Behavior Genetics, Vol. 24, No. 1, 1994

Simultaneous Genetic Analysis of Means and Covariance Structure: Pearson–Lawley Selection Rules

Conor V. Dolan,[1] Peter C. M. Molenaar,[1] and Dorret I. Boomsma[2]

Received 2 Dec. 1991—Final 4 June 1993

The object of this paper is to indicate that the Pearson—Lawley selection rules form a plausible general theory for the simultaneous genetic analysis of means and covariance structure. Models are presented based on phenotypic selection and latent selection. Previously presented quantitative genetic models to decompose means and covariance structure simultaneously are reconsidered as instances of latent selection. The selection rules are very useful in the context of behavior genetic modeling because they lead to testable models and a conceptual framework for explaining variation between and within groups by the same genetic and environmental factors.

KEY WORDS: Pearson–Lawley selection formulas; environmental factors; genetic factors; covariance structure; mean structure; phenotypic selection; latent selection.

INTRODUCTION

To decompose differences in means between groups into genetic and environmental components requires certain knowledge of the genetic and environmental influences present in each of the groups. When a difference in mean is observed with respect to a polygenetic character in human samples, such knowledge amounts to a detailed catalog of the genetic and environmental influences active in the groups under consideration. As this knowledge is typically absent or, at best, incomplete, quantitative genetic studies of human polygenetic phenotypes are limited to the sources of within-group variation. The sources of within-group variation can be identified and their contributions estimated using individuals in known genetic and environmental relationships (Neale and Cardon, 1992; Plomin *et al.*, 1990).

That the sources of between-group variation, in the circumstances mentioned, are not identified does not mean, however, that one cannot consider mean differences in the light of various assumptions and (or) circumstantial evidence. It does mean that the hypotheses following from such considerations cannot be tested rigorously and thus remain speculative. One assumption, which has been employed by Jensen (1973), is that the sources of between-group variation and the sources of within-group variation are identical. We refer to this assumption conveniently as the assumption of common causation. Recently Turkheimer (1991) has employed this assumption to model mean differences in IQ between adoptees and their parents. Both Jensen (1973, p. 135) and Turkheimer (1991, p. 393) emphasize the plausibility of this assumption without discounting its speculative nature. Jensen (1973, p. 134) states clearly, "The simple fact is that one cannot, in any strict formal sense, infer between-groups heritability from a knowledge of within-groups heritability. This is true even when the heritability ... of the trait is perfect." If the assumption of common causation holds, the mean phenotypic difference between the groups under consideration can be directly related to the decomposition of the within-group phenotypic variation (see Turkheimer, 1990, 1991; Furby, 1973; Jensen, 1973, Chap. 5; Dolan, 1992). This implies that the between-group variation can be modeled using information derived from the analysis of within-group variation.

In the present paper we indicate that the Pearson–Lawley selection rules provide a useful framework for the simultaneous genetic analysis of means

[1] Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands.
[2] Department of Experimental Psychology, Free University, De Boelelaan 1111, 1081 HV Amsterdam, The Netherlands.

0001-8244/94/0100-0017$07.00/0

and covariance structure. The Pearson–Lawley selection formulas (Lawley, 1943) can be used to assess the changes in covariance and means structure of a set of variables following selection on the basis of one or more selection variables under the assumption that the regression of the selection variables on the set is homoscedastic and linear. These formulas have proven to be useful in covariance structure modeling generally (Thomson, 1945; Meredith, 1964; Jöreskog, 1971; Sörbom, 1974, 1976; Muthén, 1989), but also in genetic covariance structure modeling, for instance, of assortative mating (Fulker, 1988) and ascertainment bias (Neale *et al.*, 1989; Neale and Cardon, 1992, pp. 360–361). The application of the selection rules to the simultaneous quantitative genetic modeling of means and covariance structure provides an explicit mechanism that leads to a situation where between-group variation and within-group variation are attributable to the same latent factors. This mechanism constitutes a useful conceptual framework for thinking about common effects on means and variance which can be evaluated on both theoretical and statistical criteria. A model presented previously by Dolan *et al.* (1992) is considered in the light of selection by positing a process of group selection on the basis of criteria related to the latent variables, i.e., the environmental and genetic factors. A second case that leads to a simultaneous genetic analysis of mean variation and individual differences is based on a process of phenotypic selection (Muthén, 1989; Thomson, 1945).

