Acta Genet Med Gemellol 36:29-39 (1987)

Constrained Maximum Likelihood Analysis of Familial Resemblance of Twins and Their Parents

D.I. Boomsma[1], P.C.M. Molenaar[2]

[1] Department of Experimental Psychology, Free University of Amsterdam, and [2] Department of Psychology, University of Amsterdam, Netherlands

Abstract. When the univariate twin design is extended by including parents of twins, it is possible to assess additive genetic effects in the presence of assortative mating and genotype-environment correlation, the effects of parental influence, as well as the extent of residual shared environmental influences. The analysis of data obtained in such an extended twin design can be carried out by means of constrained maximum likelihood confirmatory factor analysis. Specifically, the structural model underlying this design can be represented as a LISREL model with nonlinear constraints. This representation offers the possibility to consider extended multivariate twin designs involving common genetic and environmental factors. The proposed method will be illustrated with applications to simulated and real data.

Key words: Parent-offspring design, LISREL, Constrained optimization, Twins

INTRODUCTION

The model for data obtained on twins and their parents as developed by Jencks et al [6] and Eaves et al [3] and subsequently further elaborated by Fulker [4] may be formulated as a LISREL model. Briefly summarized, the parent-offspring design (Fig. 1) allows for the estimation of additive genetic effects in the presence of assortative mating, the effects of parental influences, the correlation between genotype and environment and the effects of residual shared environmental influences among the offspring [4].

In this paper we want to show how the general parent-offspring model, where in this case the offspring consists of identical and fraternal twins, can be rewritten as a LISREL model. The advantages of this undertaking are twofold: first, there is a conceptual advan-

