12 Genetic and environmental factors in a developmental perspective

PETER C. M. MOLENAAR, DORRET I. BOOMSMA AND CONOR V. DOLAN

INTRODUCTION

Developmental behavior genetics is concerned with the diverse ways in which genetic and environmental processes are involved in changes as well as continuity in development (Plomin, 1986; DeFries & Fulker, 1986). During ontogenesis, observed (phenotypic) change of a quantitative character may be due to distinct subsets of genes turning on and off, whereas continuity, on the other hand, may be caused by stable environmental causes. In contrast to the popular point of view, then, genetically determined characters are not always stable, nor are longitudinally stable characters always due to hereditary influences. Only through carefully designed longitudinal investigation of phenotypic changes in genetically related individuals can the dynamic patterns of genetic and environmental influences be disentangled.

In the following we shall mainly be concerned with a particular type of genetic model for the analysis of longitudinal phenotypic data, namely the simplex model (Jöreskog, 1970). The genetic simplex model is a genuine time series model and therefore can explain the characteristic time-dependent patternings of serial correlation (autocorrelation) as observed in longitudinal studies. It was already shown by Cronbach (1967) that common factor analysis of autocorrelation matrices will yield spurious, i.e. invalid, results. Consequently, recent efforts in the genetic modeling of longitudinal data have put particular emphasis on the elaboration of simplex models in this context (Boomsma & Molenaar, 1987*a*; Eaves, Hewitt & Heath, 1988).

Presently, we will introduce two important generalizations of the genetic simplex model. Firstly, we will consider the estimation of latent time-dependent profiles of genetic and environmental influences for each individual subject. Behavior geneticists never considered the genetic and environmental scores of single subjects. Yet in a mathematical–statistical sense, genetic single-subject scores are similar to factor scores and therefore can be obtained by means of standard techniques for the estimation of factor scores. This approach would

seem particularly worthwhile for genetic research into deviances and the consequent possibilities of carrying out preventive interventions (cf. Mednick *et al.*, 1983). In fact, Boomsma, Molenaar & Orlebeke (in press) show that for two subjects with the same phenotypic pattern of high blood pressure, the high blood-pressure of one subject is associated with a high genetic score, whereas the blood pressure of the other subject is associated with a high environmental score. Given such information, any interventions aimed at normalizing blood pressure could be entirely different for each of these subjects. We will show that the estimation of single-subject scores can be extended to the genetic simplex model, thus yielding longitudinal trajectories of intraindividual variation of genetic and environmental scores. This will be shown for the most difficult case in which only univariate phenotypic measurements are available at each time point. Not only is this the more relevant case from the application-oriented point of view, but its success will guarantee the success of analogous multivariate cases because the availability of multivariate phenotypic measurements at each time point will always yield much better conditioned estimates of the latent time-dependent profiles concerned.

Secondly, the simplex model will be generalized to include latent genetic and environmental trends. Behavior geneticists do not usually consider the role of genetic and environmental influences on both the stability and change of individual differences as well as the species-specific developmental function or average growth curve. Yet, these two aspects of longitudinal data are complementary and not necessarily independent in understanding development. This approach would seem particularly useful for genetic analyses of unstandardized longitudinal data pertaining to, for example biological or ability development. In particular when multivariate phenotypic measurements are available at each time point, this combined analysis of interindividual and intraindividual changes in the means and variation of genetic and environmental scores can yield important explanatory evidence concerning the maturational and learning processes underlying developmental trait patterns (Baltes, personal communication; see Baltes & Nesselroade, 1973). For the same reasons alluded to earlier, we will show that the simplex model for univariate phenotypic measurements can be reliably generalized to include latent genetic and environmental trends.

In the following, the validity of both the estimation of individual genetic and environmental profiles as well as the analysis of the genetic simplex model with structured means will be shown by means of simulation studies. In this way one can directly compare the results obtained with a finite sample of longitudinal data and the true model used in the simulation of the data. In the closing section, we will present

the results of an illustrative application of the genetic simplex model, including the proposed generalizations, to real data.

THE GENETIC SIMPLEX

In this section we will discuss the basic genetic simplex model for univariate repeated measurements in a heuristic, application-oriented way. After the presentation of the defining equations, the genetic simplex model for monozygotic (MZ) and dizygotic (DZ) twin data will be reformulated as a particular multigroup LISREL model (see Jöreskog & Sörbom, 1986). As the LISREL program is now widely available, this will facilitate regular applications of the proposed approaches. For ease of presentation, several simplifying assumptions will be made that enable concentration on the main issues. Specifically, interactions and covariances between genetic and environmental influences are assumed to be absent, as are interactions between alleles at loci (i.e., no genetic dominance effects) and interactions between loci (i.e., no epistasis). In addition, it is assumed that assortative mating does not occur, while only one particular type of environmental influences will be considered, namely those influences that are not shared by members of a family. These assumptions do not imply that the effects concerned cannot be detected or modeled: the analysis of genotype–environment interactions and correlations when environmental measures are available is presented in, for example, Plomin, DeFries & Fulker (1988), the application of nonlinear factor analysis to genotype–environment interaction is discussed in Molenaar & Boomsma (1987*a*) and Molenaar, Boomsma, Neeleman & Dolan (in press), a theoretical model of assortative mating and cultural transmission is given in Fulker (1988), while shared environmental influences and dominance can straightforwardly be included in the basic genetic model. Only some of the effects (e.g., epistasis) may be difficult to quantify in human research (Eaves, 1977) and represent cases where theory outruns the available data (Eaves and Young, 1981).

Before presenting our basic model we will briefly discuss its underlying assumptions. The genetic simplex model is a particular instance of the general covariance structure model (Jöreskog & Sörbom, 1986) and therefore obeys the same assumptions as the latter model. That is, for maximum likelihood (ML) estimation to apply it is assumed that the vector of repeated phenotypic observations has a multivariate normal distribution. Browne & Shapiro (1988) show that ML estimation in covariance structure models is quite robust against departures from multivariate normality. A similar result regarding the robustness of ML estimation in structural models of both covariances and means has been obtained by Gourieroux, Monfort & Trognon (1984): if the distribution

of the observed phenotypes belongs to the general class of linear exponential distributions and if the model for the means is correctly specified, then ML estimates are consistent and asymptotically normally distributed. The latter result bears on the genetic simplex model including latent genetic and environmental trends.

Another issue concerns the number of MZ or DZ twin pairs, or more generally the number of pedigrees, which is required. The minimum number of pedigrees is directly related to the number of repeated phenotypic measurements. If the number of repeated measurements increases, the dimension of the estimated matrices of mean cross-products (see below) also increases. In order to guarantee the required positive-definiteness of the latter matrices, the number of pedigrees then also has to increase to a value that is strictly larger than the number of repeated measurements. Specifically, if we have T univariate phenotypic measurements then the number of MZ twin pairs and the number of DZ twin pairs each have to be larger than T.

Turning to our basic model, if in a longitudinal design a univariate phenotype P is observed at $t = 1, \ldots, T$ time points, the following dynamic genetic model can be considered:

$$P(t) = G(t) + E(t) + e(t), \quad t = 1, \ldots, T, \tag{1}$$

where $P(t)$, $G(t)$ and $E(t)$ represent phenotypic, genetic and (non-shared) environmental time series, respectively, while $e(t)$ denotes a residual series. An advantageous parametric time series model for the genetic and environmental influences is then given by a first-order autoregression:

$$G(t) = \beta_G(t)G(t-1) + \zeta_G(t). \tag{2}$$

$$E(t) = \beta_E(t)E(t-1) + \zeta_E(t). \tag{3}$$

The autoregressive coefficient $\beta_G(t)$ (similar remarks apply to $\beta_E(t)$) is a measure of the amount of genetic variation at time point $t-1$ that is transmitted to time point t and therefore is associated with the stability (cf. Rudinger, Andres and Rietz, Chapter 13 of this volume) or, stated otherwise, the memory of the genetic process between $t-1$ and t. The so-called innovation $\zeta_G(t)$ (similar remarks apply to $\zeta_E(t)$) denotes the effects of new genes turned on at time point t and will therefore lower the stability of the genetic process between $t-1$ and t.

