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

Testing Structural Equation Models for Twin Data Using LISREL

A. C. Heath,[1] M. C. Neale,[1] J. K. Hewitt,[1] L. J. Eaves,[1] and D. W. Fulker[2]

Simple genetic models can be fitted to twin data using software packages such as LISREL (Jöreskog and Sörbom, 1986a). After discussion of data preparation and routine checks on possible violation of assumptions of the twin method, we illustrate univariate, bivariate, and multivariate genetic models which can be tested in cross-sectional twin data using LISREL. These include models for cohort or cohabitation effects, genotype × sex interaction, and certain types of genotype × environment interaction and genotype–environment correlation.

KEY WORDS: twin data analysis; LISREL; structural modeling.

INTRODUCTION

During exploratory analyses of twin data, simple genetic and environmental models can be tested using a regression approach (e.g., DeFries and Fulker, 1985). The regression approach breaks down, however, once we wish to incorporate multiple (possibly reciprocally interacting) dependent variables, or multiple family relationships, in the models. Software packages for structural equation modeling such as LISREL (Jöreskog and Sörbom, 1986a) make it possible to fit such models to summary covariance or correlation matrices by maximum-likelihood or other meth-

This research was supported in part by NIH Grants GM30250, AG04954, HD19802, and HD18426 and by ADAMHA Grants AA06781 and MH40828. This paper is modified from a presentation at a workshop on twin data analysis held at the Catholic University of Leuven, supported by NATO Grant 86/0823 and grants from the Belgian National Research Fund, the State University of Gent, and the Catholic University of Leuven.

[1] Department of Human Genetics, Medical College of Virginia, Box 33, MCV Station, Richmond, Virginia 23298.

[2] Institute for Behavioral Genetics, University of Colorado, Campus Box 447, Boulder, Colorado 80309.

0001-8244/89/0100-0009$06.00/0

ods. Such programs provide a chi-square test of the goodness of fit of a model and give estimates of the model parameters and their standard errors. In this paper we give a basic overview of the stages of model fitting and the types of models that can be tested with twin data and show how model fitting can be performed using LISREL (Jöreskog and Sörbom, 1986a). A subsequent paper gives a brief overview of hypothesis testing using LISREL (Neale *et al.*, 1989b).

TESTING THE "EQUAL-ENVIRONMENTS" ASSUMPTION

Prior to model fitting, a number of routine checks of possible violations of the assumptions of the twin method should be performed. Estimates of genetic and environmental parameters from twin data will be valid only if the trait-relevant environments of monozygotic twin pairs are no more highly correlated than the trait-relevant environments of dizygotic twin pairs, except in so far as the environments are actively created by the twins themselves (e.g., Eaves *et al.*, 1977; Plomin *et al.*, 1977; Scarr and McCartney, 1983; Martin *et al.*, 1986). Monozygotic twins may have experienced more similar environments as children in many respects (e.g., dress, sharing the same friends), but this greater similarity of experience is important *only* if these environmental features are predictive of the trait under study. The greater environmental similarity of monozygotic than dizygotic twin pairs may arise because twins actively create or select their own environments (one form of genotype–environment correlation; see below), and the type of environment created is influenced by traits which are in part genetically determined. It is only when monozygotic twin pairs are *passive recipients* of more similar environmental treatments than are dizygotic twin pairs that potential problems arise.

A simple regression analysis will detect significant associations with the trait under study of "passive" environmental variables (i.e., environmental treatments of which twin pairs are passive recipients). If there is a higher correlation between monozygotic twin pairs than between dizygotic twin pairs for such passive environmental variables, failure to adjust for their effects will inflate estimates of genetic variance. Significant effects of variables such as age, for which monozygotic and dizygotic twin pairs are equally correlated, will inflate the estimate of shared environmental variance. In either case, the effects of the environmental variables may be adjusted for by including them as covariates in a genetic analysis (Neale and Martin, 1989). With most variables (e.g., for such epidemiological risk factors as smoking, drinking, exercise, etc.), there are no a priori grounds for assuming that twins are passive recipients of an environmental feature or risk factor. For such cases a bivariate genetic

analysis will be needed to determine the direction of causation (see Bivariate Models, Causal Pathways, and Genotype–Environment Correlation, below) and adjust for the effects of the epidemiological risk factor where appropriate.

It is not feasible to assess all the possible trait-relevant passive environmental features for which monozygotic twins may be more highly correlated than dizygotic twins. A test for the absence of such effects is still feasible. We would expect greater concordance for such environmental features in twin pairs who are still living in the same household than in twin pairs who are living apart (e.g., Heath *et al.*, 1988a) and, perhaps, also greater concordance in twin pairs who have frequent social contact than in twin pairs who rarely have social contact (e.g., Kaprio *et al.*, 1987). Thus if twin similarity does *not* vary as a function of cohabitation or amount of social contact, this implies that such effects are unimportant. We compute for each twin pair, separately for each monozygotic twin group, an estimate of the within-pair variance [$(T1 - T2)^2$, where T1 is the score of the first twin and T2 the score of the second twin from the pair] and test whether this decreases with increasing amount of social contact.

If a significant association *is* found between twin similarity and amount of social contact, several possible explanations must be considered. If there are mean differences in trait level between pairs having different amounts of social contact, the association may be an artifact arising because of heteroscedasticity, i.e., differences in error variance at different points on the scale of measurement. In such a case, data transformation (see below) should reduce or eliminate the association. It may be that more dissimilar pairs choose not to live together or choose to have less social contact. Finally, it is also possible that pairs who have more frequent social contact experience more similar environments. Monozygotic twins do report more frequent social contact with their cotwin than dizygotic twins (e.g., Kendler *et al.*, 1986). Thus this third explanation would indeed suggest that the environments of monozygotic twin pairs are more highly correlated than the environments of dizygotic twin pairs. Whether more similar twin pairs choose to have more frequent social contact, or more frequent social contact leads to greater similarity, cannot in general be resolved without prospective data (e.g., Kessler and Greenberg, 1981). If it is indeed the case that differences in frequency of social contact cause differences in level of similarity, then these effects must be corrected for in the genetic analysis. A simple way of modeling cohabitation effects is presented below (see Genotype × Sex/Cohort/Environment Interaction). Hopper and Matthews (1983; Hopper and Culross, 1983) and Lange (1986) discuss more sophisticated models for

cohabitation effects, but these require fitting models to raw data and, so, cannot be used with LISREL.

NONRANDOM ASCERTAINMENT

Nonrandom sampling can also cause serious problems in twin data (Martin and Wilson, 1982; Lykken *et al.*, 1987; Neale *et al.*, 1989a). Once again the critical question is whether a twin sample can be regarded as representative of the general population with respect to the variables under study. A sample which underrepresents twins from the lower socioeconomic strata may be entirely suitable for studying dermatoglyphic traits (e.g., Holt, 1968) or personality traits (e.g., Eysenck and Eysenck, 1975), traits which vary little as a function of social class, but may give misleading results for an analysis of genetic and environmental factors in occupational achievement. Nonrandom sampling is likely to be a particularly important problem in volunteer twin studies (Lykken *et al.*, 1978, 1987).

Fortunately, twin data provide a means of detecting the effects of nonrandom sampling, a test which is not available in samples of unrelated individuals. For many variables, including cognitive ability measures and clinical end points (e.g., Cox *et al.*, 1977), individuals falling below a certain cutoff point on a trait have a reduced or zero probability of volunteering to participate in a study. Nonparticipation by either twin will exclude a twin pair from the sample. Thus for those traits for which monozygotic twins are more highly correlated than dizygotic twin pairs, there will be greater loss of dizygotic than monozygotic pairs from the sample. It has been shown that under these conditions such truncate sampling will lead to differential attenuation of the monozygotic and dizygotic twin correlations, with the latter correlation being more highly attenuated than the former (Neale *et al.*, 1989a). Failure to take account of the nonrandom sampling will thus lead to biased estimates of genetic and environmental parameters. However, the nonrandom sampling will also produce differences in mean (and variance) as a function of zygosity (Lykken *et al.*, 1987), differences which would not be predicted under any simple genetic model. Furthermore, there will be mean and variance differences between twins from pairs concordant for response and twins from pairs discordant for participation in the study. Correction for the effects of nonrandom sampling is possible but cannot be achieved using LISREL or similar software. Significant mean differences as a function of zygosity, or between twins from complete and incomplete pairs, are thus a strong indication against proceeding to the model-fitting stage, until the possible consequences of nonrandom sampling have been investigated.

DATA TRANSFORMATION

Both in regression analyses and in model-fitting analyses, parameter estimates are very sensitive to violation of the assumption of homoscedastic error variances (e.g., Eaves *et al.*, 1977). For most behavioral measures, a systematic relationship between mean and variance is found when untransformed scores are used. For measures of alcohol consumption, for example, it is found that error variance is greater for those reporting higher mean levels of consumption (Jardine and Martin, 1984). Such heteroscedasticity can be detected as a significant polynomial regression of intrapair variance on pair mean in monozygotic twin pairs. Usually it can be removed by a simple data transformation which will be suggested by the properties of the measurement scale (Bartlett, 1947). Transformation of raw data can be easily effected using PRELIS, a preprocessor for LISREL (Jöreskog and Sörbom, 1986b).

