

Specification, Testing, and Interpretation of Gene-by-Measured-Environment Interaction Models in the Presence of Gene–Environment Correlation

Paul J. Rathouz · Carol A. Van Hulle ·
Joseph Lee Rodgers · Irwin D. Waldman ·
Benjamin B. Lahey

Received: 17 April 2007 / Accepted: 22 January 2008 / Published online: 22 February 2008
© Springer Science+Business Media, LLC 2008

Abstract Purcell (Twin Res 5:554–571, 2002) proposed a bivariate biometric model for testing and quantifying the interaction between latent genetic influences and measured environments in the presence of gene–environment correlation. Purcell’s model extends the Cholesky model to include gene–environment interaction. We examine a number of closely related alternative models that do not involve gene–environment interaction but which may fit the data as well as Purcell’s model. Because failure to consider these alternatives could lead to spurious detection of gene–environment interaction, we propose alternative models for testing gene–environment interaction in the presence of gene–environment correlation, including one based on the correlated factors model. In addition, we note mathematical errors in the calculation of effect size via variance components in Purcell’s model. We propose a statistical method for deriving and interpreting variance decompositions that are true to the fitted model.

Keywords Biometric model · Cholesky · Identifiability · GxM · rGE · Variance components · Variance decomposition

Edited by Stacey Cherny.

P. J. Rathouz (✉) · C. A. Van Hulle · B. B. Lahey
Department of Health Studies, University of Chicago,
5841 S. Maryland Ave., MC 2007, Chicago, IL 60637, USA
e-mail: prathouz@uchicago.edu

J. L. Rodgers
Department of Psychology, University of Oklahoma, Norman,
OK, USA

I. D. Waldman
Department of Psychology, Emory University, Atlanta,
GA, USA

Introduction

Gene-by-environment interaction (GxE) and gene–environment correlation (rGE) have long been of concern in the field of behavior genetics (BG). Jinks and Fulker (1970), Eaves et al. (1977), and Plomin et al. (1977) describe the mechanisms by which rGE can arise, the effects of GxE and rGE in terms of bias and power in biometric modeling of data from twin and adoption designs, statistical approaches to testing GxE, and the role of measurement scale in the manifestation of GxE. In this foundational work, both genetic and environmental factors involved in GxE and rGE are unobserved.

Other work has considered designs in which both a measured aspect of the environment and the phenotype of interest are observed, but in which genetic factors are unobserved. Such designs provide an opportunity to test and model the interaction of *latent* genetic effects and *measured* environmental variables as a causal influence on the phenotype of interest. This is of great importance to the field of developmental psychopathology (Moffitt 2005; Rutter 2006) as current causal models do not adequately describe the gene–environment interactions that influence psychopathology (Lahey and Waldman 2003). The motivation behind gene-by-measured environment (GxM) models is the possibility that some genes increase a person’s sensitivity to variation in environmental factors impacting the phenotype, or alternatively, that genetic variation impacting the phenotype is greater under certain environmental conditions than under others (Eaves et al. 2003).

Purcell (2002), in a seminal and already often-cited article, set out a modeling framework on which others have relied for testing and quantifying such GxM effects. In the first two of several models described by Purcell, the

measured environment, or putative *moderator*¹ (M) is both a regressor for the phenotype (P) and also interacts with additive genetic, and shared and nonshared environmental effects to explain phenotypic variance in the residual of P after regressing out the main effects of M . That is, these models test GxM in P residualized for M . These models have been widely applied (Jang et al. 2005; Kendler et al. 2004; Kremen et al. 2005; Rice et al. 2006).

As Purcell (2002, p. 563) notes, his first two models are limited because they give a partial picture of the genetic and environmental variance in P . Purcell explains that these models could fail to detect GxM when it exists, and could under- or over-estimate heritability. This is because the genetic and environmental effects are being modeled on the residual from the regression of P on M . Therefore, any genetic or environmental effects that operate through—or are in common with— M are partialled out. Genetic effects on P that are common to M give rise to gene-measured environment correlation (rGM). The presence of rGM can therefore bias estimation of heritability and mask GxM.

These limitations motivated a new approach to modeling GxM while simultaneously allowing for rGM (Purcell 2002, p. 563). Purcell's new approach is based on a joint biometric model for the moderator and the phenotype which extends the classic Cholesky parameterization of the bivariate biometric model to include interaction terms between the moderator and genetic and/or environmental effects. The model, described below under "Purcell's GxM with rGM model" yields a potentially powerful and widely applicable approach to an important problem in behavior genetics, namely untangling the structure of gene-by-environment interplay. The range of possible fitted models has been described by Johnson (2007). Purcell's model has been used to test and quantify GxM in the possible presence of rGM in several recent papers (Burt et al. 2006; Johnson and Krueger 2005a, b, c, 2006; Timberlake et al. 2006). These papers have examined the relation of phenotypes such as Body Mass Index (BMI), physical health, and conduct disorder to moderators such as income, perceived life control, and age of menarche. It should be noted that, although Purcell's model is important in its own right, it can also be seen as the simplest of a broad and flexible class of new approaches involving two or more observed variables in joint biometric structural models, a prominent example of which was recently proposed by Eaves et al. (2003).

