Behavior Genetics, Vol. 19, No. 1, 1989

Factor and Simplex Models for Repeated Measures: Application to Two Psychomotor Measures of Alcohol Sensitivity in Twins

D. I. Boomsma,[1] N. G. Martin,[2] and P. C. M. Molenaar[3]

As part of a larger study, data on arithmetic computation and motor coordination were obtained from 206 twin pairs. The twins were measured once before and three times after ingesting a standard dose of alcohol. Previous analyses ignored the time-series structure of these data. Here we illustrate the application of simplex models for the genetic analysis of covariance structures in a repeated-measures design and compare the results with factor models for the two psychomotor measures. We then present a bivariate analysis incorporating simplex processes common and specific to the two measures. Our analyses confirm the notion that there is genetic variation affecting psychomotor performance which is "switched on" in the presence of alcohol. We compare the merits of analysis of mean products versus covariance matrices and confront some practical problems that may arise in situations where the number of subjects is relatively small and where the causal structure among the latent variables places a heavy demand on the data.

KEY WORDS: repeated measures; simplex models; twins; alcohol; psychomotor sensitivity; LISREL.

This work was aided by NATO Grant 86/0823 and grants from the Belgian National Research Fund, the State University of Gent, and the Catholic University of Leuven. The authors would very much like to thank Drs. R. Derom and R. Vlietinck for making possible the Twin Data Analysis Workshop. The collection of data was supported by a grant from the Australian Associated Brewers to N. G. Martin, J. G. Oakeshott, J. B. Gibson, and G. A. Starmer.

[1] Free University, Department of Experimental Psychology. De Boelelaan 1115, 1081 HV Amsterdam, The Netherlands.

[2] Queensland Institute of Medical Research, Bramston Terrace, Brisbane, Queensland 4006, Australia.

[3] University of Amsterdam, Department of Psychology, Weesperplein 8, 1018 XA Amsterdam. The Netherlands.

0001-8244/89/0100-0079$06.00/0

INTRODUCTION

Martin *et al.* (1985) obtained data on psychomotor performance and physiological responses from 206 twin pairs, who were measured once before and three times after ingesting a standard dose of alcohol. Their aim was to see whether there was evidence for genetic variation in psychomotor sensitivity to alcohol. To do this they fitted factor models with and without a genetic factor loading only on the three alcohol measurements. A significant increase in likelihood with this factor was taken as evidence for alcohol specific gene action, and this was detected for most of the measurements in their psychomotor battery.

They ignored, however, the temporal structure of their data, which are better suited to a time-series approach. It has recently been shown how simplex models for time-dependent processes may be incorporated in a genetic and environmental model for repeated measures in twin data (Eaves *et al.*, 1986; Boomsma and Molenaar, 1987; Hewitt *et al.*, 1988). Here we apply these models to the data of Martin *et al.* (1985) and show how a more elegant mathematical model can illuminate the physiological processes at work.

Because most psychomotor variables were independent of each other, Martin *et al.* (1985) reported only univariate analyses for these measures. We selected two variables, arithmetic computation (AKT) and motor coordination (VDA), because of their moderate correlation. These variables were first analyzed separately with factor analysis and simplex models and a hybrid model of factor and simplex structures. Then we performed a bivariate analysis of the four measures for each task in which we modeled simplex processes common and specific to arithmetic computation and motor coordination. Martin *et al.* presented extensive factor analyses of these data, so in this paper we restrict ourselves to the formal presentation of univariate and bivariate simplex models.

Some numerical problems that were encountered during minimization are discussed. These numerical problems formally are small sample problems but are, in most cases, more a problem of a large system than of a small number of observations (Theil and Laitinen, 1980). In our applications numerical problems arose when covariance matrices of Twin 1 and Twin 2 were used as input for LISREL instead of matrices of mean cross products and, more specifically, when a simplex model was fitted to these matrices. We discuss how to cope with these problems and the merits of alternative data summaries.

DATA

Twins aged 18–34 years (mean, 23 years) were trained to plateau on psychomotor apparatus, measured when sober and 1, 2, and 3 h after

drinking a standard dose of alcohol (0.75 g EtOH/kg body weight). There were 43 monozygotic (MZ) female, 42 MZ male, 44 dizygotic (DZ) female, 38 DZ male, and 39 unlike-sex twin pairs. Here we analyze the data from two of the psychomotor tasks: (i) Arbeit und Konzentration Testgrate (AKT)—the number of correct addition and subtraction computations completed in 2 min; and (ii) Vienna determination apparatus (VDA)—the number of correct responses within 1 s to 100 presentations of visual and auditory stimuli to which the subject had to give specific button or foot-pedal responses. Details of subjects and measurements are given by Martin *et al.* (1985).

MODELS

The LISREL model consists of two parts: the measurement model and the structural equation model (Jöreskog and Sörbom, 1986). The measurement model describes how latent variables are related to observed variables and can be conceived of as a confirmatory factor analysis model. In our applications to twin data we usually employ for an observed variable y, with latent variables G (genotype). E (individual environment), and C (shared environment) and measurement error ϵ, the following measurement model: $y = hG + eE + cC + \epsilon$.