DATA SUMMARY

Early papers on model-fitting analyses of twin data (e.g., Eaves, 1977; Martin and Eaves, 1977) used mean squares and products, derived from an analysis of variance, as a data summary for each twin group. Such summary statistics can be used with LISREL (e.g., Fulker *et al.*, 1983; Boomsma *et al.*, 1986; Molenaar and Boomsma, 1987). Their use is inappropriate, however, if we wish to test hypotheses about genotype × sex interaction (e.g., Eaves, 1977) or genotype × environment interaction (e.g., Eaves, 1982), where error variances may be expected to vary as a function of sex or environmental exposure, or if we wish to include data on the parents of twins (e.g., Eaves *et al.*, 1978) or the spouses of twins (e.g., Eaves, 1979; Heath and Eaves, 1985) or other relatives. For generality, therefore, we focus on the analysis of covariance matrices, which either may be computed using another statistical package, or may be computed from raw data by LISREL or its preprocessor, PRELIS (Jöreskog and Sörbom, 1986a,b).

If variables are not measured on a continuous scale but consist of two or more categories (i.e., are dichotomous or polychotomous, e.g., because they consist of responses to individual items in a questionnaire), methods used with continuous data cannot be applied. If such data are at least ordinal, it is still possible to estimate polychoric correlations [between two polychotomous variables (e.g., Olsson, 1979)] or polyserial correlations [between a polychotomous and a continuous variable (e.g., Olsson *et al.*, 1982)] using mainframe versions of LISREL or microcomputer versions of PRELIS. Estimation of polychoric or polyserial correlations implies the assumptions that the discontinuous response distri-

bution observed for each polychotomous variable is determined by an underlying continuous normal distribution with thresholds superimposed and that the joint distribution of the underlying latent variables (polychoric correlation) or the underlying latent variable and the observed variable (polyserial correlation) is bivariate normal. The polychoric correlation estimates the correlation between the underlying, normally distributed *latent* variables, not the observed discontinuous variables. Likewise, the polyserial correlation estimates the correlation between the latent variable underlying the polychotomous variable and the continuously distributed observed variable. If the data are at least trichotomous, PRELIS provides a chi-square goodness-of-fit test of this distributional assumption. Provided that chi-square values are nonsignificant, matrices of polychoric, polyserial, and product–moment correlations can be used for data summary. Models can then be fitted to these correlation matrices using LISREL. Since these matrices are often nonpositive definite, model fitting to these matrices by maximum likelihood is not possible. Models are usually fitted by unweighted least squares (Jöreskog and Sörbom, 1986a). Weighted least-squares estimation, using estimates of the variances and covariances of the correlations as weights, has been described by Browne (1984) and is expected to be included as an option in LISREL VII (Jöreskog and Sörbom, 1986b). An alternative approach, in which nonpositive definite matrices are robustified so that the maximum-likelihood procedure can be used, has also been described (Boomsma *et al.*, 1989a; Martin and Boomsma, 1989).

In the analysis of family-structured data, the family (e.g., twin pair, twin pair plus parents, twin pair plus spouses), rather than the individual respondent, becomes the basic unit or "case" for data analysis. Thus if there are k observations per respondent and m individuals per family, there will be $k \times m$ variables per case. In what follows we order the data so that the first k variables for each family are observations on the first twin, the second k variables are observations on the second twin, and so on. Data on twins from unlike-sex pairs must be reordered so that the first k variables are observations on, say, the male twin, and the next k variables observations on the female twin (so that variances and covariances are computed about the sex-appropriate means). In testing for genotype $\times$ environment interaction involving a dichotomous exposure variable (e.g., Eaves, 1982; Heath *et al.*, 1988a), zygosity groups are subdivided into concordant exposed, discordant, and concordant nonexposed groups. Twins from discordant pairs must be reordered so that the first twin is always, say, the exposed twin, and the cotwin the nonexposed twin. Twin pairs may be further subdivided on the basis of sex, in which case unlike-sex discordant pairs must be divided into two groups

according to whether the male or female twin is exposed. Summary covariance and correlation matrices are computed separately for each twin group. If raw data are used as input, LISREL expects one case (i.e., family) per record of data and requires that the data be presorted into twin groups, but PRELIS does not have these restrictions.

THE BASIC LISREL MODEL

The use of LISREL is facilitated by at least a working knowledge of path analysis (e.g., Wright, 1968). In a path diagram (Wright, 1968, p. 299),

> every included variable, measured or hypothetical, is represented by arrows as either completely determined by certain others, which may in turn be represented as similarly determined, or as an ultimate variable. Each ultimate factor in the diagram must be connected by lines with arrowheads at both ends with each of the other ultimate factors to indicate possible correlations through still more remote, unrepresented factors, except in cases in which it can safely be assumed that there is no correlation.

Figure 1 illustrates in the form of a path diagram the full model used by LISREL. (For most genetic applications only a subset of the variables in the LISREL model is used.) Variables *V*1 . . . *V*5 are latent "ultimate" or independent variables (termed KSI variables by LISREL), of which variables *V*2, *V*3, and *V*4 are represented by two-headed arrows as being correlated. Variables *W*1 and *W*2 are latent dependent variables (termed

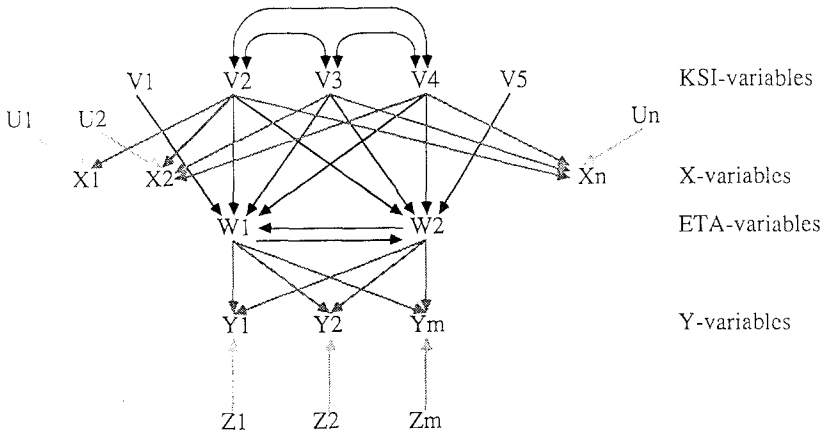

Fig. 1. The basic LISREL model. *V*1 . . . *V*5 denote latent ultimate variables, which are indexed by observed variables *X*1 . . . *Xn*; *W*1 and *W*2 are latent dependent variables, which are indexed by observed variables *Y*1 . . . *Ym*. *U*1 . . . *Un* and *Z*1 . . . *Zm* are residual variables influencing *X*1 . . . *Xn* and *Y*1 . . . *Ym*.

ETA variables by LISREL) which are completely determined by the effects of the latent ultimate variables plus the reciprocal effects of *W*1 on *W*2, and vice versa. A model of this type might be used to represent, for example, the effects of latent personality factors on two subtypes of alcohol abuse (e.g., Cloninger, 1987).

LISREL distinguishes between the "structural model," which specifies the relationships between ultimate and dependent variables (including the effects of dependent variables on other dependent variables), and the "measurement" model relating the ultimate and dependent variables to the corresponding observed variables (termed *x* variables and *y* variables, respectively, by LISREL). In the simplest genetic applications there will be a one-to-one correspondence, and an identity relationship, between the latent dependent variables (phenotypes) and observed variables. The diagram in Fig. 1, however, represents a more complicated model. Observed variables *X*1, *X*2 . . . *Xn* are completely determined by the latent independent variables *V*2, *V*3, and *V*4 plus variable-specific residual effects (*U*1, *U*2 . . . *Un*). Observed variables *Y*1, *Y*2 . . . *Ym* are completely determined by the latent dependent variables *W*1 and *W*2 plus variable-specific residual effects (*Z*1, *Z*2 . . . *Zm*). The diagram thus represents a second-order factor model, where the two first-order factors, *W*1 and *W*2 (e.g., alcohol abuse factors), indexed by measurements *Y*1 . . . *Ym*, are determined by three second-order factors *V*2, *V*3, and *V*4 (e.g., latent personality factors), indexed by measurements *X*1 . . . *Xn*, plus residual effects (*V*1, *V*5).

Table I summarizes the parameter matrices used by LISREL. In most genetic applications only a small number of these will be needed (given in the upper section of Table I). The reader is referred to Jöreskog and Sörbom (1986a) for a formal specification of the LISREL model. It should be noted that, whereas the matrices beta, gamma, lambda-*X*, and lambda-*Y* give estimates of path coefficients, the matrices theta-delta and theta-epsilon are used to specify residual *variances* in the observed *x* variables and *y* variables, respectively. This is an unfortunate inconsistency because it means that LISREL can sometimes give impossible negative estimates for these residual variances.

