STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
https://doi.org/10.1080/10705511.2020.1789465

Classical Models for Twin Data

Karl G. Jöreskog

Uppsala University

ABSTRACT

The classical models ACE and ADE were used in the 1990s to estimate heredity of a phenotype from data on monozygotic and dizygotic twins. These models are extended to a model called ACDE with four parameters instead of only three. It is showed how these models can be easily estimated by maximum likelihood. The models and methods are extended to two populations in which the heredity is the same in both populations. Examples are given to estimate the heredity of BMI using twin data from the UK and Australia.

KEYWORDS

SEM; twin models; twin data; heredity

Twin data models

Models for twin data or other types of family relationships were common in the 1980s and 1990s (see, e.g., Neale & Cardon, 1992). One has data on N_{mz} monozygotic (mz) twins and N_{dz} dizygotic (dz) twins. It is assumed that mz twins have all their genes in common and dz twins have half of their genes in common. We observe a phenotype on each twin, denoted P_1 and P_2. A phenotype may be anything that one can measure or observe on each twin, such as symptoms (e.g., allergy), illnesses (e.g., cancer, diabetes, asthma), physical measures (e.g., weight, height), personality traits (e.g., nervousness), longevity (how long you live), etc. Here we assume that P_1 and P_2 are continuous variables.

From biometric genetic theory, dating back to theories of Galton and experiments of Mendel in the nineteenth century (see, e.g., Mather & Jinks, 1971; Neale & Cardon 1992, Chapter 3), one can formulate the simplest type of twin design as follows. The observed phenotypes P_1 and P_2 of twin 1 and twin 2 in a twin pair are postulated to depend on additive genes A_1 and A_2, common environments C_1 and C_2 (environmental influences shared by twins reared in the same family), dominance genetic deviations D_1 and D_2(dominance effects of alleles at multiple loci) and unique environments E_1 and E_2:

$$P_1 = \mu_1 + h_1A_1 + c_1C_1 + d_1D_1 + E_1\ , \tag{1}$$

$$P_2 = \mu_2 + h_2A_2 + c_2C_2 + d_2D_2 + E_2\ , \tag{2}$$

where μ_1 and μ_2 are means and D_1, A_1, C_1, C_2, A_2, and D_2 are latent random variables with means zero and the correlation matrix

$$\mathbf{\Phi} = \begin{pmatrix} 1 & & & & & \\ 0 & 1 & & & & \\ 0 & 0 & 1 & & & \\ 0 & 0 & 1 & 1 & & \\ 0 & x & 0 & 0 & 1 & \\ y & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3}$$

where x and y are constants with $x = 1$ for mz, $x = \frac{1}{2}$ for dz, $y = 1$ for mz, $y = \frac{1}{4}$ for dz. E_1 and E_2 are uncorrelated random variables, uncorrelated with all the other latent variables. Since twin 1 and twin 2 are interchangeable, it is assumed that the effects of the latent variables are the same for twin 1 and twin 2, so that $h_1 = h_2 = h$, $c_1 = c_2 = c$, and $d_1 = d_2 = d$. It is also assumed that the variances of E_1 and E_2 are equal. A path diagram is shown in Figure 1, where we combine information for monozygotic (mz) and dizygotic (dz) twins. The first to derive the correlation between P_1 and P_2 was Sir Ronald A Fisher (1918).

We consider only the simplest type of models here and look at this as a statistical estimation problem.

From these assumptions, it follows that the model implied covariance matrices of P_1 and P_2 are

$$\mathbf{\Sigma}_{mz} = \begin{pmatrix} h^2 + c^2 + d^2 + e^2 & \\ h^2 + c^2 + d^2 & h^2 + c^2 + d^2 + e^2 \end{pmatrix}, \tag{4}$$

$$\mathbf{\Sigma}_{dz} = \begin{pmatrix} h^2 + c^2 + d^2 + e^2 & \\ \frac{1}{2}h^2 + c^2 + \frac{1}{2}d^2 & h^2 + c^2 + d^2 + e^2 \end{pmatrix}, \tag{5}$$

Note that the Σ's are not functions of h, c, d and e but of their squares.[1] Equations (4) and (5) can be combined as

$$\begin{pmatrix} \sigma_{11}^{(mz)} \\ \sigma_{21}^{(mz)} \\ \sigma_{22}^{(mz)} \\ \sigma_{11}^{(dz)} \\ \sigma_{21}^{(dz)} \\ \sigma_{22}^{(dz)} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ \frac{1}{2} & 1 & \frac{1}{4} & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} h^2 \\ c^2 \\ d^2 \\ e^2 \end{pmatrix}. \tag{6}$$

This represents six equations in four parameters h^2, c^2, d^2, and e^2. Obviously, only three of the six equations are linearly independent. If we retain only the equations that are linearly independent we may take

CONTACT Karl G. Jöreskog karl.joreskog@telia.com Uppsala University, Uppsala, Sweden.