In the univariate case h, e, and c are loadings of observed variables on latent factors, while in the multivariate case h, e, and c become vectors of factor loadings on common factors and ϵ becomes a vector of measurement errors. Depending on the number of observed variables, one or more common and unique factors can be specified that may be independent or correlated.

To define the units of measurement in the latent variables h, e, and c can be fixed at 1 (a one in each column of the matrix of factor loadings, Λ, in the multivariate case), so that the scale of measurement is the same as in the observed variables and the variance of the latent factors is to be estimated. Alternatively, the latent factors can be standardized to have unit variance and the factor loadings estimated.

The second part of the LISREL model is the structural equation model that causally relates latent variables to other latent variables. One example of such a structural model is the simplex model (Boomsma and Molenaar, 1987); another example is the parent–offspring model of Eaves *et al.* (1989). A simplex structure defines an autoregressive model where latent variables at time i are causally related to latent variables at time $i - 1$. For G, for example (similar structures can be specified for E and C),

$$G_i = \beta_i G_{i-1} + \zeta_i,$$

where G_i is the latent genotype at time i ($i > 0$). β_i is the regression of the latent factor on the previous latent genotype, and ζ represents a random input term (innovation) that drives the genetic process and that is uncorrelated with G_{i-1}. An important conceptual distinction is thus made between innovations of latent factors and measurement errors of observed variables. The innovations are that part of the latent factor at time i that is not caused by the latent factor at time $i - 1$ but is part of every subsequent time point $i + 1$. The ϵ terms are random errors of measurement that do not influence subsequent observed variables.

The parameters of this model are (denoted by their representation in the LISREL program and using y and η variables only)

(1) $\Psi(g)_0 = \text{var}(G_0)$, the variance of the latent factor at time $t = 0$;
(2) $\Psi(g)_i = \text{var}[\zeta(g)_i]$, the variances of the residuals or innovations at time $t > 0$;
(3) β_i, the regression of the latent factor at time i on time $i - 1$; and
(4) Θ_i, the variances of the measurement errors.

The above specifications imply that the variances of the innovations are estimated in Ψ and that the loadings of the observed variables on the latent genetic factors are fixed at 1, so that the measurement scale of the latent variables is the same as that of the observed variables. With this model the Λ matrix can be used to specify the genetic weights for MZ and DZ between and within matrices of mean cross products (Boomsma and Molenaar, 1987). Ψ and Θ are both free diagonal matrices, and the paths from latent variables at time i to time $i + 1$ are specified in B.

When covariance matrices of Twin 1 and Twin 2 are used as input matrices instead of mean cross-product matrices, the structural model of course stays the same, but the estimation using LISREL must be different. In this case the variances of the latent factors cannot be estimated in Ψ, because Ψ needs to be used to specify correlations between latent factors for Twin 1 and those for Twin 2. The scale of measurement of the latent factors is now defined by standardizing their innovations at unit variance. Only the variances of the innovations can be specified, since the covariance matrix of η factors is not a free parameter matrix in LISREL. This does not apply to the first latent factor that is itself standardized, because it has to be conceived of as an innovation (at the initial measurement occasion the first factor cannot be explained by factors associated with an earlier point in time). Factor loadings of observed on latent variables are estimated in Λ. These loadings in Λ correspond to the square roots of the Ψ variances. The estimates in B, however, have to be conceived of as scaled regression coefficients. Hence their absolute values have to

be interpreted with care. As shown below, these estimates will look quite different from those in the first estimation procedure, without, of course, affecting the goodness of fit.

ANALYSIS

Two data summaries are commonly employed and we compare their merits in this paper. If there are g groups and v variables, then we calculate either $2g$ $v \times v$ mean products matrices of g $2v \times 2v$ covariance matrices. Most work to date, including the original analysis of these data by Martin *et al.* (1985), has employed mean products matrices. This has the advantage that input and parameter matrices are smaller, and this can be important when using PC-LISREL, which has size restrictions.[4] On the other hand, it is easier conceptually to specify models for covariance matrices and these may be computed from raw data by PRELIS or using the RA card in LISREL. Nevertheless, it is possible to run into severe numerical problems with the covariance matrix formulation, as we shall see below.

Between-pair and within-pair mean products matrices were computed for all five twin groups. Each 4×4 matrix has 10 unique statistics, providing a total of $10 \times 10 = 100$ df. To the AKT data we fitted factor models, simplex models, and a hybrid model of factor and simplex structures. These models were also fitted to the same data summarized as covariance matrices of Twin 1 and Twin 2. These 8×8 matrices have 36 unique statistics, so for this analysis we have $5 \times 36 = 180$ df. Factor and hybrid models were also fitted to mean product matrices for VDA data. For the bivariate analysis, models were fitted to mean product matrices of the four AKT and four VDA measures. There were thus ten 8×8 matrices, providing a total of 360 df for the bivariate analysis.