REPRESENTING GENETIC MODELS IN LISREL

Figures 2a and b illustrate a simple univariate genetic model, for monozygotic twin pairs and for dizygotic twin pairs, respectively. The model allows for additive gene action (parameter *h*), environmental effects shared by twins (parameter *c*), and nonshared environmental effects which make one twin differ from his/her cotwin (parameter *e*). In addition,

Table I. Summary of Parameter Matrices Used to Specify Genetic and Environmental Models in LISREL

Matrix	Parameters	Typical element
PHI	Variances, covariances of ultimate variables	PH(i,j) gives covariance between ith, jth ultimate variables (= correlation for standardized variables) PH(i,i) gives variance of ith ultimate variable (= 1 if standardized)
GAMMA	Paths from ultimate to dependent variables	GA(i,j) gives path from jth ultimate to ith dependent variable
BETA	Paths from dependent variables to other dependent variables	BE(i,j) gives path from jth to ith dependent variable
LAMBDA-Y	Paths from dependent to observed (y) variables	LY(i,j) gives path from jth dependent to ith observed variable
THETA-EPSILON	Residual variances for y variables	TE(i) gives residual variance for ith observed y variable
PSI	Variances, covariances of dependent variables	PS(i,j) gives covariance between ith, jth dependent variables (= correlation for standardized variables) PS(i,i) gives variance of ith dependent variable (= 1 if standardized)
LAMBDA-X	Paths from ultimate to observed (x) variables	LX(i,j) gives path from jth ultimate to ith observed variable
THETA-DELTA	Residual variances for x variables	TD(i) gives residual variance for ith observed x variable

the model allows for reciprocal sibling interaction (parameter s), i.e., environmental effects of the first twin's phenotype on that of his/her cotwin, and vice versa (Carey, 1986). We might expect sibling interaction to be important in cases where, for example, extroversion in the first twin has an inhibitory effect on extroversion in the cotwin, and vice versa ($s < 0$), or where drug or alcohol use by one twin encourages drug or alcohol use by the cotwin, and vice versa ($s > 0$). All parameters are constrained to be the same in twins from monozygotic (MZ) and dizygotic (DZ) twin pairs, as well as in first and second twins from each twin group. Thus the model assumes that trait-relevant "passive" environmental effects are no more highly correlated in monozygotic than in dizygotic twin pairs. In terms of traditional variance components (e.g., Falconer, 1982), when $s = 0$, $VA = h^2$, $VC = c^2$, and $VE = e^2$.

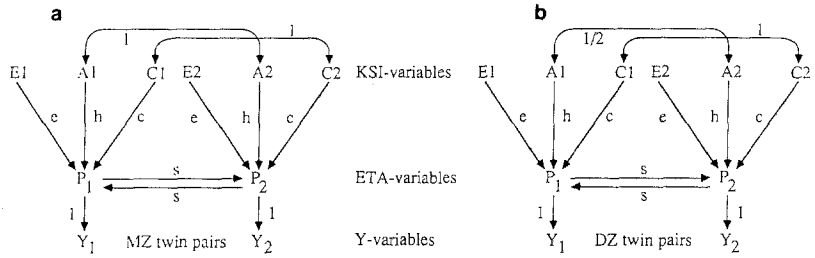

Fig. 2. Univariate genetic model. *P*1, *E*1, *A*1, and *C*1 denote the phenotype, unique environmental deviation, additive genetic deviation, and shared environmental deviation of the first twin; *P*2, *E*2, *A*2, and *C*2 denote the corresponding variables for the second twin; *Y*1 and *Y*2 denote observations on the first and second twins. See text for identification of parameters *e*, *h*, *c*, and *s*.

Parameter matrices needed to set up this job in LISREL are the 6 × 6 PHI matrix, the 2 × 6 GAMMA matrix, the 2 × 2 BETA matrix, and the 2 × 2 LAMBDA-*Y* matrix (see Table I). Taking the variables in Figs. 2a and b in the order in which they occur from left to right, these matrices will be as follows (see Table I).

$$\text{PHI} = \begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{matrix} \quad \text{(for MZs)} \qquad \text{or}$$

$$\text{PHI} = \begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & .5 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & .5 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{matrix} \quad \text{(for DZs)},$$

$$\text{GAMMA} = \begin{matrix} e & h & c & 0 & 0 & 0 \\ 0 & 0 & 0 & e & h & c \end{matrix},$$

$$\text{LAMBDA-}Y = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix},$$

$$\text{BETA} = \begin{matrix} 0 & s \\ s & 0 \end{matrix}.$$

Only the PHI matrix will differ between twin groups. The effects of genetic

dominance and shared environment are confounded in data on twin pairs reared together, the former decreasing the dizygotic correlation to less than half the monozygotic correlation, and the latter increasing the dizygotic correlation to greater than half the monozygotic correlation. If the data are consistent with genetic dominance rather than shared environmental effects, then we must estimate the parameter d ($= VD^{1/2}$) instead of c by fixing PHI (6,3) [= PHI (3,6)] for dizygotic twins to .25 rather than unity. Appendix I (Fig. A1) gives a sample LISREL job for fitting a simple univariate genetic model to twin data, allowing for additive gene action, nonshared environmental effects, plus reciprocal sibling interaction [i.e., setting $c = 0$; the joint resolution of sibling interaction and either shared environmental effects or genetic dominance will not be feasible with data on MZ and DZ twin pairs alone, for realistic sample sizes (Eaves, 1976; Jardine, 1985; Carey, 1986)]. The LISREL manual (Jöreskog and Sörbom, 1986a), should be consulted for full details of how to set up a problem run in LISREL.

Genotype × Sex/Cohort/Emvironment Interaction

The model in Fig. 2 is easily modified to allow for differences in genetic and environmental effects as a function of sex, age cohort (e.g., older versus younger twin pairs), cohabitation history (e.g., twin pairs living together versus twin pairs living apart), or environmental exposure (e.g., exposure or nonexposure to a high-risk environment). This is illustrated in the three composite diagrams in Figs. 3 and 4. Figure 3 would be appropriate for testing hypotheses about genotype × sex interaction. The same model could be used for testing hypotheses about cohort or cohabitation effects (with the younger/older or cohabiting/living-apart pairs replacing male and female like-sex pairs), except that in these cases there will be no equivalent of the unlike-sex pairs. The model allows for differences in sibling interaction effects as a function of sex, cohabitation, etc. For unlike-sex or discordant twin pairs, there are two different ways of representing sibling interaction, which differ with respect to whether the environmental effect of twin's phenotype on cotwin's phenotype is allowed to depend upon the sex of the actor or the sex of the recipient (the case represented in Fig. 3). These two different representations will generally lead to different predictions in twin data and, thus, will be testable by chi-square test of goodness of fit.

The model in Fig. 3 not only allows for differences in the magnitude of genetic and environmental effects as a function of sex (or cohabitation, etc.), but also for the possibility that those environmental effects which are shared by male twins, and those environmental effects which are

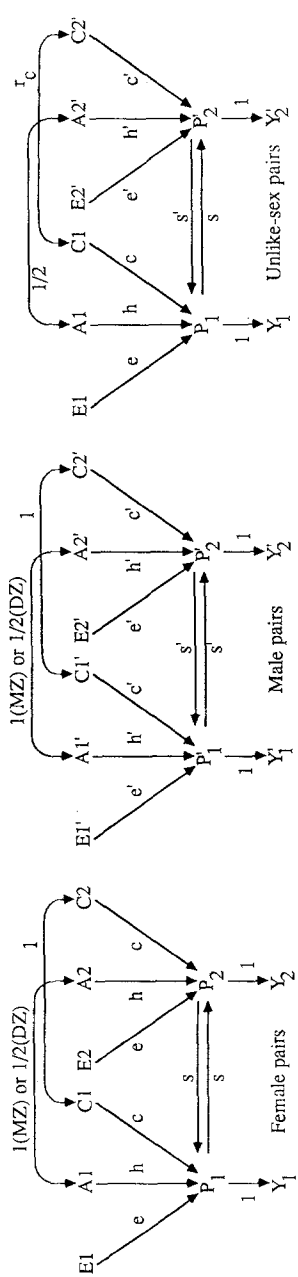

Fig. 3. Univariate genetic model allowing for genotype × sex interaction. Primes are used to distinguish variables and parameters in males from those in females. See Fig. 2 for identification of variables. See text for identification of parameters.