[1]From now on, we regard the parameters h^2, c^2, d^2, and e^2 as single entities rather than as squares. The best interpretation of these is as variances or variance contributions.

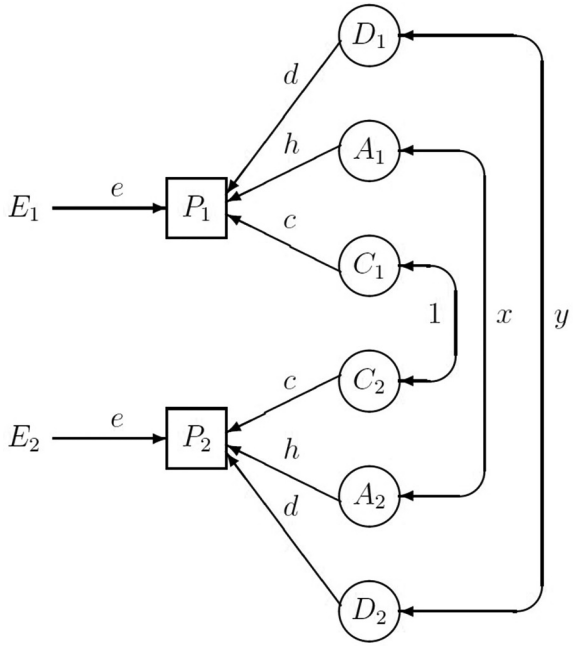

Figure 1. Simple twin model with dominance effects, additive genes, common and unique environments ($x = 1$ for mz, $x = \frac{1}{2}$ for dz, $y = 1$ for mz, $y = \frac{1}{4}$ for dz).

$$\begin{pmatrix} \sigma_{11}^{(\mathrm{mz})} \\ \sigma_{21}^{(\mathrm{mz})} \\ \sigma_{21}^{(\mathrm{dz})} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ \frac{1}{2} & 1 & \frac{1}{4} & 0 \end{pmatrix} \begin{pmatrix} h^2 \\ c^2 \\ d^2 \\ e^2 \end{pmatrix}. \tag{7}$$

For given $\sigma_{11}^{(\mathrm{mz})}, \sigma_{21}^{(\mathrm{mz})}, \sigma_{21}^{(\mathrm{dz})}$, there are infinitely many solutions of (7) in terms of h^2, c^2, d^2, e^2. However, e^2 is uniquely determined as

$$e^2 = \sigma_{11}^{(\mathrm{mz})} - \sigma_{21}^{(\mathrm{mz})} . \tag{8}$$

The usual way to resolve resolve this indeterminacy is by setting either c^2 or d^2 to 0. If $c^2 = 0$ the model is called an ADE model. If $d^2 = 0$ the model is called an ACE model.

However, the restrictions of c^2 or d^2 to 0 are undesirable. If we had an additional group of twins that were separated from birth, we could assume that $c^2 = 0$ for that group. Lacking such information, it seems reasonable to assume that there is variation in the common environment in which the twins grow up, and hence $c^2 > 0$.

We propose an alternative model which makes it possible to estimate all four parameters. This will be called an ACDE model.

Let $\mathbf{A}$ be the matrix in (7) and let $\mathbf{A}^\star$ be

$$\mathbf{A}^\star = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}. \tag{9}$$

Using paper and pencil algebra gives

$$\mathbf{A}^\star = \begin{pmatrix} 0 & \frac{1}{2} & -\frac{2}{7} \\ 0 & -\frac{1}{2} & 1\frac{3}{7} \\ 0 & 1 & -1\frac{1}{7} \\ 1 & -1 & 0 \end{pmatrix} . \tag{10}$$

The ACDE model is defined by

$$\begin{pmatrix} h^2 \\ c^2 \\ d^2 \\ e^2 \end{pmatrix} = \mathbf{A}^\star \begin{pmatrix} \sigma_{11}^{(\mathrm{mz})} \\ \sigma_{21}^{(\mathrm{mz})} \\ \sigma_{21}^{(\mathrm{dz})} \end{pmatrix} . \tag{11}$$

Heredity H is defined as

$$H = \frac{h^2}{h^2 + c^2 + d^2 + e^2} , \tag{12}$$

which is the fraction of variance in the phenotype attributable to genes alone.

