Acta Genet Med Gemellol 39:165-172 (1990)

Sixth International Congress
on Twin Studies

Simultaneous Genetic Analysis of Longitudinal Means and Covariance Structure Using the Simplex Model: Application to Repeatedly Measured Weight in a Sample of 164 Female Twins

S. Fischbein[1], P.C.M. Molenaar[2], D.I. Boomsma[3]

[1]Department of Educational Research, Institute of Education, Stockholm, Sweden; [2]Department of Psychology, University of Amsterdam, and [3]Department of Psychology, Free University, Amsterdam, The Netherlands

Abstract. The simultaneous analysis of means and covariance structures is applied to longitudinal twin data. Body weight was measured on six occasions in a sample of young female MZ and DZ twins. When average body weight at the first measurement occasion, as well as the increments in weight at later occasions, are specified in the genetic part of the model that also adequately explains the covariance structure, a good fit is obtained. In this application the increase in body weight at each occasion is weighted by the square root of the genetic variance innovation terms that represent the new genetic variance entering into the process.

Key words: Body weight, Longitudinal analysis, Means, Twin data

INTRODUCTION

In behavior genetics, human development is viewed from the perspective of individual differences. This perspective, with its emphasis on the association between phenotype and genetic and environmental differences, does not address the changes in the average growth curve, or, as it is also called, the species-specific developmental function. The emphasis on covariance structure analysis in human behavior genetics springs from the limitations inherent in studying behavior in a genetically heterogeneous population where usually no differential predictions can be made regarding first-degree statistics of relatives within or between successive generations

[7]. Certain hypotheses regarding the contribution of genetic and environmental factors to the average growth curve are however feasible and may be tested within the context of the standard analysis of covariance structure [6]. Specifically, the hypothesis can be tested that those genetic and environmental factors that account for the phenotypic variance also account for the phenotypic means. In a previous paper, this hypothesis was examined in the context of the static common factor model using multivariate twin data [2]. In the present paper we consider the hypothesis in the context of the Markov simplex model [5] using univariate twin data. The Markov simplex model is appropriate for the analysis of covariance structures characterized by time-dependent patternings of serial correlation (autocorrelation) such as typically arise in repeated measures designs.

As in Dolan et al [2], the hypothesis is tested by comparing the analysis of covariance structure with unconstrained means to the analysis with structured means. Sörbom has suggested a number of structural equation models including structured means for the comparison of groups regarding differences in latent means and covariance structure both at a single occasion [8,10,11] and longitudinally [9]. Because differences in latent means are estimated instead of absolute mean values, Sörbom's approach can be applied to measurements made on an interval scale, where the origin of measurement is arbitrary. Sörbom's models can be applied to multivariate twin data to test a number of hypotheses regarding sex-related differences in means and covariance structure. In the present paper, the discussion is limited to variables measured on a ratio scale, ie, a scale with a nonarbitrary origin.

THE GENETIC SIMPLEX MODEL

The genetic simplex model has been described by Boomsma and Molenaar [1] and in a more general approach including second-order common factors, by Eaves et al. [3]. Given a model with an additive genetic factor (G) and a specific environmental factor (E), the simplex model consists of a measurement model in which G and E are additively related to the phenotype Y. We make the nontrivial assumption that G and E are independent in their effect on Y so that the covariance of E and G is assumed to equal zero. A structural equation model defines the linear time-dependent relationship between the successive factors E and G. Given the same unit of measurement in the observed and the latent variables, the measurement model is:

$$Y_t = G_t + E_t + \epsilon_t\,, \quad (t = 1, p)\,, \tag{1}$$

where ϵ_t is a random measurement error term and p is the number of measurement occasions. The structural equation model is defined by the following equations:

$$\begin{aligned} G_t &= \beta(g)_t G_{t-1} + \zeta(g)_t\,, \quad (t = 2, p) \\ E_t &= \beta(e)_t E_{t-1} + \zeta(e)_t\,, \quad (t = 2, p)\,, \end{aligned} \tag{2}$$

https://doi.org/10.1017/S0001566000005390 Published online by Cambridge University Press

