Testing distributional assumptions in psychometric measurement models with substantive applications in psychology

Molenaar, D.

Publication date
2012

6

Using the Heteroscedastic Factor Model to Detect Specific Genotype by Environment Interactions in the Classical Twin Design.

Considerable effort has been devoted to the analysis of Genotype by Environment (GxE) interactions in various phenotypic domains, such as cognitive abilities and personality. In many studies, environmental variables were observed (measured) variables. In case of an unmeasured environment, van der Sluis et al. (2006) proposed to study heteroscedasticity in the factor model using only MZ twin data. This method is closely related to the Jinks & Fulker test for GxE (1970), but slightly more powerful. In this Chapter, we identify four challenges to the investigation of GxE in general, and specifically to the heteroscedasticity approaches of Jinks and Fulker and van der Sluis et al. We propose extensions of these approaches purported to solve these problems. These extensions comprise: 1) including DZ twin data, 2) modeling both AxE and AxC interactions; and 3) extending the univariate approach to a multivariate approach. By means of simulations, we study the power of the univariate method to detect the different GxE interactions in varying situations. In addition, we study how well we could distinguish between AxE, AxC, and CxE. We apply a multivariate version of the extended model to an empirical data set on cognitive abilities.

6.1 Introduction

The topic of Genotype by Environment (GxE) interaction has received increasing attention in the past decade in twin and family studies, and in (genome-wide) genetic association studies (GWAS). A GxE interaction denotes the degree to which the

* This chapter is published as: Molenaar, D., van der Sluis, S., Boomsma, D.I., & Dolan, C.V. (in press). Detecting Specific Genotype by Environment Interaction using Marginal Maximum Likelihood Estimation in the Classical Twin Design. *Behavior Genetics*.

phenotypic variation explained by genetic factors varies across environmental conditions, or, conversely, the degree to which phenotypic variation explained by environmental influences varies across genotypes (see Boomsma & Martin, 2002).

Using multi-group designs (Boomsma, de Geus, van Baal, & Koopmans, 1999) or the moderation model proposed by Purcell (2002), various twin- and family studies have shown that within the ACE-model, the phenotypic variance decomposition into additive genetic factors (A), common environmental factors (C) and unique environmental factors (E) varies across environmental conditions. This has been established with respect to various behavioral measures (e.g., aggression and alcohol consumption; see Kendler, 2001, for a review including more examples) and specifically with respect to cognitive ability (Bartels et al., 2009a; Grant et al., 2010; Harden et al., 2007; Johnson et al., 2009a; Turkheimer et al., 2003; van der Sluis et al., 2008), personality (Bartels et al., 2009b; Boomsma et al., 1999; Brendgen et al., 2009; Distel et al., 2010; Heath et al., 1998; Hicks et al., 2009a; Hicks et al., 2009b; Johnson et al., 2009b; Silberg et al., 2001; Tuvblad et al., 2006; Zhang et al., 2009), health-related phenotypes (Johnson & Krueger, 2005; Johnson et al., 2010; McCaffery et al., 2008; McCaffery et al., 2009), and measures of brain morphology (Lenroot et al., 2009; Wallace et al., 2006).

In these studies, the extent to which the additive genetic factor A explains phenotypic variation fluctuates as a function of a specific measured environmental variable. It has, however, proven difficult to identify the (multiple) relevant environmental conditions that moderate the influence of genetic factors (e.g., Eichler et al., 2010). In GWAS, for example, GxE interaction is usually not modeled, although in theory, the presence of unmodeled GxE may affect the power to detect genetic variants (e.g., Eichler et al., 2010; Maher, 2008; Manolio et al., 2009).

As the identification of environmental variables involved in GxE can be difficult, methods to detect GxE interactions given unmeasured genetic and environmental factors remain useful. At presence, two MZ-twin based methods are available. Letting Y_1 and Y2 denote MZ twin pair scores, Jinks & Fulker (1970) showed that GxE may be detected in the dependency of $|Y_1 - Y_2|$, a proxy for the variance of E, on Y_1+Y_2, a proxy for the level of A (see Jinks & Fulker, 1970). In a similar approach, van der Sluis et al. (2006) used Marginal Maximum Likelihood to test for heteroscedastic E variance by conditioning on A in MZ twin data (Hessen & Dolan, 2009; Molenaar, Dolan, & Verhelst, 2010). Like Jinks and Fulker (1970), these authors focused on the detection of AxE, i.e., heteroscedastic E variance as a function of A.

In the following, we use the term ‘GxE’ to refer to the general concept of ‘genotype-by-environment interaction’. In addition, we refer to specific instances of GxE that are modeled in a given statistical model (e.g., AxE in the ACE model; AxM in the moderation model of Purcell, 2002, where M is a measured variable).

6.1.1 Problems with existing heteroscedasticity approaches

The methods of Jinks and Fulker (1970) and van der Sluis et al. (2006) face a number of challenges. Here we address the following four: non-normality, conflation of AxE and CxE, heteroscedastic measurement error, and genotype-environment correlation.

Non-normality. As heteroscedasticity due to GxE results in non-normality of the observed phenotypic variable, other sources of non-normality can result in spurious GxE. These include floor and ceiling effects (see van der Sluis et al., 2006), poor scaling of the measurement (Eaves, 2006; Evans, Gillespie, & Martin, 2002) and non-linear factor-to-indicator relations (Tuker-Drob, Harden, & Turkheimer, 2009).

Heteroscedastic measurement error. As discussed by Turkheimer & Waldron (2000), the statistical 'unique environment factor', E, is not necessarily equal to the conceptual notion of environmental influences underlying phenotypic scores, as the former may for instance include measurement error (see also Loehlin & Nichols, 1976). This is a challenge as heteroscedastic measurement error may mimic GxE.

Conflation of AxE and CxE. The existing univariate approaches by Jinks and Fulker and van der Sluis utilize MZ twin data only. This precludes distinguishing between the additive genetic effects, A, and the common environment effects, C (Evans et al., 2002). It is therefore possible that an observed effect can be due to CxE rather than AxE.

Genotype-environment correlation. Measures of the environment that interact with A may themselves be affected by either the same or unique genetic influences (e.g., Turkheimer, Harden, D'Onofrio, & Gottesman, 2009). Such genotype-environment correlation is known to affect tests using measured environments, in both the case that the genetic influences are unique and common to the measured environment and the phenotype (Purcell, 2002). It is however unknown how it affects the heteroscedasticity approaches as presented above.

Note that the problems discussed above are not limited to the approaches of Jinks & Fulker and van der Sluis et al. in which the environment is unmeasured. Given measured environment, non-normality of the phenotypic variable can also result in spurious GxE (Purcell, 2002). In addition, testing for GxE in presence of a genotype-environment correlation is a challenge in the measured moderator approach as well (see van der Sluis, Posthuma, & Dolan, 2011; Rathouz et al., 2008).

6.1.2 Towards a solution

In this Chapter, we address the problems mentioned above in an extended version of the approach of van der Sluis et al. Specifically, we extend the van der Sluis et al. method to include dizygotic (DZ) twin data to avoid the conflation of the A and C components. The inclusion of DZ data has several advantages: First, one can distinguish between AxE and AxC. Second, inclusion of DZ twin data will increase the power simply due to the increase in total sample size. Third, AxE effects may be detected more readily if the C component can be isolated. Finally, as A and C are separated, we hypothesize that the presence of CxE does not result in spurious AxE.

In addition to the extension of van der Sluis et al (2006), we propose a multivariate extension. In the multivariate extension we use the common path way model to distinguish between the measurement model (a phenotypic one factor model) and the biometric model (McArdle & Goldsmith, 1984; Kendler, Heath, Martin, & Eaves, 1987; Franić et al., 2011). In this model, genetic and environmental influences contribute to the observed phenotypic variance via one common phenotypic construct. In the measurement model, the observed phenotypic variables are linked to the latent phenotypic construct. In the biometric model, the latent phenotypic construct is decomposed into the A, C, and E components. In this way we can introduce the AxE and AxC interactions at the level of the construct, instead of at the level of the observed variable. We thereby avoid the conflation of measurement error with unique environment influences, as measurement error is now explicitly modeled in the measurement part of the model, and the unique environment factor is separately modeled at the level of the latent phenotypic construct. So we can introduce heteroscedastic residuals in the measurement model to account for floor, ceiling, and/ or poor scaling effects, and test GxE at the level of the biometric model.