For the ACE model, an interesting relationship is worth noting. Since the covariance matrices of P_1 and P_2 are

$$\mathbf{\Sigma}_{\mathrm{mz}} = \begin{pmatrix} h^2 + c^2 + e^2 & \\ h^2 + c^2 & h^2 + c^2 + e^2 \end{pmatrix} , \tag{13}$$

$$\mathbf{\Sigma}_{\mathrm{dz}} = \begin{pmatrix} h^2 + c^2 + e^2 & \\ \frac{1}{2}h^2 + c^2 & h^2 + c^2 + e^2 \end{pmatrix} , \tag{14}$$

the correlation between two mz twins is

$$\rho_{\mathrm{mz}} = \frac{h^2 + c^2}{h^2 + c^2 + e^2} \tag{15}$$

and the correlation between two dz twins is

$$\rho_{\mathrm{dz}} = \frac{\frac{1}{2}h^2 + c^2}{h^2 + c^2 + e^2} , \tag{16}$$

so that

$$H = 2(\rho_{\mathrm{mz}} - \rho_{\mathrm{dz}}) , \tag{17}$$

suggesting that H could be estimated by

$$\hat{H} = 2(r_{\mathrm{mz}} - r_{\mathrm{dz}}) , \tag{18}$$

where the r's are sample correlations.

In this paper, we use the parameters α, β, and γ defined by $\alpha = h^2 + c^2 + d^2 + e^2$, $\beta = h^2 + c^2 + d^2$, and $\gamma = xh^2 + yd^2 + c^2$, so that

$$\mathbf{\Sigma}_{\mathrm{mz}} = \begin{pmatrix} \alpha & \\ \beta & \alpha \end{pmatrix} \quad \mathbf{\Sigma}_{\mathrm{dz}} = \begin{pmatrix} \alpha & \\ \gamma & \alpha \end{pmatrix} .$$

For the ACE model, there is a one-to-one relationships between α, β, γ and h^2, c^2, e^2:

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ \frac{1}{2} & 1 & 0 \end{pmatrix} \begin{pmatrix} h^2 \\ c^2 \\ e^2 \end{pmatrix}$$
$$\begin{pmatrix} h^2 \\ c^2 \\ e^2 \end{pmatrix} = \begin{pmatrix} 0 & 2 & -2 \\ 0 & -1 & 2 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \tag{19}$$

Similarly, for the ADE model, there is a one-to-one relationships between α, β, γ and h^2, d^2, e^2:

$$\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix}=\begin{pmatrix}1 & 1 & 1\\ 1 & 1 & 0\\ \frac{1}{2} & \frac{1}{4} & 0\end{pmatrix}\begin{pmatrix}h^2\\ d^2\\ e^2\end{pmatrix}$$
$$\begin{pmatrix}h^2\\ d^2\\ e^2\end{pmatrix}=\begin{pmatrix}0 & -1 & 4\\ 0 & 2 & -4\\ 1 & -1 & 0\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix} \tag{20}$$

For the ACDE model no such one-to-one relationships exist, but

$$\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix}=\begin{pmatrix}1 & 1 & 1 & 1\\ 1 & 1 & 1 & 0\\ \frac{1}{2} & 1 & \frac{1}{4} & 0\end{pmatrix}\begin{pmatrix}h^2\\ c^2\\ d^2\\ e^2\end{pmatrix}$$
$$\begin{pmatrix}h^2\\ c^2\\ d^2\\ e^2\end{pmatrix}=\begin{pmatrix}0 & \frac{1}{2} & -\frac{2}{7}\\ 0 & -\frac{1}{2} & 1\frac{3}{7}\\ 0 & 1 & -1\frac{1}{7}\\ 1 & -1 & 0\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix} \tag{21}$$

The matrix in the bottom of (21) is the Moore-Penrose inverse of the matrix in the top, also called a right inverse, see, for example, Magnus and Neudecker (1999, pp. 32).

It is assumed that

$$0<\gamma\le\beta\le\alpha\,. \tag{22}$$

If $\gamma>\beta$, the correlation for dz would be greater than that of mz. If $\beta>\alpha$, the covariance for mz would be larger than the variances for mz.