where the $\beta(g)_t$, $\beta(e)_t$, are autoregressive coefficients and the residual terms $\zeta(g)_t$ and $\zeta(e)_t$ are uncorrelated with the E_{t-1} and the G_{t-1}. On the first occasion the variances of G_1 and E_1 are estimated as independent parameters. The variances of G_t and E_t on all subsequent occasions are in part attributable to the immediately preceding occasion and in part independent. The stability arising from continuity in the effects of genetic and environmental factors on individual differences is modeled by the regression coefficients $\beta(g)_t$, and $\beta(e)_t$. The variances of the residual terms $\zeta(g)_t$ and $\zeta(e)_t$ are referred to as innovation variances, as they represent the changes in individual differences from occasion to occasion due to the inception of new genetic and environmental influences.

In most applications of the simplex model, the vector of phenotypic means is unconstrained, as the measurements are taken from the phenotypic means. All means, therefore, can be taken to equal zero:

$$E[G_t] = E[E_t] = E[Y_t] = 0\,, \quad (t = 1, p) \tag{3}$$

The description of the genetic simplex model has been restricted to an additive genetic and unshared environmental series to ease presentation. Nonadditive genetic or shared environmental series can be described in an analogous manner.

The Genetic Simplex Model with Structured Means

Structured means are introduced into the genetic simplex model as follows: $E[Y_t]$, the phenotypic mean at time t is the sum of the latent genetic and environmental means at t.

$$E[Y_t] = E[G_t] + E[E_t] \tag{4}$$

At the start of the time series, the latent means $E[G_1]$ and $E[E_1]$ are estimated independently. Subsequent latent means are partly attributable to the immediately preceding occasion $t - 1$ and partly independent:

$$\begin{aligned} E[G_t] &= \beta(g)_t E[G_{t-1}] + w_{gt} G_\Delta\,, \quad (t = 2, p) \\ E[E_t] &= \beta(e)_t E[E_{t-1}] + w_{et} E_\Delta\,, \quad (t = 2, p)\,. \end{aligned} \tag{5}$$

In equation (5) the autoregressive coefficients $\beta(g)_t$ and $\beta(e)_t$ now account for both continuity in the mean and the stability of individual differences. The terms G_Δ and E_Δ represent time invariant (hence unsubscripted) independent input at each occasion analogous to the random variance innovation terms. These parameters are weighted at each occasion by the terms w_{gt} and w_{et}. These weights cannot be estimated as free parameters because this would render model void as regards the means structure: the number of parameters associated directly with the means structure would then exceed the number of phenotypic means. The weights will be

assumed to be known for now. The question of the choice of weights is taken up below.

The question of the identification of the parameters associated with the means trend, $E[G_1]$, $E[E_1]$, G_Δ and E_Δ will be addressed by considering the linear equations relating the phenotypic means to the latent means. Those equations for (say) 5 occasions are (see also the Figure):

$$E[Y_1] = E[G_1] + E[E_1]$$

$$E[Y_2] = \beta g_1 E[G_1] + w_{g2} G_\Delta + \beta_{e1} E[E_1] + w_{e2} E_\Delta$$

$$E[Y_3] = \beta_{g2}\beta_{g1} E[G_1] + (w_{g2}\beta_{g2} + w_{g3}) G_\Delta + \beta_{e2}\beta_{e1} E[E_1] + (w_{e2}\beta_{e2} + w_{e3}) E_\Delta$$

$$E[Y_4] = \beta_{g3}\beta_{g2}\beta_{g1} E[G_1] + (w_{g2}\beta_{g3}\beta_{g2} + w_{g3}\beta_{g3} + w_{g4}) G_\Delta + \beta_{e3}\beta_{e2}\beta_{e1} E[E_1]$$
$$+ (w_{e2}\beta_{e3}\beta_{e2} + w_{e3}\beta_{e3} + w_{e4}) E_\Delta$$

$$E[Y_5] = \beta_{g4}\beta_{g3}\beta_{g2}\beta_{g1} E[G_1] + (w_{g2}\beta_{g4}\beta_{g3}\beta_{g2} + w_{g3}\beta_{g4}\beta_{g3} + w_{g4}\beta_{g4} + w_{g5}) G_\Delta$$
$$+ \beta_{e4}\beta_{e3}\beta_{e2}\beta_{e1} E[E_1] + (w_{e2}\beta_{e4}\beta_{e3}\beta_{e2} + w_{e3}\beta_{e4}\beta_{e3} + w_{e4}\beta_{e4} + w_{e5}) E_\Delta$$