APPLICATION OF THE PEARSON–LAWLEY SELECTION RULES IN COVARIANCE STRUCTURE MODELING

In the present section we summarize published results obtained by applying the selection rules within the context of the common factor model. These results are subsequently applied to metric characters characterized by a simple genetic model. Meredith (1964) considered the consequences of selection based on the latent variable(s) and Muthén (1989; see also Thomson, 1945) recently considered the consequences of selection based on the observed variable(s) for the common factor structure. The former is referred to as latent selection; the latter, as phenotypic selection. The common factor model, which includes various genetic covariance structure models as special cases (Martin and Eaves, 1977; Neale and Cardon, 1992), is described briefly. Subsequently, the selection rules are applied to the factor model.

Let $\mathbf{y}$ denote a p-dimensional vector of multinormal observed variables with mean $E[\mathbf{y}] = \nu$. Let $\boldsymbol{\eta}$ denote a zero-mean q-dimensional vector of common latent variables and $\boldsymbol{\epsilon}$ a zero-mean p-dimensional vector of residuals. In quantitative genetic applications the components of $\boldsymbol{\eta}$ may include additive genetic, dominance, and environmental factors. The relationship between these variables in given by Eq. (1a), where $\boldsymbol{\Lambda}$ is a $(p \times q)$ matrix containing the regression coefficients of the regressions of $\mathbf{y}$ on $\boldsymbol{\eta}$.

$$\mathbf{y} = \nu + \Lambda\eta + \epsilon \tag{1a}$$

$$E[\mathbf{y}] = \nu \tag{1b}$$

$$E[(\mathbf{y} - \nu)(\mathbf{y} - \nu)'] = \Sigma_{yy} = \Lambda\Psi\Lambda' + \Theta \tag{1c}$$

The mean vector and the covariance matrix of the observed variables expressed as a function of the latent and residual variables are given in Eqs. (1b)–(1c). In Eq. (1c), we define $E[\epsilon\epsilon'] = \Theta$ and $E[\eta\eta'] = \Psi$. In view of Eq. (1b), finally, $E[\epsilon] = E[\eta] = 0$.

We may select a subsample from a well-defined reference population on the basis of a (possibly multivariate) metric criterion $\mathbf{s}$, a predictor of $\mathbf{x}$. The regression of $\mathbf{x}$ on $\mathbf{s}$ is linear and homoscedastic. A case is selected if the score $\mathbf{s}$ exceeds (c.q. lies below) some predetermined cutoff score. Under the assumption of linearity and homoscedasticity, we deduce the covariance and means structure in the phenotypically selected sample using the selection formulas. The selection formulas are

$$E[\mathbf{x}]^* = E[\mathbf{x}] + \Sigma_{xs}\Sigma_{ss}^{-1}(E[\mathbf{s}]^* - E[\mathbf{s}]) \tag{2a}$$

$$\Sigma_{xx}^* = \Sigma_{xx} + \Sigma_{xs}\Sigma_{ss}^{-1}(\Sigma_{ss}^* - \Sigma_{ss})\Sigma_{ss}^{-1}\Sigma_{xs}' \tag{2b}$$

$$\Sigma_{xs}^* = \Sigma_{xs}\Sigma_{ss}^{-1}\Sigma_{ss}^* \tag{2c}$$

where $\boldsymbol{\Sigma}_{xs}$ is the covariance matrix of $\mathbf{x}$ and $\mathbf{s}$, $E[\mathbf{s}]$ is the mean of the criterion variable s, etc. The symbols with a superscript asterisk represent values in the selected sample; e.g., $E[\mathbf{s}]^*$ is the mean of the criterion variable in the selected group, $\boldsymbol{\Sigma}_{xx}^*$ is the covariance matrix of $\mathbf{x}$ in the selected sample, etc.

Latent Selection. Meredith (1964; Lawley and Maxwell, 1971; Muthén and Jöreskog, 1983) applied the Pearson–Lawley selection formulas to investigate the effects of latent selection in the context of the common factor model. We reproduce his results by introducing $\mathbf{s}$ as an r-dimensional random

variable such that the regression of each component of η on $\mathbf{s}$ is homoscedastic and linear: $\eta = \mathbf{as} + \delta$ (Fig. 1). The $(q \times r)$ matrix $\mathbf{a}$ contains regression coefficients, and δ is a q-dimensional vector of residuals. The regression of $\mathbf{y}$ on η is given above.