Below, we first shortly introduce the univariate method discussed by van der Sluis et al. (2006) to detect AxE interactions in MZ twin data. Next, we extend this model to an ACE-model with both AxE and AxC interactions. We then investigate the extended model in simulation studies. We investigate whether the method can properly distinguish the different interactions. In addition, we compare the power to detect the various interactions of the extended method to the power of the van der Sluis et al. approach. We also investigate whether we can distinguish between AxE / AxC on the one hand and CxE on the other hand. Furthermore, we compare the present method with unmeasured C and E factors to the approach of Purcell (2002) that makes use of measured environment variables. Next, we discuss an extension of the method to include multivariate data, and apply the multivariate extension to an IQ data set (Osborne, 1980). We conclude the Chapter with a short discussion.

6.2 The Univariate Case

6.2.1 Van der Sluis' model: AE

Van der Sluis et al. (2006) was limted to the AE model. Specifically, given N twin pairs:

$$Y_j = \upsilon + a \times A_j + e \times E_j \quad (1)$$

where Y_j denotes the phenotypic score of the j-th twin member (j = 1,2), and A_j and E_j denote the zero mean additive genetic and unshared environmental factor, respectively. The parameter υ is the intercept (phenotypic mean) and a and e are regression coefficients (factor loadings).

Given the usual assumptions of the twin method, the MZ covariance matrix includes the elements:

$$\text{var}(Y_1) = \text{var}(Y_2) = \sigma_A^2 + \sigma_E^2 \quad (2)$$

$$\text{cov}(Y1, Y2) = \sigma_A^2 \quad (3)$$

To test for a possible AxE interaction, van der Sluis et al. (2006) proposed to test for heteroscedasticity of σ_E^2, by testing whether σ_E^2 varied systematically over the values of factor A. They specified a parametric function between σ_E^2 and the score of the twins on A, i.e.,

$$\sigma_E^2 | A = \exp(\beta_0 + \beta_1 A). \quad (4)$$

where 'σ_E^2| A' denotes 'σ_E^2 conditional on the level of A'. The exponential function, exp(.), is used to avoid negative variances (see also Bauer & Hussong, 2009; Hessen & Dolan, 2009; Molenaar, Dolan & Verhelst, 2010). In the equation, β_0 is a baseline parameter and β_1 is a heteroscedasticity parameter, which models the dependency of σ_E^2 on A. If $\beta_1 = 0$, the model reduces to the standard AE-model. The model may be extended to accommodate more complicated relations between σ_E^2 and A, e.g., $\sigma_E^2 | A = \exp(\beta_0+\beta_1 A+\beta_2 A^2)$.

To fit the model to data, van der Sluis et al. used Marginal Maximum Likelihood (Bock & Aitkin, 1981). As $A_1 = A_2 = A$, the marginal log likelihood function contains a single integral over A, which may be approximated using an one-dimensional Gauss-Hermite quadrature approximation, i.e.,

$$\ell = \log \int_{-\infty}^{\infty} f(y_1, y_2; \mu \mid A, {\sigma_E}^2 \mid A) g(A) dA \approx \log \sum_{g=1}^{Q} W_g \times f(y_1, y_2; \mu \mid N_g, {\sigma_E}^2 \mid N_g) \quad (5)$$

where $g(.)$ is the normal density for factor A, $f(.)$ is the bivariate normal density function for y_1 and y_2, conditional on the level of A, with $\mu|A = \nu + aA$, and $\sigma_E^2|$ A given by Equation 4, and $cor(y_1,y_2)|A = 0$. W_g and N_g are the g-th weight and node in the Gauss-Hermite quadrature approximation (e.g., Stroud and Secrest, 1966). Van der Sluis et al (2006) showed that the model performed well in terms of statistical power to detect the AxE interaction. Below we extend this model by the addition of the DZ twins.

6.2.2 ACE-model

In the classical twin model, including MZ and DZ twins, the phenotypic covariance matrix of the ACE model includes the elements:

$$var(Y_1) = var(Y_2) = \sigma_A^2 + \sigma_C^2 + \sigma_E^2 \quad (6)$$

$$cov(Y_1,Y_2) = \rho_A * \sigma_A^2 + \sigma_C^2 \quad ,$$

where σ_C^2 is the shared environmental variance and ρ_A is 1 (MZ) or .5 (DZ). We now consider both AxE and AxC interactions. To introduce the AxE interaction, we proceed as above, i.e.,

$$\sigma_E^2 | A_j = \exp(\beta_0 + \beta_1 A_j) \quad (7)$$

We now include the subscript *j* because A of twin 1 and twin 2 are distinct in DZ twins. We model AxC interaction as heteroscedastic C variance, conditional on A:

$$\sigma_C^2 | A_j = \exp(\gamma_0 + \gamma_1 A_j) \quad (8)$$

with

$$cov(C_1,C_2) = \sqrt{{\sigma_C}^2 \mid A_1 \times {\sigma_C}^2 \mid A_2} \quad (9)$$

where γ_0 and γ_1 are the baseline and heteroscedasticity parameter, respectively (as β_0 and β_1 in Equation 7). If AxC is present, the covariance between C_1 and C_2 will vary as

a function of A_1 and A_2. However, as required, the correlation between C_1 and C_2 will be 1 for every level of both A_1 and A_2. We model these AxC and AxE simultaneously, i.e., we estimate β_1 and γ_1 simultaneously. In the standard ACE-model without GxE, the distribution of the phenotypic scores of the twins and their co-twins is assumed to be a bivariate normal distribution (Figure 6.1a). In case of GxE, the bivariate distribution of the data becomes skewed due to AxC (Figure 6.1b) or AxE (Figure 6.1c). As can be seen, the two types of interactions result in specific violations of bivariate normality. Specifically, the presence of a positive AxC interaction ($\gamma_1 >0$; C variance is increasing across A) results in an observed distribution that is skewed to the right, see Figure 6.1b. Similarly for positive AxE, see Figure 6.1c. In addition, a negative AxC interaction ($\gamma_1<0$) or a negative AxE interaction ($\beta_1<0$) results in left skew.

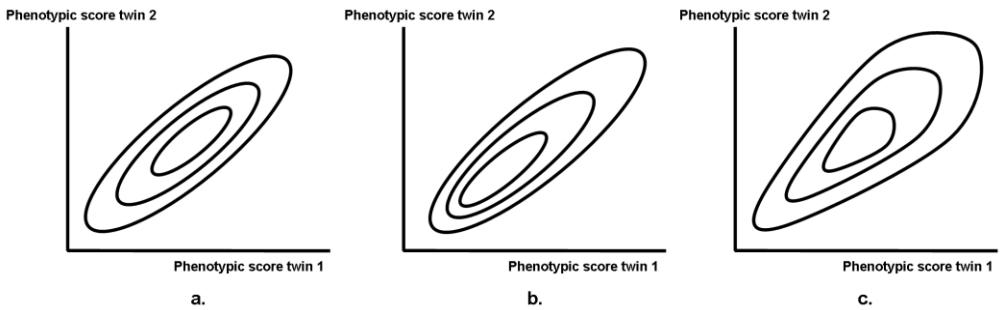

Figure 6.1. Schematic representation of the implied bivariate distribution of the twin data in case of a) the standard ACE-model; b) an ACE model with positive AxC ($\gamma_1>0$); and (c) an ACE-model with positive AxE ($\beta_1>0$).

In this approach of modeling GxE we choose to model σ_E^2 and σ_C^2 as a function of a latent A factor. This is different from Purcell (2002) who modeled the factor loading of A as a function of observed E or C. We choose the former option as it connects better to the framework of Jinks & Fulker (1970) who define GxE as heteroscedastic E with respect to A (see also Evans et al., 2002).

With MZ and DZ twin data, the marginal log likelihood involves a double integral (i.e., over A_1 and A_2), which can be approximated using multivariate Gauss-Hermite quadratures. As we have 2 dimensions now, we have two sets of nodes, N_{1g} and N_{2h}, where $g = 1, \ldots, Q$ and $h = 1, \ldots, Q$ (the total number of nodes is therefore Q^2).