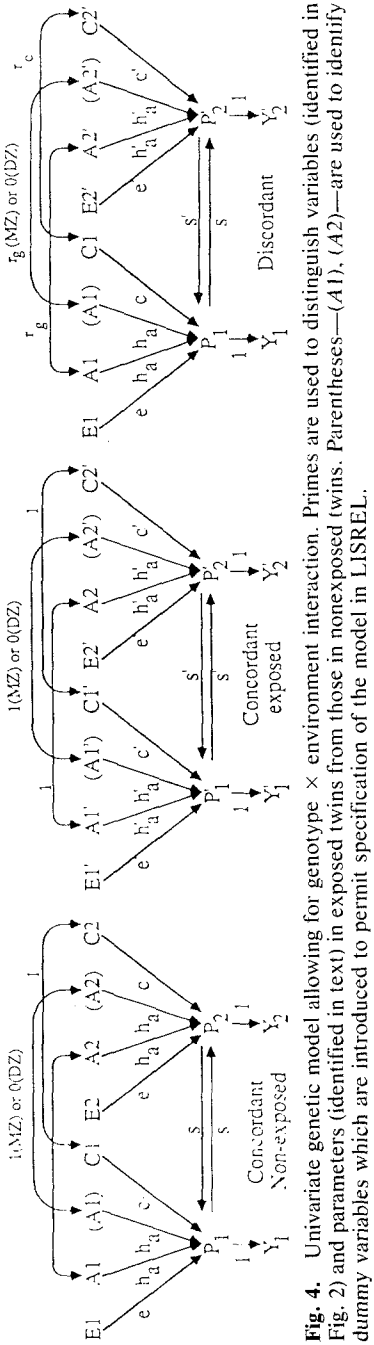

Fig. 4. Univariate genetic model allowing for genotype × environment interaction. Primes are used to distinguish variables (identified in Fig. 2) and parameters (identified in text) in exposed twins from those in nonexposed twins. Parentheses—(A1), (A2)—are used to identify dummy variables which are introduced to permit specification of the model in LISREL.

shared by female twins, are imperfectly correlated in twins from unlike-sex pairs (i.e., $r_C < 1$). With twin data, when testing for genotype × sex interaction, it is not possible to test simultaneously for imperfect correlations in both gene effects and shared environmental effects in the two sexes, since there are no unlike-sex monozygotic twin pairs! However, when testing for genotype × environment interaction, there will be both monozygotic and dizygotic pairs discordant for exposure, so we will be able to test whether the correlation between gene effects in exposed versus nonexposed twins, r_g, is less than unity. This cannot be achieved directly using the model in Fig. 3, since there is no way of indicating in LISREL that the element of the PHI matrix in discordant dizygotic twins which gives the correlation betwen $A1$ and $A2$ should be one-half the corresponding element of the PHI matrix for discordant monozygotic pairs. However, we can reparameterize the model in Fig. 3 to achieve this effect, as illustrated in Fig. 4. Here we are estimating parameters h_a and h_a' instead of h and h', where the genetic variance in exposed twins is given by $VA = h^2 = 2\,h_a^2$, and the genetic variance in nonexposed twins by $VA' = h'^2 = 2h_a'^2$. After model fitting, estimates of h_a and h_a' should be transformed to estimates of the genetic parameters h and h' (or VA and VA'). This model permits us to estimate r_g, the correlation between gene effects under the two conditions of environmental exposure. It can also be used to represent genotype × sex interaction where the correlation between gene effects in the two sexes is less than unity. Appendix II (Fig. A1) gives a sample LISREL job which fits a genotype × environment interaction model to twin data.

Bivariate Models, Causal Pathways, and Genotype–Environment Correlation

With twin data on two or more variables, we can test hypotheses about the contribution of genetic and environmental factors to the *covariation* of variables. A general bivariate model for the resemblance of twin pairs for two traits (or the same trait measured on two occasions) is presented in Fig. 5 (cf. Eaves and Eysenck, 1975). (To simplify the diagram, shared environmental effects have been omitted.) The full model estimates separate genetic and environmental parameters for each trait (h, h', etc.), together with correlations between gene effects, shared environmental effects, and nonshared environmental effects for the two traits (r_G, r_C, and r_E). The parameters i and i' are redundant in the full model. However, sometimes we may wish to test a more restrictive hypothesis, that trait PA is one of the *causes* of variation in trait PB ($i > 0$), against the more general model, and the alternative hypothesis that

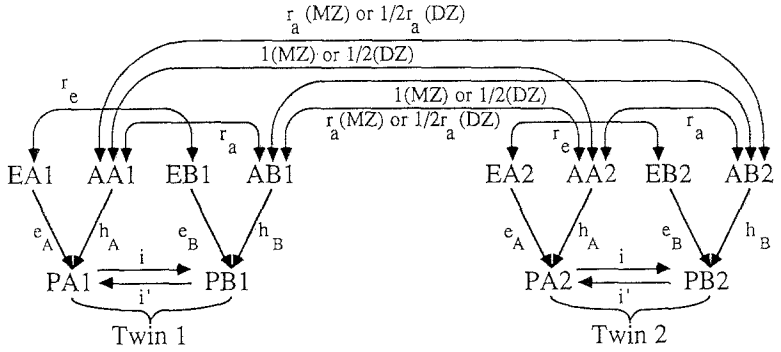

Fig. 5. Bivariate genetic model allowing for reciprocal interaction between variables. *PA*1, *EA*1, and *AA*1 and *PA*2, *EA*2, and *AA*2 denote the phenotypic deviation, unique environmental deviation, and additive genetic deviation for the first trait in the first and second twins: *PB*1, *EB*1, and *AB*1 and *PB*2, *EB*2, and *AB*2 denote the corresponding variables for the second trait; r_a and r_e denote the correlations between additive genetic deviations and unique environmental deviations for the first and second traits; subscripts A and B distinguish genetic and environmental parameters for the first versus second trait; paths i and i' represent direct effects of the first trait on the second trait, and vice versa.

trait *PB* causes variation in trait *PA* ($i' > 0$). Even with *cross-sectional* bivariate twin data, such causal pathway models can be specified and tested, just as can be achieved with *prospective* data on samples of unrelated individuals (e.g., Kessler and Greenberg, 1981), provided that twin correlations differ in magnitude for the two variables *PA* and *PB*. In testing these alternative submodels, we set $r_A = r_C = r_E = 0$ and either $i = 0$ (*PB* is a cause of *PA*) or $i' = 0$ (*PA* is a cause of *PB*). Once again, the full model cannot be directly represented in LISREL—because the cross-correlation between AA_1 and AB_2 or between AB_1 and AA_2 is r_G in MZ pairs but $\frac{1}{2}r_G$ in DZ pairs—but must be reparameterized by including the variables AA_1 and AB_1 twice for each twin. Since the same principle applies as in the reparameterized model in Fig. 4, we do not give the LISREL representation here.

The bivariate model represented in Fig. 5 can be used to test for one type of genotype–environment correlation (Eaves *et al.*, 1977), where genetic liability is positively correlated with exposure to environmental risk factors because individuals high on genetic liability (e.g., to depression) are more likely to expose themselves to high-risk environments [e.g., stressful life events (cf. Brown and Harris, 1978)]. Thus the model can be used to determine whether, for example, stressful life events are a cause of depression or depression is a cause of self-exposure to stressful life events. The more general hypothesis, that there is reciprocal inter-

action between *PA* and *PB* (i.e., $i \neq 0$ and $i' \neq 0$, e.g., because stressful life events are a cause of depression, which in turn increases self-exposure to stressful life events) cannot in general be resolved with *cross-sectional* twin data.

Multivariate Genetic Models

In analyses involving three or more variables per twin, more elaborate hypotheses about the structure of genetic and environmental effects can be conducted within the framework of multivariate genetic analysis (e.g., Martin and Eaves, 1977; Kendler *et al.*, 1987; Silberg *et al.*, 1987). Multivariate genetic analysis is a generalization of factor analysis, in which we use not only phenotypic covariances or correlations between variables, but also covariances and cross-variable covariances between MZ and DZ twin pairs, to estimate loadings on separate genetic, shared environmental, and nonshared environmental common factors. The simplest possible model would combine any of the "structural" models in Figs. 2–5 with a factor "measurement" model to specify the relationship between *P*1 (or *PA*1), etc., and the observed variables *Y*1, *Y*2 . . . *Yn*, allowing also for residual environmental effects on the *y* variables. However, this model implies the very strong assumption that all variable-specific effects are uncorrelated over twin pairs (i.e., that there are no trait-specific genetic or shared environmental effects), an assumption which is rarely satisfied in real data.

Figure 6 presents a somewhat more general version of such a model, which has sometimes been described as the "psychometric" model (McArdle and Goldsmith, 1984) or "common pathway" model (Kendler *et al.*, 1987), which does allow for variable-specific genetic and environmental effects. The diagram has been simplified by the omission of shared environmental effects. The model illustrated is termed the single-common factor model, since the correlations between observed variables *Y*1, *Y*2 . . . *Yn* are directly determined by one latent variable, *PA*. It is not possible to estimate the absolute magnitudes of the parameters *h* and *e* in Fig. 6, only their relative magnitude, so one of these parameters (or *c*, in models allowing for shared environmental effects) must be fixed to unity. If this model fits the data, the heritability of the latent phenotype (or phenotypes) can be computed in the usual fashion, as $h^2/(h^2 + c^2 + e^2)$. The path coefficients corresponding to the paths from *PA* to *P*1, *P*2 . . . *Pn* are the factor loadings of these variables on the latent phenotype *PA*. The paths from the variables *A*1, *E*1, etc., are the variable-specific genetic and environmental loadings. Multivariate genetic models can be represented in LISREL in a variety of different ways. Appendix III (Fig.