It is particularly easy to estimate α, β, and γ and their covariance matrix, see Section Estimation. The estimates of h^2, c^2, d^2, e^2 are obtained as linear combinations of the estimates of α, β, and γ, using Equations (19)–(21). From (12), one finds that

$$H=\frac{h^2}{h^2+c^2+d^2+e^2}=\frac{p\beta+q\gamma}{\alpha}\,, \tag{23}$$

where p and q are constants which vary across models, such that $p=2,q=-2$ for ACE, $p=-1,q=4$ for ADE, and $p=\frac{1}{2}$,$q=-\frac{2}{7}$ for ACDE.

Estimation

To estimate the models, all we need are the two sample covariance matrices $\mathbf{S}_{\text{mz}}$ and $\mathbf{S}_{\text{dz}}$ and their sample sizes N_{mz} and N_{dz}.

Intuitive simple estimates of α, β, and γ are

$$\hat\alpha=\frac{1}{4}(s_{11}^{(\text{mz})}+s_{22}^{(\text{mz})}+s_{11}^{(\text{dz})}+s_{22}^{(\text{dz})}),\quad \hat\beta=s_{21}^{(\text{mz})},\quad \hat\gamma=s_{21}^{(\text{dz})}. \tag{24}$$

Assuming bivariate normality and ignoring means, the logarithm of the likelihood function is

$$\ln L=-\frac{N_{\text{mz}}}{2}\{\ln|\boldsymbol{\Sigma}_{\text{mz}}|+\text{tr}(\mathbf{S}_{\text{mz}}\boldsymbol{\Sigma}_{\text{mz}}^{-1})\}-\frac{N_{\text{dz}}}{2}\{\ln|\boldsymbol{\Sigma}_{\text{dz}}|+\text{tr}(\mathbf{S}_{\text{dz}}\boldsymbol{\Sigma}_{\text{dz}}^{-1})\}\,. \tag{25}$$

Let

$$\mathbf{S}_{\text{mz}}=\begin{pmatrix}a & \\ b & c\end{pmatrix},\ \mathbf{S}_{\text{dz}}=\begin{pmatrix}d & \\ e & f\end{pmatrix}. \tag{26}$$

Then maximizing $\ln L$ in (25) "boils down" to minimizing the function

$$f(\alpha,\beta,\gamma)=k_1\left[\ln(\alpha^2-\beta^2)+\frac{(a+c)\alpha-2b\beta}{\alpha^2-\beta^2}\right]+k_2\left[\ln(\alpha^2-\gamma^2)+\frac{(d+f)\alpha-2e\gamma}{\alpha^2-\gamma^2}\right], \tag{27}$$

with respect to α, β, and γ. Here $k_1=N_{\text{mz}}/N$ and $k_2=N_{\text{dz}}/N$, where $N=N_{\text{mz}}+N_{\text{dz}}$.

Using (24) as starting values, the function (27) can be minimized numerically using first and second derivatives. At the solution, one can obtain an estimate of the asymptotic covariance matrix of $\hat\alpha$, $\hat\beta$, and $\hat\gamma$ from the information matrix $\hat{\mathbf{E}}$ which is the inverse of the expected Hessian at the solution point.

Let $\hat f$ be the minimum value of f. Then

$$c=N\left[\hat f-k_1(2+\ln(ac-b^2))-k_2(2+\ln(df-e^2))\right], \tag{28}$$

can be used as a chi-square statistic with three degrees of freedom for testing the fit of the model. This test statistic is the same as the log likelihood ratio test statistic for testing the hypothesis that the variances are equal within and between groups.

ACE model

For the ACE model, we have

$$\begin{pmatrix}\hat h^2\\ \hat c^2\\ \hat e^2\end{pmatrix}=\begin{pmatrix}0 & 2 & -2\\ 0 & -1 & 2\\ 1 & -1 & 0\end{pmatrix}\begin{pmatrix}\hat\alpha\\ \hat\beta\\ \hat\gamma\end{pmatrix}=\mathbf{B}\begin{pmatrix}\hat\alpha\\ \hat\beta\\ \hat\gamma\end{pmatrix}. \tag{29}$$

with estimated asymptotic covariance matrix

$$\text{ACov}\begin{pmatrix}\hat h^2\\ \hat c^2\\ \hat e^2\end{pmatrix}=\mathbf{B}\hat{\mathbf{E}}\mathbf{B}'. \tag{30}$$