RESULTS

Arithmetic Computation (AKT)

Table I shows some results of model fitting to the AKT data using mean product matrices as input. The first factor model (a) consists of a general genetic and a general environmental factor with unique environmental factors. Which LISREL matrix was used for estimation is indicated in Table I. In this case all factor loadings were estimated in Λ and Ψ was standardized and used for the weighting of the genetic and envi-

[4] All analyses reported in this paper were carried out with PC-LISREL on a personal computer (AT).

Table I. Factor, Simplex, and Hybrid Models Fitted to Mean Product Matrices for Arithmetic Computation (AKT) in All Five Twin Groups[a]

(a) Basic factor model (from Martin *et al., 1985)*

E		*G*,
General Λ	Unique Λ	General Λ
6.40	7.76	15.12
6.16	6.40	15.15
6.34	5.43	15.11
7.47	5.91	14.80

$\chi^2(88) = 96.33, p = .255$

(b) a plus alcohol genetic factor (from Martin *et al.*, 1985)

E		*G*	
General Λ	Unique Λ	General Λ	Alcohol Λ
6.57	6.37	15.75	—
6.19	6.34	14.50	4.43
6.34	5.38	14.44	4.49
7.40	6.00	14.31	3.81

$\chi^2(85) = 85.08, p = .477$

(c) Simplex models for *E* and *G*

E, simplex		*G*, simplex	
Ψ	β	Ψ	β
62.52	—	278.51	—
65.18	.384	8.21	.916
47.26	.474	.0[b]	1.000
46.50	.457	.0[b]	1.035

$\chi^2(88) = 148.52, p = .000$

(d) c plus unique *E* variance

E, simplex		*E*, unique	*G*, simplex	
Ψ	β	Θ	Ψ	β
54.43	—	32.92	243.73	—
4.53	.776	32.92	19.29	.943
.0[b]	1.035	32.92	3.58	.971
.0[b]	1.202	32.92	2.32	.978

$\chi^2(87) = 88.35, p = .440$

(e) Hybrid model factor structure for *E* and simplex for *G*

E		*G* simplex	
General Λ	Unique Ψ	Ψ	β
6.78	40.19	244.94	—
6.14	39.18	18.32	.933
6.22	29.18	2.20	.991
7.61	29.58	4.73	.971

$\chi^2(85) = 84.26, p = .502$

[a] The four variables are measurements before ingesting alcohol and 1, 2, and 3 h after ingestion.
[b] Constrained at zero.

ronmental factors. Λ thus is a 4 × 6 matrix, consisting of factor loadings of the four observed variables on the general genetic and environmental factors and the four unique environmental factors associated with each variable. Ψ is the 6 × 6 correlation matrix of these latent factors and its diagonal elements are used to specify the weighting of the genetic and environmental factors:

MZB: Ψ = dia [2 1 1 1 1 1],
MZW: Ψ = dia [0 1 1 1 1 1],
DZB: Ψ = dia [1.5 1 1 1 1 1],
DZW: Ψ = dia [0.5 1 1 1 1 1].

Martin *et al.* added to this model a second genetic factor. This model (b), with two general genetic factors, is identified because the first loading on the genetic alcohol factor is constrained at zero. Martin *et al.* (1985) interpreted the significant likelihood-ratio chi-square (11.25 for 3 df) from the comparison of these two models as evidence for alcohol-specific gene action. However, the constraint on the first loading of the "alcohol genetic factor" could also be applied to any of the other loadings (i.e., at any other time) without changing the fit of the model, although the interpretation of these other models would be complicated.

No such indeterminacy or rotation problems arise with a simplex analysis, because than a natural causal structure among the latent variables obtains.

Model c, with a simplex structure for both *G* and *E*, does not fit the data, however. This is because no allowance has been made for measurement errors or unique E factors. If we allow for measurement errors in Θ_ϵ (model d) which is the variance/covariance matrix of measurement errors, we get a proper fit. Error variances were constrained to be equal across measurements, because there is an indeterminacy associated with the "outer" variables in a simplex model (in this case times 1 and 4). Because preceding (at time 1) and subsequent (at time 4) observations are lacking for the outer variables, additional constraints are needed for identification of parameters (Jöreskog and Sörbom, 1986, p. III.74). In a repeated-measures design the most natural way to eliminate these indeterminacies is to constrain error variances to be equal (although it would be sufficient to set $\theta_1 = \theta_2$ and $\theta_3 = \theta_4$). For both simplex models some parameters were constrained at zero because the unconstrained models gave small, negative variance estimates.

The zero residuals for the environmental simplex structure indicate that *E* tends toward a common factor structure. Table I (e) shows the result of fitting a hybrid model where the *E* simplex structure has been replaced by a general environmental factor. Unique environmental variance for each variable and a simplex structure for the genetic influences are also specified. This last part of the model shows most clearly how the genetic variance after alcohol ingestion is composed of a part that is shared with the latent *G* factor at time 1 and a part that comes into play after the ingestion of alcohol.