Standard two-dimensional Gauss-Hermite quadrature approximation assumes both dimensions (here A_1 and A_2) to be uncorrelated. We therefore transform the nodes N_{1g} and N_{2h} into N_{1g}^* and N_{2h}^* so that these transformed nodes have the proper correlations (i.e., 1 for MZ twins and .5 for DZ twins). Thus for the MZ twins we use

$$N_{1g}^{*} = N_{1g} \tag{10}$$

$$N_{2h}^{*} = N_{1h} \tag{11}$$

and for the DZ twins:

$$N_{1g}^{*} = N_{1g} \tag{12}$$

$$N_{2h}^{*} = .5 \times N_{1g} + \sqrt{1-0.5^2} \times N_{2h} \tag{13}$$

The likelihood function of the model is now given by

$$\ell = \log \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2; \mu \mid A_1, \mu \mid A_2; \sigma^2 \mid A_1, \sigma^2 \mid A_2) h(A_1, A_2) dA_1 dA_2$$

$$\approx \log \sum_{g=1}^{Q} \sum_{h=1}^{Q} W_g W_h \times f(y_1, y_2; \mu \mid N_{1g}^{*}, \mu \mid N_{2h}^{*}; \sigma^2 \mid N_{1g}^{*}, \sigma^2 \mid N_{2h}^{*})$$

where $h(.)$ is the multivariate normal distribution for A_1 and A_2, $f()$ is the bivariate normal distribution of Y_1 and Y_2 with $\mu|A_j = \nu + \sigma_A \times A_j$ and

$$\sigma^2|A_j = \sigma_E^2|A_j + \sigma_C^2|A_j. \tag{14}$$

W_g and W_k are the same weights as in the AE model (see above). The conditional correlation between y_1 and y_2 is

$$\text{cor}(y_1,y_2)|A_1^{*},A_2^{*} = \frac{\sqrt{\sigma_C^2 \mid A_1 \times \sigma_C^2 \mid A_2}}{\sqrt{\sigma_C^2 \mid A_1 + \sigma_E^2 \mid A_1}\sqrt{\sigma_C^2 \mid A_2 + \sigma_E^2 \mid A_2}}. \tag{15}$$

6.3 Simulation Study 1

With the present models in place, we studied how well we can detect the various types of interactions, and how well we can distinguish between them. In addition we investigated whether the presence of a CxE interaction will influence the detection of AxE and/or AxC.

6.3.1 Design

We simulated data according to 3 scenarios. In all scenario's A, C, and E are continuous variables. In scenario I, named 'A predominant', explained phenotypic variances by the A, C, and E factor equaled approximately 50%, 25% and 25% respectively (in the absence of any GxE interaction). In scenario II, named 'AC predominant', explained variances equaled approximately 40%, 40%, and 20% for the A, C, and E factors respectively. Finally, in scenario III, named 'C predominant', explained variances equaled 20%, 60%, and 20%

Within each scenario we simulated five different data sets. The first data set included an AxE interaction. The second data set included an AxC interaction. The third data set was simulated with both interactions (AxC and AxE) in the same direction, the fourth data set was simulated with both interactions in opposite direction, and the fifth data set included a CxE interaction. For each scenario, we additionally simulated a data set with no effect, i.e., according to the standard homoscedastic ACE-model. All data sets including an interaction effect were simulated to include either a small, a medium, or a large effect. We considered an interaction 'small' when the percentage of variance explained by the environmental factor in question increased with 3% to 4% for each standardized unit of A within the [-3; 3] interval. In the 'medium' condition, explained variance increased with 4% to 5% over the levels of A. In the 'large' condition explained variance increased with 5% to 6% over the levels of A. See Table 6.1 for the true values of the heteroscedasticity parameter, β_1 and γ_1. The other parameters equaled: $\sigma_A^2 = 4$, β_0=0.45, and γ_0=0.45 (scenario I), $\sigma_A^2 = 4$, β_0=0.65, and γ_0=1.40 (scenario II), and $\sigma_A^2 = 2$, β_0=0.65, and γ_0=1.70 (scenario III). See Figure 6.2 for a graphical representation of the effect sizes across the scenarios.

Table 6.1

Mean, standard deviation and percent bias of the parameter estimates in simulation study 1 for the GxE parameters.

Effect	Scenario	Size	AxE parameter β_1				AxC parameter γ_1			
			true	mean	SD	%bias	true	mean	SD	%bias
AxC	I	small	-	-	-	-	**0.20**	0.17	0.15	-15.11
		medium	-	-	-	-	**0.25**	0.21	0.17	-14.20
		large	-	-	-	-	**0.30**	0.25	0.15	-18.19
	II	small	-	-	-	-	**0.15**	0.11	0.08	-26.11
		medium	-	-	-	-	**0.20**	0.15	0.08	-26.41
		large	-	-	-	-	**0.25**	0.18	0.08	-26.80
	III	small	-	-	-	-	**0.15**	0.11	0.07	-26.83
		medium	-	-	-	-	**0.20**	0.14	0.07	-27.76
		large	-	-	-	-	**0.25**	0.18	0.07	-28.49
AxE	I	small	**0.20**	0.22	0.09	11.67	-	-	-	-
		medium	**0.25**	0.28	0.09	13.97	-	-	-	-
		large	**0.30**	0.34	0.09	13.36	-	-	-	-
	II	small	**0.20**	0.22	0.10	10.08	-	-	-	-
		medium	**0.25**	0.28	0.10	13.24	-	-	-	-
		large	**0.30**	0.34	0.10	12.82	-	-	-	-
	III	small	**0.20**	0.21	0.12	3.50	-	-	-	-
		medium	**0.25**	0.27	0.11	6.54	-	-	-	-
		large	**0.30**	0.32	0.11	7.93	-	-	-	-
Opp	I	small	**0.20**	0.27	0.10	35.51	**-0.20**	-0.24	0.15	22.01
		medium	**0.25**	0.33	0.10	31.54	**-0.25**	-0.29	0.13	14.17
		large	**0.30**	0.39	0.09	31.46	**-0.30**	-0.33	0.13	8.36
	II	small	**0.20**	0.27	0.12	36.29	**-0.15**	-0.16	0.09	6.81
		medium	**0.25**	0.34	0.11	34.33	**-0.20**	-0.20	0.08	1.48
		large	**0.30**	0.40	0.11	34.54	**-0.25**	-0.24	0.08	-2.42
	III	small	**0.20**	0.24	0.15	21.95	**-0.15**	-0.14	0.09	-9.26
		medium	**0.25**	0.31	0.14	24.79	**-0.20**	-0.18	0.08	-10.25
		large	**0.30**	0.38	0.13	26.28	**-0.25**	-0.22	0.08	-11.67
Same	I	small	**0.20**	0.25	0.12	26.19	**0.20**	0.07	0.19	-65.05
		medium	**0.25**	0.32	0.11	26.82	**0.25**	0.10	0.19	-59.16
		large	**0.30**	0.37	0.10	21.92	**0.30**	0.14	0.20	-54.17
	II	small	**0.20**	0.24	0.14	22.20	**0.15**	0.07	0.11	-55.94
		medium	**0.25**	0.31	0.13	22.82	**0.20**	0.10	0.11	-50.30
		large	**0.30**	0.35	0.13	18.00	**0.25**	0.13	0.11	-47.06
	III	small	**0.20**	0.20	0.18	1.85	**0.15**	0.10	0.11	-36.17
		medium	**0.25**	0.26	0.18	5.25	**0.20**	0.13	0.11	-35.13
		large	**0.30**	0.31	0.17	3.89	**0.25**	0.16	0.11	-37.51

Note. 'same dir.' means that both an AxE and an AxC interaction are in the data in the same direction. 'opp. dir.' means that they are in opposite direction.