From the definition of H, the estimate $\hat H$ is obtained as

$$\hat H=\frac{p\hat\beta+q\hat\gamma}{\hat\alpha}. \tag{31}$$

Using a Taylor expansion of H around the true values gives $\hat H$ approximately as

$$\hat H\approx H-\frac{p\beta+q\gamma}{\alpha^2}(\hat\alpha-\alpha)+\frac{p}{\alpha}(\hat\beta-\beta)+\frac{q}{\alpha}(\hat\gamma-\gamma) \tag{32}$$

So the asymptotic variance of $\hat H$ is

$$\mathrm{AVar}(\hat{H}) \approx \alpha^{-4}(-(p\beta+q\gamma)\ \ p\alpha\ \ q\alpha)\,\hat{\mathbf{E}}\begin{pmatrix}-(p\beta+q\gamma)\\ p\alpha\\ q\alpha\end{pmatrix}. \tag{33}$$

Plugging in the estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, and using $p = 2, q = -2$ we obtain the estimate of $\mathrm{AVar}(\hat{H})$.

ADE model

The ADE model can be estimated in the same way as the ACE model. The parameters α, β, and γ need not be reestimated.

For the ADE model, we have

$$\begin{pmatrix}\hat{h}^2\\ \hat{d}^2\\ \hat{e}^2\end{pmatrix} = \begin{pmatrix}0 & -1 & 4\\ 0 & 2 & -4\\ 1 & -1 & 0\end{pmatrix}\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\\ \hat{\gamma}\end{pmatrix} = \mathbf{B}\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\\ \hat{\gamma}\end{pmatrix}. \tag{34}$$

with estimated asymptotic covariance matrix

$$\mathrm{ACov}\begin{pmatrix}\hat{h}^2\\ \hat{d}^2\\ \hat{e}^2\end{pmatrix} = \mathbf{B}\hat{\mathbf{E}}\mathbf{B}'. \tag{35}$$

To obtain the estimate $\hat{H}$ and $\mathrm{AVar}(\hat{H})$ for the ADE model we use (32) and (33) with $p = -1$, $q = 4$.

ACDE model

For the ACDE model, we use the matrix $\mathbf{A}^\star = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ as $\mathbf{B}$ matrix and obtain the estimates as

$$\begin{pmatrix}\hat{h}^2\\ \hat{c}^2\\ \hat{d}^2\\ \hat{e}^2\end{pmatrix} = \begin{pmatrix}0 & \frac{1}{2} & -\frac{2}{7}\\ 0 & -\frac{1}{2} & 1\frac{3}{7}\\ 0 & 1 & -1\frac{1}{7}\\ 1 & -1 & 0\end{pmatrix}\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\\ \hat{\gamma}\end{pmatrix} = \mathbf{B}\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\\ \hat{\gamma}\end{pmatrix}. \tag{36}$$

with estimated asymptotic covariance matrix

$$\mathrm{ACov}\begin{pmatrix}\hat{h}^2\\ \hat{c}^2\\ \hat{d}^2\\ \hat{e}^2\end{pmatrix} = \mathbf{B}\hat{\mathbf{E}}\mathbf{B}'. \tag{37}$$

This matrix is singular, positive semidefinite and of rank 3. Nevertheless, its diagonal elements can be used to obtain the estimated standard errors of $\hat{h}^2$, $\hat{c}^2$, $\hat{d}^2$, and $\hat{e}^2$.

To obtain the estimate $\hat{H}$ and $\mathrm{AVar}(\hat{H})$ for the ACDE model we use (32) and (33) with $p = \frac{1}{2}$, $q = -\frac{2}{7}$.

Examples: Estimating the heredity of BMI

Example 1: The UK data

We use raw data from the UK twin register maintained by the Department of Twin Research and Genetic Epidemiology, King's College, London, UK. The data on BMI were extracted in January 2016. There are 1552 twin pairs. The oldest twin pair was born 1924 and the youngest twin pair was born 1991. Among the 1552 twin pairs, there were 794 mz twins and 758 dz twins.

The sample covariance matrices are

$$\mathbf{S}_{\mathrm{mz}} = \begin{pmatrix}a & \\ b & c\end{pmatrix} = \begin{pmatrix}24.366 & \\ 18.797 & 23.587\end{pmatrix}, \tag{38}$$

$$\mathbf{S}_{\mathrm{dz}} = \begin{pmatrix}d & \\ e & f\end{pmatrix} = \begin{pmatrix}28.379 & \\ 12.657 & 25.751\end{pmatrix}, \tag{39}$$

The value of c in (28) is 6.84. As a chi-square with three degrees of freedom it has P-value of 0.08. The estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and their estimated covariance matrix are given in Table 1.

