tr}(\Delta_{ga} yy') & \frac{1}{2}\operatorname{tr}(\Delta_{ad} yy') & \operatorname{tr}(E_a yy') \end{pmatrix},$$

$z_{..} = \sum_{i=1}^{u} z_{ii}/u$, and $y_{..} = \sum_{i\neq j}^{u} y_{ij}/[u(u-1)]$. The likelihood function (2.5) is the product of four separate likelihood functions, corresponding to the parameters λ_r, λ_s, Λ_{gda} and $(\Lambda_{zy}, \xi^{\text{off}}, \xi^{\text{diag}})$, respectively. All of them should be sufficiently familiar so that (2.5) can be viewed as readily applicable in practical situations. Note that the parameters cannot be estimated via maximizing the likelihood (2.5), because the multiplicative component of the likelihood involving Λ_{zy} effectively corresponds to a single realization of a bivariate normal random vector. Since Λ_{zy} is 2×2, we need at least 3 independent replicates of x in order to obtain the maximum likelihood estimates for all the parameters. If $(\xi^{\text{off}}, \xi^{\text{diag}})$ is constrained to be $(0, 0)$, then we only need at least 2 independent replicates of x. Those conclusions can also be reached by invoking the general theorems in AM (1998) using the concept of structure constants. In the next section we look at a submodel of the dyadic symmetry model in which the parameters can be estimated without independent replicates, and provide arithmetic expressions for the quadratic sufficient statistics.

3. A useful submodel. To motivate the submodel, compare the "off-diagonal" part of (2.1) to Model (b) in Cockerham and Weir [(1977), page 188]. Note that the former becomes the latter if μ_y is set to 0. If we accordingly set μ_z to be 0, then (2.1) becomes

$$\begin{aligned} y_{ij} &= \xi^{\text{off}} + g_i + g_j + s_{ij} + d_i - d_j + r_{ij}, \\ z_{ii} &= \xi^{\text{diag}} + a_i, \end{aligned} \tag{3.1}$$

which can be regarded as an extension of Cockerham and Weir's (1977) Model (b) "by including diagonal measurements." When μ_z and μ_y are set to be 0, the covariance parameters $\sigma_y^2\rho_{y0}$, $\sigma_z^2\rho_{z0}$ and $\sigma_y\sigma_z\rho_{zy0}$ are all equal to 0, and this is the submodel to be considered in this section.

From (2.5), the maximum likelihood estimators for ξ^{off} and ξ^{diag} are obviously $y_{..}$ and $z_{..}$, respectively.

The parameters in the covariance structure, which are usually the focus of interest, can be estimated via the marginal likelihood (or REML likelihood) [Searle, Casella and McCulloch (1992), page 323] obtained by integrating out ξ^{off} and ξ^{diag}, as given below:

$$\begin{aligned} &l(\lambda_s, \lambda_r, \Lambda_{gda} \mid y, z) \\ &\quad\propto \lambda_r^{-(u-1)(u-2)/4} \lambda_s^{-u(u-3)/4} |\Lambda_{gda}|^{-(u-1)/2} \\ &\qquad\times \exp\left(-\frac{\operatorname{tr}(E_r yy')}{2\lambda_r}\right) \exp\left(-\frac{\operatorname{tr}(E_s \mathbf{y}y')}{2\lambda_s}\right) \exp\left(-\frac{1}{2}\operatorname{tr}\Lambda_{gda}^{-1} S\right). \end{aligned} \tag{3.2}$$

Since $\sigma_y^2\rho_{y0}$, $\sigma_z^2\rho_{z0}$ and $\sigma_y\sigma_z\rho_{zy0}$ are all equal to 0, there is a one-to-one linear relation between the eight canonical parameters in the above marginal likelihood and the eight nonzero covariance parameters. Consequently, inference on all the nonzero covariance parameters can be based on the above marginal likelihood, from which we obtain the following mutually independent sufficient statistics and their distributions:

$$
\begin{aligned}
SS_s &= y'E_s y = \operatorname{tr}(E_s yy') \sim \lambda_s \chi^2_{u(u-3)/2},\\
SS_r &= y'E_r y = \operatorname{tr}(E_r yy') \sim \lambda_r \chi^2_{(u-1)(u-2)/2},
\end{aligned}
$$