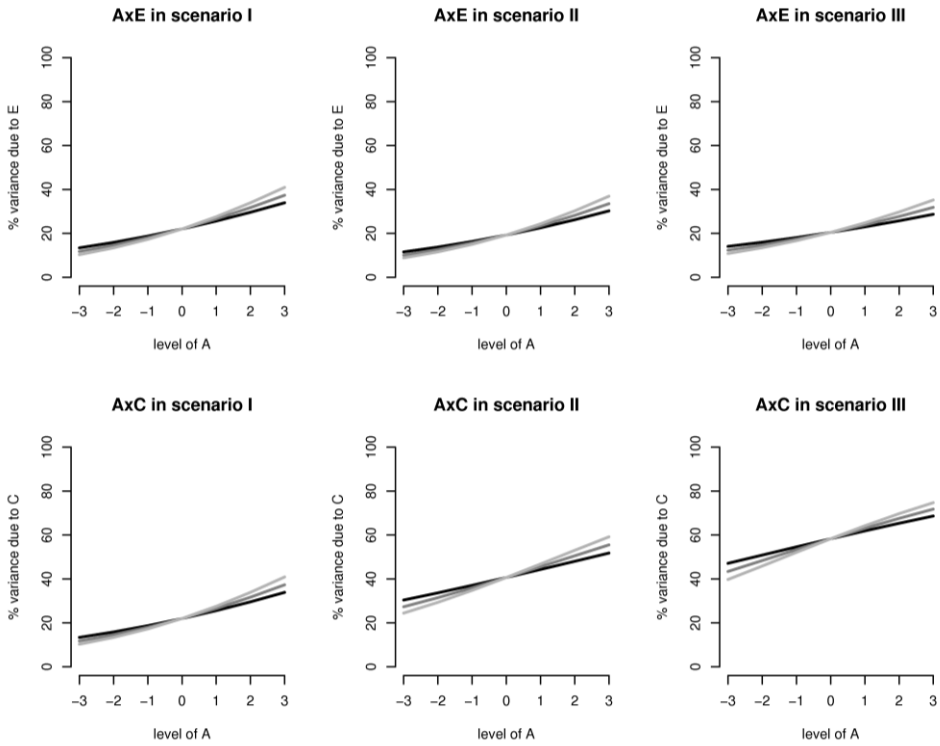

Figure 6.2. A graphical representation of the effect sizes as used in simulation study 1. Depicted is – for each scenario - the percentage of total variance explained by E (top graphs) and C (bottom graphs) as a function of the level of A.

For each condition in the design of the simulation study we simulated 1000 data sets with 500 MZ and 500 DZ twin pairs. To each of these data sets, we fitted an ACE model: 1) with AxE interaction (ACE-AxE), 2) with AxC interaction (ACE-AxC), 3) with an AxE and an AxC interaction simultaneously (ACE-AxE-AxC), and 4) with AxE interaction using the MZ twin data only (AE-AxE). For each model, we calculated the power of the likelihood ratio test to detect the effects in the model (see Saris & Satorra 1993; Satorra & Saris, 1985). See Molenaar, Dolan, & Wicherts (2009) for an easy step-by-step illustration. All models were fitted in the freely available software package Mx (Neale, Boker, Xie, & Maes, 2002). We used Marginal Maximum Likelihood estimation (Bock & Aitkin, 1981) with 100 multivariate Gauss-Hermite quadrature points (i.e., 10 for each dimension) to approximate both integrals in the likelihood function as discussed above. In case of the AE-model, we used 10 quadrature points as the likelihood function of this model only includes a single integral. Power was calculated using a .05 level of significance. All Mx input scripts are available from www.dylanmolenaar.nl.

6.3.2 Results

In Table 6.1, parameter recovery is summarized for the cases in which the true model is fitted to the data (e.g., ACE-AxE when the data contains an AxE effect and ACE-AxE-AxC when the data contains both effects). In the Table, average parameter estimates of the GxE parameters, β_1 and γ_1 are shown together with their true values, standard deviation, and bias (which is defined as the difference between the average estimate and the true value divided by the true value). As appears from the Table, in case of an AxC effect in the data, the AxC parameter γ_1 is somewhat underestimated within the ACE-AxC with percent bias between 15 and 29 percent in the three scenarios. In case of only an AxE effect in the data, the AxE parameter, β_1, of the ACE-AxE is hardly biased with bias between 3 and 14 percent. In the case that both effects are in the opposite direction in the data, β_1 is overestimated (bias between 20 and 37 percent), but γ_1 is reasonably unbiased (bias between -11 and 22 percent). In the case that both effects are in the same direction in the data, β_1 is somewhat biased in scenario I and II, but not biased in scenario III, and γ_1 is severely biased in scenario I and II. The latter suggests that when both effects are in the same direction in scenario I and II, the AxC effect is absorbed to some degree by the AxE parameter β_1.

Table 6.2 shows the power of the different models to detect the effects in scenario I ('A predominant'). We only focus on scenario I to save space (as tables get really large) and because the main conclusions are the same for all scenario's. However, power results of scenario II and III are available from www.dylanmolenaar.nl. As can be seen in Table 6.2, in the absence of an effect, power coefficients approximately equal the level of significance (0.05). For example, when only an AxE is in the data, power to detect AxC should equal 0.05, as ideally the AxE effect in the data should not be detected as an AxC interaction. For all such cases, power coefficients are underlined in Table 6.2.

The underlined power coefficients in the table show that for each effect size, false positives are largely absent. That is, all power coefficients are close to .05 in the absence of an effect. Furthermore it can be concluded from the power coefficients that in the ACE-AxE-AxC model, the distinct interaction effects (AxE versus AxC) are generally not confounded. However, in the ACE- AxE and ACE-AxC models, there is an increased risk on false positives. Specifically, the ACE-AxE model has an increased power to detect the AxC effect, and the ACE-AxC model has an increased power to detect the AxE effect.

If we consider the power to detect the effects that are actually in the data (i.e., the power coefficients that are not underlined in Table 6.2), we can conclude that within the ACE-models, the power to detect an AxE interaction is generally acceptable. For the ACE- AxE-AxC model, power is good for a large effect size (.92), power is

acceptable for a medium effect size (.81), and moderate for a small effect size (.61). Power to detect AxC interaction using the different models is far lower than the power to detect AxE. That is, large sample sizes are needed to detect the AxC effect. For the ACE-AxE-AxC model, power to detect AxC is at most . 32 in case of a large effect size, while it is .92 for AxE. However, if the AxC interaction is accompanied by an AxE interaction in the opposite direction, effects are somewhat easier to resolve with power of at most .70.

Table 6.2
Power to detect AxC and AxE using different models in scenario I.

effect	**data**	**ACE-AxE-AxC**			**ACE-AxC**	**ACE-AxE**	**AE 500**	**AE 1000**
	power to detect:	*AxC*	*AxE*	*both*	*AxC*	*AxE*	*AxE*	*AxE*
no GxE		0.06	0.05	0.05	0.05	0.05	0.05	0.05
small	**AxE**	0.08	0.61	0.26	0.72	0.64	0.42	0.70
	AxC	0.16	0.05	0.24	0.11	0.18	0.05	0.05
	same dir.	0.05	0.54	0.70	0.90	0.83	0.47	0.76
	opp. dir.	0.35	0.67	0.06	0.44	0.57	0.35	0.60
	CxE	0.07	0.30	0.13	0.36	0.29	0.20	0.36
medium	**AxE**	0.09	0.81	0.40	0.90	0.85	0.63	0.90
	AxC	0.24	0.05	0.34	0.15	0.27	0.05	0.05
	same dir.	0.07	0.73	0.90	0.98	0.97	0.72	0.95
	opp. dir.	0.50	0.85	0.09	0.65	0.79	0.52	0.81
	CxE	0.07	0.42	0.17	0.50	0.42	0.32	0.56
large	**AxE**	0.11	0.92	0.55	0.97	0.95	0.77	0.97
	AxC	0.32	0.05	0.46	0.19	0.36	0.05	0.05
	same dir.	0.10	0.85	0.97	1.00	1.00	0.84	0.99
	opp. dir.	0.70	0.96	0.12	0.82	0.93	0.68	0.93
	CxE	0.11	0.56	0.24	0.66	0.59	0.43	0.72

Note. 'same dir.' means that both an AxE and an AxC interaction are in the data in the same direction. 'opp. dir.' means that they are in opposite direction. Underlined values consider the cases in which the fitted model includes none of the effects in the data. The results for the AE-model are based on the MZ-twins only. AE 500 and AE 1000 means that the analyses are based on 500 MZ and 1000 MZ twins respectively.

We now compare the results of the models including data for both MZ and DZ twins with the AE-model, which includes data of MZ twins only. As the previous analysis involved a total of 1000 subjects, we calculated the power of the AE-model to detect the interactions in the data in case of 1000 MZ twins. In this case, power is approximately equal to the ACE-AxE model.