Table I shows how the same data can be fitted almost equally well by different models. The reason that a factor model fits these data as well

as a simplex (or a combination of both) is that a factor model can accommodate a simplex structure if the number of repeated observations is small. In fact, with four observed variables the fit of a two-factor model is identical to the fit of a simplex model. Although the present genetic application is more complex, the small number of repeated observations leads to a convergence of the factor model to the simplex. Alternatively, a simplex model may be seen as a factor model with as many correlated factors as observed variables. When the correlation between these factors is one, we have the same factor across all occasions and thus a common-factor model.

Next we tried to repeat the above analyses using covariance matrices of Twin 1 and Twin 2 as input for LISREL but we encountered serious problems during minimization when we tried to fit a simplex or hybrid model. Before discussing the reasons and possible solutions for these problems, we first illustrate for female twin pairs the use of covariance matrix input and discuss differences in estimates from those obtained in the mean product formulation of the problem.

Table II shows the LISREL estimates (a) for the hybrid model fitted to input matrices of mean cross products between and within female MZ and DZ twin pairs and (b) for the same data input as covariance matrices of Twin 1 and Twin 2. It is clear that the squared factor loadings of b are the same as the variances of the latent factors in a. Notice, however, that estimates in Beta are highly dissimilar, whereas the standardized estimates are the same in a and b. Both estimation procedures must, of course, lead to the same conclusions, and the explanation for this difference is as follows. For both formulations the total genetic variance at time point i, $i > 0$, is computed as $\text{var}(G_i) = \lambda_i^2([\beta_i^2 \text{ var}(G_{i-1}) + \text{var}(\zeta_i)]$. In formulation a, λ is fixed at one, so this expression reduces to $\text{var}(G_i) = \beta_i^2 \text{ var}(G_{i-1}) + \psi_i$, where ψ_i is the variance of the innovations. This leads to total genetic variances $\text{var}(G_1) = 227.25$, $\text{var}(G_2) = 0.915^2 * 227.25 + 18.03 = 208.30$, $\text{var}(G_3) = 0.972^2 * 208.30 + 3.34 = 200.14$, and $\text{var}(G_4) = 1.022^2 * 200.14 + 4.64 = 213.68$.

For formulation b, the total genetic variance is computed as $\text{var}(G_i) = \lambda_i^2[\beta_i^2 \text{ var}(G_{i-1}) + 1]$, since the variance of the innovations was standardized; the term in parentheses represents the variance of G at time i. So we get $\text{var}(G_1) = 15.08^2 = 227.41$ and $\text{var}(G_2) = 4.21^2 (3.27^2 * 1 + 1) = 17.72 * 11.69 = 207.24$, because the variance of the first latent genetic factor was 1. The genetic variance at the third and fourth measurement points is now $\text{var}(G_3) = 1.71^2 (2.41^2 * 11.69 + 1) = 2.92 * 68.89 = 201.15$ and $\text{var}(G_4) = 2.20^2 (0.79^2 * 68.89 + 1) = 4.84 * 43.99 = 212.96$.

When the hybrid model was fitted to the male covariance matrices and to all five twin groups, we found it impossible to obtain a proper fit,

Table II. Comparison of Hybrid Model Fitted to Mean Product Versus Covariance Matrices of Arithmetic Computation (AKT) Data

(a) Fitted to two 4 × 4 mean product matrices (females)

E		G, simplex		
General Λ	Unique Ψ	Ψ	β	β (standardized)
4.57	44.08	227.25	—	—
6.29	36.53	18.03	.915	.956
6.80	31.53	3.34	.972	.992
5.70	29.85	4.64	1.022	.989

$\chi^2(25) = 17.54, p = .861$

(b) Fitted to two 8 × 8 covariance matrices (females)

E		G, simplex		
General Λ	Unique Λ	Λ	β	β (standardized)
4.43	6.65	15.08	—	—
6.25	6.08	4.21	3.27	.956
6.75	5.66	1.71	2.41	.993
5.84	5.47	2.20	.79	.989

$\chi^2(57) = 46.48, p = .893$

(c) Fitted to 8 × 8 covariance matrices for all 5 twin groups using robustification technique as described in text

E		G, simplex		
General Λ	Unique Λ	Λ	β	β (standardized)
3.90	7.10	16.06	—	—
5.45	6.23	4.04	3.70	.965
6.28	5.27	1.52	2.61	.995
6.06	5.80	1.92	.80	.992

$\chi^2(165) = 183.81, p = .150$

as unreasonable parameter estimates were obtained after 250 iterations. Since we already know from Table I that the hybrid model will fit these data when they are summarized as mean product matrices, this state of affairs can be due only to numerical problems with the optimization of the likelihood function when the data are summarized as covariance ma-

trices. One obvious difference is that the dimensions of covariance matrices are double those of the corresponding mean product matrices. More specifically, the following problems can be discerned.