Identification of the genetic simplex model defined by equations (1)–(3) requires data from genetically related individuals, such as twins. The genetic relations between family members provide information with respect to the latent factor series $G(t)$ and $E(t)$. In particular, monozygotic (MZ) twin pairs have all their genetic material in common and hence $\mathrm{cor}[G_1(t), G_2(t)] = 1$. The instantaneous genetic correlation

for dizygotic (DZ) twins is 0.5 on average, i.e. $\text{cor}[G_1(t), G_2(t)] = 0.5$, as is the genetic correlation among ordinary siblings. This information can be used to arrive at an identified model with more than one latent process, even if the observed phenotypic time series is univariate. Specifically, longitudinal data from twins open up the possibility of decomposing a univariate phenotypic series into a genetic and an environmental series which may each behave quite differently in time. This may not always be immediately evident from the observed series (cf. Heath, Jardine, Eaves & Martin, 1988). For example, variance due to $G(t)$ may increase over time, while variance due to $E(t)$ may decrease. Increased genetic variance could be due to the amplification of existing genetic differences, as expressed by large values of $\beta_G(t)$, or to new genetic variance coming into play (new genes 'turned on'), as expressed by large values of $\zeta_G(t)$.

We will now reformulate the genetic simplex model defined by equations (1)–(3) as a LISREL model, while restricting attention to longitudinal MZ and DZ twin data. Furthermore, it is assumed for ease of presentation that the residual series $e(t)$ in equation (1) is absent. First we introduce the $(T \times T)$ diagonal matrices Ψ_G and Ψ_E with time-dependent variances $\Psi_G(t, t) = \text{var}[\zeta_G(t)]$ and $\Psi_E(t, t) = \text{var}[\zeta_E(t)]$ as the tth diagonal elements, respectively $(t = 1, \ldots, T)$. Notice that the first diagonal elements of Ψ_G (similar remarks apply to Ψ_E) is special in that no autoregression for $G(1)$ can be formulated at time point 1. Consequently, at $t = 1$ equation (2) becomes simply $G(1) = \zeta_G(1)$, and $\Psi_G(1, 1) = \text{var}[\zeta_G(1)]$ now denotes the genetic variance at time point 1. Only for $t > 1$ do the diagonal elements of $\Psi_G(t, t)$ denote the variances of the genetic innovations. Next, the $T \times T$ matrices B_G^* and B_E^* are introduced. For instance, B_G^* has the following pattern:

$$B_G^* = \begin{bmatrix} 0 & & \cdots & & 0 & 0 \\ \beta_G(2) & 0 & \cdots & & 0 & 0 \\ 0 & \beta_G(3) & 0 & \cdots & & \\ \vdots & 0 & & \ddots & & \vdots \\ & \vdots & \vdots & & 0 & \\ 0 & 0 & \cdots & 0 & \beta_G(T) & 0 \end{bmatrix}$$

where $\beta_G(2), \ldots, \beta_G(T)$ are the autoregressive coefficients in equation (2) B_E^* has the same pattern. Reformulation of the genetic simplex model for MZ and DZ twin pairs is now accomplished by specification of the expected structure of the $(T \times T)$ matrices of mean cross-products between and within these twin pairs. Let Σ_{MZB} and Σ_{MZW} denote the expected structure of the matrices of mean cross-products between and within MZ pairs, respectively. Before specifying the expected structure concerned, we will first consider their respective estimates S_{MZB} and S_{MZM} in order to elucidate the nature of these matrices of mean

cross-products. For example, S_{MZB} is obtained as the hypothesis matrix from a one-way multivariate analysis of variance, where the factor has as many levels as there are MZ twin pairs and where the $(T \times 1)$ vector $P = [(P(1), \ldots, P(T)]'$ derived from equation (1) is the dependent variable (the prime $'$ denotes the transpose). S_{MZB} is obtained as the error matrix from the same multivariate analysis of variance. Accordingly, the longitudinal data from the MZ and DZ twin groups are summarized in the four matrices S_{MZB}, S_{MZW}, S_{DZB} and S_{DZW}. The expected structure of these matrices for the genetic simplex model is:

$$\left.\begin{aligned} \Sigma_{MZB} &= \Lambda_{MZB} B_G \Psi_G B_G' \Lambda_{MZB}' + B_E \Psi_E B_E' \\ \Sigma_{MZW} &= B_E \Psi_E B_E' \\ \Sigma_{DZB} &= \Lambda_{DZB} B_G \Psi_G B_G' \Lambda_{DZB}' + B_E \Psi_E B_E' \\ \Sigma_{DZW} &= \Lambda_{DZW} B_G \Psi_G B_G' \Lambda_{DZW}' + B_E \Psi_E B_E' \end{aligned}\right\} \quad (4)$$

In these expressions, $B_G = (I - B_G^*)^{-1}$, where I denotes the $(T \times T)$ unity matrix. The general structure of the above expectations is firmly based on the genetical foundation underlying the biometric model of continuous variation (Mather & Jinks, 1977). In particular, the $(T \times T)$ matrices Λ_{MZB}, Λ_{DZB} and Λ_{DZW} are diagonal matrices with fixed loadings derived from the biometrical model: $\sqrt{2}$ for Λ_{MZB}, $\sqrt{1.5}$ for Λ_{DZB}, and $\sqrt{0.5}$ for Λ_{DZW}. In view of the absence of genetic influences within MZ twins, Λ_{MZW} is a zero matrix.

The four matrix equations given in (4) are fitted to the associated estimates of the matrices of mean cross-products using a four-group LISREL design with parameters to be estimated invariant across groups (for further details, see Boomsma & Molenaar, 1987*a*). In addition to the estimates of all the parameters in Ψ_G, B_G^*, Ψ_E and B_E^*, the standard LISREL solution also gives the matrices of genetic and environmental correlations between time points. Moreover, LISREL supplies the so-called factor scores regression matrix which can be used to estimate the latent $G(t)$ and $E(t)$ time series for each individual subject. In the next section we will discuss a simulation study of the validity of individual $G(t)$ and $E(t)$ trajectories thus obtained.

A SIMULATION STUDY OF INDIVIDUAL GENETIC PROFILES

In order to study the validity and reliability of estimates of the $G(t)$ and $E(t)$ time series for individual subjects, the genetic simplex model will be applied to four sets of simulated data. Each of the factor scores regression matrices thus obtained will be applied to the corresponding data set and the estimated individual $G(t)$ and $E(t)$ trajectories will then

be compared with the associated 'true' trajectories used in the simulation. Each longitudinal data set comprises univariate phenotypic observations at $T = 10$ time points for $N = 100$ MZ twin pairs and $N = 100$ DZ twin pairs. At each time point the variance of $G(t)$ as well as that of $E(t)$ was fixed at $\text{var}[G(t)] = \text{var}[E(t)] = 100$, whence the heritability at each point equals $h^2(t) = 0.5$, $t = 1, \ldots, T$. The four data sets are further characterized as follows:

1. $\beta_G(t) = 0$ and $\beta_E(t) = 0$; consequently, the phenotypic series $P(t)$ lacks autocorrelation.
2. $\beta_G(t) = 0.8$ and $\beta_E(t) = 0$; the genetic series $G(t)$ is autocorrelated, whereas $E(t)$ is not autocorrelated.
3. $\beta_G(t) = 0$ and $\beta_E(t) = 0.8$; $G(t)$ is not autocorrelated whereas $E(t)$ is autocorrelated.
4. $\beta_G(t) = 0.8$ and $\beta_E(t) = 0.8$; both $G(t)$ and $E(t)$ are autocorrelated.