Finally, from Table 6.2 we conclude that the presence of a CxE interaction results in an increased false positive rate in detecting AxE. Specifically, given a small effect size, the ACE-AxC-AxE model has a power of .30 to detect an AxE interaction, while a CxE interaction is in the data. This power coefficient could be compared to the case that there truly is an AxE interaction in the data. In that comparison, this model has a power of .61 to detect an AxE effect. Thus, from Table 6.2 it can be seen that in scenario I for all effect sizes, power to detect AxE is larger when AxE is present than when CxE is present, which is reasonably acceptable. However, with respect to scenario II and III (not tabulated), results are somewhat different: In scenario II, where C explains more variance, the power to detect an AxE interaction is about equal when AxE is in the data and when CxE is present, for all effect sizes. In scenario III where C is the predominant factor, power to detect an AxE interaction is even larger when CxE is present than when AxE is in the data.

6.3.3 Conclusion

Overall, the power to detect an AxE interaction is acceptable. In contrast, to detect an AxC interaction, large sample sizes are needed as the power is low. This appears to be mainly due to underestimation of the AxC parameter, particularly in the case that AxC and AxE effects are both present in the same direction. However, results show that it could be important to take the AxC effect into account as it will increase the power to detect an AxE interaction. Within the ACE model, it is thus advisable to use the ACE-AxE-AxC model when one has no idea whether the interaction is AxE or AxC. Using the ACE-AxE or the ACE-AxC model can lead to an increased false positive rate (i.e., an AxC may be detected as an AxE, while AxE is absent).

Besides the underestimation of AxC, it appeared that the AxE effect could in some cases be somewhat overestimated. However, this is not a main problem as it appeared from the power study that the AxE effect is not associated with false positives. That is, when there is no AxE effect in the data, no spurious AxE effect arise.

From the simulation it is also clear that one can distinguish relatively well between AxE and AxC. However it is difficult to distinguish between AxE and CxE, particularly when C is a relatively large source of variation. If a CxE interaction is present, it may be mistakenly detected as an AxE interaction. We return to this point in the discussion.

6.4 Simulation Study 2

In simulation study 2 we investigate the relation between the present approach with unmeasured environment, and the GxE approach where the environment is measured (Purcell, 2002). First, it is interesting to see how interactions between genotypes and measured environment are detected in the ACE-AxE-AxC model, and second it is interesting to see how the ACE-AxE-AxC model deals with genotype by environment interactions where the environment is open to genetic influences as well. To investigate this, we simulated data according to an ACE-model in which the A component is moderated by a measured environment variable. We distinguish between two cases 1) *univariate moderation,* in which the environment moderates the genetic variance unique to the phenotype of interest (i.e., the moderator may be influenced by genes, but these genes are not shared with the phenotype of interest); 2) *bivariate moderation,* in which the environment moderates the genetic variance common to the moderator and the phenotype of interest (i.e., the moderator is influenced by the same genes as the phenotypic variable resulting in a genotype by environment correlation). Purcell (2002) proposed a model for both cases, which we refer to as the univariate and bivariate moderation model, respectively. We considered both the univariate model and the bivariate model, and fitted the ACE-AxE-AxC model to it to see whether the moderation effects are detected and how the gene by environment correlation influences the results.

6.4.1 Design univariate moderation

We simulated data according to an ACE-model, in which the A component was moderated by an external variable, M, i.e., (omitting subject and twin subscript)

$$Y = \mathrm{m} \times M + (\mathrm{a}_0 + \mathrm{a}_1 M) \times A + \mathrm{c} \times C + \mathrm{e} \times E \qquad (16)$$

where M is the (mean-centered) moderator, i.e., a measure of the environment, a_0 is the baseline parameter, a_1 is the moderation parameter, and parameter *m* takes into account the main effect of *M* (which is advisable when modeling interactions, see Nelder, 1994). If a_1 departs from 0, A is moderated by M, which amounts to an AxE interaction. In the present simulation study we choose: $\mathrm{a}_0 = \mathrm{c} = \mathrm{e} = 1$. In addition, we choose the main effect of the moderator to be to be either small (m = 0.5), medium (m = 0.75), or large (m=1.0). Note that the main effect of the moderator is the same across the MZ and DZ twins (i.e., a C moderator). In addition, we chose the degree of moderation, to be small ($\mathrm{a}_1 = 0.5$), medium ($\mathrm{a}_1 = 0.75$), or large ($\mathrm{a}_1 = 1$). Finally, we manipulated the within twin correlation of M to be either 0, 0.5, 0.7, or 1.0). As we are not interested in

the exact power of the ACE-AxE-AxC model to detect the effects, effect sizes do not necessary reflect realistic effect sizes. The main aim of this simulation study is to see whether the moderation effects are detected by the ACE-AxE-AxC model. Note that we simulated the data using the observed moderator variable, but in fitting the ACE-AxE-AxC model, we do not use this variable.

6.4.2 Results univariate moderation

Table 6.3 shows the power to detect AxE in the presence of AxC and the power to detect AxC in the presence of AxE. Given these results, we note that when the within twin correlation of the moderator is 0 or 0.5, power to detect AxE is generally large, while the power to detect AxC is small. This indicates that the moderation effect in the data is generally detected as AxE. When the correlation increases to 0.7 or 1.0, power to detect AxE is small, and power to detect AxC is large, i.e., in this case the moderation effect in the data is generally detected as AxC. These results hold irrespective of the size of the main effect of the moderator. Power of the Purcell model equaled 1 in nearly all simulated scenarios (not tabulated). Power of the Purcell model is thus larger than the power in the ACE-AxE-AxC model, but this is not surprising as this approach uses the information available in the moderator variable.

6.4.3 Design bivariate moderation

As noted in Purcell (2002) the moderator could share genetic influences with the phenotypic variable, we denote these common influences, A_c. Purcell proposes the following model for the mean-centered M and Y:

$$Y = (a_0 + a_1 \times M) \times A_c + c_c \times C_c + e_c \times E_c + a_u \times A_u + c_u \times C_u + e_u \times E_u \quad (17)$$

$$M = a_m \times A_c + c_m \times C_c + e_m \times E_m \quad (18)$$

i.e., the phenotypic variance is decomposed into A_c, C_c, and E_c components which are shared with the moderator variable, and into A_u, C_u, and E_u components which are unique to the phenotypic variable. Note that the model could be extended to introduce moderation of the C_c and E_c. When only the A_u component is moderated, the univariate moderation model from Equation 16 will suffice.

Table 6.3.

Power to detect AxE in the presence of AxC, and power to detect AxC in the presence of AxE when data is simulated under Purcell's univariate moderation model.

Cor. within Twins	Main effect mod	Effect GxE	ACE-AxE-AxC	
			AxE	*AxC*
0	**small**	**small**	0.92	0.24
		medium	0.79	0.39
		large	0.72	0.76
	medium	**small**	0.99	0.21
		medium	0.99	0.21
		large	0.98	0.05
	large	**small**	1.00	0.19
		medium	1.00	0.23
		large	1.00	0.22
0.5	**small**	**small**	0.51	0.22
		medium	0.37	0.57
		large	0.49	0.92
	medium	**small**	0.72	0.14
		medium	0.60	0.40
		large	0.52	0.46
	large	**small**	0.85	0.15
		medium	0.77	0.25
		large	0.76	0.38
0.7	**small**	**small**	0.29	0.31
		medium	0.26	0.73
		large	0.44	0.94
	medium	**small**	0.33	0.33
		medium	0.29	0.74
		large	0.29	0.78
	large	**small**	0.44	0.44
		medium	0.31	0.72
		large	0.29	0.86
1	**small**	**small**	0.22	0.30
		medium	0.54	0.75
		large	0.75	0.93
	medium	**small**	0.18	0.66
		medium	0.47	0.95
		large	0.49	0.98
	large	**small**	0.15	0.94
		medium	0.37	1.00
		large	0.46	1.00

Note. 'cor within twins' refers to the correlation between the twin 1 and twin 2 scores on the moderator variable.