(1) Covariance matrices associated with highly intercorrelated repeated measures can become nearly singular. This means that the smallest eigenvalue divided by the average of all eigenvalues is almost zero. If the ratio in question is too small (where the bounds are determined by the precision of the floating-point representation of reals during computation), then numerical optimization of the likelihood is ill conditioned. One way out of this problem is to robustify the estimation of covariances (cf. Jöreskog and Sörbom, 1986, p. I29), preferably by invoking a robust estimator that is specially devised for the employment in structural analysis of covariance (e.g., Theil and Laitinen, 1980). Another, more direct way to tackle this problem is to perturb the smallest eigenvalue(s) slightly away from zero and then reconstruct the covariance matrix (Boomsma *et al.*, 1989). However, even if one of these procedures leads to better-conditioned numerical optimization, one would still expect the chi-square statistic to be positively biased (in contrast to parameter estimates, which are generally unbiased). The reason is that the chi-square statistic depends on the minimum of the fitting function F, where F is defined by

$$F = \ln | \Sigma | + \text{trace}[S\Sigma^{-1}] - \ln | S | - p,$$

where S and Σ denote the observed covariance matrix and its expectation, respectively. Roughly speaking, the determinant $| \Sigma |$ depends on the product of the largest eigenvalues of S, whereas the determinant $| S |$ equals the product of all eigenvalues of S. Hence, if S has been only slightly perturbed away from singularity (the mentioned procedures should not distort the available information too much), the difference in $\ln | \Sigma | - \ln | S |$ is large, as is the resulting chi-square statistic.

(2) Even if there are no singularity problems, the doubling of dimensions inherent in a covariance analysis entails the addition to the likelihood function of several nuisance parameters which relate to the population variances of the observed variables (in an analysis of mean cross-products matrices, the latter are assumed to be equal within twin pairs). This basically nonessential increase in the dimensionality of parameter space can give rise to a serious flattening of the likelihood function (Box and Tiao, 1973), thus leading to indeterminacies in the computation of local gradients during the iterative search for its maximum value. The situation here resembles the occurrence of collinearity in regression analysis, which usually is counteracted by the invocation of ridge regression (e.g., Mardia *et al.*, 1982, pp. 253–254). Loosely speaking, in ridge regression the projection onto the parameter space is brought into sharper focus (at the cost

of slightly biasing the obtained estimates) by means of a small perturbation of the diagonal of the projection operator. A similar approach can be used in LISREL by (i) the addition of a small positive constant to the diagonal of the observed covariance matrix and (ii) the model correction for this perturbation by fixing the diagonal of Θ at the same positive constant; this is the approach we have taken.

In the present application the covariance matrices do not appear to be near-singular, so we tried to counteract the numerical problems by increasing the variances for all groups by 2 (i.e., adding 2 to the diagonal elements of all input covariance matrices). At the same time we specified Θ as a diagonal matrix with variances fixed at 2, in order to correct for the perturbation in the data matrices. Using Θ in this way to ensure a numerical solution means that measurement errors have to be accounted for differently; the unique variances normally estimated in the diagonal Θ matrix can be estimated instead as loadings on unique Λ factors. The estimates obtained by this method are shown in Table II (c). In contrast to the alternative estimates from the female data (a and b), these estimates for all five twin groups differ considerably from their counterparts in Table I (e) even after one has taken squares or square roots to convert appropriately.

The difference is most likely caused by the DZ male data, in which the differences in variances between Twin 1 and Twin 2 were largest and where, in addition, these variances were also larger than in any of the other groups. Fitting the hybrid model to all groups except the DZ males resulted in a χ^2 of 128.5 (df = 129, p = .496) and estimates that resembled those in Table I (e) much more closely.

VDA DATA

To avoid the problems encountered above we restrict ourselves to input of matrices of mean cross products in the analysis of the motor coordination data for all five twin groups. These data showed a marked increase in variance after alcohol ingestion. Table III shows (a) the final factor model of Martin *et al.* (1985) and (b) a model where the common genetic factors have been replaced by correlated genetic influences across time. Although both models seem to fit the data equally well, they offer different interpretations for the increase in variance. According to the factor model the increase in variance at time 2 (1 h after ingesting alcohol) is mainly environmental, while the simplex model indicates that all of the increase is genetic and is accompanied by a decrease in shared environmental variance. The estimate of 1.248 for β implies that part of this increase is due to amplification of genetic variance at the first measure-

Table III. Factor and Hybrid Models Fitted to Motor Coordination (VDA) Data Mean Product Matrices for All Five Twin Groups

(a) Factor model of Martin *et al.* (1985)					Total variance	
E		*C*,	*G*			
General Λ	Unique Λ	General Λ	General Λ	Alcohol Λ	Total genetic variance	Total variance
6.45	9.23	9.01	13.96	—	194.88	402.86
10.31	9.50	10.96	12.97	6.48	210.21	526.88
12.28	5.32	7.91	15.74	6.19	286.06	527.73
9.52	7.70	6.93	15.81	4.72	272.23	470.18
$\chi^2(81) = 79.40, p = .53$						
(b) Hybrid model with genetic simplex process					Total variance	
E		*C*,	*G*, simplex			
General Λ	Unique Ψ	General Λ	Ψ	β	Total genetic variance	Total variance
6.88	78.47	10.07	176.03	—	176.03	403.08
9.70	79.71	6.05	38.55	1.248	312.71	523.10
12.12	26.32	8.61	19.69	.910	278.64	526.00
9.49	59.79	9.02	2.66	.919	237.99	469.20
$\chi^2(81) = 77.11, p = .602$						

ment occasion, while a smaller part (38.55, or 12%) is caused by the innovation term. For the other time points, more or less the same conclusions for the division of the total variances follow from the factor and the simplex model.