A detailed description of the simulation algorithm is given in Molenaar & Boomsma (1987*b*).

The fit of the genetic simplex model to each of the longitudinal data sets is good: chi-squared goodness-of-fit is 190.37 ($p = 0.320$) for data set 1, 171.07 ($p = 0.709$) for data set 2, 187.70 ($p = 0.370$) for data set 3, and 182.41 ($p = 0.477$) for data set 4 (d.f. = 182 in each case). As the combination of an autocorrelated series and an uncorrelated series yields a phenotypic series that has an intricate pattern of autocorrelation (cf. Granger & Morris, 1976), and because this particular combination has been the subject of many theoretical approaches in mathematical signal analysis, we will present more detailed results for data set 2. Table 12.1(a) shows the estimates of $\beta_G(t)$, $\beta_E(t)$, $\Psi_G(t, t)$ and $\Psi_E(t, t)$ for each time point t. Notice that the true values are: $\beta_G(t) = 0.8$, $\beta_E(t) = 0$, $\Psi_G(1, 1) = 100$ at $t = 1$, $\Psi_G(t, t) = 36$ for $t > 1$, and $\Psi_E(t, t) = 100$ for each t. The associated standard errors are, on the average 0.117 for $\beta_G(t)$, 0.72 for $\beta_E(t)$, 12.143 for $\Psi_G(t, t)$ and 11.925 for $\Psi_E(t, t)$.

Using the factor scores regression matrices for data set 2, the $G(t)$ and $E(t)$ time series for each of the 400 individuals subjects in this sample can be estimated and compared with the true individual series used in the simulation. We can now answer several important questions. Firstly, what is the reliability of the estimates of individual $G(t)$ and $E(t)$ series? The answer can be obtained from Table 12.1(b) which shows the standard errors of the estimated $G(t)$ and $E(t)$ series for MZ and DZ twins. Remember that the variance of the true $G(t)$ and $E(t)$ scores is: $\text{var}[G(t)] = \text{var}[E(t)] = 100$. As the standard error of estimated $G(t)$ scores of MZ twins at each time point is about 5, this implies that $G(t)$ scores which differ by at least 1 standard deviation in the original metric can be reliably distinguished at usual significance levels. Similar remarks apply to the reliability of the $E(t)$ scores of the MZ twins as well as the scores of DZ twins (the standard errors with the DZ twins are only slightly larger). Secondly, we can answer the question whether estimated $G(t)$ and $E(t)$ scores are valid indicators of the associated true scores. The answer can be obtained from Table 12.1(c) which shows for each time point t the correlation between estimated and true $G(t)$ scores as well as $E(t)$ scores for both MZ and DZ twins. It turns out that these correlations all lie in the neighbourhood of 0.8 and therefore estimated scores can be considered to yield valid indications of the corresponding true scores at each time point. Thirdly, we can answer the question whether the dynamic structure of the true $G(t)$ and $E(t)$ series is faithfully reflected by the estimated individual trajectories. The answer to this question can be obtained from Table 12.1(d) which shows the autocorrelation between times 1 and 2, times 2 and 3, etc., of the estimated individual $G(t)$ and $E(t)$ trajectories of MZ and DZ twins. The true autocorrelation between these pairs of neighboring time points is 0.8 for the $G(t)$ series and 0 for the $E(t)$ series. Table 12.1(d) shows that this dynamic structure is indeed recovered by the estimated individual trajectories. Finally, we would like to know the cross-correlation between estimated $G(t)$ and $E(t)$ scores. According to the genetic simplex model this cross-correlation is expected to be zero. On the other hand, however, each *univariate* $P(t)$ series is decomposed into two trajectories: an estimated $G(t)$ and $E(t)$ series. Consequently, the effective cross-correlation between the latter series will deviate from zero. Indeed, it is found that the cross-correlation between estimated $G(t)$ and $E(t)$ series, averaged over 10 time points, is 0.35 for MZ twins and 0.58 for DZ twins.

In conclusion, the results of our simulation study show that on the basis of the genetic simplex model for univariate phenotypic time series, the latent genetic and environmental series of each individual subject can be reliably and validly estimated. The detailed results for data sets 1, 3 and 4 are similar to the results presented above for data set 2. The genetic simplex model has been fitted by means of the LISREL program, which also yields the factor scores regression matrices used in the estimation of the individual genetic and environmental trajectories. Hence, the proposed approach can be readily implemented for the kinds of longitudinal research purposes alluded to earlier.

BEHAVIOR GENETIC SIMPLEX WITH STRUCTURED MEANS

So far the focus has been on detecting the genetic and environmental influences on the stability and change of inter-individual differences during development. The role of genetic and environmental influences on the species-specific function or average growth curve does not

Table 12.1. *Results with data set 2*

(a) *Parameter estimates*

t(ime):	1	2	3	4	5	6	7	8	9	10
$\beta_E(t)$		0.799	0.879	0.778	0.843	0.784	0.839	0.879	0.771	0.727
$\beta_E(t)$		0.040	0.034	0.079	0.079	0.096	0.157	0.090	−0.104	0.090
$\Psi_G(t,t)$	99.89	33.22	40.16	33.72	25.59	30.36	53.83	31.12	17.74	31.52
$\Psi_E(t,t)$	102.0	86.28	104.42	98.89	95.76	113.1	92.4	114.1	107.3	106.9

(b) *Standard errors of estimated G(t) and E(t)*

t(ime):	1	2	3	4	5	6	7	8	9	10
MZ	5.81	4.96	5.30	5.09	4.88	5.09	5.48	5.39	4.48	4.92
DZ	6.86	5.98	6.32	6.02	5.77	5.91	6.53	6.42	5.31	5.64

(c) *MZ and DZ correlations true-estimated*

t(ime):	1	2	3	4	5	6	7	8	9	10
MZ										
G(*t*)	0.783	0.879	0.851	0.847	0.868	0.858	0.848	0.832	0.815	0.822
E(*t*)	0.783	0.826	0.852	0.837	0.831	0.857	0.838	0.835	0.854	0.798
DZ										
G(*t*)	0.759	0.813	0.832	0.813	0.834	0.780	0.807	0.813	0.789	0.776
E(*t*)	0.734	0.747	0.809	0.793	0.780	0.777	0.806	0.835	0.753	0.781

(d) *Recovered lag 1 autocorrelations of G(t) and E(t)*

lag:	*t*1 − *t*2	*t*2 − *t*3	*t*3 − *t*4	*t*4 − *t*5	*t*5 − *t*6	*t*6 − *t*7	*t*7 − *t*8	*t*8 − *t*9	*t*9 − *t*10
MZ twins									
G(*t*)	0.763	0.788	0.828	0.858	0.845	0.730	0.855	0.894	0.831
E(*t*)	−0.021	0.002	0.088	0.051	0.131	0.176	0.153	−0.140	0.073
DZ twins									
G(*t*)	0.755	0.788	0.832	0.864	0.859	0.729	0.838	0.886	0.835
E(*t*)	0.089	−0.013	0.056	0.116	0.048	0.206	0.039	−0.109	0.119

usually feature in behavior genetic studies of development (see McArdle, 1986, for one attempt to analyze such influences by means of a common-factor model). These two aspects of longitudinal data, however, are complementary and not necessarily independent in understanding development (McCall, 1981). Indeed, Thomas (1980) has noted strikingly large (usually positive) correlations between repeatedly measured means and standard deviations indicating that changes in one are usually accompanied by similar changes in the other. These correlations, which are observed for a variety of physical and psychological variables, encourage the idea that means and variances may in some fashion be related.