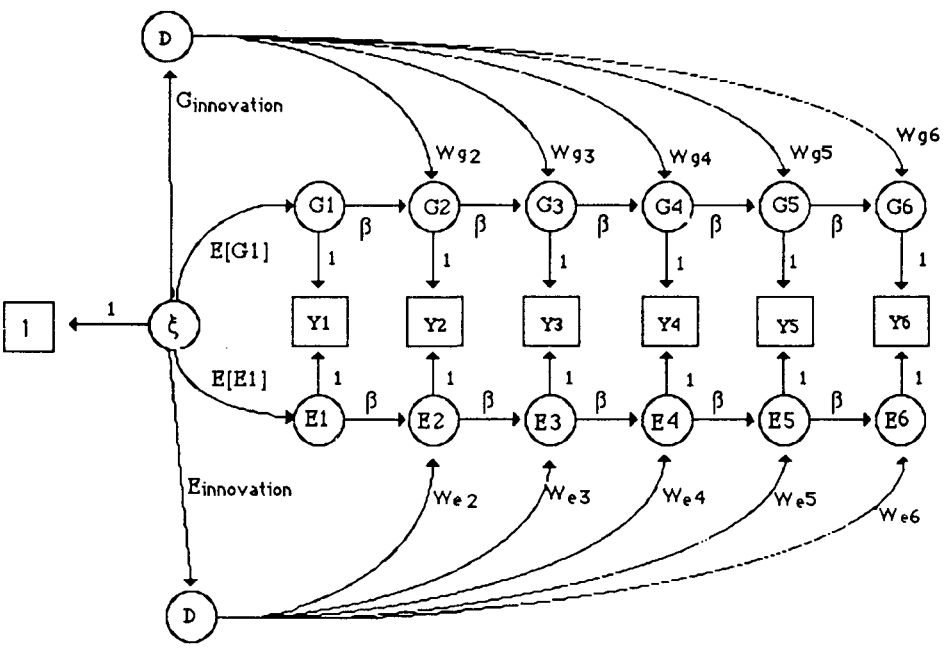

Figure. Graphic representation of a longitudinal simplex model where the same parameters account for both the covariance and means structure. See text for explanation of parameters.

https://doi.org/10.1017/S0001566000005390 Published online by Cambridge University Press

These equations can be rewritten in matrix notation:

$$y = Ac, \tag{6}$$

where, given a time series of length 5, A is a (5×4) matrix of known coefficients consisting of the sums of products of the autoregressive coefficients, c is a (4×1) vector consisting of the unknown parameters $E[G_1]$, $E[E_1]$, G_Δ and y is the (5×1) vector of known phenotypic means.

This formulation is helpful because it draws attention to the conditions that have to be met for the identification of the parameters in c: the number of equations must exceed the number of unknowns and the system of equations must be consistent. The first condition implies that overidentification of the four unknowns associated with the means requires a time series exceeding four occasions. Because the number of unknowns increases with the number of latent series, increasingly longer series is required for the overidentification of the parameters associated with mean structure. Generally, the length of the series must exceed the number of latent times two.

The second condition concerns the generalized inverse of the matrix A. If its inverse, $(A'A)^{-1}A$ does not exist, the system of equation is inconsistent and cannot be solved for the unknown parameters. Assuming the requirements relating to the length of the series have been met, the existence of the inverse of A depends on the values of the autoregressive coefficients, $\beta(g)_t$ and $\beta(e)_t$ and on the values of the weights w_{gt} and w_{et}. For instance, let us assume that the weights w_{gt} and w_{et} are fixed to equal 1. Then A can be partitioned into two $(p \times 2)$ submatrices, $A = [A1A2]$. When $A1 = A2$, the (4×4) matrix $A'A$ consists of 4 identical (2×2) submatrices and is obviously singular. Hence the autoregressive coefficients should not equal each other across the series given weights equal to one.