We simulated data according to the bivariate moderation model. We manipulated the effect size of the GxE effect into no effect ($a_1 = 0$), a small effect ($a_1 = 0.5$), a medium effect ($a_1 = 0.75$), and a large effect ($a_1 = 1.0$). In addition, we manipulated the size of the genotype by environment correlation, into 0.3 (i.e., $a_m = 0.5$), 0.4 ($a_m = 0.75$) and 0.5 ($a_m = 1$). We simulated an 'E moderator', that is, besides the effects of A, the moderator was influenced by E but not by C ($c_m = 0$, $e_m = 1$). The other parameters equaled $c_c = e_c = c_u = a_0 = a_u = e_u = 1$. We note again that the chosen effect sizes are not necessarily realistic as we are only interested in how the 'Purcell' effects are detected in the ACE-AxE-AxC model.

6.4.4 Results bivariate model

Table 6.4 shows the power of the ACE-AxE-AxC model to detect AxE and AxC effects in the data under the different scenarios. We see that the moderation effect in the data is mainly detected as AxE (i.e., power of AxE effect is large, power of AxC effect is small). This was to be expected as the moderator was not influenced by C. In addition, we see that in case of no moderation in the data, no GxE is detected (i.e, power approaches 0.05 in all these cases). Thus, the genotype by environment correlation does not appear to cause spurious interactions.

Table 6.4

Power to detect AxE in the presence of AxC, and power to detect AxC in the presence of AxE when data is simulated under Purcell's bivariate moderation model.

rGE	**GxE effect**	**ACE-AxE-AxC**	
	power to detect:	*AxE*	*AxC*
.3	none	0.05	0.05
	small	0.84	0.05
	medium	0.87	0.05
	large	0.91	0.05
.4	none	0.05	0.05
	small	0.66	0.05
	medium	0.70	0.05
	large	0.84	0.05
.5	none	0.06	0.06
	small	0.51	0.20
	medium	0.53	0.45
	large	0.95	0.65

Note. rGE denotes the size of the genotype by environment correlation due to shared genes between the moderator and the phenotypic variable.

6.4.5 Conclusion and discussion

This second simulation study showed two important results. First, a correlation between phenotype and environment due to shared genes does not affect the results concerning tests on GxE in the ACE-AxE-AxC model. Second, interactions between observed measures of the environment and the additive genetic factor, A, can in principle be detected using the ACE-AxE-AxC model. Depending on the within twin correlation of the moderator, the interaction will arise as an AxE or AxC. Of course power is an issue here, as small effects will possibly remain undetected. However, given a sufficiently large sample size, phenotypic variables can be screened on GxE when no explicit hypotheses exist on which measures of the environment will interact with genetic influences of the phenotype, or when the relevant environment measures are not available (e.g., an IQ datasets which lacks a measure of SES).

6.4.6 Application

We applied the univariate GxE model to the Osborne data (Osborne, 1980), which comprise scores of 477 twin pairs on various tests of cognitive ability. We analyzed the scores of the twin pairs on the first-principal component of 13 cognitive ability tests from the Osborn data. We found the ACE-AxE model to provide the best model fit, indicating that an AxE interaction is present in these data. We do not present the detailed results in this Chapter to save space, and because we apply the multivariate model to these data below. However, a small report of this application is available from www.dylanmolenaar.nl.

6.5 The Multivariate Case

In this section, we introduce a multivariate approach in which we distinguish between a measurement model and a biometrical model (the common pathway model). In the biometrical part of the model, we introduce the AxC and AxE effects, and in the measurement model we introduce heteroscedastic residuals to account for possible heteroscedastic measurement error, and/or floor, ceiling, and poor scaling effects. In addition, we show how one can test for non-linear factor loadings within the multivariate approach. We outline the multivariate approach below.

Let $\mathbf{y_1}$ denote the N x p-dimensional matrix of the scores of the N twin 1 members on p phenotypic scores, and let $\mathbf{y_2}$ denote the scores of the twin 2 members. These scores are submitted to a k dimensional factor model which is referred to as the measurement model. In the measurement model, the observed variables are linked to a

(set of) phenotypic construct(s). Specifically, the covariance matrix $\mathbf{\Sigma}_{y1,y2}$ of the horizontally stacked matrices $\mathbf{y}_1,\mathbf{y}_2$ is modeled as

$$\mathbf{\Sigma}_{y1,y2} = \mathbf{\Lambda}\,\mathbf{\Sigma}_{\eta}\,\mathbf{\Lambda} + \mathbf{\Sigma}_{\theta} \tag{19}$$

where $\mathbf{\Lambda}$ are the factor loadings, $\mathbf{\Sigma}_{\eta}$ is the covariance matrix of the phenotypic constructs, and $\mathbf{\Sigma}_{\theta}$ is the covariance matrix of the residuals. The structure of the factor loading matrix, $\mathbf{\Lambda}$, may be derived from theory, such as the *general intelligence* theory by Spearman (1904), or the Big Five personality theory (Digman, 1990). In principle, $\mathbf{\Lambda}$ can be submitted to a Cholesky decomposition to test for general and specific genetic and environmental contributions, however then, the measurement model is not separated from the biometric model anymore. Here, we focus on a theory based factor model, but we return to the Cholesky decomposition in the discussion.

As an illustration, we consider general intelligence or *g* (Spearman, 1904). According to *g* theory, a single phenotypic latent construct underlies all scores of a given intelligence test. That is, in both the twin 1 and twin 2 samples, we postulate 1 common factor. Given 4 observed cognitive variables, we have the following factor loading matrix:

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ \lambda_1 & 0 \\ \lambda_2 & 0 \\ \lambda_3 & 0 \\ 0 & 1 \\ 0 & \lambda_1 \\ 0 & \lambda_2 \\ 0 & \lambda_3 \end{bmatrix}. \tag{20}$$

where the factor loadings of the first variables of each twin are fixed to 1 for identification purposes.

In the biometric model, the 2 x 2 covariance matrix of the phenotypic constructs, $\mathbf{\Sigma}_{\eta}$, is decomposed as follows

$$\mathbf{\Sigma}_{\eta} = \mathbf{\Sigma}_{A} + \mathbf{\Sigma}_{C} + \mathbf{\Sigma}_{E} \tag{21}$$

i.e., the covariance matrix of the general intelligence factor underlying the twin 1 and twin 2 subtest data is modeled as a function of the A, C, and E factors.

To model AxC and AxE interactions, we can apply the univariate method from Equation 7 and 8 to the matrices $\mathbf{\Sigma_C}$, and $\mathbf{\Sigma_E}$, i.e.,

$$\mathbf{\Sigma_E} \mid A_1, A_2 = \begin{bmatrix} \exp(\beta_0 + \beta_1 A_1) & \\ 0 & \exp(\beta_0 + \beta_1 A_2) \end{bmatrix} \tag{22}$$

and

$$\mathbf{\Sigma_C} \mid A_1, A_2 = \begin{bmatrix} \exp(\gamma_0 + \gamma_1 A_1) & \\ \sqrt{\exp(\gamma_0 + \gamma_1 A_1)\exp(\gamma_0 + \gamma_1 A_2)} & \exp(\gamma_0 + \gamma_1 A_2) \end{bmatrix}. \tag{23}$$

where '$|A_1,A_2$' means that the corresponding covariance matrix is conditional on both A_1 and A_2. The term on the off-diagonal of $\Sigma_C|A_1,A_2$ ensures that the correlation between factor C_1 and factor C_2 remains equal to 1. For the general intelligence factor, we thus have two heteroscedasticity parameters, β_1 and γ_1 for the AxE and AxC interaction, respectively. Note that when there are multiple factors (e.g., in applications to Big Five personality data), each factor is associated with it's own β_1 and γ_1 parameters.