These estimates satisfy the inequality (22).

Estimated parameters and their standard errors are given in Table 2.

Is age a biasing factor?

Age is not included in the UK dataset but the year of birth is. So we can use 2016 – birth year as a proxy for age. Figure 2 shows the age distribution for all 1552 twins. It is seen that there are not many young twins but some very old ones.

Age and sex accounts for less than 3% of the variance of BMI. Nevertheless, we will investigate the effect of age on the estimates. We do this by estimating the bivariate regression of BMI on age under the constraint that the regression coefficients are the same for each twin in a twin pair. This can be tested with one degree of freedom. Then we saved the residual covariance matrix in each regression. The resulting residual covariance matrices are as follows. Note that residual variances are only slightly smaller than the original variances.

$$\mathbf{S}_{\mathrm{mz}} = \begin{pmatrix}a & \\ b & c\end{pmatrix} = \begin{pmatrix}23.837 & \\ 18.128 & 22.782\end{pmatrix}, \tag{40}$$

$$\mathbf{S}_{\mathrm{dz}} = \begin{pmatrix}d & \\ e & f\end{pmatrix} = \begin{pmatrix}27.445 & \\ 11.892 & 25.152\end{pmatrix}, \tag{41}$$

Table 1. The UK data: estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and their covariance matrix $\hat{\mathbf{E}}$.

		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
$\hat{\alpha}$	25.723	0.59465		
$\hat{\beta}$	20.499	0.57965	0.63311	
$\hat{\gamma}$	11.647	0.44680	0.43561	0.79396

Table 2. The UK data: parameter estimates (0* indicates a fixed value) and standard errors.

Parameter	ACE	ADE	ACDE
$\hat{h}^2$	17.705 (1.491)	26.088 (3.139)	6.922 (0.314)
$\hat{c}^2$	2.794 (1.438)	0*	6.389 (1.075)
$\hat{d}^2$	0*	−5.582 (2.875)	7.189 (0.821)
$\hat{e}^2$	5.224 (0.262)	5.224 (0.262)	5.224 (0.262)
$\hat{H}$	0.688 (0.057)	1.014 (0.110)	0.269 (0.009)

Note that the estimate $\hat{d}^2$ is negative but not significant for the ADE model. We could set $\hat{d}^2 = 0$ which essentially makes the model an AE model with a very large component $\hat{h}^2$. Note also that the ACE model gives a large value of $\hat{H}$ at the expense of a small and non-significant value of $\hat{c}^2$. The results for ACDE seem more reasonable, where all parameter estimates are statistically significant. Heredity is estimated at 0.27 indicating that 27% of the variance in the phenotype is due to genes. Note that the standard error of $\hat{H}$ is very small indicating that this estimate is very precise.

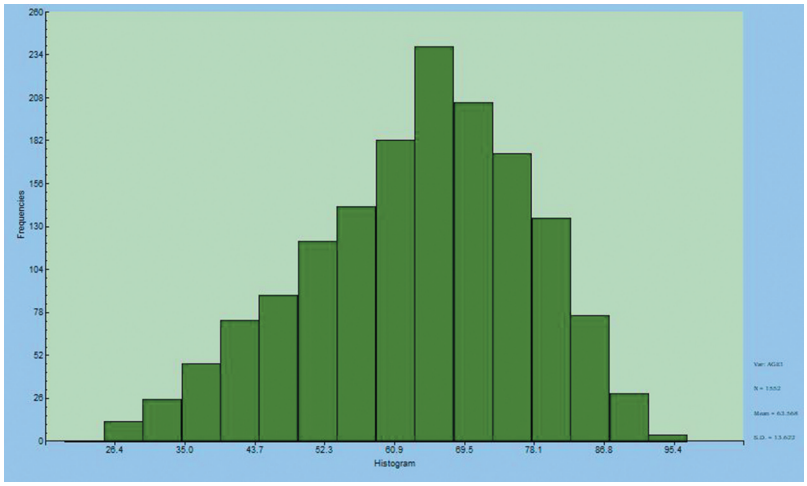

Figure 2. The UK data: age distribution.

The estimate $\hat{e}^2$ is the same for all models because

$$\hat{e}^2 = \hat{\alpha} - \hat{\beta}$$

for all models.