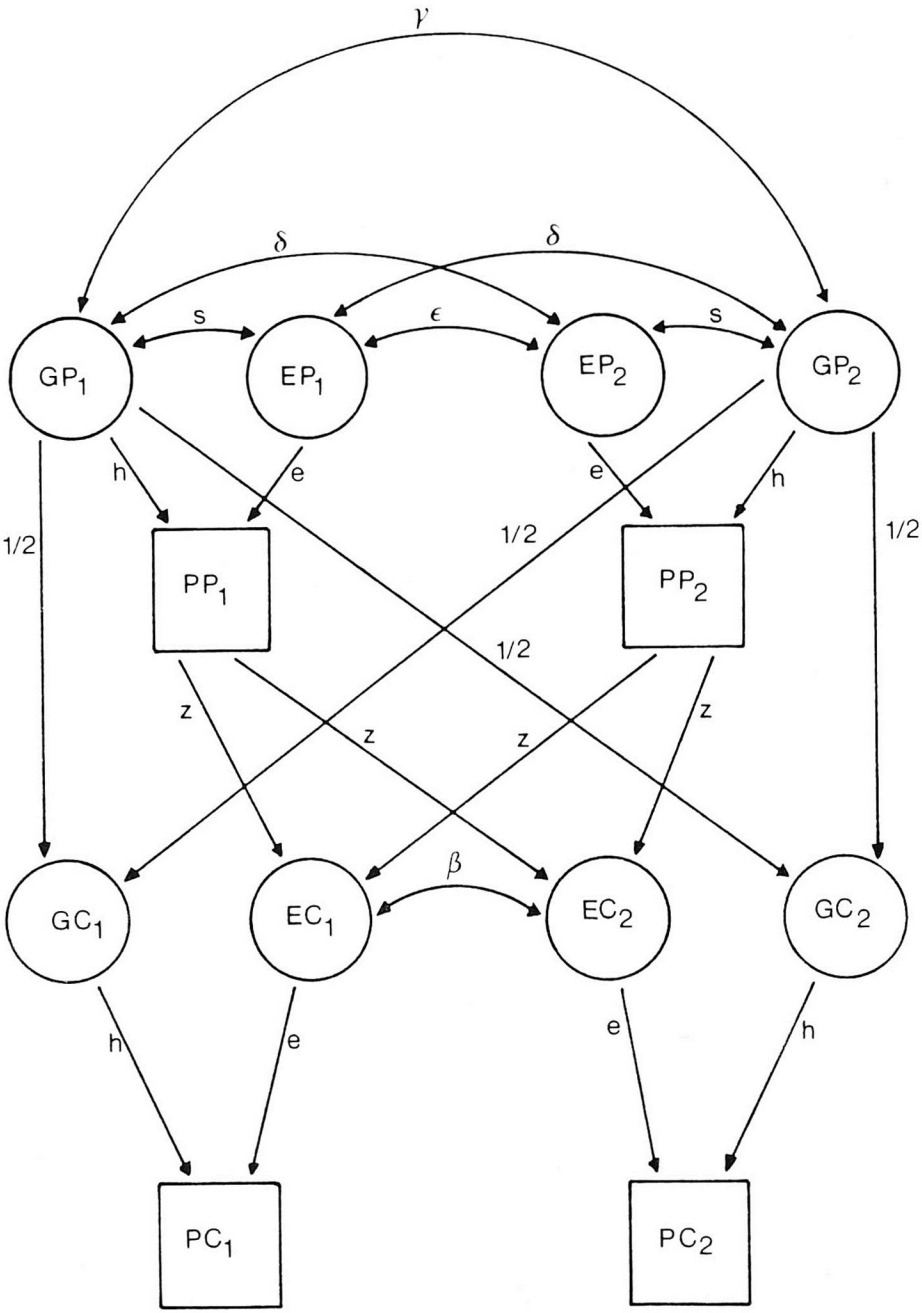

Fig. 1 - Parent-offspring model. PP_1 and PP_2 are parental phenotypes, PC_1 and PC_2 are children's phenotypes. G and E represent genetic and environmental influences. γ, ϵ and δ represent correlations induced by assortative mating and s is the correlation between G and E. Influence of parental phenotype on child's environment is z, and residual shared environment among offspring is β.

tage, when we want to analyze alternative models and also when the univariate model is extended to the case where parents and their offspring have been measured on more than one variable. Second, there is a numerical advantage, in terms of computer time and precision of the solution obtained. The univariate LISREL model will be illustrated with simulated and real data. Finally, it is shown, once the univariate LISREL model is specified, how this model can be extended to handle multivariate data sets.

The LISREL model was originally developed by Jöreskog [7] and is used to describe and estimate the unknown coefficients in a set of linear structural equations. The variables in the model are either directly observed or are unmeasured latent factors. In its most general form the model assumes that there is a causal structure among a set of latent variables and that the latent variables are underlying causes of the observed ones. Within a quantitative genetics framework, we assume that an observed phenotype is causally related to an underlying genotype and an environment that are not measured, and that there are relationships among parental genotypes and environments and the latent genotypes and environments of their children. Thus, relations among observed variables, expressed as correlations among parents and their offspring, are explained by an underlying model of latent factors. In LISREL this correlation matrix may be expressed as the product of several parameter matrices as follows (we here use a subset of the general LISREL model where $y = \Lambda\eta$ and $\eta = B\eta + \zeta$):

$$\Sigma = \Lambda B^{-1} \psi B^{-1\prime} \Lambda',$$

where Λ is a matrix containing the factor loadings of the observed on the latent variables, B contains the direct effects of latent variables on other latent variables and ψ contains the variances of the residuals of the latent factors and the covariances between these residuals. In the following the application of these parameter matrices to parent-offspring models is explained in more detail.

Λ contains the factor loadings of the observed phenotypes on the latent genotypic and environmental factors:

	GP_1	EP_1	GP_2	EP_2	GC_1	EC_1	GC_2	EC_2
PP_1	h	e	o	o	o	o	o	o
PP_2	o	o	h	e	o	o	o	o
PC_1	o	o	o	o	h	e	o	o
PC_2	o	o	o	o	o	o	h	e

From this matrix it can be seen that the phenotype of each subject is defined in terms of G and E by the structural equation:

$$P = hG + eE,$$

where h is the factor loading on the genotype and e on the environment.

B contains the coefficients that represent the influences of parental genotype on the genotype of the child and the influences of parental genotype and environment on the environment of the child:

	GP_1	EP_1	GP_2	EP_2	GC_1	EC_1	GC_2	EC_2
GP_1	1	o	o	o	o	o	o	o
EP_1	o	1	o	o	o	o	o	o
GP_2	o	o	1	o	o	o	o	o
EP_2	o	o	o	1	o	o	o	o
GC_1	0.5	o	0.5	o	1	o	o	o
EC_1	zh	ze	zh	ze	o	1	o	o
GC_2	0.5	o	0.5	o	o	o	1	o
EC_2	zh	ze	zh	ze	o	o	o	1

From this matrix it can be seen that the path from parental genotype to the genotype of the child is 0.5. The path from parental genotype to the environment of the offspring equals zh and from parental environment to the environment of the child ze. Thus the path from parental phenotype to child's environment equals z as required.