Now, in the measurement model, we introduce heteroscedastic residual variances in Σ_θ to account for heteroscedasticity that is specific to the observed phenotypic variables and not due to heteroscedasticity of E or C on the level of the latent phenotypic construct, thus:

$$\mathbf{\Sigma}_\theta \mid A_1, A_2 = \begin{bmatrix} \exp(\delta_{01} + \delta_{11} A_1) & & & & & & \\ 0 & \ddots & & & & & \\ 0 & 0 & \exp(\delta_{04} + \delta_{14} A_1) & & & & \\ \sigma_{\theta 1} \mid A_1, A_2 & 0 & 0 & \exp(\delta_{01} + \delta_{11} A_2) & & & \\ 0 & \ddots & 0 & 0 & \ddots & & \\ 0 & 0 & \sigma_{\theta 4} \mid A_1, A_2 & 0 & 0 & \exp(\delta_{04} + \delta_{11} A_2) \end{bmatrix}. \tag{24}$$

In this equation, δ_{01} is the baseline parameter for phenotypic variable 1, δ_{04} is the baseline parameter for phenotypic variable 4, δ_{11} is the heteroscedasticity parameter for phenotypic variable 1, etc. In addition, $\sigma_{\theta 1}|A_1,A_2$ is the conditional residual covariance between the scores of twin 1 and twin 2 on phenotypic variable 1, and $\sigma_{\theta 4}|A_1,A_2$ is the conditional residual covariance between the scores of twin 1 and twin 2 on phenotypic variable 4. These conditional covariances account for possible genetic and environment influences on the level of the residuals. These covariances could in principle be submitted to an ACE-decomposition, including AxE and/or AxC effects on the level of

the individual variable. This would enable a test on whether GxE occurs at the level of the phenotypic construct or at the level of the individual variable. However, these GxE tests on the level of the variable are vulnerable to problems like poor scaling. For present purposes (testing GxE on the level of the phenotypic construct to avoid problems like poor scaling) we do not distinguish between ACE-components on the level of the variable. Instead, we account for similarities between twins of the same twin pair by conditional covariances between the residuals as introduced in Equation 24. The conditional covariances are calculated as follows, e.g., for variable 1,

$$\sigma_{\theta 1} \mid A_1, A_2 = \frac{\rho_1}{\sqrt{\exp(\delta_{01} + \delta_{11} A_1)\exp(\delta_{01} + \delta_{11} A_2)}}, \tag{25}$$

where ρ_1 is the residual correlation between the twin 1 and twin 2 scores on variable 1 after the phenotypic construct is taken into account. Note that this correlation is constant across A_1 and A_2. Thus, to conclude, in the measurement model 15 parameters are estimated: λ_1,λ_2, λ_3, and δ_{01} to δ_{04}, δ_{11} to δ_{14}, and ρ_1 to ρ_4.

In the model above, we introduced heteroscedasticity in the biometric model to model AxE and AxC and we introduced heteroscedasticity in the measurement model to model heteroscedastic residuals. As the GxE effects are modeled on the factor that is common to all phenotypic variables (i.e., the phenotypic construct), the AxE and AxC effects capture the heteroscedasticity that is common to all variables of the construct. Variable specific heteroscedasticity (i.e., not shared among all variables) is captured by the heteroscedastic residuals. In doing so, confounds specific to the variables - like poor scaling - are absorbed by the heteroscedastic residuals. The GxE effects that arise on the level of the construct can therefore be more confidently interpreted as such. However, as Eaves (2006) pointed out, the same artifacts of scale could be present in all variables in a GxE study. In the present approach, this may give rise to spurious GxE on the level of the construct.

6.5.1 Testing for spurious GxE due to non-linearity

The measurement model in Equation 19 is based on the premise that the observed phenotypic scores are linearly predicted from the latent phenotypic construct. Tucker-Drob et al. (2009) showed that when the relation between the observed phenotypic variables and the latent phenotypic construct is non-linear, this can result in spurious GxE. To exclude possible spurious GxE we can test the factor loadings on non-linearity. Note that we test for non-linearity in the measurement model, but still retain the ACE decomposition in the biometric model. Testing for non-linearity of the factor

loadings is straightforward in Mx (Neale et al., 2006; see Molenaar et al., 2010 for an Mx example) and Mplus (Muthén & Muthén, 2007; see Tucker-Drob, 2009 for an Mplus example).

6.6 Application

6.6.1 Data

We analyzed the Osborne data (Osborne, 1980), which include the scores of 328 Caucasian twin pairs and 149 Afro-American twin pairs on various tests of cognitive abilities. As sample size within both groups is insufficient, we analyzed both groups together for illustrational purposes. The 477 twin pairs included 247 MZ twins (110 males, 137 females), and 230 DZ twins, of which 180 were same sex twins (65 male-male, 115 female-female) and 50 were opposite sex twins. Mean age was 15.30 (sd: 1.55; min: 12; max: 20).

From the Osborne data, we selected 4 subtests, the Mazes Test, Object Apeture Test, Simple Arithmetic Test, and New Castle Spatial Test that fitted a one-factor model well. To the scores of the twin 1 and twin 2 samples on these subtests, we fitted a one-factor model representing the general intelligence factor. The variance of this latent phenotypic factor was decomposed into an A, C, and E component, with AxE and AxC interactions, as in Equations 22 and 23. In the full sample (i.e., MZs and DZs together), the scores were standardized to have variances equal to 4 to facilitate parameter estimation. See Table 6.5 for the correlation matrices in the MZ and DZ samples. The baseline model without AxE and AxC interaction fitted adequately compared to the saturated model [$\chi^2(66)= 51.57$]. In this model, the phenotypic factors correlated .76 (SE = 0.04) between the members of the DZ twins, and .95 (SE = 0.01) between the members of the MZ twins.

Table 6.5.
MZ (below the diagonal) and DZ twin correlations for the twin 1 and twin 2 samples

	MT1	**OA1**	**AR1**	**NS1**	**MT2**	**OA2**	**AR2**	**NS2**
MT1	1	0.38	0.38	0.50	0.42	0.31	0.3	0.39
OA1	0.34	1	0.49	0.72	0.26	0.47	0.4	0.57
AR1	0.24	0.46	1	0.65	0.28	0.34	0.63	0.46
NS1	0.36	0.7	0.56	1	0.37	0.49	0.46	0.67
MT2	0.59	0.43	0.33	0.45	1	0.43	0.35	0.46
OA2	0.28	0.66	0.48	0.68	0.44	1	0.49	0.72
AR2	0.21	0.45	0.85	0.57	0.32	0.52	1	0.54
NS2	0.33	0.67	0.52	0.86	0.51	0.73	0.57	1

Note. CT: Calendar Test; CC: Cube Comparison Test ; WV: Wide Range Vocabulary Test; SD: Surface Development Test; AR: Simple Arithmetic Test; FB: Form Board; SV: Self-Judging Vocabulary Test; PF: Paper Folding Test; OA: Object Aperture Test; IP: Identical Pictures Test; NS: Newcastle Spatial Test; SA: Spelling Achievement Test; MT: Mazes Test. The number after the abbreviation denotes the twin member. (e.g., MT1 refers to the scores of the twin 1 sample on the Mazes Test; MT2 refers to the scores of his/her cotwin on this test).

6.6.2 Results

First, we tested the factor loadings in the measurement model for non-linearity. We did this using Mplus (Muthén & Muthén, 2007). Parameter estimates and model fit statistics are in Table 6.6. According to the AIC, BIC, and LRT, the model with non-linear factor loadings fitted best. However, only subtest OA is associated with a non-linear factor loading. As the effect concerns only a single variable, we continue our analysis assuming linearity for all variables for illustrational purposes. However, we stress that in practice one should be cautious drawing conclusion on GxE in the presence of unmodeled non-linearity. We return to this point in the discussion.

Table 6.6
Parameter estimates of the non-linear multivariate ACE model.

Parameter	Variable	Model	
		Quadratic λ	*Linear λ*
λ_{lin}	MT	1.00	1.00
	OA	1.57 (0.12)	1.56 (0.12)
	AR	1.25 (0.11)	1.25 (0.11)
	NS	1.83 (0.14)	1.82 (0.14)
λ_{quad}	MT	0.01 (0.07)	-
	OA	0.15 (0.05)	-
	AR	0.06 (0.07)	-
	NS	0.18 (0.05)	-
σ_A^2	*g*	0.40 (0.09)	0.39 (0.09)
σ_C^2	*g*	0.56 (0.12)	0.56 (0.12)
σ_E^2	*g*	0.05 (0.02)	0.05 (0.02)
		Model fit statistics	
	$\chi^2(4)$	-	15.33
	AIC	14395.86	14402.42
	BIC	14508.38	14498.27

Note. MT: Mazes Test; OA: Object Aperture Test; NS: Newcastle Spatial Test; AR: Simple Arithmetic Test. λ_{lin} is the baseline factor loading which models the linear relation between the phenotypic construct and the observed variables. λ_{quad} is the non-linearity parameter which accounts for the non-linearity in the relation between the phenotypic construct and the observed variable. $\chi^2(4)$ is a Satorra-Bentler corrected likelihood ratio test between the model with non-linear and linear factor loadings.