BIVARIATE MODEL

The AKT and VDA variables are correlated about .3 at each measurement occasion. Since genetic simplex processes underlie the correlation of successive measurements for both AKT and VDA, we wanted to see the extent to which the same simplex process underlies the genetic continuity of the two variables. A bivariate model that specifies a simplex process common to both AKT and VDA is depicted in Fig. 1. In addition to a common underlying genetic time series, where both VDA and AKT load on the genotype specific to each time point, separate time series are specified for both variables. There are two separate individual environmental factors, one on which the four AKT measures load and a second

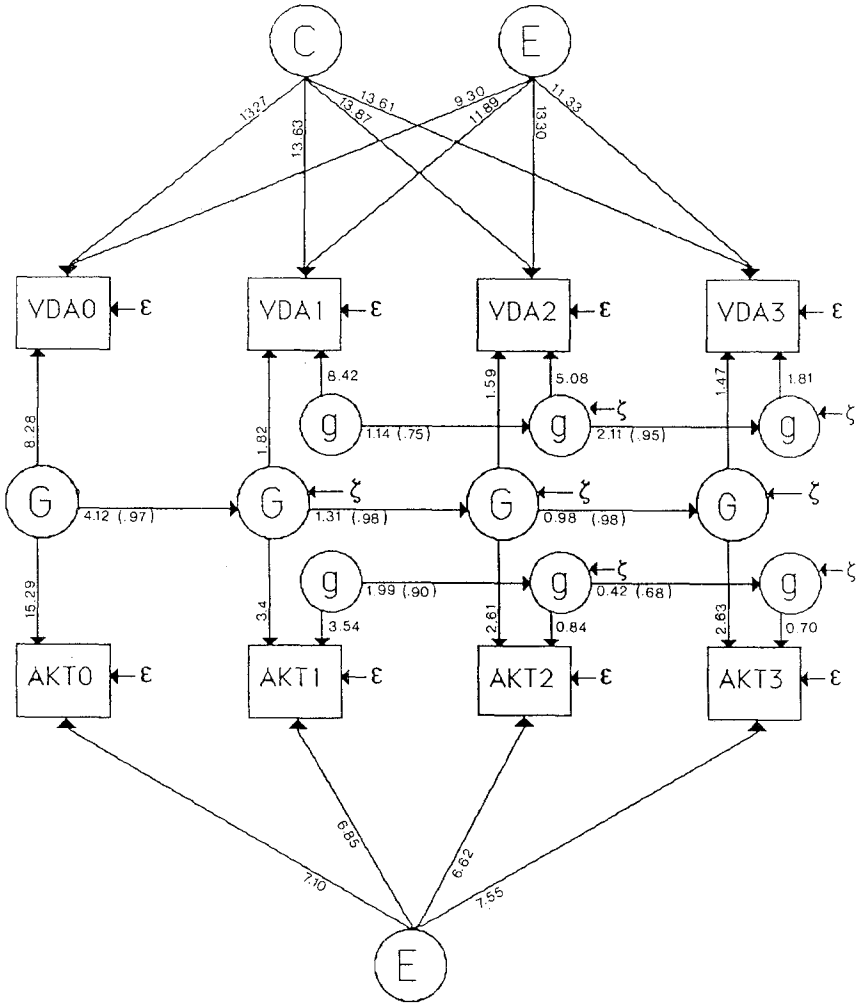

Fig. 1. Bivariate model and LISREL estimates for VDA and AKT data. At each measurement occasion both variables load on a common genetic series (G), on a genetic series (g) that is specific to each variable, and on general individual environmental factors (E) that are specific to each variable. VDA also has loadings on a general common environmental factor (C). ζ's are the innovations of the genetic process that are standardized to have unit variance. The ϵ's are measurement errors/unique environmental influences. Variances of $\epsilon_{0.1.2.3}$ were 78, 75, 26, and 58 for VDA and 41, 31, 28, and 27 for AKT. Values of transmission parameters in parentheses are standardized regression coefficients.

one for VDA. The four VDA measures also have loadings on a common environmental factor. The correlation between VDA and AKT at each point in time thus is explained by their loadings on the same genetic series and the genetic variance that is not shared between them by the unique genetic series. For the identification of the model these unique series have to start at the second measuring point, because their effects cannot be disentangled from the common series at the first occasion. Although each autoregressive process in the model is assumed to occur at each occasion, the mere fact that measuring has to start at a particular point in time precludes the complete identification of all these processes.