Finally, the diagonal elements in ψ contain the variances of the residuals of the latent factors:

	GP_1	EP_1	GP_2	EP_2	GC_1	EC_1	GC_2	EC_2
GP_1	1							
EP_1	s	1						
GP_2	γ	δ	1					
EP_2	δ	ϵ	s	1				
GC_1	o	o	o	o	$1-(0.5+0.5\gamma)$			
EC_1	o	o	o	o	o	$1-(2z^2(1+\mu))$		
GC_2	o	o	o	o	α	o	$1-(0.5+0.5\gamma)$	
EC_2	o	o	o	o	o	β	o	$1-(2z^2(1+\mu))$

These variances are one for the latent factors of the parents (assuming that all observed and latent variables are standardized with means zero and unit variance) since parental genotype and environment are not explained by any other factors in the model. The covariances between the latent factors in the parents, that is, s, γ, δ and ϵ, are also in ψ. The variances of the latent factors of the children are a little more complicated. Since the genotype of the children is a function of the genotype of the parents, the variance of the latent genotype of the offspring is built up as follows:

$$GC_1 = 0.5\ GP_1 + 0.5\ GP_2 + \text{genetic residual}$$
$$GC_2 = 0.5\ GP_1 + 0.5\ GP_2 + \text{genetic residual}$$

$$\text{Var (GC)} = 0.5 + 0.5\ \gamma + \text{Var (genetic residual)}$$

so that the variance of the genetic residuals equals

$$\text{Var (genetic residual)} = 1 - 0.5\ (1 + \gamma)$$

For identical twins the correlation between the genetic residuals is one, since their genotypes are identical. Thus, the covariance between their genetic residuals is equal to the variance of these residuals. The correlation of the genetic residuals of fraternal twins is zero, since their genetic resemblance is fully explained by the relationship with the parents.

The same reasoning applies to the variances and the correlations of the environmental residuals in the offspring. Part of the correlation between the environments of the children is explained by parental influences:

$$EC_1 = z\,PP_1 + z\,PP_2 + \text{environmental residual}$$
$$EC_2 = z\,PP_1 + z\,PP_2 + \text{environmental residual}$$
$$\text{Var (EC)} = 2z^2 + 2z^2\mu + \text{Var (environmental residual)}$$

Thus, the variances of the environmental residuals equal:

$$\text{Var (environmental residual)} = 1 - 2z^2(1 + \mu)$$

The covariance between the environmental residuals is represented by β, that is, the covariance between the childrens' environments after allowing for parental effects.

The LISREL model was developed by Jöreskog along with a computer program. We cannot, however, use the original LISREL program, since estimation of the unknown parameters in the parent-offspring model involves a set of nonlinear constraints (Table 1).

We therefore developed our own program, using subroutines written by the Numerical Algorithms Group [10]. (See [1] for a discussion of this constrained optimization technique). The program uses first derivatives of the constraints and of the likelihood function with respect to the parameters in the model. The use of such derivatives results in a good conditioned and faster converging optimization. Jöreskog [8] has published the explicit equations for the derivatives of the likelihood function, which therefore can be determined in a fully automatized manner.

Table 1. Constraints

$h^2 + e^2 + 2hes = 1$
$zh/h = ze/e$
$s = zh(1 + \gamma) + ze(s + \delta)$
$\text{Var (genetic residuals)} = 0.5 - 0.5\gamma$
$\text{Var (environmental residuals)} = 1 - 2(z^2 + z^2\mu) =$
$= 1 - 2[(zh)^2(1 + \gamma) + (ze)^2(1 + \epsilon) + 2(zh)(ze)(s + \delta)]$
$\mu = \gamma/(h + se)^2 = \epsilon/(e + sh)^2 = \delta/(h + se)(e + sh)$

SIMULATION

Correlation matrices for MZ and DZ families were simulated in two ways: by computing the exact expected correlation matrices and by generating data for 100 MZ and 100 DZ families, using IMSL subroutine FTGEN [5]. True parameter values are shown in Table 2.