Using the residual covariance matrices for analysis, we obtain the estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and their estimated covariance matrix given in Table 3. Estimated parameters and their standard errors are given in Table 4.

Comparing Tabels 2 and 4, it is seen that the differences of the parameter estimates between the analysis before and after controlling for age are very small but the large differences between models remain. One can conclude that age has no effect on heredity which is as it should be.

Example 2: Australian data

Table 8.1, p. 151 in Neale and Cardon (1992) gives means, variances and correlations for mz and dz twins categorized by age (young = 30 years or less, and old = 31 years or older) and sex (males and females). This gives four groups of mz and dz twins. We pooled all groups into one group separately for mz and dz twins using the formulas

$$N_T = \sum_g N_g \,, \tag{42}$$

$$\bar{\mathbf{x}}_T = \frac{\sum_g N_g \bar{\mathbf{x}}_g}{N_T} \,, \tag{43}$$

$$\mathrm{SSCP}_g = N_g(\mathbf{S}_g + \bar{\mathbf{x}}_g \bar{\mathbf{x}}_g^{'}) \,, \tag{44}$$

$$\mathrm{SSCP}_T = \sum_g \mathrm{SSCP}_g \,, \tag{45}$$

$$\mathbf{S}_T = \frac{\mathrm{SSCP}_T}{N_T} - \bar{\mathbf{x}}_T \bar{\mathbf{x}}_T^{'} \,, \tag{46}$$

where $\bar{\mathbf{x}}$ is a 2×1 sample mean vector for twin 1 and twin 2, $\mathbf{S}$ is the corresponding 2×2 sample covariance matrix and SSCP is a 2×2 matrix of sums of squares and cross-products. Subscript g is a group index, $g = 1, 2, 3, 4$ and T denotes the total group.

The resulting sample covariance matrices are

$$\mathbf{S}_{\mathrm{mz}} = \begin{pmatrix} a & \\ b & c \end{pmatrix} = \begin{pmatrix} 9.939 & \\ 7.594 & 10.203 \end{pmatrix}, \tag{47}$$

$$\mathbf{S}_{\mathrm{dz}} = \begin{pmatrix} d & \\ e & f \end{pmatrix} = \begin{pmatrix} 10.002 & \\ 3.806 & 11.141 \end{pmatrix}, \tag{48}$$

based on $N_{\mathrm{mz}} = 1703$ and $N_{\mathrm{dz}} = 1029$.[2]

Using the Australian covariance matrices for analysis, we obtain the estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and their estimated covariance matrix shown in Table 5. Estimated parameters and their standard errors are given in Table 6.

The estimates for Australia data are much smaller that those of the UK data in absolute sense but taking standard error into account they are similar in relative sense. The estimate $\hat{c}^2$ for the ACE model is negative but non-

Table 3. The UK residual data: estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and their covariance matrix $\hat{\mathbf{\Sigma}}$.

		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
$\hat{\alpha}$	25.003	0.55685		
$\hat{\beta}$	19.776	0.54188	0.59542	
$\hat{\gamma}$	10.940	0.40898	0.39798	0.75326

Table 4. The UK residual data: parameter estimates and standard errors.

Parameter	ACE	ADE	ACDE
$\hat{h}^2$	17.673 (1.487)	23.982 (3.076)	6.763 (0.311)
$\hat{c}^2$	2.103 (1.430)	0*	5.740 (1.057)
$\hat{d}^2$	0*	−4.206 (2.840)	7.247 (0.818)
$\hat{e}^2$	5.227 (0.262)	5.227 (0.262)	5.227 (0.262)
$\hat{H}$	0.707 (0.058)	0.959 (0.112)	0.270 (0.009)

[2]Although the means and max/min values given in Table 8.1, p. 151 in Neale and Cardon (1992) are of the same order of magnitude as in the UK data, the variances reported are about 40% less than those in the UK data. The reason for this is unknown.

Table 5. Australia data: estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and their covariance matrix $\hat{\mathbf{E}}$.

		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
$\hat{\alpha}$	10.300	0.053181		
$\hat{\beta}$	7.816	0.051216	0.056446	
$\hat{\gamma}$	3.636	0.033385	0.032151	0.09161

Table 6. Australia data: parameter estimates and standard errors.