The LISREL set up for this analysis is given in the Appendix (Fig. A1). In Λ are the loadings on the latent genetic and environmental factors. Ψ is a diagonal matrix that contains the variances of the three environmental factors and, as the genetic factors are correlated across time, the variances of the genetic innovations. In this analysis Ψ was standardized and used for the weighting of the latent factors. In Θ is that part of the environmental variance that is unique to each variable as well as to each measuring occasion (and is also uncorrelated within twin pairs). The B matrix, finally, gives the estimates of the influence of latent genotypes on subsequent latent genotypes for the common genetic and the two unique genetic series. As Ψ was used to standardize the variances of the genetic innovations, the unstandardized estimates of the regression coefficients again have to be interpreted as scaled estimates (see above).

In Fig. 1 are the estimates of the transmission coefficients (with their standardized values in parentheses) and estimates of the factor loadings. To obtain the total amount of environmental variance at each occasion, these factor loadings are squared and added to the unique environmental variances estimated in Θ. The genetic variance at $i > 0$ (for the common genetic process) and at $i > 1$ (for the unique genetic processes) is again computed as $\text{var}(G_i) = \lambda_i^2[\beta_i^2 \text{ var}(G_{i-1}) + 1]$. The standardized β's for the genetic process that is common to both VDA and AKT approach unity and the innovations of this process are small. In contrast, the β's of the unique series go up for VDA and go down for AKT. The chi-square for this model was 351 with 319 df ($p = .1$).

DISCUSSION

The examples given above demonstrate how LISREL can be used to represent and compare different models for the same data set. These different LISREL models can be employed in the genetic analysis of repeated-measures or longitudinal data. When analyzing longitudinal data that are not as close together in time as the repeated observations that

were analyzed in this study, it is likely that factor models will no longer fit the data very well and that an autoregressive model is needed for an adequate representation. More importantly, also when the number of repeated observations increases, factor models will not fit this type of data because a factor model does not recognize the time-dependent structure of the data, wheras a simplex model does. In general, with a small number of observations that are positioned closely together, there is not enough time for the process to develop.

As we saw in the analysis of the VDA data, factor and simplex models can offer radically different interpretations, while fitting the data equally well. Therefore, criteria other than goodness of fit have to be considered when choosing between these models. As a simplex model specifies a time-dependent structure, an explanation of the increase in total variance after alcohol ingestion by an increase in the genetic variance seems to be the more likely alternative.

Our last analysis shows how, for multivariate time series, more elaborate models can be fitted. As most longitudinal studies usually collect multivariate data sets, this kind of analysis of developmental patterns seems worthwhile to consider.

APPENDIX

```
MZBETWEEN FEMALES VDA & AKT C=FACTOR E1=FACTOR + SPECIFIC, G=SIMPLEX
DA NG=10 NI=8 NO=42 MA=CM
LA
*
'VDA0' 'VDA1' 'VDA2' 'VDA3' 'AKT0' 'AKT1' 'AKT2' 'AKT3'
CM SY
<DATA>
MO NY=8 NE=13 PS=DI,FI LY=FU,FI BE=FU,FI TE=DI,FR
LE
*
'G1BOTH' 'G2BOTH' 'G3BOTH' 'G4BOTH' 'CVDA' 'EVDA' 'EAKT'
'G2VDA'  'G3VDA'  'G4VDA'  'G2AKT'  'G3AKT' 'G4AKT'
ST 2.0 PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(8,8) PS(9,9)
ST 2.0 PS(10,10) PS(11,11) PS(12,12) PS(13,13)
ST 2.0 PS(5,5)
ST 1.0 PS(6,6) PS(7,7)
FR BE(2,1) BE(3,2)  BE(4,3)
FR BE(9,8) BE(10,9) BE(12,11) BE(13,12)
FR LY(1,1) LY(2,2) LY(3,3) LY(4,4)
FR LY(5,1) LY(6,2) LY(7,3) LY(8,4)
FR LY(1,5) LY(2,5) LY(3,5) LY(4,5)
FR LY(1,6) LY(2,6) LY(3,6) LY(4,6)
FR LY(5,7) LY(6,7) LY(7,7) LY(8,7)
FR LY(2,8) LY(3,9) LY(4,10)
FR LY(6,11) LY(7,12) LY(8,13)
ST 9.0 ALL
ST .9 BE(2,1) BE(3,2) BE(4,3)
ST .9 BE(9,8) BE(10,9) BE(12,11) BE(13,12)
OU NS SS RS TM=5000*
GROUP2 MZWITHIN FEMALES
DA NO=43
CM SY
<DATA>
MO NY=8 NE=13 LY=IN PS=DI,FI BE=IN TE=IN
ST 1.0 PS(6,6) PS(7,7)
OU
GROUP3 MZBETWEEN MALES
DA NO=41
CM SY
<DATA>
MO NY=8 NE=13 LY=IN PS=DI,FI BE=IN TE=IN
ST 2.0 PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(8,8) PS(9,9)
ST 2.0 PS(10,10) PS(11,11) PS(12,12) PS(13,13)
ST 2.0 PS(5,5)
ST 1.0 PS(6,6) PS(7,7)
OU
GROUP4 MZWITHIN MALES
DA NO=42
CM SY
<DATA>
MO NY=8 NE=13 LY=IN PS=DI,FI BE=IN TE=IN
ST 1.0 PS(6,6) PS(7,7)
OU
GROUP5 DZBETWEEN FEMALES
DA NO=43
CM SY
<DATA>
MO NY=8 NE=13 LY=IN PS=DI,FI BE=IN TE=IN
ST 1.5 PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(8,8) PS(9,9)
ST 1.5 PS(10,10) PS(11,11) PS(12,12) PS(13,13)
ST 2.0 PS(5,5)
ST 1.0 PS(6,6) PS(7,7)
OU
```