For both cases several sets of starting values were considered. In the first case true parameter values were recovered exactly and χ^2 was zero. Input correlation matrices for the second simulation are in Table 2. The solution to these data was obtained by starting h at 0.3 and e at 0.9. As can be seen, the solution is close to the true parameter values.

Table 2. Simulated data

True and Estimated Parameter Values		
h	0.707	0.684
e	0.5	0.507
zh	0.236	0.248
ze	0.167	0.184
s	0.354	0.397
γ	0.156	0.185
δ	0.133	0.163
ϵ	0.113	0.143
β	0.025	0.096
res(g)	0.422	0.408
res(e)	0.733	0.675

Observed (upper triangle) and Expected (lower triangle) Correlations

MZ (n=100)				DZ (n=100)			
PP_1	PP_2	PC_1	PC_2	PP_1	PP_2	PC_1	PC_2
1.0	0.2097	0.6884	0.6885	1.0	0.2608	0.5338	0.6239
0.2360	1.0	0.5043	0.5074	0.2360	1.0	0.6739	0.5930
0.6014	0.6014	1.0	0.8516	0.6014	0.6014	1.0	0.6588
0.6014	0.6014	0.8512	1.0	0.6014	0.6014	0.6604	1.0

SHYNESS DATA

We also applied the program to shyness data collected by Dr. Rose in 144 MZ and 106 DZ twin families. These are the same data that were analyzed in Fulker's article. Table 3 shows the LISREL estimates and the estimates obtained by Fulker, as well as the observed and expected correlation matrices.

The two sets of parameter estimates are fairly similar, with the possible exception of the estimate for β. The total β is almost the same in both estimation procedures. But in

Table 3. Shyness data

Parameter Estimates		
	Fulker (1982)	LISREL
h	0.84	0.785
e	0.74	0.759
z_1	−0.29	−0.20
z_2	−0.19	−0.20
s	−0.20	−0.161
γ	0.096	0.091
δ	0.079	0.087
c	0.065	0.083
μ	0.200	0.208
β(total)	0.160	0.172
β	0.016	0.076

Observed (upper triangle) and Expected (lower triangle) Correlations

MZ (n=144)				DZ (n=106)			
PP_1	PP_2	PC_1	PC_2	PP_1	PP_2	PC_1	PC_2
1.0	0.18	0.1523	0.1723	1.0	0.2367	0.0387	0.0323
0.208	1.0	0.0597	0.1478	0.208	1.0	0.262	0.1771
0.1296	0.1296	1.0	0.525	0.1296	0.1296	1.0	0.2433
0.1296	0.1296	0.5245	1.0	0.1296	0.1296	0.2444	1.0

our case a smaller part of this correlation between the environments of the twins is induced by parental influences. Notice also that in our estimation procedure we did not separate the parental influences into maternal and paternal influences. When we did, however, this did not seem to result in a better fit. χ^2 with 8 df was 2.65. However, the χ^2 should be regarded as an approximation to the exact sampling distribution of the likelihood-ratio test. In constrained problems such as these, the precision of this test is unknown [eg, 2].

With this LISREL model it now has become quite easy to analyze parent-offspring data with different models, or to analyze data from more than two generations. For example: analyzing a model in which there is a relationship between parental environment and environment of the child instead of a path from parental phenotype to the environment of the offspring [eg 11] only involves changing the coefficients in the B matrix (and of course also changing the constraint for s).

MULTIVARIATE MODELS

Now that we have formulated a univariate LISREL model the extension to the multivariate case may be readily accomplished. We may, for example, consider a factor model in which the correlation between different variables is explained by their loadings on

common genetic and common environmental factors. For the correlations among twins we then get the model as implied in Fig. 2, where factor loadings on common factors are represented by capital letters, and loadings on the unique factors by lower case letters.