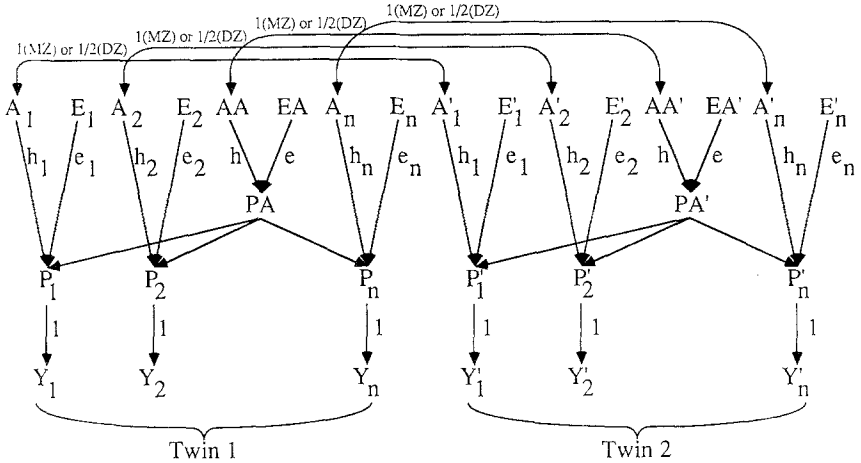

Fig. 6. Latent phenotype multivariate genetic model. P_1, P_2 . . . P_n and Y_1, Y_2 . . . Y_n denote the phenotype and observation for traits 1 . . . n; A_1 . . . A_n and E_1 . . . E_n denote corresponding trait-specific genetic and environmental effects. *PA* denotes the intervening latent phenotype, and *AA* and *EA* denote the corresponding common genetic and environmental effects. Primes are used to distinguish variables of the second twin from those of the first twin.

A1) illustrates one possible way of representing the model in Fig. 6 in LISREL.

A more general multivariate genetic model is illustrated, for the case of two latent genetic common factors and two latent nonshared environmental common factors (by convention a "two-factor" model), in Fig. 7. To simplify, we give the diagram for a single twin, rather than for both members of a twin pair, and omit variable-specific genetic and environmental effects. Under this model, which has sometimes been described as the "independent pathway" model (Kendler *et al.*, 1987) or "biometrical" model (McArdle and Goldsmith, 1984), no latent phenotype intervenes between the common genetic and environmental factors and the observed variables. Thus the genetic factor structure may be quite different from the environmental factor structure, as indeed has usually been found in multivariate analyses of affective variables (e.g., Kendler *et al.*, 1987; Silberg *et al.*, 1987). The genetic and environmental common factors in Fig. 7 are orthogonal (i.e., *AA*, *AB*, *EA*, and *EB* are all uncorrelated), although this constraint can be relaxed when assortative mating or joint cultural and biological inheritance generate correlations between the common factors (cf. Rao *et al.*, 1976; Cloninger *et al.*, 1979; Carey, 1987). Appendix IV (Fig. A1) illustrates a LISREL job for fitting the "one-

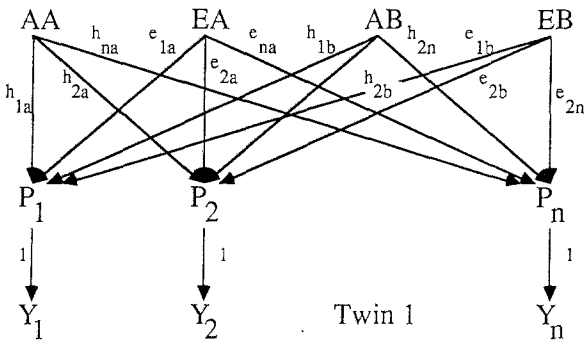

Fig. 7. General multivariate genetic model. To simplify, only variables for the first twin from a pair are represented, and trait-specific influences are omitted. *AA* and *AB* denote the first and second latent genetic common factors; *EA* and *EB* denote the first and second latent environmental common factors.

factor" version of the model in Fig. 7. The model in Fig. 6 is a special case of the latter model, where loadings on the common genetic or common shared environmental factors are constrained to be constant multiples of loadings on the corresponding common unique environmental factors. The goodness of fit of the restricted model can thus be compared to that of the more general model by likelihood-ratio chi-square test (e.g., Jöreskog, 1978; Neale *et al.*, 1989b), to test for differences in genetic and environmental factor structures.

When two or more common factors are estimated for a given source (i.e., two or more genetic common factors, or two or more shared environmental common factors, or two or more unique environmental common factors), the problem of factor rotation (e.g., Harman, 1976) often arises. An infinite number of equivalent solutions may exist, so that constraints must be imposed to ensure that LISREL converges on one of the possible solutions. In LISREL this may be achieved by arbitrarily fixing to zero, separately for each source, one loading on the second common factor, two loadings on the third common factor, and so on. Programs for factor rotation (e.g., SAS Institute, 1985) may then be used to rotate separately the genetic loadings, the shared environmental loadings, and the unique environmental loadings, to maximize conformity with traditional criteria for "simple structure" (Harman, 1976). In versions of the common pathway model in Fig. 6 which estimate two or more latent phenotypes (*PA*, *PB*, etc.), the rotation problem can sometimes be avoided (Heath *et al.*, 1988b). This is possible, however, only when there are differences in genetic architecture between latent phenotypes, such

that there is only a single genetic dominance common factor corresponding to one latent phenotype, a single shared environmental common factor corresponding to a second latent phenotype, a single additive genetic common factor exhibiting genotype × sex interaction, and so on.

Multivariate genetic models can be elaborated exactly as in the univariate case, allowing for genotype × age/environment/cohort interactions, causal pathway models, genotype–environment correlation, etc. However, if we wish to test whether some genetic or environmental effects are sex specific or specific to one condition of environmental exposure, instead of estimating correlations between latent genetic and environmental factors (as in Figs. 4 and 5), we now estimate additional common genetic and environmental factors which are specific to only one sex or exposure condition.

Other Applications

We have considered here only the analysis of cross-sectional twin data. Hewitt *et al.* (1988) and Boomsma and Molenaar (1986; Boomsma *et al.*, 1989b) illustrate the analysis of developmental or time-series data using LISREL. The power of the twin design is greatly enhanced by obtaining data on the relatives of twins, especially their parents (e.g., Eaves *et al.*, 1978; Fulker, 1982; Vogler and Fulker, 1983; Heath *et al.*, 1985) and spouses (e.g., Eaves, 1979; Heath and Eaves, 1985; Heath, 1987). Some simple ways of modeling cultural and biological inheritance in twin-family data, in the presence of assortative mating, are discussed by Eaves *et al.* (1989).

CONCLUSIONS

It is important to be aware of the problems that cannot be handled using LISREL. If twin pairs have been ascertained because at least one twin is affected by a disease, and so do not constitute a random sample from the population, the use of LISREL will be inappropriate. A general treatment of assortative mating (e.g., Carey, 1987) is not possible using LISREL. Joint cultural and biological inheritance (e.g., Rao *et al.*, 1976; Cloninger *et al.*, 1979), when this gives rise to genotype–environment correlation, cannot easily be represented in LISREL. LISREL will not handle very large problems. Even moderately small multivariate genetic problems are laborious to set up in LISREL because there is no efficient way of specifying parameters which are constrained to be equal (e.g., corresponding genetic and environmental loadings for first and second twins from each group). Current versions of LISREL will not easily han-

dle problems where the number of variables per case differs between groups. Thus the analysis of twin-family data where the family structure is variable (e.g., because only some twins are married or because parental data are missing from some families) is better achieved using purpose-designed programs. In analyses of discontinuous variables, the current versions of LISREL or PRELIS will not permit threshold values to be constrained to be equal across groups. It is sometimes necessary to fit models directly to raw data (e.g., Lange *et al.*, 1976), perhaps because genetic or environmental effects vary as functions of continuously distributed variables such as age or measured environmental risk factors. For such applications the user is forced to write his/her own software. For the simple applications which we have discussed in this article, however, the use of a software package for structural equation modeling such as LISREL will allow rigorous testing of a wide range of genetic and environmental hypotheses.