In the present section an extension of the genetic simplex model is presented to analyze phenotypic means and covariance structure simultaneously within the context of developmental behavior genetics. The extended simplex is suggested as a way to examine the relationship between means and covariance structure by modeling these with a common set of parameters. This implies that the objective of the simultaneous analysis of means and covariance structure is to test the hypothesis that they can be attributed to a common underlying process. The extension of the double simplex can be detailed briefly as follows: let $E[P(t)]$ denote the expectation of the phenotypic mean at occasion t ($t = 1$ to T) which is, as a basic assumption of the twin method, assumed to be identical for MZ and DZ twins (Mather & Jinks, 1977). $E[P(t)]$ at occasion t is the sum of the latent genetic and environmental means:

$$E[P(t)] = E[G(t)] + E[E(t)], \quad (t = 1 \text{ to } T). \tag{5}$$

The latent means at each occasion t ($t \neq 1$) are in part attributable to the immediately preceding occasion $t - 1$ and in part independent thereof:

$$E[G(t)] = \beta_G(t)E[G(t-1)] + G_\Delta \quad (t = 2 \text{ to } T) \tag{6}$$

$$E[E(t)] = \beta_E(t)E[E(t-1)] + E_\Delta \quad (t = 2 \text{ to } T). \tag{7}$$

The autoregressive coefficients $\beta_G(t)$ and $\beta_E(t)$ now account for both the stability of individual differences and the continuity in the mean. The terms G_Δ and E_Δ represent a time-invariant (hence unsubscripted) independent input at each occasion analogous to random innovation terms ζ (see equations (2) and (3)). The latent means ($E[G(1)]$, $E[E(1)]$) are estimated independently at the start of the time series. The addition of structured means does not alter the simplex model for the covariance structure. However, this extension may result in a rejection of the model whenever the continuity in the mean and the stability of individual differences can not both be explained by the common set of autoregressive coefficients.

It can be shown (cf. Dolan, Molenaar & Boomsma, submitted) that the identification of the parameters associated with the mean trend, $E[G(1)]$, $E[E(1)]$, G_Δ and E_Δ, requires longitudinal data for at least at $T = 5$ time points. That is, T must be greater than the number of parameters associated with the means.

Specification as a LISREL VI model

In specifying the extended simplex model as a LISREL model the following has to be taken into account. Firstly we require a method to estimate factor means. Usually this is accomplished by introducing a

unit variable and a dummy factor. The parameters of the latent means are then estimated by regressing the latent factors on this dummy factor. As can be seen in Figure 12.1, the unit variable loads on the dummy factor with a fixed loading equal to 1.0 so that the dummy factor ξ, has a mean equal to 1.0. The latent means at the first occasion and the subsequent means innovations are estimated by regressing the first latent variables $G(1)$ and $E(1)$ and the subsequent $G(t)$ and $E(t)$ on the dummy factor.

Secondly, the data have to be summarized in a fashion that retains the phenotypic means. Generally simultaneous structural equation modeling of means and covariance structure is carried out over the so-called augmented (AM) moment matrix, i.e. the matrix of the uncentered moments when a variable equaling one for every sample unit has been added as the last variable (see Jöreskog & Sörbom, 1986). The addition of the phenotypic means to the cross-product matrices is done as follows. Let $S_B(T \times T)$ represent the between cross-product matrix and let $m(T \times 1)$ be the vector of phenotypic means. Then the augmented matrix, $S_{AM}(T+1, T+1)$, is:

$$S_{AM} = \begin{bmatrix} (S_B + mm') & m \\ m' & 1.0 \end{bmatrix}$$

where 1.0 is the variance of the unit variable which will be specified to load with a fixed loading of 1.0 on the dummy factor. Although there are two (between and within) input matrices for MZ and DZ twins, it is sufficient to add the means to just one matrix associated with each zygosity. As the factor loadings of the phenotypic variables on the genetic series are zero in the MZ within-matrix, the means are added to the between-matrices.

A complicating aspect of the genetic simplex in the present context is the presence of the fixed genetic weights in the matrices of factor loadings (see equation (4)). Because these weights are different for each group, it is necessary to introduce a second dummy factor to compensate for the effect of these weights on the estimation of the genetic mean trend. Figure 12.1 shows how the presence of the genetic weights (w) as fixed loadings would weight the contributions of G_Δ at each occasion. As these weights differ across groups it is not possible to constrain the estimates of the latent mean parameters across groups in accordance with the assumption of equal MZ and DZ means. The second dummy factor, denoted D in Figure 12.1, allows one to weigh the latent means parameters with the inverse of the genetic weight in each group, thus canceling out the effect of the genetic weights.

The phenotypic mean at the first occasion is then modeled as:

$$E[P(1)] = wE[G(1)] + E[E(1)], \tag{8}$$

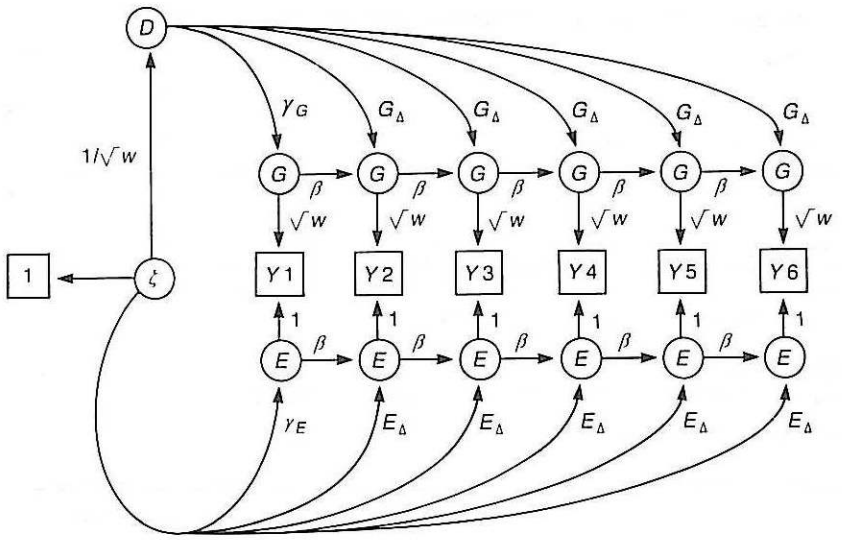

Figure 12.1. The genetic simplex with structured means. γ_G is the estimate of $E[G(1)]$. γ_E is the estimate of $E[E(1)]$. G_Δ and E_Δ are the time invariant means innovations terms.

where

$$E[G(1)] = \gamma_G(1/w) \tag{9}$$

and

$$E[E(1)] = \gamma_E. \tag{10}$$

The subsequent phenotypic means are modeled as:

$$E[P(t)] = wE[G(t)] + E[E(t)] \quad (t = 2 \text{ to } T), \tag{11}$$

where

$$E[G(t)] = (\beta_G(t)E[G(t-1)] + G_\Delta)(1/w) \tag{12}$$

and

$$E[E(t)] = \beta_E(t)E[E(t-1)] + E_\Delta. \tag{13}$$

The terms G_Δ and E_Δ are the fixed innovation terms.