It follows from the Pearson–Lawley selection formulas that the covariance matrix of $\mathbf{y}$ in a group i, derived from the reference population by selection on $\mathbf{s}$, equals

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} + \boldsymbol{\Lambda}\mathbf{a}(\boldsymbol{\Sigma}_{s_i} - \boldsymbol{\Sigma}_s)\mathbf{a}'\boldsymbol{\Lambda}' \tag{3a}$$

and the means

$$E[\mathbf{y}_i] = E[\mathbf{y}] + \boldsymbol{\Lambda}\mathbf{a}(E[\mathbf{s}_i] - E[\mathbf{s}]) \tag{3b}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{y}$ in the reference population, $\boldsymbol{\Sigma}_s$ and $E[\mathbf{s}]$ are the covariance matrix and mean of $\mathbf{s}$ in the reference population, and $\boldsymbol{\Sigma}_{s_i}$ and $E[\mathbf{s}_i]$ are the covariance matrix and mean of $\mathbf{s}$ in the selected subsample, respectively. Similarly,

$$\boldsymbol{\Psi}_i = \boldsymbol{\Psi} + \mathbf{a}(\boldsymbol{\Sigma}_{s_i} - \boldsymbol{\Sigma}_s)\mathbf{a}' \tag{4a}$$

$$E[\eta_i] = E[\eta] + \mathbf{a}(E[\mathbf{s}_i] - E[\mathbf{s}]) \tag{4b}$$

So we can write

$$\begin{aligned}\boldsymbol{\Sigma}_i &= \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}\mathbf{a}(\boldsymbol{\Sigma}_{si} - \boldsymbol{\Sigma}_s)\mathbf{a}'\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \\ &= \boldsymbol{\Lambda}[\boldsymbol{\Psi} + \mathbf{a}(\boldsymbol{\Sigma}_{si} - \boldsymbol{\Sigma}_s)\mathbf{a}']\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \\ E[\mathbf{y}_i] &= \nu + \boldsymbol{\Lambda}E[\eta] + \boldsymbol{\Lambda}\mathbf{a}(E[\mathbf{s}_i] - E[\mathbf{s}]) \\ &= \nu + \boldsymbol{\Lambda}[E[\eta] + \mathbf{a}(E[\mathbf{s}_i] - E[\mathbf{s}])]\end{aligned}$$

or, in view of Eqs. (4a) and (4b),

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}\boldsymbol{\Psi}_i\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \tag{5a}$$

$$E[\mathbf{y}_i] = \nu + \boldsymbol{\Lambda}E[\eta_i] \tag{5b}$$

in the selected sample i compared to Eqs. (1b) and (1c) in the original sample. Thus given selection on a variable related directly to η, the matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Theta}$ and the vector $\boldsymbol{\nu}$ remain invariant, while changes are expressed in the covariance matrix, $\boldsymbol{\Psi}$, and the mean vector $\mathbf{E}[\eta]$.

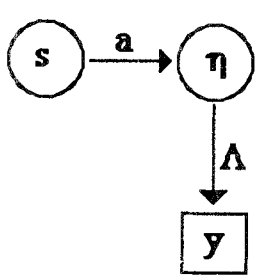

Fig. 1. The relationship between the selection variable $\mathbf{s}$ and the variables in the factor model in the case of latent selection.

Meredith (1964, pp. 184–185) points out that it is not necessary actually to know the selection variable, as it does not appear in Eqs. (5a) and (5b). This implies that one need not identify one group as the original (reference) population and one as the selected sample. The parameters of the model for the phenotypic means equals $p+q$ (dimension of ν plus dimension of $E[\eta_i]$), whereas the number of means equals the number of groups, ng, times the number of variables: $ng*p$. Thus multiple congeneric indicators of a given phenotype are required [i.e., $(p+q)<2*p$] for the means model to be testable. This is by no means a trivial requirement, as in genetic covariance structure modeling test-specific variance is often found to contain systematic environmental and genetic components (Martin and Eaves, 1977; Neale and Cardon, 1992).

Phenotypic Selection. So far we have considered selection based on a possibly unknown variable which is directly related to the latent variables. We now consider phenotypic selection. In the context of factor analysis, this was considered by Thomson (1945) and more recently by Muthén (1989). Fulker (1988) reached similar results in modeling the effects of phenotypic assortative mating on the genetic and environmental covariance between spouses.

Muthén (1989) demonstrates how the mean and covariance structure of the latent variables η observed in a reference sample change following selection of individuals on the bases of a phenotypic variable representing a linear combination of the observed variables: $s = \mathbf{w}'\mathbf{y}$, where $\mathbf{w}$ is a vector of known weights. In the selected group, indicated by the subscript j, we have

$$E[\eta_j] = \boldsymbol{\Psi}\boldsymbol{\Lambda}'\mathbf{w}\kappa_j \tag{6a}$$

$$E[\epsilon_j] = \boldsymbol{\Theta}\mathbf{w}\kappa_j \tag{6b}$$

$$E[\eta_j\,\eta_j'] = \boldsymbol{\Psi} + \boldsymbol{\Psi}\boldsymbol{\Lambda}'\mathbf{w}\omega_j\mathbf{w}'\boldsymbol{\Lambda}\boldsymbol{\Psi} \tag{6c}$$

and phenotypically

$$E[\mathbf{y}_j] = \nu + \boldsymbol{\Sigma}\mathbf{w}\kappa_j \tag{7a}$$

$$\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{w}\omega_j\mathbf{w}'\boldsymbol{\Sigma} \tag{7b}$$