APPENDIXES I–IV

Appendix I

This LISREL program fits a simple genetic model to twin data, allowing for additive genetic effects, unique environmental effects and sibling interaction (see Heath et al., 1988a, for further information about the data-set). The MO cards have been split over two lines; concatenation of these lines is necessary for LISREL input.

```
Oz Alcohol - Young MZF pairs cohabiting
DA NG=2 NI=2 NO=171 MA=CM
LA
*
'ALC1' 'ALC2'
CM FU
*
1.280 0.766
0.766 1.194
MO NY=2 NE=2 NK=6 GA=FU,FR BE=FU,FI
LY=ID PH=SY,FI PS=ZE TE=ZE
LK
*
'E1' 'A1' 'C1' 'E2' 'A2' 'C2'
LE
*
'P1' 'P2'
ST 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
ST 1.0 PH(6,6) PH(2,5) PH(3,6)
FR BE(1,2) BE(2,1)
EQ BE(1,2) BE(2,1)
ST 0.00 BE(1,2)
PA GA
*
1 1 0 0 0 0
0 0 0 1 1 0
EQ GA(1,1) GA(2,4)
EQ GA(1,2) GA(2,5)
ST 0.3 GA(1,1)-GA(2,6)
OU TM=600 ND=5
Young DZF pairs cohabiting
DA NI=2 NO=101 MA=CM
LA
*
'ALC1' 'ALC2'
CM FU
*
1.077 0.463
0.463 0.962
MO GA=IN PS=IN LY=IN PH=SY,FI TE=IN
BE=IN
LK
*
'E1' 'A1' 'C1' 'E2' 'A2' 'C2'
LE
*
'P1' 'P2'
ST 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
ST 1.0 PH(6,6) PH(3,6)
ST 0.5 PH(2,5)
OU TM=60 ND=5 SE TV PC
```

Appendix II

This LISREL program fits a model which allows for the interaction of genetic and environmental effects with exposure or non-exposure to an environmental risk factor (c.f. Heath et al., 1988a). The MO cards have been split over two lines; concatenation of these lines is necessary for LISREL input.

```
Alcohol - Concordant exposed MZF's
DA NG=6 NI=2 NO=36 MA=CM
LA
*
'ALC1' 'ALC2'
CM FU
*
1.080 0.771
0.771 1.061
MO NY=2 NE=2 NK=8 GA=FU,FR BE=ZE LY=ID
PH=SY,FI PS=ZE TE=ZE
LK
*
'E1' 'A1' 'DUMMY1' 'C1' 'E2' 'A2' 'DUMMY2' 'C2'
LE
*
'P1' 'P2'
ST 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
ST 1.0 PH(6,6) PH(7,7) PH(8,8)
ST 1.0 PH(2,6) PH(3,7) PH(4,8)
PA GA
*
1 1 1 1 0 0 0 0
0 0 0 0 1 1 1 1
EQ GA(1,1) GA(2,5)
EQ GA(1,2) GA(2,6) GA(1,3) GA(2,7)
EQ GA(1,4) GA(2,8)
ST 0.3 GA(1,1)-GA(2,8)
OU TM=600 ND=5
Concordant exposed DZF
DA NI=2 NO=18 MA=CM
LA
*
'ALC1' 'ALC2'
CM FU
*
1.214 0.106
0.106 1.536
MO NY=2 NE=2 NK=8 GA=FU,FR BE=ZE LY=ID
PH=SY,FI PS=ZE TE=ZE
LK
*
'E1' 'A1' 'DUMMY1' 'C1' 'E2' 'A2' 'DUMMY2' 'C2'
LE
*
'P1' 'P2'
ST 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
ST 1.0 PH(6,6) PH(7,7) PH(8,8)
```

Fig. A1

```
ST 1.0 PH(2,6) PH(4,8)
PA GA
*
1 1 1 1 0 0 0 0
0 0 0 0 1 1 1 1
EQ GA(1,1,1) GA(1,1) GA(2,5)
EQ GA(1,1,2) GA(1,2) GA(2,6) GA(1,3) GA(2,7)
EQ GA(1,1,4) GA(1,4) GA(2,8)
OU TM=600 ND=5
Concordant non-exposed MZF
DA NI=2 NO=391 MA=CM
LA
*
'ALC1' 'ALC2'
CM FU
*
1.262 0.719
0.719 1.203
MO NY=2 NE=2 NK=8 GA=FU,FR BE=ZE LY=ID
PH=SY,FI PS=ZE TE=ZE
LK
*
'E1' 'A1' 'DUMMY1' 'C1' 'E2' 'A2' 'DUMMY2' 'C2'
LE
*
'P1' 'P2'
ST 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
ST 1.0 PH(6,6) PH(7,7) PH(8,8)
ST 1.0 PH(2,6) PH(3,7) PH(4,8)
PA GA
*
1 1 1 1 0 0 0 0
0 0 0 0 1 1 1 1
EQ GA(1,1) GA(2,5)
EQ GA(1,2) GA(2,6) GA(1,3) GA(2,7)
EQ GA(1,4) GA(2,8)
ST 0.3 GA(1,1)-GA(2,8)
OU TM=600 ND=5
Concordant non-exposed DZF
DA NI=2 NO=217 MA=CM
LA
*
'ALC1' 'ALC2'
CM FU
*
1.212 0.456
0.456 1.290
MO NY=2 NE=2 NK=8 GA=FU,FR BE=ZE LY=ID
PH=SY,FI PS=ZE TE=ZE
LK
*
'E1' 'A1' 'DUMMY1' 'C1' 'E2' 'A2' 'DUMMY2' 'C2'
LE
*
'P1' 'P2'
ST 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
ST 1.0 PH(6,6) PH(7,7) PH(8,8)
ST 1.0 PH(2,6) PH(4,8)
PA GA
*
1 1 1 1 0 0 0 0
0 0 0 0 1 1 1 1
EQ GA(3,1,1) GA(1,1) GA(2,5)
EQ GA(3,1,2) GA(1,2) GA(2,6) GA(1,3) GA(2,7)
EQ GA(3,1,4) GA(1,4) GA(2,8)
OU TM=600 ND=5
Exposed/non-exposed MZF
DA NI=2 NO=113 MA=CM
LA
*
'AGE1' 'AGE2'
CM FU
*
1.557 0.642
0.642 1.348
MO NY=2 NE=2 NK=8 GA=FU,FR BE=ZE LY=ID
PH=SY,FI PS=ZE TE=ZE
LK
*
'E1' 'A1' 'DUMMY1' 'C1' 'E2' 'A2' 'DUMMY2' 'C2'
LE
*
'P1' 'P2'
ST 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
ST 1.0 PH(6,6) PH(7,7) PH(8,8)
ST 1.0 PH(4,8)
FR PH(2,6) PH(3,7)
EQ PH(2,6) PH(3,7)
ST 0.75 PH(2,6)
PA GA
*
1 1 1 1 0 0 0 0
0 0 0 0 1 1 1 1
EQ GA(1,1,1) GA(1,1)
EQ GA(3,1,1) GA(2,5)
EQ GA(1,1,2) GA(1,2) GA(1,3)
EQ GA(3,1,2) GA(2,6) GA(2,7)
EQ GA(1,1,4) GA(1,4)
EQ GA(3,1,4) GA(2,8)
OU TM=600 ND=5
(Exposed/non-exposed) DZF
DA NI=2 NO=89 MA=CM
LA
*
'ALC1' 'ALC2'
CM FU
*
1.213 0.088
0.088 1.081
MO NY=2 NE=2 NK=8 GA=FU,FR BE=ZE LY=ID
PH=SY,FI PS=ZE TE=ZE
LK
*
'E1' 'A1' 'DUMMY1' 'C1' 'E2' 'A2' 'DUMMY2' 'C2'
LE
*
'P1' 'P2'
ST 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
ST 1.0 PH(6,6) PH(7,7) PH(8,8)
ST 1.0 PH(4,8)
```

Fig. A1 (Continued)

```
FR PH(2,6)
ST 0.75 PH(2,6)
PA GA
*
1 1 1 1 0 0 0 0
0 0 0 0 1 1 1 1
EQ GA(1,1,1) GA(1,1)
EQ GA(3,1,1) GA(2,5)
EQ GA(1,1,2) GA(1,2) GA(1,3)
EQ GA(3,1,2) GA(2,6) GA(2,7)
EQ GA(1,1,4) GA(1,4)
EQ GA(3,1,4) GA(2,8)
OU TM 600 ND=5 SE TV PC
```