Each observed phenotype X and Y has loadings on a common genetic factor and on a unique genetic factor. The same is true for the environmental factors. At the bottom of the diagram are the unique factors for each variable. This part of the model is just a repetition of the univariate models for X and Y. The upper part of the diagram repeats this model for the common factors. The variance of X and the twin correlation become:

$$\text{Var}(X) = H_X^2 + E_X^2 + 2sH_X E_X + h_X^2 + e_X^2 + 2s_X h_X e_X$$

$$\text{Cor}(PC_1 X, PC_2 X) = \alpha H_X^2 + \beta E_X^2 + 2sH_X E_X + \alpha_X h_X^2 + \beta_X e_X^2 + 2s_X h_X e_X$$

The phenotypic correlation between X and Y and the twin cross-correlations are a function of the loadings of X and Y on the common factors:

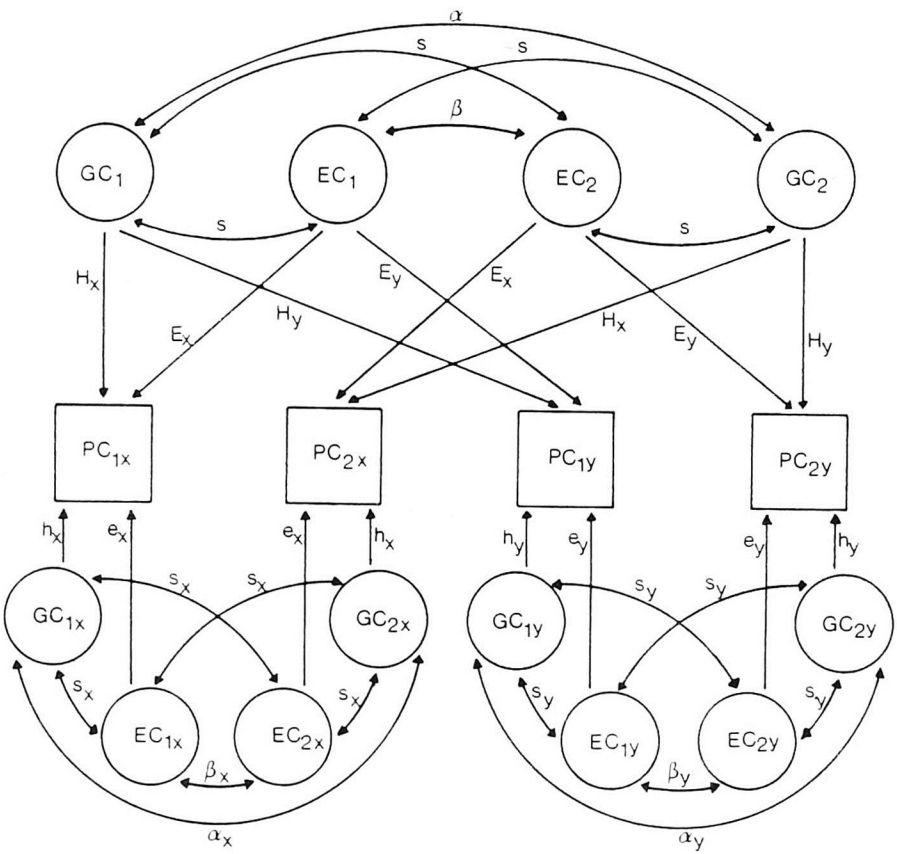

Fig. 2 - Bivariate factor model for twin correlations.

$$\text{Cor}(X,Y) = H_x H_y + E_x E_y + sH_x E_y + sH_y E_x$$

$$\text{Cor}(PC_1 X, PC_2 Y) = \alpha H_x H_y + \beta E_x E_y + sH_x E_y + sH_y E_x$$

where α is 1 for MZ and $0.5 + 0.5\gamma$ for DZ twins. The factor model for the relationship between parents is seen in Fig. 3.

Again the unique factors for each variable are at the bottom of the diagram and the common genetic and environmental factors in the upper part. The spouse correlation for X now may be expressed as follows:

$$\mu_{xx} = \gamma H_x^2 + \epsilon E_x^2 + (\delta_1 + \delta_2) H_x E_x + \gamma_x h_x^2 + \epsilon_x e_x^2 + 2\delta_x h_x e_x$$

and the spouse cross-correlations:

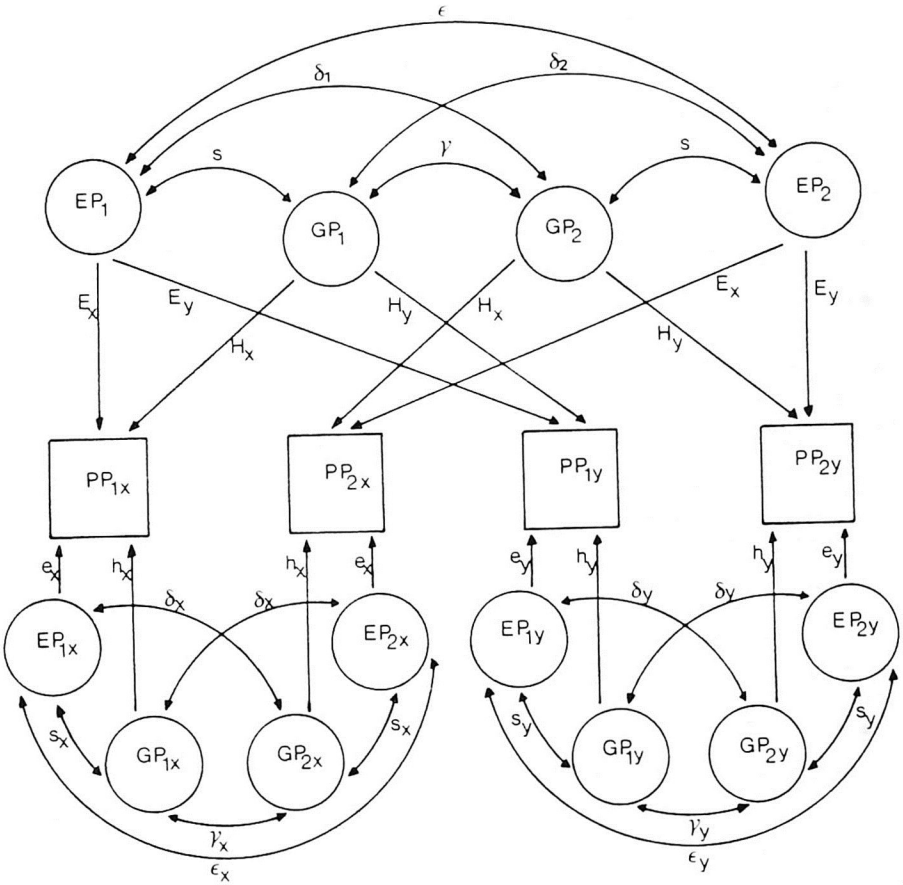

Fig. 3 - Bivariate factor model for spouse correlations.

$$\mu_{xy} = \gamma H_x H_y + \epsilon E_x E_y + \delta_2 H_x E_y + \delta_1 H_y E_x$$

$$\mu_{yx} = \gamma H_x H_y + \epsilon E_x E_y + \delta_1 H_x E_y + \delta_2 H_y E_x$$

Finally, the last two path diagrams are combined in Fig. 4 where the relationship of parents with one child is shown.

This last figure is the factor model as proposed by Martin and Eaves [9], but now extended to the parent offspring case. For example, the correlation between parent 1 and offspring for X becomes:

$$\text{Cor}(PP_1X, PC_1X) = (E_x Z_x + e_x z_x)(1 + \mu_x) + E_x Z_y(\text{cor}(X,Y) + \mu_{xy}) +$$
$$+ 1/2H_x^2(1 + \gamma) + 1/2h_x^2(1 + \gamma_x) + 1/2H_x E_x(s + \delta_1) + 1/2h_x e_x(s_x + \delta_x)$$

and the parent offspring cross-correlation between trait X in parent 1 and trait Y in the child:

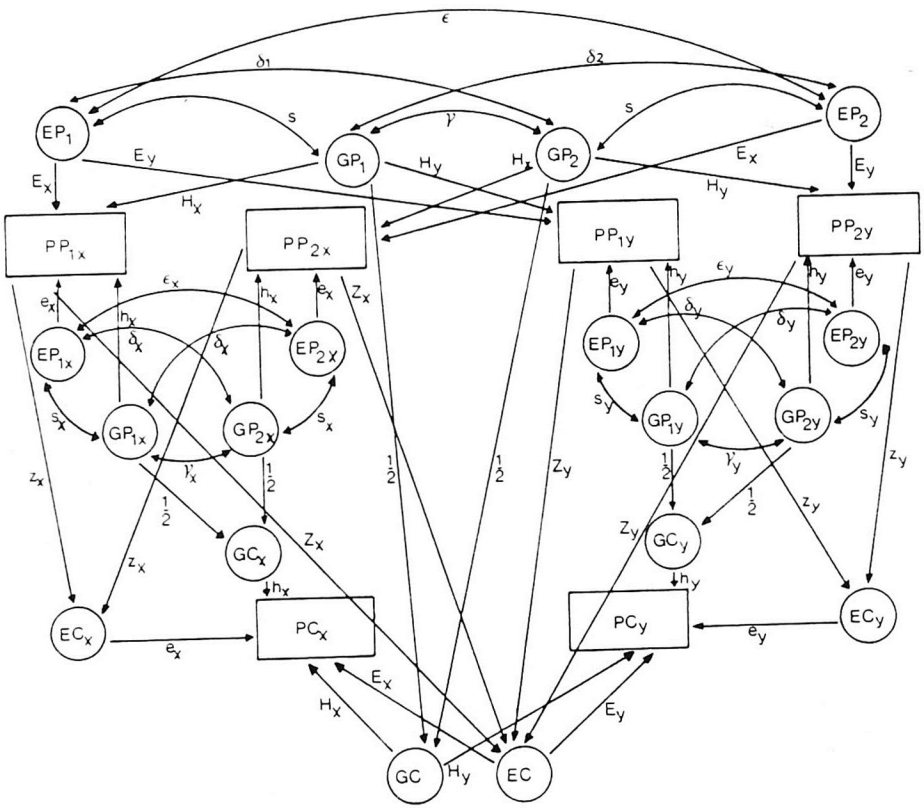

Fig. 4 - Bivariate factor model for parent-offspring correlations.

$$\text{Cor}(PP_1X, PC_1Y) = E_yZ_x(1 + \mu_x) + (E_yZ_y + e_yz_y)(\text{cor}(X,Y) + \mu_{xy}) +$$
$$+ 1/2H_xH_y(1 + \gamma) + 1/2H_yE_x(s + \delta_1)$$

The estimation of these parameters may be accomplished with the LISREL model as follows: In Λ are the factor loadings on the common and unique genetic and environmental factors, and in B the influences of parental genotype and environment on all latent genetic and environmental factors of the twins. ψ again contains the variances of the residuals of the latent factors and the covariances between these residuals.

In conclusion, the parent-offspring design together with the LISREL model just proposed, offers extensive possibilities for uni- and multivariate quantitative genetics analysis.

REFERENCES

1. Bertsekas DP (1982): Constrained Optimization and Lagrage Multiplier Methods. New York: Academic Press.
2. Box GEP, Tao GC (1973): Bayesian Inference in Statistical Analysis. Reading, Massachusetts: Addison-Wesley Publishing Company.
3. Eaves LJ, Last KA, Young PA, Martin NG (1978): Model-fitting approaches to the analysis of human behavior. Heredity 41:249-320.
4. Fulker DW (1982): Extension of the classical twin method. In Bonné-Tamir B (ed): Human Genetics: Part A, The Unfolding Genome. New York: Alan R. Liss, pp 395-406.
5. IMSL Inc. (1979): IMSL Library Reference Manual Edition 7. Houston Texas: IMSL Inc.
6. Jencks C, Smith M, Acland H, Bane MJ, Cohen D, Gintis H, Heyns B, Michelson S (1972): Inequality: A Reassessment of the Effect of Family and Schooling in America. New York: Basic-Books.
7. Jöreskog KG (1973): A general method for estimating a linear structural equation system. In Goldberger AS, Duncan OD (eds): Structural Equation Models in the Social Sciences. New York: Seminar Press, pp 85-112.
8. Jöreskog KG (1977): Structural equation models in the social sciences: specification, estimation and testing. In Krishnaiah PR (ed): Applications of Statistics. Amsterdam: North Holland Publishing Co, pp 265-287.
9. Martin NG, Eaves LJ (1977): The genetical analysis of covariance structure. Heredity 38:79-95.
10. Numerical-Algorithms-Group (1974): E04VAF in NAG Library Manual: Mark IV. Oxford: NAG Central Office Oxford University.
11. Vogler GP, Fulker DW (1983): Family resemblance for educational attainment. Behav Genet 13: 341-354.

Correspondence: Dorret Boomsma, Department of Experimental Psychology, Free University, De Boelelaan 1115, 1081 HV Amsterdam, Netherlands.