The symbols ω_j and κ_j represent

$$\omega_j = \mathrm{var}[s]^{-1}\{\mathrm{var}[s_j] - \mathrm{var}[s]\}\mathrm{var}[s]^{-1}$$

$$\kappa_j = \mathrm{var}[s]^{-1}(E[s_j] - E[s])$$

where var[s] and $E[s]$ are the variance and mean of the selection variable s in the reference sample (var[s_j] and $E[s_j]$ defined analogously in the selected sample j). Note that the error terms in the selected sample, in contrast to those in the reference group, no longer have zero means, i.e., $E[\epsilon_j] \neq 0$. This is taken up below.

Thus phenotypic selection leads to a multigroup model: Eqs. (7a) and (7b) in the phenotypically selected subsample and Eqs. (1a)–(1c) in the representative reference sample. Here, in contrast to latent selection, we must identify the samples explicitly and distinguish between the selected and the reference sample.

A Special Case Where Phenotypic and Latent Selection Are Indistinguishable. Before considering simple quantitative genetic applications, we note that the models following from latent and phenotypic forms of selection are indistinguishable when the vector $\mathbf{w}$ equals the factor score regression matrix. Because the linear combination $\mathbf{w}'\mathbf{y}$ yields factor scores in this special case, phenotypic selection will coincide with latent selection. The factor score regression matrix calculated according to the regression method (Lawley and Maxwell, 1971, Chap. 8) equals $\Psi\Lambda'\Sigma^{-1}$. Substituting $\Psi\Lambda'\Sigma^{-1}$ for $\mathbf{w}'$, we arrive at the following expression for the covariance matrix and mean vector in the selected sample:

$$\begin{aligned}\Sigma_j &= \Sigma + \Sigma(\Sigma^{-1}\Lambda\Psi\omega_j\Psi\Lambda'\Sigma^{-1})\Sigma \qquad (8a)\\ &= \Lambda\Psi\Lambda' + \Theta + \Lambda\Psi\omega_j\Psi\Lambda'\\ &= \Lambda(\Psi + \Psi\omega_j\Psi)\Lambda' + \Theta\end{aligned}$$

$$\begin{aligned}E[y_j] &= \nu + \Sigma\Sigma^{-1}\Lambda\Psi\kappa_j \qquad (8b)\\ &= \nu + \Lambda\Psi\kappa_j\end{aligned}$$

Substituting Ψ_j for $(\Psi + \Psi\,\omega_j\,\Psi)$, we obtain the Merdith–Jöreskog–Sörbom model $\Sigma_j = \Lambda\Psi_j\Lambda' + \Theta$. Other methods of constructing factor scores yield a similar reduction of $\Sigma_j = \Sigma + \Sigma\,\mathbf{w}\omega_j\mathbf{w}'\Sigma$ to $\Sigma_j = \Lambda\Psi_j\Lambda' + \Theta$. For instance, in the case of the Bartlett method of calculating factor scores (Lawley and Maxwell, 1971, Chap. 8), $(\Lambda'\Theta^{-1}\Lambda)^{-1}\Lambda'\,\Theta^{-1}$ is substituted for $\mathbf{w}'$. We find $\Psi_j = [\Psi + \Psi + [\Lambda'\Theta^{-1}\Lambda]^{-1}\omega_j[\Lambda'\Theta^{-1}\Lambda]^{-1} + \Psi]$ so that again we may write $\Sigma_j = \Lambda\Psi_j\Lambda' + \Theta$. The means are then modeled as $E[\mathbf{y}_j] = \nu + \Lambda\,(\Psi + (\Lambda'\Theta^{-1}\Lambda)^{-1})\,\kappa_j$. This result holds for the various other methods of calculating factor scores. The reader is referred to Saris *et al.* (1978) for a review of methods for factor scores calculation and to Boomsma *et al.* (1990) for a quantitative genetic application of factor scores.

That the effects of latent and phenotypic selection converge as the vector $\mathbf{w}$ approaches one of the factor score matrices has two implications. First, it implies that processes that are modeled by means of either latent or phenotypic selection may be hard to distinguish empirically. One such process is assortative mating, which, depending on the theory of assortment, can be modeled as either a phenotypic selection process (Fulker, 1988) or a latent selection process (Falconer, 1990, page 176). Second, it implies that a multigroup factor model (e.g., Sörbom, 1974), which usually involves fitting Eqs. (1b) and (1c) in one group and Eqs. (5a) and (5b) in the other, can be specified more parsimoniously. Specifically one can retain Eqs. (1b) and (1c) for one group and fit $\Sigma_j = \Lambda\,\Psi_j\,\Lambda' + \Theta$ in the other, with Ψ_j constrained to equal $(\Psi + \Psi\,\omega_j\Psi)$. A clear advantage of this approach is the reduction in parameters from $q^*(q+1)/2$ parameters [i.e., Ψ_i in Eq. (5a)] to a single parameter (*viz.*, ω_j; of course this applies only if $q>1$). As a test of strict latent selection, this approach is more parsimonious and therefore more powerful (see Dolan and Molenaar, In Press).