Table 6.7.
Model fit statistics for the different models in the multivariate illustration

Model	Fit indeces		
	AIC	*BIC*	*LRT*
1: ACE-AxE-AxC-het	6224.81	-4649.59	-
2: ACE-AxE-het	6224.32	-4651.93	2 vs 1: $\chi^2(1) = 1.50$
3: ACE-het	6231.55	-4650.40	3 vs 2: $\chi^2(1) = 9.23$
4: ACE-AxE	6222.475	-4661.181	4 vs 2: $\chi^2(4) = 6.158$

Note. For the AIC and BIC, best values are underlined. 'AxE-AxE-AxC-het' is the ACE-AxE-AxC model with heteroscedastic residual variances. The LRT concerns a likelihood ratio test between the models mentioned.

In Table 6.7, the results of the multivariate analyses are summarized. We started with the full model, the ACE-AxE-AxC-het, where 'het' denotes that heteroscedastic residuals are present (δ_{11} to δ_{14} are estimated). In this model, the AxE and AxC effects are on the level of the general intelligence factor. From the model we dropped the AxC interaction. All model fit indices indicated that the model fit improved, indicating that an AxC interaction was absent [$\chi^2(1) = 1.50$]. Next, we dropped the AxE interaction from the model (resulting in an ACE-het model). All fit statistics indicated that the model fit deteriorated [$\chi^2(1) = 9.23$]. We thus concluded that the ACE-AxE-het model was a better fitting model. Parameter estimates of this model are in Table 6.8. As can be seen, the heteroscedasticity parameters of the residuals (δ_{11} to δ_{14}) did not differ significantly form 0, as judged by their confidence intervals. We therefore dropped these parameters, resulting in an ACE-AxE model. According to a likelihood ratio test, this model fitted better than a model with heteroscedastic residuals [$\chi^2(4) = 6.158$], this was confirmed by the AIC and BIC (see Table 6.7). Parameter estimates of the ACE-AxE, are in Table 6.8. It appears that dropping the heteroscedastic residuals (parameter δ_{11} to δ_{14}) hardly affected the AxE parameter, β_1. The estimate of β_1 changed from 1.40 to 1.38. As the estimate of β_1 was larger than zero, the variance of factor E increases with increasing levels of factor A. Thus, for increasing genetic levels (i.e., for an increasing position on the additive genetic factor, A), differences between twins in phenotypes are larger because differences in environments increase. Note that this is consistent with the notion of ability differentiation in which the general intelligence factor is hypothesized to be a weaker source of individual differences at higher levels of this factor (Deary, et al. 1996). This is similar to what we found in the univariate application where we used PC1 scores (as described shortly above). However, the advantage of the multivariate approach is that it enables us to show that the AxE effect involves the common phenotypic factor and is not due to heteroscedastic residuals.

6.7 Conclusion

In this Chapter we identified four challenges to the detection of GxE using the existing univariate heteroscedastic approaches of Jinks & Fulker (1970) and van der Sluis et al (2006); non-normality, conflation of AxE and CxE, heteroscedastic measurement error, and gene by environment correlation. We presented an extension of the heteroscedasticity approach meant to overcome these problems. Specifically, we presented a univariate method suitable to study the presence of AxC and AxE interactions using both MZ and DZ twin data. In this approach, we explicitly distinguished between the A and C component so as to avoid the conflation of A and C. We showed that AxE and AxC interactions are well separable, but it turned out that

AxE analyses are still influenced by the presence of CxE. One might argue that this problem could be solved by constructing a model that incorporates both AxE and CxE interaction simultaneously, so that the effects can be disentangled. We considered such a model, in which the variance of E was modeled as a function of both A and C. (Note that this simultaneous modeling of AxE and CxE requires an extension of the ACE-model that is not covered by the equations in the present Chapter.) Simulations demonstrated that, although the extended model could be specified and fit without problems, AxE and CxE could not be distinguished. Specifically, when the simulated effect, e.g., AxE, was dropped, the likelihood hardly changed because the effect was almost fully absorbed by the CxE effect. Details about this extended model and the simulations are in the Appendix.

The difficulty of distinguishing AxE and CxE is related to the well known problem that A and C are less well resolvable compared to A and E, or C and E (Martin et al., 1978). The simulations that we presented show that the presence of CxE will bias tests of AxE, depending on the strength of C as a source of individual differences. For some phenotypic measures, it is known that the strength of C is negligibly small, specifically in cognitive abilities from adolescence onwards (see Boomsma, et al., 2002). In these cases, AxE interactions may arguably be interpreted as such. In cases that C is substantial (i.e., situations comparable to scenario II and III from the simulations), one should be more careful in interpreting a significant AxE interaction, as the effect could indicate the presence of CxE rather than AxE. In such cases, it seems wise to interpret AxE as the interaction between familiarity factors and environmental factors, as in the analysis of MZ twin data only (as in Jinks & Fulker, 1970, and van der Sluis et al., 2006). That is, one leaves unresolved the exact dimension across which the strength of the environmental factor increase, i.e., A or C. A possible solution proposed by Jinks & Fulker (1970) is to consider twin data that includes MZ twins who are reared apart. In theory this improves the distinction of A and C. However, in practice such data are scarce. Nevertheless, the model could be useful as an explorative tool to screen phenotypic variables on GxE when no ideas exist (yet) on what measures to include in a Purcell (2002) type of analysis.

Extending the univariate approach of van der Sluis et al. (2006) to include DZ twins did not solve the conflation of AxE with CxE. However, this does not disqualify our new model as an approach of testing GxE. We think that the new method has some clear advantages over existing approaches. First, in our new method we can distinguish between AxE and AxC (although large samples or large effect sizes are needed to detect AxC). Second, because of the increased sample size due to the addition of the DZ twin data, power to detect AxE is increased as compared to the van der Sluis et al and Jinks and Fulker model. Third, in both the simulation and application we showed

that taking into account AxC interaction which is possible due to the DZ twin data, may be beneficial in terms of the power to detect the AxE effect.

Appendix: Distinction between CxE and AxE within an ACE-AxE-CxE model

In this appendix we show that within an ACE-AxE-CxE model the AxE and CxE effects are empirically unidentified. The results below are obtained under circumstances similar to scenario I from the Chapter. Table 9 depicts the power of an ACE-AxE-CxE model to detect CxE and AxE.

Table 9.

Power to detect AxE and AxC in an ACE-AxE-CxE model

		Model fitted			
effect in data	**effect size**	**ACE-AxE-CxE**		**ACE-AxE-**	**ACE- CxE**
	power to detect:	*CxE*	*AxE*	*AxE*	*CxE*
none		0.05	0.05	0.05	0.05
AxE	*small*	0.09	0.08	0.81	0.82
	medium	0.10	0.09	0.93	0.94
	large	0.16	0.14	0.99	0.99
CxE	*small*	0.08	0.05	0.44	0.47
	medium	0.15	0.08	0.57	0.62
	large	0.15	0.07	0.75	0.79

The Table shows that when an AxE is in the data, the effect is not detected as AxE and not as CxE within the ACE-AxE-CxE model. This is because the AxE effect can arise as both CxE and as AxE in the model. Thus, consider the case that AxE is in the data, and the ACE-AxE-CxE is fitted. Dropping the AxE effect from the model shows no deterioration in model fit (while the AxE effect *is* in the data) as the AxE effect is fully absorbed in the CxE parameter. When the AxE parameter is freed, and the CxE parameter is dropped, model fit again shows no deterioration as the AxE effect is now be absorbed by the AxE parameter. Same holds for the case when a CxE effect is in the data.

When the ACE-AxE and the ACE-CxE models are considered, power is always large irrespective of the exact effect that is in the data. For instance, power of ACE-AxE is large when an AxE is in the data *and* when an AxC is in the data. From the above we conclude that the AxE parameter can not be distinguished from the CxE parameter (the parameters are highly correlated) under reasonable circumstances (i.e., the chosen effect and sample sizes).