Less obviously, singularity also arises when the autoregressive coefficients are unequal across the series, but equal within each series, ie, $\beta(g)_t = x$, $\beta(e)_t = y$ and $x \neq y$, $t = 2, p$. This is not the only situation where singularity arises. However, given variable autoregressive coefficients and/or weights, the equations are likely to be consistent.

The Choice of Weights

There are a number of possibilities regarding the choice of the weights w_{gt} and w_{et}. As mentioned above, the weights can be set equal to one. This would mean that independent increments in the means are constant throughout the time series and that the link between the phenotypic means and covariance structure consists of the dual function of the autoregressive coefficients. This choice of weights will render the power of the model dependent upon the variation in the autoregressive coefficients. For example, given fairly constant autoregressive coefficients, the model will not be able to distinguish between the situation were the phenotypic mean trend is attributable to the environmental series and the situation were it is attributable

Table 1 - Description of sample

Mean of age (yr) at time of measurement (sd = 0.37)						
t	1	2	3	4	5	6
Mean	11.5	12.0	12.5	13.0	13.5	14.0
Covariance matrix, DZ twins (N = 50)						
	24.27					
	24.68	26.61				
	26.25	28.13	32.96			
	26.17	28.20	32.17	34.03		
	26.17	28.18	32.55	33.69	36.13	
	25.93	27.62	31.98	33.25	35.42	37.52
Mean	35.4	37.6	40.3	42.6	45.2	47.4
Covariance matrix, MZ twins (N = 32)						
	36.77					
	38.80	42.74				
	39.58	43.16	46.52			
	40.32	44.30	48.38	52.42		
	40.75	44.78	48.78	53.12	56.14	
	40.09	43.63	46.99	50.81	53.77	54.35
Mean	36.4	38.7	41.4	43.7	46.0	48.1

Table 2 - Structural equation modeling results

Model	df	χ^2	p
1 Covariance no means	(62)	65.97	0.341
2 Covariance means in E and G	(71)	71.33	0.466
3 Covariance means in G	(73)	71.58	0.525
4 Covariance means in E	(73)	81.52	0.168

Parameters related to covariance and means structure and standard errors (model 3)

$\beta(g)$1.03 (0.005)	1.03 (0.007)	1.02 (0.006)	1.02 (0.006)	1.01 (0.006)	
$\beta(e)$0.93 (0.050)	1.03 (0.084)	0.84 (0.072)	0.81 (0.071)	0.95 (0.085)	
var($G1$) and variances of genetic innovations					
21.60 (3.57)	1.37 (0.24)	2.31 (0.45)	1.37 (0.35)	1.71 (0.37)	2.16 (0.43)
var($G1$) and variances of environmental innovations					
3.45 (0.75)	0.32 (0.084)	0.87 (0.216)	0.91 (0.214)	0.72 (0.177)	0.87 (0.207)
h^2	0.86	0.88	0.87	0.88	0.91

Parameters related to means

$E(G1) = 35.77$ (1.296)
G-innovation (G_Δ) 1.006 (0.044).

https://doi.org/10.1017/S0001566000005390 Published online by Cambridge University Press

to the genetic series. Another possibility is to let the weights reflect differences in intertest interval, assuming a correspondence between chronological time and developmental time. Below, we have chosen as weights the standard deviations of the innovation terms $\zeta(e)_t$ and $\zeta(g)_t$. This choice of weights strengthens the association between the mean structure and the covariance structure and thus leads to a stronger hypothesis. Changes in the variances reflect developmental change which can be related to the introduction of new genetic and environmental influences. It seems reasonable that, given the hypothesis of a common process underlying means and covariance structure, these changes will have a proportionate effect on the average growth-curve.