SIMPLE QUANTITATIVE GENETIC APPLICATIONS

Biometric Decomposition of Phenotypic Means Following Phenotypic Selection. The application of the selection rules to the common factor model has an important consequence for quantitative genetic modeling of continuously varying characters in human samples. We arrive at models which incorporate the notion of common causation of within- and between-group variation in a precise and empirically testable manner. In the present section, phenotypic selection is applied and the resulting biometric decomposition of mean variation [i.e., Eqs. (1a)–(1c) vs. Eqs. (7a) and (7b)] explicated.

We assume that the environmental and genetic sources (G and E) of individual differences in a phenotypic variable y are known: $y = \nu + \lambda g\,G + \lambda e\,E$ or $y = \nu + \Lambda\eta$ with $\Lambda = [\lambda g\ \lambda e]$, $\eta' = [G,E]$ and $E[y] = \nu$. The variance is decomposed as follows: $E[(y-\nu)(y-\nu)'] = \Sigma_{yy} = \Lambda\Psi\Lambda'$, where Ψ equals a 2×2 identify matrix $\mathbf{I}$. Independence of the latent variables G and E is not a necessary assumption, but we do assume that genotype $\times$ environment interaction is absent. We simply select

individuals directly on the basis of their observed scores (i.e., $\mathbf{w}=1$) on the basis of some cutoff score, say $E[y]+sd$. The difference between the mean in the population and that in the phenotypically selected sample can then be expressed as a function of the differences in latent means in the representative sample, i.e., $E[G]$ and $E[E]$, and the latent means in the selected sample j, $E[E_j]$ and $E[G_j]$. Applying Eqs. (6a) and (6b), we thus find that

$$\begin{aligned} E[\mathrm{G}_j] &= \mathrm{var}\,[s]^{-1}\,\lambda_g\,(E[s_j] - E[s_j]) \\ &= \mathrm{var}[s]^{-1}\lambda_g\,E\,[s_j] \\ E[E_j] &= \mathrm{var}\,[s]^{-1}\,\lambda_e\,(E[s_j] - E[s]) \\ &= \mathrm{var}[s]^{-1}\lambda_e\,E\,[s_j] \end{aligned}$$

and

$$E[y_j] = \mathrm{v} + \lambda_g E[G_j] + \lambda_e E[E_j]$$

while in the reference sample we have simply $E[G] = E[E] = 0$. The covariance matrix in the selected sample, finally, is a matrix that will certainly no longer be orthogonal: $\mathbf{\Psi}_j = \mathbf{I}+\mathbf{\Lambda}'\omega_j\mathbf{\Lambda}$. As mentioned above, an error term can no longer be expected to have a zero means [see Eq. (6b)] following phenotypic selection. In the present simple quantitative genetic model, the error term is not defined explicity; rather it features implicitly as a part of the unshared environmental variance (Neale and Cardon, 1992, p. 14). It is important to realize therefore that the term $\lambda_e E[E_j]$ will be biased by the contribution of the nonzero mean of the error variance. Correction can be made if an estimate of the error variance in the reference population is available.

Explicit phenotypic selection leads to a biometric decomposition of phenotypic means based on knowledge of the biometric decomposition of individual differences, i.e., a situation where the decomposition of individual differences is informative for the decomposition of group differences. Muthén (1989) derived Eqs. (6a)–(7b) as a means to investigate the factor structure in samples selected on the basis of a linear combination of test scores on, for instance, an admission test. The factor structure in the selected group is of interest because it reveals the effect of selection: "... If a certain factor mean increases relatively little when moving from the general population to the top group, this indicates that the items measuring the factor have little power in discriminating between average and successful testers" (Muthén, 1989, p. 82). Clearly, if the genetic covariance structure is known in the reference population, the mean differences between the reference population and any phenotypically selected group can be broken down into genetic and environmental components.

We have considered the most simple case with one phenotypic variable and direct selection ($w=1$). In the case of a multivariate phenotype, one can use Eqs. (16a)–(17b) to study the effects of phenotypic selection on the basis of a variety of phenotypic criteria such as the average of the phenotypes or a single component of the phenotypic vector. In the latter case, the indirect effects of the selection on a second correlated phenotypic component can be evaluated.