$$
\begin{pmatrix} SS_g & SC_{gd} & SC_{ag} \\ SC_{gd} & SS_d & SC_{ad} \\ SC_{ag} & SC_{ad} & SS_a \end{pmatrix} = \begin{pmatrix} \operatorname{tr}(E_g yy') & \frac{1}{2}\operatorname{tr}(\Delta_{gd} yy') & \frac{1}{2}\operatorname{tr}(\Delta_{ga} y\mathbf{y}') \\ \frac{1}{2}\operatorname{tr}(\Delta_{gd} yy') & \operatorname{tr}(E_d yy') & \frac{1}{2}\operatorname{tr}(\Delta_{ad} y\mathbf{y}') \\ \frac{1}{2}\operatorname{tr}(\Delta_{ga} yy') & \frac{1}{2}\operatorname{tr}(\Delta_{ad} yy') & \operatorname{tr}(E_a yy') \end{pmatrix} \tag{3.3}
$$

$$
\sim W(\Lambda_{gda}, u-1).
$$

Since these statistics are quadratic functions of data that can be arranged into matrices having Wishart distributions, including the one-dimensional special case of χ^2 distributions, we call them sums of squares (*SS*, for diagonal terms in a Wishart matrix) or cross products (*SC*, for off-diagonal terms in a Wishart matrix), making use of the existing terminology in analysis of variance and multivariate analysis. The maximum marginal likelihood (or REML) estimators for the parameters λ's, which are also unbiased, are the corresponding sums of squares or cross products divided by their degrees of freedom, which will naturally be called mean squares or cross products. Finally, this system of nomenclature leads to the λ's being called expected mean squares or cross products. The following arithmetic expressions for the sufficient statistics provide further justification for the nomenclature, since the averaging operations involved are exactly those usually found in (multivariate) analysis of variance:

$$
\begin{aligned}
SS_g &= \sum_{i\neq j}\left\{\frac{1}{2(u-2)}\left(y_{i+} + y_{+i} + y_{j+} + y_{+j} - \frac{4}{u}y_{++}\right)\right\}^2,\\
SS_d &= \sum_{i\neq j}\left\{\frac{1}{2u}(y_{i+} - y_{+i} - y_{j+} + y_{+j})\right\}^2,\\
SS_s &= SS_m - SS_g,\\
SS_r &= SS_{\text{total},y} - SS_g - SS_s - SS_d,\\
SC_{gd} &= \sum_{i\neq j}\frac{1}{2u\sqrt{u(u-2)}}(y_{i+} - y_{+i} - y_{j+} + y_{+j})(y_{i+} - y_{+j}),\\
SS_a &= \sum_i\left(z_{ii} - \frac{1}{u}z_{++}\right)^2,
\end{aligned} \tag{3.4}
$$

$$SC_{ag} = \frac{1}{\sqrt{2(u-2)}} \sum_i \left(z_{ii} - \frac{1}{u} z_{++} \right) (y_{i+} + y_{+i}),$$

$$SC_{ad} = \frac{1}{\sqrt{2u}} \sum_i z_{ii} (y_{i+} - y_{+i}),$$

where SS_m is $\sum_{i \neq j} [(y_{ij} + y_{ji})/2]^2 - y_{++}^2/u(u-1)$, and $SS_{\text{total},y}$ is the usual total sum of squares of the off-diagonal observations. The $+$ sign in the subscripts indicates sum. Thus y_{i+} is a "row sum," y_{+j} is a "column sum," and y_{++} is the grand sum.

It should be noted that, while from the maximum likelihood estimates of the canonical parameters we can obtain those of the covariance parameters via the linear relations displayed in Table 1, there is no guarantee that those of the parameters in the latent variable model (3.1) can also be obtained in this way. The reason is that the parameter space of model (2.1) is a proper subset of that of the dyadic symmetry model, and likewise the parameter space of model (3.1) is a proper subset of that of the submodel defined by the three zero constraints. This gives rise to what is essentially the problem of "negative estimates of variance components." Various strategies have been proposed to address this problem, and most can be applied here. Given the theme of this paper, however, we will not launch an in-depth discussion. The readers are referred to Box and Tiao (1973), Searle, Casella and McCulloch (1992) and Rao (1997).