ANALYSIS OF REPEATEDLY MEASURED WEIGHT IN A SAMPLE OF 164 FEMALE TWINS

The model described was applied to twin data which comprise repeatedly measured height and weight on 13 occasions in a sample of about 300 Swedish twin pairs [4]. We analyze a subset of the sample consisting of repeatedly measured weight on 6 occasions in a sample of 82 (50 DZ and 32 MZ) female twin pairs. Table 1 gives sample characteristics. Manova showed no effect of zygosity on weight, nor of zygosity × occasion.

The structural equation modeling results are detailed in Table 2. It turns out that a genetic simplex model comprising an additive genetic series and an unshared environmental series gives an adequate description of the covariance structure ($\chi^2 = 65.95$ df $= 62$, $p < 0.341$). After inclusion of mean trends in the G and E series, the fit again appears to be adequate ($\chi^2 = 71.33$, df $= 71$, $p < 0.46$). Next, two restricted models, one in which only G has a mean trend and one in which only E has a mean trend, were compared. The chi-squares, both on 73 degrees of freedom, equal 71.58 ($p < 0.52$) and 84.52 ($p < 0.16$), respectively. The model where the mean structure is modeled solely in G provides the best description of the data. The fit obtained in this case is only slightly worse than for the full model. Parameters estimates are given in Table 2. In all these analyses, the weights are obtained as the square roots of the genetic variance innovations. The remaining part of Table 2 gives parameter estimates for the most parsimonious model. Heritability is high at all occasions and the contribution of new variance entering into the process (the genetic innovations) is small.

DISCUSSION

The simultaneous analysis of means and covariance structure, as presented in this paper, is based upon the assumption that the genetic and environmental factors which contribute to individual differences also make a large contribution to the phenotypic means. The test of this assumption consist of comparing the results

from the analysis with and without means structure. In the illustration given above, this assumption appears to be tenable. The statistical power, however, is low in view of the relatively small sample of twins. Also, as mentioned above, the overidentification of the parameters associated solely with the mean structure is based in part on the length of the time series. Viewed in this light, a time series consisting of six occasions is on the short side. The estimation of absolute latent mean values requires variables measured on a scale with a meaningful origin, such as weight or height. The application of the present model to data measured on, eg, an interval scale, may require a scale parameter to accommodate arbitrary changes in measurement origin.

REFERENCES

1. Boomsma DI, Molenaar PCM (1987): The genetic analysis of repeated measures I: Simplex models. Behav Genet 17:111-123.
2. Dolan CV, Molenaar PCM, Boomsma DI (1989): LISREL analysis of twin data with structured means. Behav Genet 19:51-62.
3. Eaves LJ, Long J, Heath AC (1986): A theory of developmental change in quantitative phenotypes applied to cognitive development. Behav Genet 16:143-162.
4. Fischbein S (1977): Intra-pair similarity in physical growth of monozygotic and of dizygotic twins during puberty. Ann Hum Biol 4:417-430.
5. Jöreskog KG (1970): Estimation and testing of simplex models. Br J Math Stat Psychol 23:121-145.
6. Martin NG, Eaves LJ (1977): The genetical analysis of covariance structure. Heredity 38:79-95.
7. Mather K, Jinks JL (1977): Introduction to Biometrical Genetics. New York: Cornell University Press.
8. Sörbom D (1974): A general method for studying differences in factor means and factor structures between groups. Br J Math Stat Psychol 27:229-239.
9. Sörbom D (1976): A statistical model for the measurement of change in true scores. In DNM deGruijter, JLTh vanderKamp (eds): Advances in Pyschological and Educational Measurement. New York: Wiley, pp 159-169.
10. Sörbom D (1978): An alternative to the methodology for analysis of covariance. Psychometrika, 43:381-396.
11. Sörbom D (1981): Structural equation models with structured means. In KG Jöreskog, H Wold (eds): Systems under Indirect Observation: Causality, Structure and Prediction. Amsterdam: North-Holland Publishing Co.

Correspondence: Dr. Siv Fischbein, Department of Educational Research, Institute of Education, Box 34103, S-100 26 Stockholm, Sweden.