Biometric Decomposition of Phenotypic Means Based on Latent Selection. A model based on the Meredith–Jöreskog–Sörbom application of the selection rules is presented by Dolan *et al.* (1992) to obtain a decomposition of gender-related variation in phenotypic mean blood pressure into genetic and environmental components. In that application, the model deviates slightly (for details see Dolan *et al.*, 1992) from the model following from a strict application of latent selection, i.e., the model given by Eqs. (5a) and (5b). In the present section we consider the latter model.

In Fig. 2, the (at least) trivariate phenotype y is regressed on an additive genetic and environmental factor, denoted G and E. The variables G and E in turn are regressed on the selection variable s, which may be univariate, combining both the additive genetic and the environmental factors (Fig. 2B), or multivariate (Fig. 2A). The covariance

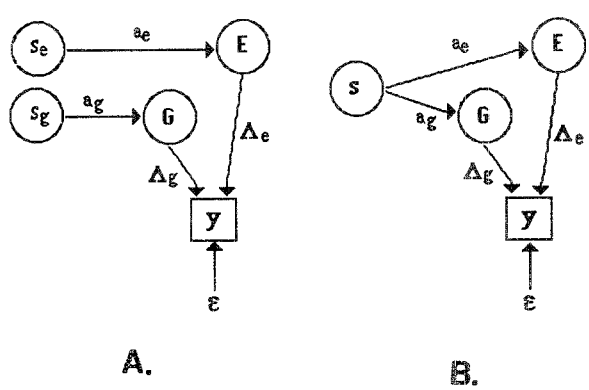

Fig. 2. The relationship between the selection variables s and the variables in a simple genetic model in the case of latent selection. The phenotype y is multivariate; the symbols Λ_g and Λ_e represent vectors containing factor loadings.

structure of $\mathbf{y}$ is $\Sigma = \Lambda I \Lambda' + \Theta$, where $\Lambda = [\Lambda_g, \Lambda_e]$ and $\Theta = [\text{var}(\epsilon)]$. Let $E[\mathbf{y}] = \nu$, so that prior to selection $E[G] = E[E] = 0$. We assume that the variables G and E are orthogonal.

Applying the selection rules as given above [Eqs. (4a) to (5b)], we find the covariance and mean structure of $\mathbf{y}$ in following selection of the basis of $\mathbf{s}$. For Fig. 2A, we have

$$\mathrm{a} = \begin{bmatrix} a_g & 0 \\ 0 & a_e \end{bmatrix}, \Sigma_s = \begin{bmatrix} \text{var}[s_g] & 0 \\ 0 & \text{var}[s_e] \end{bmatrix},$$

$$\Sigma_{si} = \begin{bmatrix} \text{var}[s_{gj}] & 0 \\ 0 & \text{var}[s_{ej}] \end{bmatrix}$$

and

$$\Psi_i = \mathrm{I} + a(\Sigma_{sj} - \Sigma_s)a^t = \begin{bmatrix} 1 + a_g^2(\text{var}[s_{gj}] - (\text{var}[s_g]) & 0 \\ 0 & 1 + a_e^2(\text{var}[s_{ej}] - (\text{var}[s_e]) \end{bmatrix}$$

where $\text{var}[s_g]$ and $\text{var}[s_{gj}]$ represent the variance of s_g in the selected sample and the reference population, respectively ($\text{var}[s_e]$ and $\text{var}[s_{ej}]$ are defined analogously). The phenotypic means in the selected sample equal $E[y_j] = \nu + \Lambda_g a_g E[s_{gj}] + \Lambda_e a_e E[s_{ej}]$. For Fig. 2B, we have $\mathbf{a}^t = [a_g, a_e]$. Assuming that they were in the original or reference population, the factors G and E in group j are no longer orthogonal in the case of Fig. 2A:

$$\Psi_j = \mathrm{I} + a(\text{var}[s_j] - \text{var}[s])a^t = \begin{bmatrix} 1 + a_g^2(\text{var}[s_j] - \text{var}[s]) & a_g a_e(\text{var}[s_j] - \text{var}[s]) \\ a_g a_e(\text{var}[s_j] - \text{var}[s]) & 1 + a_e^2\,(\text{var}[s_j] - \text{var}[s]) \end{bmatrix}$$

The expression for the means is $E[y_j] = \nu + [\Lambda_g a_g + \Lambda_e a_e]\, E[s_j]$.