Illustration

The extended double simplex model will be demonstrated with simulated data. To this end $T = 6$ repeated measures were simulated for 100 MZ and 100 DZ twin pairs according to the model described above with an additive genetic and a (specific or non-shared) environmental series.

Table 12.2. *Analysis of cross-product structure with structured means*

	t:2	3	4	5	6
True	0.9	0.8	0.5	0.4	0.3
Est.	0.920(0.038)	0.810(0.054)	0.403(0.073)	0.275(0.087)	0.261(0.092)
$\beta_G(t)$					
True	0.5	0.5	0.5	0.5	0.5
Est.	0.550(0.082)	0.552(0.071)	0.696(0.095)	0.634(0.078)	0.584(0.058)
	$E[G(1)]$	G_Δ	$E[E(1)]$	E_Δ	
True	25	21	35	12	
Est.	22.89(10.5)	19.99(5.2)	36.08(10.6)	10.82(7.3)	

$\chi^2(71) = 79.25 (p = 0.23)$

The true parameter values pertaining to the means structure are given in Table 12.2. The simultaneous analysis of means and covariance structure yielded a chi-square value of 79.25 (d.f. = 71, $p = 0.21$). The parameter estimates and standard errors are given in Table 12.2. It can be seen that the latent means at the first occasion and the fixed means innovation terms are correctly recovered.

Using the factor scores regression matrices associated with the augmented moment matrices for MZ and DZ twins, it is possible to estimate the individual $G(t)$ and $E(t)$ trajectories. Notice that, for example, the estimated $G(t)$ trajectory for each single subject consists of a stochastic (autocorrelated) series, specific to this subject, superimposed on the genetic mean trend common to all subjects. Hence it is possible to estimate the mean genetic and environmental trends by averaging the respective individual trajectories of all twins. Figure 12.2 shows the true and estimated mean genetic and environmental trends thus obtained. It can be seen that the estimated mean trends closely correspond to the true trends used in the simulation.

To demonstrate briefly the discriminatory ability of the model with regard to the relationship between the means and cross-product structure, the analysis was repeated after a constant of 10 had been alternately added to and subtracted from phenotypic means at successive time points. The chi-square value increased considerably to 112.29 which is, given 71 degrees of freedom, significant ($p < 0.01$). This result indicates that the extended simplex model adequately discriminates between means structures which can and cannot be explained by the same parameters as the cross-product structure.

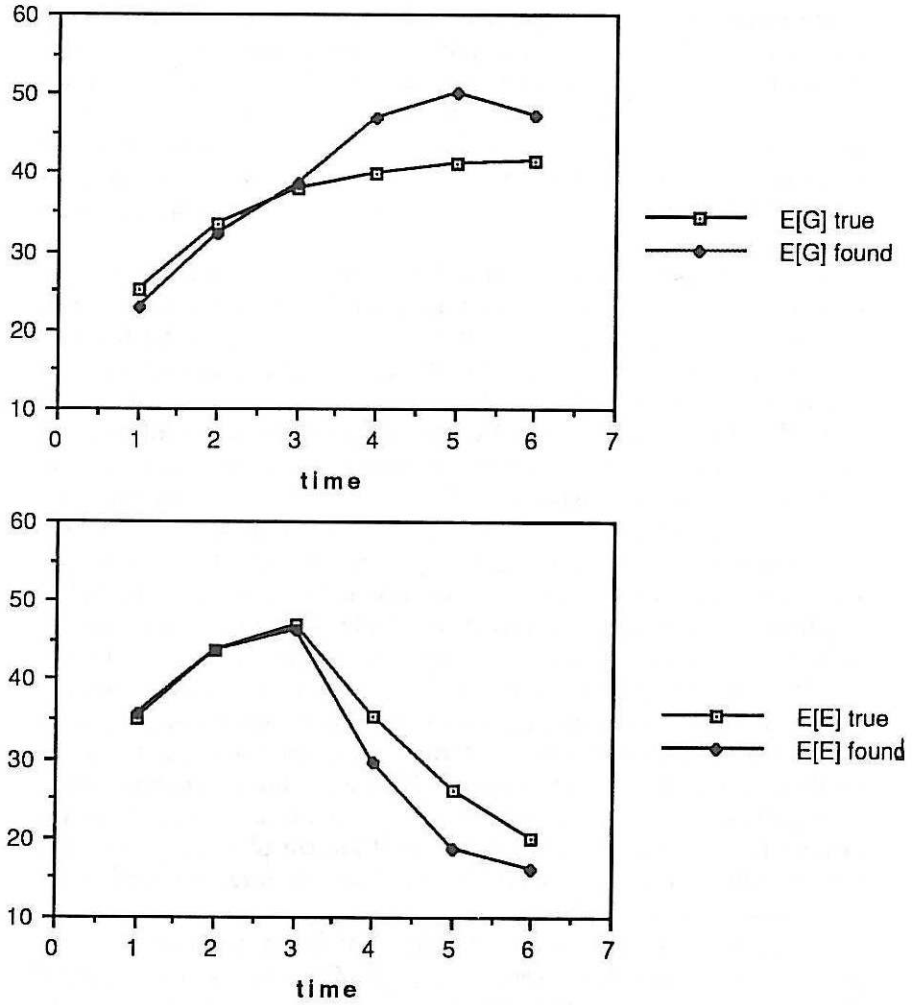

Figure 12.2. Top: the true and estimated mean trend of the genetic series. Bottom: the true and estimated mean trend of the environmental series.

AN APPLICATION TO REAL DATA

In this section we will present some results of an application of the genetic simplex model, including the extensions described earlier, to part of a longitudinal data set provided by Dr Siv Fischbein of the Stockholm Institute of Education. The analysis of the complete data set, involving longitudinal measurements of height and weight of MZ and DZ twins of both sexes at 13 equidistant time points covering seven years, will be presented elsewhere (Fischbein, Dolan, Molenaar &

Boomsma, in preparation). Here, we will restrict attention to the weight data of girls only, at $T = 6$ time points covering three years. At $t = 1$, the mean age of the girls is 11.5 years; at $t = 6$ the mean age is 14 years (standard deviation (s.d.) = 0.37). The means and variances of weight at each time point for MZ (32 pairs) and DZ (100 pairs) twins are presented in Table 12.3(a). Formal tests of the equality of variances between MZ and DZ groups at each time point are not significant at alpha = 0.01.

Firstly, the genetic simplex model is fitted to the matrices of mean cross-products (see equation (4)), thus discarding information concerning the phenotypic mean trend. It turns out that the model fits well (chi-square = 65.97, d.f. = 62, $p = 0.34$). The parameter estimates in this model, presented in Table 12.3(b), indicate the presence of an invariantly high transmission of genetic and environmental information between consecutive time points: all estimates of the autoregressive coefficients lie in the neighbourhood of 1.0. Furthermore, the estimates of the variances of genetic and environmental innovations (reflecting the inception of the effects of new genes or influences) all differ significantly from zero and do not show a decreasing trend with time. Taken together, the parameter estimates in Table 12.3 imply that (a) the stability of intraindividual differences in weight is height, (b) the variance of the genetic series is a nonstationary, almost linearly increasing function of time, (c) the variance of the environmental series shows the same nonstationary pattern as the genetic series, but is much smaller, hence (d) the heritability is high and almost constant across time points. Interestingly, both the $G(t)$ and the $E(t)$ series have the characteristic of Brownian motion, a well-known physical process (cf. Cox & Miller, 1965) that could be employed to interpret further the dynamic influences at hand.