Appendix III

This LISREL program fits a one-factor 'common pathway' multivariate genetic model, allowing for genetic and unique environmental common and specific factors, to Australian twin data on Neuroticism, symptoms of anxiety, and symptoms of depression (see Jardine et al., 1985, for further details of data-set). The MO cards have been split over two lines; concatenation of these lines is necessary for LISREL input.

```
Common pathway model Female MZ
DA NG=2 NI=6 NO=1231 MA=CM
LA
*
'N1' 'ANX1' 'DEP1' 'N2' 'ANX2' 'DEP2'
CM
*
0.0854 0.1184 0.4135 0.1032 0.2746
0.3710 0.0415 0.0636 0.0562 0.0801
0.0662 0.1637 0.1358 0.1144 0.4003
0.0575 0.1280 0.1356 0.1037 0.2709 0.3583
MO NY=6 NE=8 NK=16 GA=FU,FR LY=FU,FI
PH SY,FI PS=ZE TE=ZE BE=FU,FI
LK
*
'A1' 'E1' 'A2' 'E2' 'AA' 'EA' 'A3' 'E3' 'A1:TW2'
'E1:TW2' 'A2:TW2' 'E2:TW2' 'AA:TW2' 'EA:TW2'
'A3:TW2' 'E3:TW2'
LE
*
'P1' 'P2' 'PA' 'P3' 'P1:TW2' 'P2:TW2' 'PA:TW2'
'P3:TW2'
ST 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
ST 1.0 PH(6,6) PH(7,7)
ST 1.0 PH(8,8) PH(9,9) PH(10,10) PH(11,11)
ST 1.0 PH(12,12) PH(13,13)
ST 1.0 PH(14,14) PH(15,15) PH(16,16)
ST 1.0 PH(1,9) PH(3,11) PH(5,13) PH(7,15)
FR BE(1,3) BE(2,3) BE(4,3) BE(5,7) BE(6,7) BE(8,7)
EQ BE(1,3) BE(5,7)
EQ BE(2,3) BE(6,7)
EQ BE(4,3) BE(8,7)
ST 0.5 BE(1,3)-BE(8,7)
ST 1.0 LY(1,1) LY(2,2) LY(3,1) LY(4,5) LY(5,6)
LY(6,8)
PA GA
*
1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
EQ GA(1,1) GA(5,9)
EQ GA(1,2) GA(5,10)
EQ GA(2,3) GA(6,11)
EQ GA(2,4) GA(6,12)
EQ GA(3,5) GA(7,13)
EQ GA(4,7) GA(8,15)
EQ GA(4,8) GA(8,16)
ST 1.0 GA(3,6) GA(7,14)
ST .2 GA(1,1)-GA(6,18)
OU TM=1200 ND 7
Female DZ
DA NO=749 MA=CM
LA
*
'N1' 'ANX1' 'DEP1' 'N2' 'ANX2' 'DEP2'
CM
*
0.0754 0.1024 0.3895 0.0894 0.2540
0.3498 0.0210 0.0383 0.0334 0.0891
0.0295 0.0761 0.0655 0.1270 0.4341
0.0342 0.0727 0.0785 0.1241 0.3261 0.4529
MO NY=6 NE=8 NK=16 GA=IN LY=IN PH=SY,FI
PS=ZE TE=ZE BE=IN
LK
*
'A1' 'E1' 'A2' 'E2' 'AA' 'EA' 'A3' 'E3' 'A1:TW2'
'E1:TW2' 'A2:TW2' 'E2:TW2' 'AA:TW2' 'EA:TW2'
'A3:TW2' 'E3:TW2'
LE
*
'P1' 'P2' 'PA' 'P3' 'P1:TW2' 'P2:TW2' 'PA:TW2'
'P3:TW2'
ST 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
ST 1.0 PH(6,6) PH(7,7)
ST 1.0 PH(8,8) PH(9,9) PH(10,10) PH(11,11)
ST 1.0 PH(12,12) PH(13,13)
ST 1.0 PH(14,14) PH(15,15) PH(16,16)
ST 0.5 PH(1,9) PH(3,11) PH(5,13) PH(7,15)
OU SE TV TM=1200 ND 7
```

Appendix IV

This LISREL program fits a general one-factor multivariate genetic model, allowing for genetic and unique environmental common and specific factors, to the same data-set used in Appendix 3. The MO cards have been split over two lines; concatenation of these lines is

Fig. A1 (Continued)

```
necessary for LISREL input.

General multivariate model MZF
DA NG=2 NI=6 NO=1233 MA=CM
LA
*
'N1' 'ANX1' 'DEP1' 'N2' 'ANX2' 'DEP2'
CM
*
0.0854 0.1184 0.4135 0.1032 0.2746
0.3710 0.0415 0.0636 0.0562 0.0801
0.0662 0.1637 0.1358 0.1144 0.4003
0.0575 0.1280 0.1356 0.1037 0.2709 0.3583
MO NY=6 NE=6 NK=16 GA=FU,FR LY=ID
PH=SY,FI PS=ZE TE=ZE
LK
*
'AA' 'EA' 'A1' 'A2' 'A3' 'E1' 'E2' 'E3' 'AA:TW2'
'EA:TW2' 'A1:TW2' 'A2:TW2' 'A3:TW2' 'E1:TW2'
'E2:TW2' 'E3:TW2'
LE
*
'V1' 'V2' 'V3' 'V1:TW2' 'V2:TW2' 'V3:TW2'
ST 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
ST 1.0 PH(6,6) PH(7,7) PH(8,8) PH(9,9) PH(10,10)
ST 1.0 PH(11,11) PH(12,12)
ST 1.0 PH(13,13) PH(14,14) PH(15,15) PH(16,16)
ST 1.0 PH(1,9) PH(3,11) PH(4,12) PH(5,13)
PA GA
*
1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0
1 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0
1 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 1 1 0 0 1 0 0
0 0 0 0 0 0 0 0 1 1 0 1 0 0 1 0
0 0 0 0 0 0 0 0 1 1 0 0 1 0 0 1
EQ GA(1,1) GA(4,9)
EQ GA(1,2) GA(4,10)
EQ GA(1,3) GA(4,11)
EQ GA(1,6) GA(4,14)
EQ GA(2,1) GA(5,9)
EQ GA(2,2) GA(5,10)
EQ GA(2,4) GA(5,12)
EQ GA(2,7) GA(5,15)
EQ GA(3,1) GA(6,9)
EQ GA(3,2) GA(6,10)
EQ GA(3,5) GA(6,13)
EQ GA(3,8) GA(6,16)
VALUE .2 GA(1,1)-GA(6,16)
OU TM=1200 ND=7
General multivariate model DZF
DA NO=749 MA=CM
LA
*
'N1' 'ANX1' 'DEP1' 'N2' 'ANX2' 'DEP2'
CM
*
0.0754 0.1024 0.3895 0.0894 0.2540
0.3498 0.0210 0.0383 0.0334 0.0891
0.0295 0.0761 0.0655 0.1270 0.4341
0.0342 0.0727 0.0785 0.1241 0.3261 0.4529
MO GA=IN PS=IN BE=IN LY=IN PH=ST,FI
TE=IN
LK
*
'AA' 'EA' 'A1' 'A2' 'A3' 'E1' 'E2' 'E3' 'AA:TW2'
'EA:TW2' 'A1:TW2' 'A2:TW2' 'A3:TW2' 'E1:TW2'
'E2:TW2' 'E3:TW2'
LE
*
'V1' 'V2' 'V3' 'V1:TW2' 'V2:TW2' 'V3:TW2'
ST 1.0 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5)
ST 1.0 PH(6,6) PH(7,7) PH(8,8) PH(9,9) PH(10,10)
ST 1.0 PH(11,11) PH(12,12)
ST 1.0 PH(13,13) PH(14,14) PH(15,15) PH(16,16)
ST 0.5 PH(1,9) PH(3,11) PH(4,12) PH(5,13)
OU SE TV TM=1200 ND=7
```

Fig. A1 (Continued)

ACKNOWLEDGMENTS

We are grateful to Drs. R. Vlietinck and R. Derom for their organization of the workshop on twin data analysis and to Dr. J. McArdle for suggesting to us ways in which some of the shortcomings of LISREL can be sidestepped. We thank Drs. N. G. Martin and D. Boomsma for helpful comments on an early draft of the manuscript.

REFERENCES

Bartlett, M. S. (1947). The use of transformations. *Biometrics* **3**:39–52.

Boomsma, D. I., and Molenaar, P. C. M. (1986). Using LISREL to analyze genetic and environmental covariance structure. *Behav. Genet.* **16**:237–250.

Boomsma, D. I., and Molenaar, P. C. M. (1987). The genetic analysis of repeated measures. I. Simplex models. *Behav. Genet.* **17**:111–124.

Boomsma, D. I., Martin, N. G., and Molenaar, P. C. M. (1989a). Resemblances of parents and twins in sports participation and heart rate. *Behav. Genet.* **19**:123–141.

Boomsma, D. I., van den Bree, M. B. M., Orbeleke, J. F., and Molenaar, P. C. M. (1989b). Factor and simplex models for repeated measures: Application to two psychomotor measures of alcohol sensitivity in twins. *Behav. Genet.* **19**:79–96.

Brown, G. W., and Harris, T. (1978). *Social Origins of Depression,* London, Tavistock.

Browne, M. W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *Br. J. Math. Stat. Psychol.* **37**:62–83.

Carey, G. (1986). Sibling imitation and contrast effects. *Behav. Genet.* **16**:343–354.

Carey, G. (1987). A general multivariate approach to linear modelling in human genetics. *Am. J. Hum. Genet.* **39**:775–786.