How reasonable it is to assume that either $a_g = 0$ or $a_e = 0$ depends on the nature of the groups and the phenotypes under consideration. For instance, the hypothesis that differences in mean IQ between offspring and their adopted siblings is due to differences in environment would imply that $a_g = 0$ (e.g., Turkheimer, 1990, 1991; Jensen, 1972, p. 15). When comparing parents and their offspring with respect to mean blood pressure, on the other hand, the constraint $a_g = 0$ may be unacceptable because it implies that the same genes are operating in the same manner regardless of the difference in age (e.g., Hewitt, 1990). It appears to us that many hypotheses concerning differences between groups in genetic and/or environmental influences become explicit and empirically testable when formulated in terms of latent selection.

DISCUSSION

The multigroup covariance structure model of Jöreskog and Sörbom (Jöreskog, 1971, 1977; Sörbom, 1974, 1976) includes the model based directly on latent selection as a special case. Their approach, however, offers many more possibilities than follow strictly from latent selection [i.e., Eqs. (5a) and (5b)]. As Jöreskog (1971, p. 410) states, "Firstly, the method may be used regardless of whether the populations are derived by selection or not. The only requirement is that the populations be clearly defined and the samples independent. Secondly, the method is capable of dealing with any degree of invariance ...." Thus the mechanism of a strict selection process can be abandoned while retaining testable models (see, e.g., Dolan *et al.*, 1991). The assumption of common causation of within-group and between-group variance of course remains crucial. Comparing Eqs. (1a)–(1c) to Eqs. (5a) and (5b), we find that latent selection leads to differences between the groups with respect to the factor means and variances ($E[\eta_i]$ and Ψ_i). Using programs that fit multigroup covariance structure models such as LISREL VII (Jöreskog and Sörbom, 1988) and Mx (Neale, 1991), one may investigate differences between the groups with respect to any parameters in the covariance structure model as long as the matrix Λ, which forms the link between the phenotypes and the latent variables, remains invariant over the groups. For instance, one may investigate the the proposition that, in addition to parameters contained in $E[\eta_i]$ and Ψ_i, the error variances in Θ [in Eq. (1c)] are of a different magnitude in the groups under consideration. This is the case given by Dolan *et al.* (1992), where gender-related mean differences in blood pressure are decomposed into genetic and environmental components in a sample of twin data. The reader is referred to Byrne *et al.* (1989) for an extensive discussion and illustration of such possibilities in the common factor model.

In presenting the various simple quantitative genetic models above, we have assumed tactily that genotype–environment ($G \times E$) interaction is absent. The method suggested by Molenaar *et al.* (1990) based on fourth-order statistics of the genetic and environmental factor scores can be applied in the case of latent selection, as the method of Molenaar *et al.* (1991), like the model for latent selection presented above, requires multiple indicators. A second approach to the detection of G × E

interaction is provided by Molenaar and Boomsma (1987). In this approach, interaction between the unshared environment factor and the additive genetic factor gives rise, at the level of the covariance structure, to a second unshared environment factor. How exactly such a second factor contributes to the phenotypic means, given the assumption of common causation, and whether such a contribution can be estimated will be investigated in the future.

The central assumption of common causation can be tested by means of a likelihood-ratio test given multinormal data. Meredith and Tisak (1990) give a derivation of the maximum-likelihood function with structured means. The covariance structure model with structured means [see Eqs. (5a) and (5b)] is nested under the model with unconstrained phenotypic means $E[\mathbf{y}_i] = \nu_i$. Such a test of the model for the mean structure, regardless of the nature of the application, is crucial because it relates directly to what we have called the assumption of common causation. To be sure, there are potential causes of between-group variation imaginable which are completely unrelated to sources of within-group variation. Some of the causes may be quite trivial (see Cook and Campbell, 1979, Chap. 2).

ACKNOWLEDGMENTS

We thank Drs. Joanne Meyer and Niels Waller for their comments on an early version and Dr. Lon Cardon for his comments on a later version of this paper.

REFERENCES

Boomsma, D. I., Molenaar, P. C. M., and Orlebeke, J. K. (1990). Estimation of individual genetic and environmental factor scores. *Genet. Epidemiol.* **7**:83–91.

Byrne, B. M., Shavelston, R. J., and Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychol. Bull.* **105**:456–466.

Cook, T. D., and Campbell, T. D. (1979). *Quasi-Experimentation, Design and Analysis Issues for Field Settings*, Rand McNally, Chicago.

Dolan, C. V. (1992). *Biometric Decomposition of Phenotypic Means in Human Samples,* Unpublished doctoral dissertation, Department of Psychology, University of Amsterdam, Amsterdam.

Dolan, C. V., Molenaar, P. C. M., and Boomsma, D. I. (1991). Simultaneous genetic analysis of longitudinal means and covariance structure in the simplex model using twin data. *Behav. Genet.* **21**:49–65.