Parameter	ACE	ADE	ACDE
$\hat{h}^2$	8.358 (0.578)	6.729 (1.127)	2.869 (0.111)
$\hat{c}^2$	−0.543 (0.541)	0*	1.287 (0.393)
$\hat{d}^2$	0*	1.086 (1.082)	3.660 (0.320)
$\hat{e}^2$	2.485 (0.085)	2.485 (0.085)	2.485 (0.085)
$\hat{H}$	0.811 (0.054)	0.653 (0.105)	0.278 (0.008)

significant. Otherwise, all estimates are significant except $\hat{d}^2$ for the ADE model. The estimate of heredity H is fairly close to the same as for the UK data.

Two populations

Suppose we have two populations, for example, the UK and Australia. For each population, we have two groups of twins, one mz group and one dz group as in previous sections.

We regard H as a universal parameter, which is an attribute of the phenotype we are measuring, for example, BMI. So we wish to test the hypothesis that H is the same in both populations. Note that this is not the same as testing that h^2 is the same in both populations which is not likely to hold, since, as already noted, the variances of the phenotypes are very different in the UK and Australia.

We estimate α, β, γ separately for each population as described in Section Estimation. We save the information matrix $\hat{\mathbf{E}}_1$ and $\hat{\mathbf{E}}_2$ for each population. We also estimate H and $\mathrm{AVar}(\hat{H})$ for each population.

The hypothesis to be tested is $H_1 = H_2$ or from (23)

$$0 = H_1 - H_2 = \frac{p\beta_1 + q\gamma_1}{\alpha_1} - \frac{p\beta_2 + q\gamma_2}{\alpha_2} \tag{49}$$

where p and q depends on the model, ACE, ADE, or ACDE as before.

To test the hypothesis, $H_1 = H_2$ one can use a Wald statistic W which comes out as

$$W = \left(\hat{H}_1 - \hat{H}_2\right)^2 \hat{\alpha}_1^{-4} \left(-\left(p\hat{\beta}_1 + q\hat{\gamma}_1\right) p\hat{\alpha}_1 q\hat{\alpha}_1\right) \hat{\mathbf{E}}_1 \begin{pmatrix} -(p\hat{\beta}_1 + q\hat{\gamma}_1) \\ p\hat{\alpha}_1 \\ q\hat{\alpha}_1 \end{pmatrix} + \left(\hat{H}_1 - \hat{H}_2\right)^2 \hat{\alpha}_2^{-4} \left(-\left(p\hat{\beta}_2 + q\hat{\gamma}_2\right) \; p\hat{\alpha}_1 \; q\hat{\alpha}_1\right) \hat{\mathbf{E}}_2 \begin{pmatrix} -(p\hat{\beta}_2 + q\hat{\gamma}_2) \\ p\hat{\alpha}_2 \\ q\hat{\alpha}_2 \end{pmatrix} = \left(\hat{H}_1 - \hat{H}_2\right)^2 \left[\mathrm{AVar}(\hat{H}_1) + \mathrm{AVar}(\hat{H}_2)\right]$$

If the hypothesis holds, W is asymptotically distributed as χ^2 with one degree of freedom.

W is easily calculated from previous results. For the ACDE model we have

$$(0.269 - 0.278)^2 (0.009^2 + 0.008^2) = 0.000117 \, . \tag{50}$$

Hence, it is obvious that the hypothesis cannot be rejected. There is strong evidence that heredity of BMI is the same in the UK and Australia.

Summary and conclusion

We considered the classical models ACE and ADE for analysis of mz and dz twins and suggested an alternative model ACDE which has four components of variance: additive genes, dominance genetic deviations, common environments, and unique environments. We showed that all three models can be estimated by maximum likelihood from linear combinations of the implied variances and covariances of the phenotypes. Based on data on BMI from both the UK and Australia, it seems that the ACDE model works much better than the ACE and ADE model. The ACDE model gives stable and significant estimates of all four parameters, whereas ACE and ADE often give negative and non-significant estimates of one or the other parameter. The heredity of BMI is very precisely estimated as 27% in the UK and 28% in Australia, whereas ACE and ADE suggest much higher values. We also showed how to test the hypothesis that heredity is the same in two populations. For the UK and Australia data, there is strong evidence that heredity of BMI is the same.

References

Fisher, R. L. (1918). The correlations between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society*, *52*, 399–433. Edinburgh.

Magnus, J. R., & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. Revised Addition. Wiley.

Mather, K., & Jinks, J. L. (1971). *Biometrical genetics*. Chapman and Hall.

Neale, C. M., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Kluwer Academic Publishers.