Fig. A1. Appendix. Bivariate model.

```
GROUP6 DZWITHIN FEMALES
DA NO=44
CM SY
<DATA>
MO NY=8 NE=13 LY=IN PS=DI,FI BE=IN TE=IN
ST 0.5 PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(8,8) PS(9,9)
ST 0.5 PS(10,10) PS(11,11) PS(12,12) PS(13,13)
ST 1.0 PS(6,6) PS(7,7)
OU
GROUP7 DZBETWEEN MALES
DA NO=37
CM SY
<DATA>
MO NY=8 NE=13 LY=IN PS=DI,FI BE=IN TE=IN
ST 1.5 PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(8,8) PS(9,9)
ST 1.5 PS(10,10) PS(11,11) PS(12,12) PS(13,13)
ST 2.0 PS(5,5)
ST 1.0 PS(6,6) PS(7,7)
OU
GROUP8 DZWITHIN MALES
DA NO=38
CM SY
<DATA>
MO NY=8 NE=13 LY=IN PS=DI,FI BE=IN TE=IN
ST 0.5 PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(8,8) PS(9,9)
ST 0.5 PS(10,10) PS(11,11) PS(12,12) PS(13,13)
ST 1.0 PS(6,6) PS(7,7)
OU
GROUP9 DZBETWEEN OP-SEX
DA NO=38
CM SY
<DATA>
MO NY=8 NE=13 LY=IN PS=DI,FI BE=IN TE=IN
ST 1.5 PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(8,8) PS(9,9)
ST 1.5 PS(10,10) PS(11,11) PS(12,12) PS(13,13)
ST 2.0 PS(5,5)
ST 1.0 PS(6,6) PS(7,7)
OU
GROUP10 DZWITHIN OP-SEX
DA NO=38
CM SY
<DATA>
MO NY=8 NE=13 LY=IN PS=DI,FI BE=IN TE=IN
ST 0.5 PS(1,1) PS(2,2) PS(3,3) PS(4,4) PS(8,8) PS(9,9)
ST 0.5 PS(10,10) PS(11,11) PS(12,12) PS(13,13)
ST 1.0 PS(6,6) PS(7,7)
OU

*Longer times lead to drifting solutions,
 because of finite arithmetic precision
```

Fig. A1. (*Continued*)

REFERENCES

Boomsma, D. I., van den Bree, M. B., Orlebeke, J. F. and Molenaar, P. C. M. (1989). Resemblances of parents and twins in sports participation and heart rate. *Behav. Genet.* **19**:123–141.

Boomsma, D. I., and Molenaar, P. C. M. (1987). The genetic analysis of repeated measures. I. Simplex models. *Behav. Genet.* **17**:111–123.

Box, G. E. P., and Tiao, G. C. (1973). *Bayesian Inference in Statistical Research,* Addison–Wesley, Reading, Mass.

Eaves, L. J., Fulker, D., and Heath, A. (1989). The effects of social homogamy and cultural inheritance on the covariances of twins and their parents. A LISREL model. *Behav. Genet.* **19:**113–122.

Eaves, L. J., Long, J., and Heath, A. C. (1986). A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behav. Genet.* **16:**143–162.

Hewitt, J. K., Eaves, L. J., and Neale, M. C. (1988). Resolving causes of developmental continuity or "tracking": Longitudinal twin studies during growth. *Behav. Genet.* **18:**135–151.

Jöreskog, K. G., and Sörbom, D. (1986). *LISREL: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood,* National Educational Resources, Chicago.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1982). *Multivariate Analysis,* Academic Press, London.

Martin, N. G., Oakeshott, J. G., Gibson, J. B., Starmer, G. A., Perl, J., and Wilks, A. V. (1985). A twin study of psychomotor and physiological responses to an acute dose of alcohol. *Behav. Genet.* **15:**305–347.

Theil, H., and Laitinen, K. (1980). Singular moment matrices in applied econometrics. In Krishnaiah, P. R., (ed.), *Multivariate Analysis V,* North-Holland, Amsterdam, pp. 629–649.