4. Concluding remarks. Results described in this paper are of interest from both applied and theoretical perspectives. Theoretically, they demonstrate the benefits of the precise mathematical formulation of a statistical model. Many important features of our model are immediate consequences of the minimal and logical assumption of exchangeability among units, and the group symmetry it implies. For example, no matter how similar or different the "diagonal" measurements on pairs of identical units are relative to the "off-diagonal" measurements on pairs of distinct units, exactly the same group symmetry is generated, leading to the same covariance structure and the same symmetry-based analyses. To be more concrete, suppose the "off-diagonal" measurements are weights, symmetry-based statistical modeling would be the same whether the "diagonal" measurements are heights or weights. That is why we chose to use different letters to denote the "diagonal" and the "off-diagonal" measurements. In fact, the same symmetry-based model applies even when "diagonal" measurements are made on individual units instead of pairs; of course in this case we would probably want to change the subscript "ii" to "i" in the arithmetic expressions. Although those features seem to be platitude once they are clearly spelled out, and may even be intelligible without using the concept of symmetry, just as the additivity of sums of squares in an ordinary analysis of variance can be explained in terms of the nonmathematical notion of sources of variability, it is difficult to envision a better way to establish those properties than starting from group symmetry based on exchangeability. As Dawid (1988) has pointed out, sometimes "... mere specification of relevant symmetry represents a premodeling phase from which many important consequences flow ... one should consider carefully one's attitudes to the whole collection of observables and, in particular, consider what symmetries these attitudes should incorporate. This specification alone is sufficient to support an extensive body of theory and methodology for data-analysis and (if desired) appropriate model-building." Dawid (1988) used a standard analysis of variance to illustrate this point. This paper also serves to bear it out, through a data structure which, although rather nonstandard relative to any analysis of variance framework, is a typical linear group symmetry model, and hence obeys the general theorems in AM (1998) and Andersson (1975). It is straightforward to show that, under the basis formed by adjoining the orthonormal bases of the irreducible G-subspaces, the covariance matrix (2.2) assumes the block diagonal form given in Perlman (1987).

From the applied perspective, this paper extends some earlier results obtained in several specific contexts substantively unrelated to each other. It extends Cockerham and Weir's (1977) results in at least two ways:

(i) by making the modeling and method of analysis applicable to the set of measurements including both the "diagonal" ones and the "off-diagonal" ones, and

(ii) by providing a closed form expression for the (residual) likelihood and the exact joint distribution of a set of minimal sufficient statistics under the assumption of multivariate normality. The inclusion of "diagonal" measurements can be a significant extension, especially considering what has been noted above about the flexibility in the interpretation of "diagonal" measurements in the group invariance framework.

The closed form likelihood function would provide a basis for statistical computations needed to address some of the complicating issues such as missing data or constraints on the covariance parameters imposed by latent variable models, by using the EM algorithm [Dempster, Laird and Rubin (1977)], its recent extensions and improvements [Meng and Rubin (1991, 1993), Liu and Rubin (1994), van Dyk, Meng and Rubin (1995), Meng and van Dyk (1997, 1998), Liu, Rubin and Wu (1998), Oakes (1999)] or MCMC methods [e.g., Cappé and Robert (2000)].

The exact joint distribution would make statistical inference concerning certain parameters more accurate. It should be noted that the point estimates for the covariance parameters associated with the "off-diagonal" measurements agree with those provided in Cockerham and Weir (1977). Similar comments as above can be made with regard to previous works in various areas in psychology such as Lev and Kinder (1957), Bechtel (1967, 1971) and Warner, Kenny and Stoto (1979), which was later developed into what is now called the Social Relations Model [Kenny (1994)]. In general nonmathematical terms, the data structure considered in this paper may arise in dealing with any phenomena which could be characterized as dyadic, mutual, reciprocal, or relational; and whenever the basic data structure is conceptually relevant, its variations may help meet practical constraints or increase design efficiency. What we have derived in the previous sections could serve as prototypes for later development.

Perlman (1987) pointed out that specifying classes of statistical models based on precise mathematical structure can be a fruitful endeavor. As an example he cited Andersson's characterization of analysis of variance models in terms of the underlying lattice structure published in Andersson (1990). This paper can be viewed as a fruit borne by another such specification, that of the GS and LGS models [AM (1998)]. We have identified a model that has already been used in practice and studied it at a level that does not seem to have been reached previously, with primary results derived directly from the inherent group symmetry in the model when it is related to an LGS model. In particular, we have reaped new ANOVA-type and MANOVA-type arithmetics. The widespread application of analysis of variance is often believed to be accounted for to some extent by its instructive arithmetics, and the attendant regularities that can be understood intuitively and foster a very powerful system of thinking not tied to any specific substantive field. Results in this paper suggest that the above features could be shared by some of the GS and LGS models not falling into any current conceptualizations of analysis of variance. Considerable augmentation to this collection of "standard'" statistical methods may be possible for substantive areas to resort to in conceiving their scientific questions and implementing their scientific investigations.