To illustrate the estimation of individual latent trajectories in the genetic simplex model, we selected a single MZ twin pair and estimated the $G(t)$ and $E(t)$ series for each member of this pair by means of the factor scores regression matrix. The obtained trajectories, together with the associated phenotypic series of each girl, are shown in Figure 12.3. It can be seen that for one girl the environmental influences have an augmenting effect that decreases with time, whereas for the other girl the environmental influences have a suppressing effect that also decreases with time.

Secondly, the extended genetic simplex is fitted to the augmented moment matrices, thus including information concerning the phenotypic mean trend. The parameter estimates thus obtained for the structured means model are presented in Table 12.4. Estimates of $\beta_G(t)$, $\beta_E(t)$, var$[\zeta_G(t)]$ and var$[\zeta_E(t)]$ are almost equal to those presented in Table 12.3. If the complete structured means model is fitted, including

Table 12.3(a). *Means and variances of the monozygotic and dizygotic twin samples*

t	1	2	3	4	5	6
Mean of age (in years) at time of measurement (s.d. = 0.37)						
	11.5	12.0	12.5	13	13.5	14
Covariance matrix DZ twins (N of pairs = 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 matriz MZ twins (N of pairs = 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 12.3(b). *Results with mean cross-products of weight data (standard errors in parentheses)*

t:	2	3	4	5	6	
Autoregressive coefficients						
$\beta_G(t)$	1.05(0.027)	1.04(0.033)	1.02(0.028)	1.02(0.027)	0.99(0.027)	
$\beta_E(t)$	0.91(0.053)	1.05(0.090)	0.82(0.074)	0.82(0.073)	0.99(0.090)	
Variances of $G(1)$ and $E(1)$ and of innovations at $t > 1$						
var$[G(1)]$	23.41(3.46)					
var$[E(1)]$	3.44(0.82)					
var$[\zeta_G(t)]$	1.33(0.23)	2.27(0.44)	1.36(0.35)	1.80(0.37)	1.97(0.41)	
var$[\zeta_E(t)]$	0.33(0.08)	0.93(0.22)	0.93(0.21)	0.72(0.17)	0.93(0.22)	
Overall goodness-of-fit						
$\chi^2(62) = 65.97$ ($p = 0.341$)						
Derived statistics at each time						
var$[G(t)]$	23.41	27.59	32.12	35.04	38.26	38.49
var$[E(t)]$	3.44	3.20	4.48	4.01	3.43	4.30
$h^2(t)$	0.87	0.89	0.87	0.89	0.91	0.89

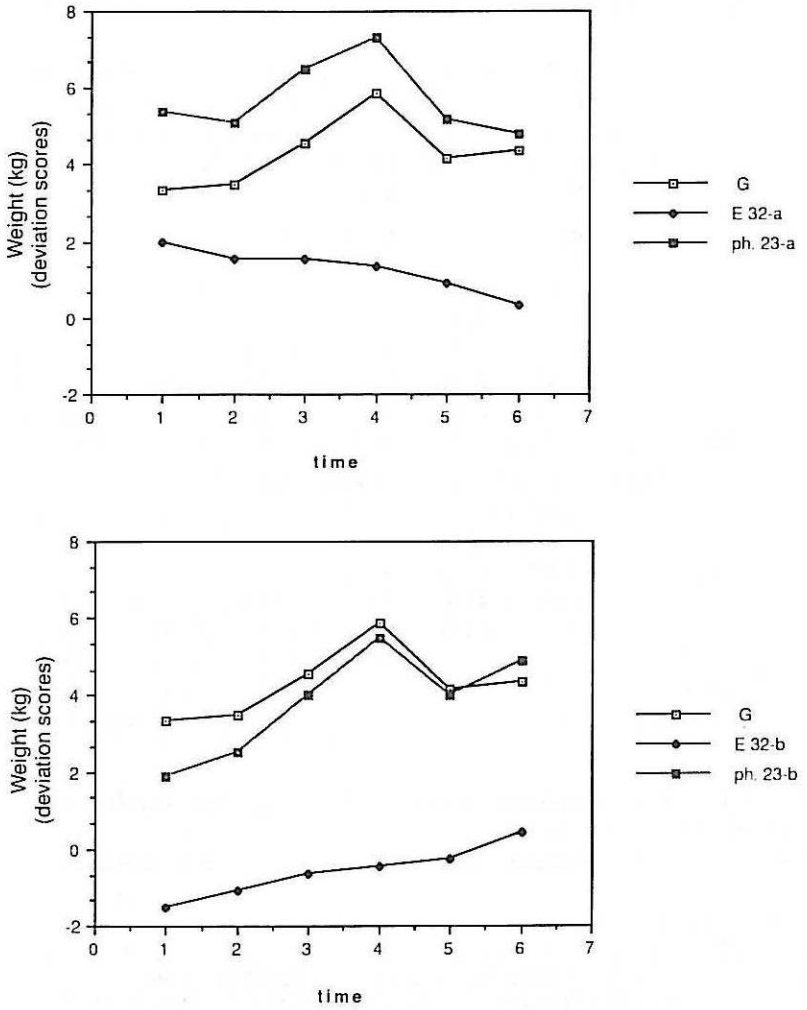

Figure 12.3. Individual phenotypic, genetic and environmental series of two monozygotic twins obtained with mean cross-products.

the parameters $E[G(1)]$, G_Δ, $E[E(1)]$, E_Δ, then the fit of the genetic simplex model thus extended is good: chi-square = 67.33, d.f. = 71, $p = 0.63$. It turns out, however, that the estimate of E_Δ and $E[E(1)]$ are not significant and therefore a restricted model is fitted in which first E_Δ is fixed at zero. It then appears that the restricted model still fits well (chi-square = 69.65, d.f. = 72, $p = 0.56$), while the estimate of $E[E(1)]$ still does not differ significantly from zero. Consequently, a second restriction was introduced by fixing $E[E(1)]$ at zero. Again the fit of this

Table 12.4. *Results with augmented moments of weight data (standard errors in parentheses)*

Model 1 means estimated in E and G series	
$E[G(1)] = 25.84(7.48)$	$G_\Delta = 5.59\ (2.78)$
$E[E(1)] = 9.95(7.45)$	$E_\Delta = -4.064\ (2.71)$
Chi-square (71) = 67.33 ($p = 0.63$)	
Model 2 means estimated at E(1) and in G series	
$E[G(1)] = 32.34\ (6.23)$	$G_\Delta = 1.68(0.77)$
$E[E(1)] = 3.45\ (6.19)$	$E_\Delta = 0.0(0.0)$
Chi-square (72) = 69.65 ($p = 0.56$)	
Model 3 means estimated in G series	
$E[G(1)] = 35.76\ (1.07)$	$G_\Delta = 1.34\ (0.507)$
Chi-square (73) = 69.79 ($p = 0.59$)	

model is good: chi-square = 69.79, d.f. = 73, $p = 0.59$. In sum, these results suggest that in the present sample of Swedish twins (a) both the stability and the change of the interindividual and intraindividual differences in weight as well as the average growth curve can be explained by the same simplex model involving genetic and non-shared environmental influences, while (b) the average growth curve appears to be solely under genetic control.