Cloninger, C. R. (1987). Neurogenetic adaptive mechanisms in alcoholism. *Science* **236**:410–416.

Cloninger, C. R., Rice, J., and Reich, T. (1979). Multifactorial inheritance with cultural transmission and assortative mating. II. A general model of combined polygenic and cultural inheritance. *Am. J. Hum. Genet.* **31**:135–145.

Cox, A., Rutter, M., Yule, B., and Quinton, D. (1977). Bias resulting from missing information: Some epidemiological findings. *Br. J. Prev. Soc. Med.* **31**:131–136.

DeFries, J., and Fulker, D. W. (1985). Multiple regression analysis of twin data. *Behav. Genet.* **15**:467–474.

Eaves, L. J. (1976). A model for sibling effects in man. *Heredity* **36**:205–214.

Eaves, L. J. (1977). Inferring the causes of human variation. *J. Roy. Stat. Soc. Ser. B* **140**:324–355.

Eaves, L. J. (1979). The use of twins in the analysis of assortative mating. *Heredity* **43**:399–409.

Eaves, L. J. (1982). The utility of twins. In Anderson, V. E. (ed.), *Genetic Basis of the Epilepsies,* Raven, New York.

Eaves, L. J., and Eysenck, H. J. (1975). The nature of extraversion: A genetical analysis. *J. Personal. Soc. Psychol.* **32**:102–112.

Eaves, L. J., Last, K. A., Martin, N. G., and Jinks, J. L. (1977). A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *Br. J. Math. Stat. Psychol* **30**:1–42.

Eaves, L. J., Last, K. A., Young, P. A., and Martin, N. G. (1978). Model fitting approaches to the analysis of human behaviour. *Heredity* **41**:249–320.

Eaves, L. J., Fulker, D. W., 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.

Eysenck, H. J., and Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire,* Hodder and Stoughton, London.

Falconer, D. S. (1982). *Introduction to Quantitative Genetics,* Oliver and Boyd, Edinburgh.
Fulker, D. W. (1982). Extensions of the classical twin method. *Proceedings of the 1981 International Congress of Human Genetics, Jerusalem,* Alan R. Liss, New York.
Fulker, D. W., Baker, L. A., and Bock, R. D. (1983). Estimating components of covariance using LISREL. *Data Anal. Comm. Comp. Data Anal.* **1**:5–8.
Harman, H. H. (1976). *Modern Factor Analysis.* University of Chicago Press, Chicago.
Heath, A. C. (1987). The analysis of marital interaction in cross-sectional twin data. *Acta Genet. Med. Gemellol.* **36**:41–49.
Heath, A. C., and Eaves, L. J. (1985). Resolving the effects of phenotype and social background on mate selection. *Behav. Genet.* **15**:15–30.
Heath, A. C., Kendler, K. S., Eaves, L. J., and Markell, D. (1985). The resolution of cultural and biological inheritance: Informativeness of different relationships. *Behav. Genet.* **15**:439–465.
Heath, A. C., Jardine, R., Martin, N. G. (1988a). Interactive effects of genotype and social environment on alcohol consumption in female twins. *J. Stud. Alcohol* (in press).
Heath, A. C., Jardine, R., Evans, L. J., and Martin, N. G. (1988b). The genetic structure of personality I. *Personal. Indiv. Diff.* **9**:59–67.
Hewitt, J. K., Eaves, L. J., Neale, M. C., and Meyer, J. M. (1988). Resolving the causes of developmental continuity or "tracking." I. Longitudinal twin studies during growth. *Behav. Genet.* **18**:133–151.
Holt, S. (1968). *The Genetics of Dermal Ridges,* Thomas, Springfield, Ill.
Hopper, J. L., and Culross, P. R. (1983). Covariation between family members as a function of cohabitation history. *Behav. Genet.* **13**:459–471.
Hopper, J. L., and Mathews, J. D. (1983). Extensions to multivariate normal models for pedigree analysis. II. Modeling the effect of shared environment in blood lead levels. *Am. J. Epidemiol.* **117**:344–355.
Jardine, R. (1985). *A Twin Study of Personality, Social Attitudes and Drinking Behavior.* Unpublished Ph.D. thesis, Australian National University.
Jardine, R., and Martin, N. G. (1984). Causes of variation in drinking habits in a large twin sample. *Acta Genet. Med. Gemellol.* **33**:435–450.
Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika* **36**:409–426.
Jöreskog, K. G., and Sörbom, D. (1986a). *LISREL VI,* Scientific Software, Mooresville, Ind.
Jöreskog, K. G., and Sörbom, D. (1986b). *PRELIS: A Preprocessor for LISREL,* Scientific Software, Mooresville, Ind.
Kaprio, J., Koskenvuo, M. D., Langinvainio, H., Romanov, K., Sarna, S., and Rose, R. (1987). Genetic influences on use and abuse of alcohol: A study of 5638 adult Finnish brothers. *Alc. Clin. Exp. Res.* **11**:349–356.
Kendler, K. S., Heath, A. C., Martin, N. G., and Eaves, L. J. (1986). Symptoms of anxiety and depression in a volunteer twin population: The etiologic role of genetic and environmental factors. *Arch. Gen. Psychiat.* **43**:213–221.
Kendler, K. S., Heath, A. C., Martin, N. G., and Eaves, L. J. (1987). Symptoms of anxiety and symptoms of depression: Same genes, different environments? *Arch. Gen. Psychiat.* **44**:451–460.
Kessler, R., and Greenberg, D. F. (1981). *Linear Panel Analysis: Models of Quantitative Change,* Plenum Press, New York.
Lange, K. (1986). Cohabitation, convergence and environmental covariances. *Am. J. Med. Genet.* **24**:483–491.
Lange, K., Westlake, J., and Spence, M. A. (1976). Extensions to pedigree analysis. III. Variance components by the scoring method. *Ann. Hum. Genet.* **39**:485–491.
Lykken, D. T., Tellegen, A., and DeRubeis, R. (1978). Volunteer bias in twin research: The rule of two-thirds. *Soc. Biol.* **25**:1–9.
Lykken, D. T., McGue, M., and Tellegen, A. (1987). Recruitment bias in twin research: The rule of two-thirds reconsidered. *Behav. Genet.* **17**:343–362.

Martin, N. G., and Boomsma, D. I. (1989). Willingness to drive when drunk and personality: A twin study. *Behav. Genet.* **19:**97–111.
Martin, N. G., and Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity* **28:**79–95.
Martin, N. G., Eaves, L. J., Heath, A. C., Jardine, R., Feingold, L. M., and Eysenck, H. J. (1986). Transmission of social attitudes. *Proc. Natl. Acad. Sci.* **83:**4364–4368.
Martin, N. G., and Wilson, S. R. (1982). Bias in the estimation of heritability from truncated samples of twins. *Behav. Genet.* **12:**467–472.
McArdle, J., and Goldsmith, H. H. (1984). Structural equation modeling applied to the twin design: Comparative multivariate models of the WAIS. *Behav. Genet.* **14:**609.
Molenaar, P. C. M., and Boomsma, D. I. (1987). Application of nonlinear factor analysis to genotype-environment interaction. *Behav. Genet.* **17:**71–80.
Neale, M. C., and Martin, N. G. (1989). The effects of age, sex, and genotype on self-report drunkenness following a challenge dose of alcohol. *Behav. Genet.* **19:**63–78.
Neale, M. C., Eaves, L. J., Hewitt, J. K., and Kendler, K. S. (1989a). Bias in correlations from truncated samples of twins. *Behav. Genet.* **19:**(in press).
Neale, M. C., Heath, A. C., Hewitt, J. K., Eaves, L. J., and Fulker, D. W. (1989b). Fitting genetic models with LISREL: Hypothesis testing. *Behav. Genet.* **19:**37–49.
Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* **44:**443–460.
Olsson, U., Drasgow, F., and Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika* **47:**337–347.
Plomin, R., DeFries, J. C., and Loehlin, J. L. (1977). Genotype-environment interaction and correlation in the analysis of human variation. *Psychol. Bull.* **84:**309–322.
Rao, D. C., Morton, N. E., and Yee, S. (1976). Resolution of cultural and biological inheritance by path analysis. *Am. J. Hum. Genet.* **28:**228–242.
SAS Institute (1985). *SAS User's Guide: Statistics, Version 5 Edition*, SAS, Cary, N.C.
Scarr, S., and McCartney, K. (1983). How people make their own environments: A theory of genotype-environment effects. *Child. Dev.* **54:**424–435.
Silberg, J. L., Martin, N. G., and Heath, A. C. (1987). Genetic and environmental factors in primary dysmenorrhea and its relationship to anxiety, depression, and neuroticism. *Behav. Genet.* **17:**363–383.
Vogler, G. P., and Fulker, D. W. (1983). Familial resemblance for educational attainment. *Behav. Genet.* **13:**341–354.
Wright, S. (1968). *Evolution and the Genetics of Populations, Vol. 1*, University of Chicago Press, Chicago.