Dolan, C. V., Molenaar, P. C. M., and Boomsma, D. I. (1992). Decomposition of multivariate phenotypic means in multi-group genetic covariance structure analysis. *Behav. Genet.* **22**:319–335.

Dolan, C. V. and Molenaar, P. C. M. (in press). Testing specific hypotheses concerning latent group differences in multi-group covariance structure analysis with structured means. *Multivariate Behavioral Research*.

Falconer, D. S. (1990). *Introduction to Quantitative Genetics*, Longman Scientific and Technical, New York.

Fulker, D. W. (1988). Genetic and cultural transmission in human behavior. In Weir, B. S., Eisen, E. J., Goodman, M. M., and Namkoong, G. (eds.), *Proceedings of the Second International Conference on Quantitative Genetics*, Sinuaer Associates, Sunderland, MA.

Furby, L. (1973). Implications of within-group heritabilities for sources of between-group differences: IQ and racial differences. *Dev. Psychol.* **9**:28–37.

Heath, A. C., Neale, M. C., Hewitt, J. K., Eaves, L. J., and Fulker, D. W. (1989). Testing structural equation models for twin data using LISREL. *Behav. Genet.* **19**:9–38.

Hewitt, J. K. (1990). Changes in genetic control during learning, development, and aging. In Hahn, M. E., Hewitt, J. K., Henderson, N. D., and Benno, R. H. (eds.). *Development Behavior Genetics: Neural, Biometrical and Evolutionary Approaches*, Oxford University Press, Oxford.

Jensen, A. R. (1972). *Genetics and Education*, Harper and Row, New York.

Jensen, A. R. (1973). *Educability and Group Differences*, Mathuen, London.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika* **36**:409–426.

Jöreskog, K. G. (1977). Structural equation models in the social sciences: Specification, estimations and tesing. In Krishaiah, P. R. (ed.), *Applications of Statistics*, North-Holland, Amsterdam.

Jöreskog, K. G., and Sörbom, D. (1988). *LISREL. VII. A Guide to the Program and Applications*, SPSS, Chicago.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proc. Roy. Soc. Edinburgh A* **62**:28–30.

Lawley, D. N., and Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method*, Butterworth, London.

Martin, N. G., and Eaves, L. J. (1977). The genetic analysis of covariance structures. *Heredity* **38**:79–95.

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika* **29**:177–185.

Meredith, W., and Tisak, J. (1990). Latent growth analysis. *Psychometrika* **55**:107–122.

Molenaar, P. C. M., and Boomsma, D. I. (1987). Application of non-linear factor analysis to genotype–environment interaction. *Behav. Genet.* **17**:71–80.

Molenaar, P. C. M., Boomsma, D. I., Neeleman, D., and Dolan, C. V. (1991). Using factor scores to detect G × E interactive origin of "pure" genetic or environmental factors obtained in genetic covariance structure analysis. *Genet. Epidemiol.* **7**:93–100.

Muthén, B., and Jöreskog, K. G. (1983). Selectivity problems in quasi-experimental studies. *Eval. Rev.* **7**:139–174.

Muthén B. O. (1989). Factor structure in groups selected on observed scores. *Br. J. Math. Stat. Psychol.* **42**:81–90.

Neale, M. C. (1991). *Mx: Statistical Modeling*, Department of Genetics, Medical College of Virginia, Richmond.

Neale, M. C., and Cardon, L. R. (1992). *Methodology for Genetic Studies of Twins and Families. NATO ASI Series D, Vol. 67*, Kluwer Academic, Dordrecht.

Neale, M. C., Eaves, L. J., Kendler, K. S., and Hewitt, J. K. (1989). Bias in correlations from selected samples of relatives: The effects of soft selection. *Behav. Genet.* **19**:163–169.

Plomin, R., DeFries, J. C., and McClearn, G. E. (1990). *Behavioral Genetics. A Primer*, 2nd ed., Freeman, New York.

Saris, W. E., de Pijper, M., and Mulder, J. (1978). Optimal procedures for estimation of factor scores. *Sociol. Methods Res.* **7**:85–106.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *Br. J. Math. Stat. Psychol.* **27**:229–239.

Sörbom, D. (1976). A statistical model for the measurement of change in true scores. In de Gruyter, D. N. M., and van der Kamp, L. J. T., (eds.), *Advances in Psychological and Educational Measurement*, John Wiley and Sons, New York.

Thomson, G. H. (1945). *The Factorial Analysis of Human Ability*, 2nd ed., University of London Press, London.

Turkheimer, E. (1990). On the alleged independence of variance components and group differences. *Eur. Bull. Cognit. Psychol.* **10**:686–690.

Turkheimer, E. (1991). Individual and group differences in adoption studies of IQ. *Psychol. Bull.* **110**:392–405.

Edited by N.G. Martin