Using the factor scores regression matrix, we estimated the individual $G(t)$ and $E(t)$ trajectories. As was pointed out earlier, it is then possible to estimate the mean genetic trend by averaging the individual genetic trajectories of all twins. Figure 12.4 shows the estimated mean genetic

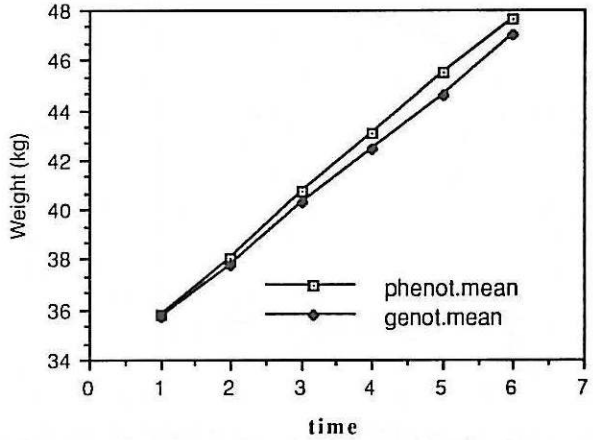

Figure 12.4. Phenotypic and estimated genetic mean trends.

trend thus obtained, together with the phenotypic mean trend. This figure reinforces our conclusion that the average phenotypic growth curve appears to be solely under genetic control.

To illustrate further the estimation of individual latent trajectories in the extended simplex model under consideration, we selected the same MZ twin pair for which the results were shown in Figure 12.3 and again estimated the $G(t)$ and $E(t)$ series for each member of this pair. Figure 12.5 shows the effect of including information concerning the

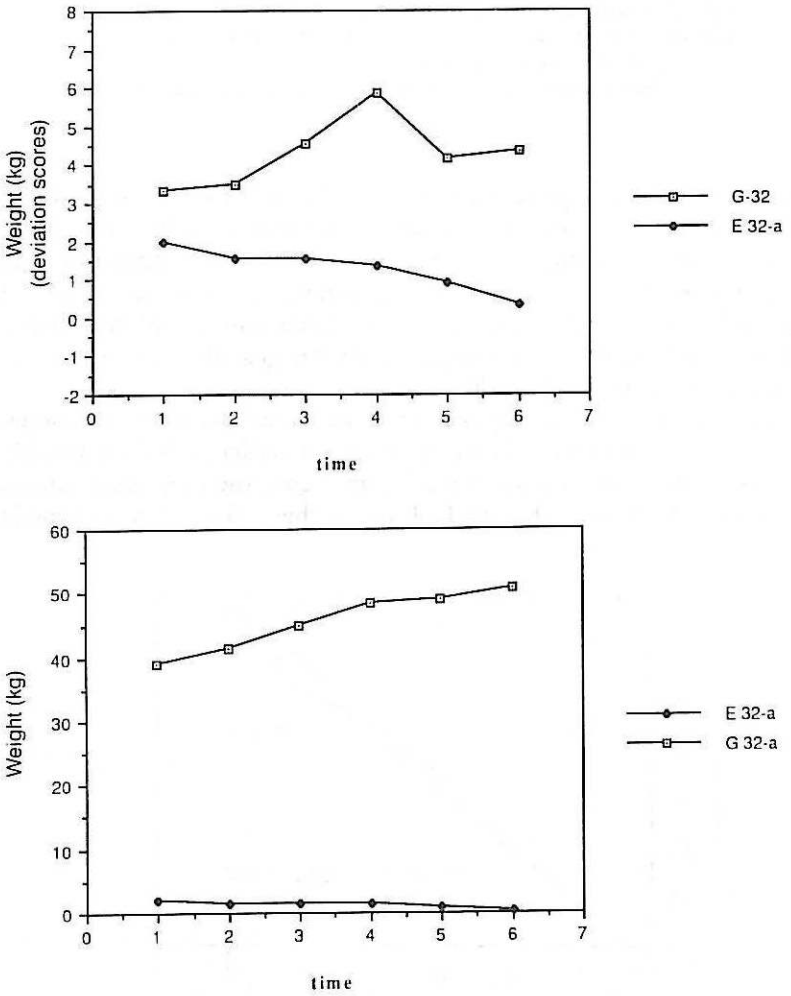

Figure 12.5. Individual genetic and environmental series of one monozygotic twin obtained with mean cross-products and augmented moments.

phenotypic mean trend on these estimates. The first graph of Figure 12.5 is a replica of the $G(t)$ and $E(t)$ series of one of the girls already shown in Figure 12.3, i.e., as determined on the basis of the covariance structure. The second graph depicts the estimated $G(t)$ and $E(t)$ series of this same girl, now determined on the basis of both the phenotypic mean trend as well as the covariance structure. Clearly, the raw longitudinal weight series of this girl appears to be almost completely due to genetic influences; a conclusion that might be missed if information concerning mean trends is not considered. Of course, this conclusion can only be drawn because both interindividual and intraindividual variation as well as mean growth can be explained by the same genetic simplex model.

DISCUSSION AND CONCLUSION

The genetic simplex model for longitudinal phenotypic data can be generalized in a number of ways. Firstly, the order of the autoregression describing $G(t)$ and/or $E(t)$ may be increased, for instance to accommodate the presence of time-dependent oscillations. Secondly, the random innovations $\zeta_G(t)$ and/or $\zeta_E(t)$ can be taken to be autocorrelated, thus leading to a consideration of more involved latent process models of autoregressive-moving average type. Thirdly, the genetic time-series model given by equation (1) can be extended by the inclusion of a common environmental process which is shared by members of the same family. Fourthly, the simplex model can be generalized to enable the decomposition of a multivariate phenotypic series into common and specific genetic and environmental processes. In a way, however, these generalizations turn out to be straightforward extensions of the genetic simplex model discussed earlier.

There have been some criticisms of simplex modeling in the recent literature. Notwithstanding the fundamental critique of applying common-factor analysis techniques to autocorrelated matrices of longitudinal data (see Wohlwill, 1973: 270–272), it has been suggested that both simplex and common-factor models are equivalent models in case variances and means are included in the analysis or a confirmatory analysis is carried out. A definite refutation of such suggestions, however, can be given on the basis of the following mathematical–statistical result: the spectrum of a Toeplitz (autocorrelation) matrix is continuous, whereas a factor model presupposes a discrete or at least mixed spectrum (cf. Grenander & Szegö, 1958). This implies, among other things, that when the number T of time points increases, an increasing number of common factors is needed in order to describe the autocorrelation structure of a univariate simplex. Cronbach (1967) already showed that the common factors so obtained are spurious. Another criticism of simplex modeling has been raised by Rogosa &

Willett (1985), who showed by means of a simulation experiment that a simplex model may yield a satisfactory fit to a covariance structure associated with an entirely different (linear random coefficient) growth model. However, we repeated this simulation experiment and found that the results of Rogosa & Willett are related to a particular instance of the linear random coefficient growth model. Specifically, if the number of time points ($T = 5$ in the original simulation experiment) is slightly increased (e.g. $T > 6$) then the simplex model no longer fits the obtained covariance structure. A full discussion of these new results will be given in a separate publication. It is concluded that the mentioned criticisms of simplex modeling of longitudinal data do not appear to be serious. This conclusion is reinforced by the fact that autoregressive models such as those considered in this chapter occupy a very prominent place in the mathematical–statistical theory of time-dependent processes (Hannan & Deistler, 1988).

Whereas the genetic simplex model can be conceived as a parametric model, there now exist non-parametric approaches to the genetic analysis of stretches of time-dependent data of arbitrary length (Molenaar & Boomsma, 1987*b*,*c*). In these non-parametric approaches, each autocorrelated phenotypic series is transformed into a sequence of uncorrelated variables, either in the time domain (Karhunen–Loève transformation) or in the frequency domain (Fourier transformation), thus enabling the application of standard biometrical analysis techniques to each transformed variable separately. These approaches, called genetic signal analyses, can lead to interesting applications in, for example, psychophysiological studies.