Acknowledgments. The author thanks Eric Loken, Robert Rosenthal and Donald Rubin for helping with this research in various ways, at various points in time. Thanks also to a referee and Associate Editor for their constructive comments and suggestions.

Part of the research was presented at the VII Latin American Congress on Probability and Mathematical Statistics (CLAPEM 98) thanks to an NSF travel grant.

REFERENCES

ANDERSSON, S. A. (1975). Invariant normal models. *Ann. Statist.* **3** 132–154.

ANDERSSON, S. A. (1990). The lattice structure of orthogonal linear-models and orthogonal variance component models. *Scand. J. Statist.* **17** 287–319.

ANDERSSON, S. A. and MADSEN, J. (1998). Symmetry and lattice conditional independence in a multivariate normal distribution. *Ann. Statist.* **26** 525–572.

BECHTEL, G. G. (1967). The analysis of variance and pairwise scaling. *Psychometrika* **32** 47–65.

BECHTEL, G. G. (1971). A dual scaling analysis for paired compositions. *Psychometrika* **36** 135–154.

BOX, G. E. P. and TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA.

CAPPÉ, O. and ROBERT, C. P. (2000). Markov chain Monte Carlo: 10 years and still running! *J. Amer. Statist. Assoc.* **95** 1282–1286.

COCKERHAM, C. C. and WEIR, B. S. (1977). Quadratic analyses of reciprocal crosses. *Biometrics* **33** 187–204.

DAVID, H. A. (1988). *The Method of Paired Comparisons*, 2nd ed. Griffin, London.

DAWID, A. P. (1988). Symmetry models and hypotheses for structured data layouts (with discussion). *J. Roy. Statist. Soc. Ser. B* **50** 1–34.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

KENNY, D. A. (1994). *Interpersonal Perception*. Guilford, New York.

LEV, J. and KINDER, E. (1957). New analysis of variance formulas for treating data from mutually paired subjects. *Psychometrika* **22** 1–15.

LI, H. (2000). Comment on "Invariance and factorial models," by P. McCullagh. *J. Roy. Statist. Soc. Ser. B* **62** 250–251.

LIU, C. and RUBIN, D. B. (1994). The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* **81** 633–648.

LIU, C., RUBIN, D. B. and WU, Y. N. (1998). Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika* **85** 755–770.

MCCULLAGH, P. (2000). Invariance and factorial models (with discussion). *J. Roy. Statist. Soc. Ser. B* **62** 209–256.

MENG, X.-L. and RUBIN, D. B. (1991). Using the EM algorithm to obtain asymptotic variance-covariance matrices: The SEM algorithm. *J. Amer. Statist. Assoc.* **86** 899–909.

MENG, X.-L. and RUBIN, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80** 267–278.

MENG, X.-L. and VAN DYK, D. A. (1997). The EM algorithm—an old folk song sung to a fast new tune (with discussion). *J. Roy. Statist. Soc. Ser. B* **59** 511–567.

MENG, X.-L. and VAN DYK, D. A. (1998). Fast EM-type implementations for mixed-effects models. *J. Roy. Statist. Soc. Ser. B* **59** 559–578.

OAKES, D. (1999). Direct calculation of the information matrix via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **61** 479–482.

PERLMAN, M. D. (1987). Group symmetry models. Comment on "A review of multivariate analysis," by M. J. Schervish. *Statist. Sci.* **2** 421–425.

RAO, P. S. R. S. (1997). *Variance Components Estimation*. Chapman and Hall, London.

SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components*. Wiley, New York.

VAN DYK, D. A., MENG X.-L. and RUBIN, D. B. (1995). Maximum likelihood estimation via the ECM algorithm: computing the asymptotic variance. *Statist. Sinica* **5** 55–75.

WARNER, R. M., KENNY, D. A. and STOTO, M. (1979). A new round robin analysis of variance for social interaction data. *J. Personality Soc. Psych.* **37** 1742–1757.

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF ROCHESTER
601 ELMWOOD AVENUE, BOX 630
ROCHESTER, NEW YORK 14642
E-MAIL: liheng@bst.rochester.edu