The possibility of estimating the time course of $G(t)$ and $E(t)$ for each individual subject can also lead to interesting applications in genetic counseling and epidemiology. In particular, this enables one to determine the impact of genetic and environmental processes on specific subjects or on groups of subjects who suffer from deviant development. Estimation of individual $G(t)$ and $E(t)$ trajectories is also possible in genetic signal analysis. Of course, these possibilities may provide important information concerning the ways in which remedial treatment is to be carried out.

The inclusion of mean genetic and environmental trend functions in a single simplex process model for the analysis of phenotypic cross-products has important theoretical implications. It is now possible to test whether genetic processes underlying individual differences can also account for mean phenotypic growth within a particular population. A sufficient number of constraints has to be imposed on the structured means model to ensure its identifiability. Here we considered a constrained means model given by equations (6) and (7) which closely resembles a first-order autoregression. In contrast, we could have

derived a means model from theoretical considerations concerning behavioral growth (Guire & Kowalski, 1979). In general, such theoretical growth models are nonlinear and can only be identified through the imposition of nonlinear constraints. A general LISREL model for the analysis of structural models with nonlinear constraints is presented in Boomsma & Molenaar (1987*b*).

In conclusion we have indicated the potential of genetic simplex models for the analysis of longitudinal data. It can be expected that regular application of this type of models will lead to significant progress in the field of developmental behavior genetics.

ACKNOWLEDGEMENTS

We wish to thank Professors P. B. Baltes, J. de Leeuw, and D. Magnusson, who provided valuable comments that led to this formulation. We also wish to thank Dr S. Fischbein for generously providing her longitudinal twin data.

REFERENCES

Baltes, P. B. & Nesselroade, J. R. (1973). The developmental analysis of individual differences on multiple measures. In J. R. Nesselroade & H. W. Reese (eds.), *Life-span developmental psychology: Methodological issues.* New York: Academic Press.

Boomsma, D. I. & Molenaar, P. C. M. (1987*a*). The genetic analysis of repeated measures. I. Simplex models. *Behavior Genetics, 17,* 111–123.

Boomsma, D. I. & Molenaar, P. C. M. (1987*b*). Constrained maximum likelihood analysis of familial resemblance of twins and their parents. *Acta Geneticae Medicae Gemellologicae, 36,* 29–39.

Boomsma, D. I., Molenaar, P. C. M. & Orlebeke, J. F. (in press). Estimation of individual genetic and environmental factor scores. *Genetic Epidemiology.*

Browne, M. W. & Shapiro, A. (1988). Robustness of normal theory methods in the analysis linear latent variable models. *British Journal of Mathematical and Statistical Psychology, 41,* 193–208.

Cox, D. R. & Miller, H. D. (1965). *The theory of stochastic processes.* London: Methuen.

Cronbach, L. J. (1967). Year-to-year correlations of mental tests: A review of the Hofstaetter analysis. *Child Development, 38,* 283–289.

DeFries, J. C. & Fulker, D. W. (1986). Multivariate behavioral genetics and development. *Behavior Genetics, 16,* 1–10.

Dolan, C. V., Molenaar, P. C. M. & Boomsma, D. I. (Submitted) Simultaneous genetic analyis of longitudinal means and covariance structure in the simplex model. *Behavior Genetics.*

Eaves, L. J. (1977). Inferring the causes of human variation. *Journal of the Royal Statistical Society A, 140,* 324–355.

Eaves, L. J. & Young, P. A. (1981). Genetical theory and personality differences. In R. Lynn (ed.), *Dimensions of personality*. Oxford: Pergamon Press.

Eaves, L. J., Hewitt, J. K. & Heath, A. C. (1988). The quantitative genetic study of human development change: A model and its limitations. In B. S. Weir, E. J. Eisen, M. M. Goodman & G. Namkoong (eds.), *Proceedings of the second international conference on quantitative genetics* (pp. 297–311). Sunderland, Mass: Sinauer Associates, Inc.

Fischbein, S., Dolan, C. V., Molenaar, P. C. M. & Boomsma, D. I. (in preparation) Analysis of repeatedly measured height and weight in a sample of Swedish twins.

Fulker, D. W. (1988). Genetic and cultural transmission in human behavior. In B. S. Weir, E. J. Eisen, M. M. Goodman & G. Namkoong (eds.), *Proceedings of the second international conference on quantitative genetics* (pp. 297–311). Sunderland, Mass.: Sinauer Associates, Inc.

Granger, C. W. J. & Morris, M. S. (1976). Time series modelling and interpretation. *Journal of the Royal Statistical Society A, 139*, (2), 246–257.

Grenander, U. & Szegö, G. (1958). *Toeplitz forms and their applications*. Berkeley: University of California Press.

Gourieroux, C., Monfort, A. & Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica, 17*, 287–304.

Guire, K. E. & Kowalski, C. J. (1979). Mathematical description and representation of developmental change functions on the intra- and interindividual levels. In J. R. Nesselroade & P. B. Baltes (eds.). *Longitudinal research in the study of behavior and development*. New York: Academic Press.

Hannan, E. J. & Deistler, M. (1988). *The statistical theory of linear systems*. New York: Wiley.

Heath, A. C., Jardine, R., Eaves, L. J. & Martin, N. G. (1988). The genetic structure of personality I. Phenotypic factor structure of the EPQ in an Australian sample. *Personality and Individual Differences, 9*, 59–67.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology, 26*, 121–145.

Jöreskog, K. G. & Sörbom, D. (1986). LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables and least squares methods, University of Uppsala, Department of Statistics, PO Box 513, s-751 20 Uppsala, Sweden.

Mather, K. & Jinks, J. L. (1977). *Introduction to biometrical genetics*. London: Chapman and Hall.

McArdle, J. J. (1986). Latent-variable growth within behavior genetic models. *Behavior Genetics, 16*, 79–95.

McCall, R. B. (1981). Nature-nurture and the two realms of development: A proposed integration with respect to mental development. *Child Development, 5*, 1–12.

Mednick, S. A., Moffitt, T. E., Pollock, V., Talovic, S., Gabrielli, W. F. & van Dusen, K. T. (1983). The inheritance of human deviance. In D. Magnusson & V. L. Allen (eds.), *Human development: An interactional perspective*. New York: Academic Press.

Molenaar, P. C. M. & Boomsma, D. I. (1987*a*). Application of nonlinear factor analysis to genotype-environment interaction. *Behavior Genetics, 17*, 71–80.

Molenaar, P. C. M. & Boomsma, D. I. (1987*b*). The genetic analysis of repeated measures II. Karhunen–Loève transformation. *Behavior Genetics, 17*, 229–242.

Molenaar, P. C. M. & Boomsma, D. I. (1987*c*). Spectral analysis of twin time series designs. *Acta Geneticae Medicae Gemellologicae, 36*, 51–59.

Molenaar, P. C. M., Boomsma, D. I., Neeleman, D. & Dolan, C. (in press). Using factor scores to detect $G \times E$ interactive origin of 'pure' genetic or environmental factors obtained in genetic covariance structure analysis. *Genetic Epidemiology*.

Plomin, R. (1986). *Development, genetics and psychology*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Plomin, R., DeFries, J. C. & Fulker, D. W. (1988). *Nature and nurture during infancy and early childhood*. Cambridge University Press.

Rogosa, D. & Willett, J. B. (1985). Satisfying a simplex structure is simpler than it should be. *Journal of Educational Statistics, 10* (2), 99–107.

Thomas, H. (1980). A theory of growth. In R. H. Kluwe & H. Spada (eds.) *Developmental models of thinking*. New York: Academic Press.

Wohlwill, J. F. (1973). *The study of behavioral development*. New York: Academic Press.