Despite its potential power, Purcell's new model requires care in its use for testing and quantifying GxM and

rGM. Whereas this was clearly noted by Purcell (2002), the purpose of the present article is to consider additional potential pitfalls of specification, identifiability, testing and interpretation of the parameters in this and related models for GxM that account for rGM. We consider equivalent and closely related alternatives to Purcell's model in our section on "Sub-models and equivalent models". These models are instances (sub-models) of a more general model containing multiplicative latent genetic and latent environmental effects. We describe some models that do not contain GxM but that may explain the data as well as Purcell's GxM model. In the section on "Correlated factors model for GxM with rGM", we propose an alternative model for testing and quantifying GxM in the presence of rGM based on the correlated factors model (Loehlin 1996), which we believe is simpler and avoids several pitfalls with the Cholesky model.

Purcell (2002) suggested computing separate genetic and environmental components of variance of P as functions of M while allowing for rGM. We show under "Misrepresentation of variance components in Purcell's model" that this idea is mathematically incorrect and we demonstrate via example under "Hypothetical examples" that it may yield incorrect conclusions about the true importance of GxM effects in certain contexts. The central problem lies in the difficulty of expressing the model for P in terms of uncorrelated effects and in quantifying the size of those effects in terms of components of variance. This derivation is more complicated in the presence of nonlinear terms such as those required in models for GxM than in linear models such as the classical ACE biometric model. In "A method for variance decomposition" we describe derivation of correct variance decompositions in the context of any structural model and apply this approach to obtain statistically meaningful decompositions in GxM with rGM models. We illustrate the ideas in the section entitled "Hypothetical examples".

Purcell's GxM with rGM model

Purcell's (2002) model for testing and quantifying GxM with rGM treats the measured putative moderating environmental variable and the response variable jointly in a bivariate model. This is an extension of the classical Cholesky parameterization of the bivariate ACE model (Loehlin 1996). In what follows, we restate the Purcell model in simplified form, leaving out the shared environmental component C for clarity of exposition. The key ideas advanced in the present article do not depend on including or excluding C , but generalize to the full bivariate ACE model. We denote the "environmental" variable by M as a

¹ The candidate moderator need not be an "environmental" variable in the narrow sense of the word. It could represent an environmental variable such as parenting practices or neighborhood crime, or it could also represent another phenotype being modeled as a precursor to the phenotype of interest.

mnemonic both for *measured* environment and *moderator* in the GxM part of the model. The response variable is P , for *phenotype*. For example, M could be family income and P could be a measure of antisocial behavior.

The standard biometric model for M is given by

$$M = \mu_M + a_M A_M + e_M E_M, \quad (1)$$

where A_M and E_M are each standard (mean 0, variance 1) normal latent random variables, uncorrelated with one another, μ_M is the mean of M , and, without loss of generality, $a_M \geq 0$ and $e_M \geq 0$. Here, A_M and E_M represent additive genetic and nonshared environmental influences on M , so that a_M^2 is the additive genetic variance component of the variance of M , e_M^2 is the nonshared environmental variance component of the variance of M , and the total variance of M is $\text{var}(M) = a_M^2 + e_M^2$.

Purcell's model for phenotype P which allows for rGM and GxM specifies

$$P = \mu_P + (a_C + \alpha_C M)A_M + (e_C + \varepsilon_C M)E_M + (a_U + \alpha_U M)A_U + (e_U + \varepsilon_U M)E_U, \quad (2)$$

where A_U and E_U are again standard normal random variables independent of each other and of (A_M, E_M) . Here, a_C and α_C represent additive genetic effects on P that are common with those on M (subscript C denoting *common*), e_C and ε_C represent nonshared environmental effects on P that are common with those on M . Corresponding quantities a_U , α_U , e_U , and ε_U represent effects that are *unique* to P , i.e., unrelated to M . In (2), the Greek coefficients α_C , ε_C , α_U , and ε_U capture the interaction of moderator M with the various genetic and environmental factors that act on P . In particular, the magnitude of α_C (α_U) captures the GxM interaction of M with common (unique) genetic factors A_M (A_U) in determining P . Models (1) and (2) together specify Purcell's bivariate biometric GxM with rGM model (2002, pp. 563–566). Figure 1a contains a path diagram describing model (1–2), and Table 1 summarizes structural components and parameters figuring in the model.

Model (2) does not explicitly contain a “main effect” of M on P . Nonetheless, (2) does capture main effects of M on P indirectly through effects a_C and e_C on common genetic and environmental influences A_M and E_M . Additionally, the model contains as a sub-model a regression with direct effects of M on P . To see this, suppose that

$$\begin{aligned} (a_C/a_M) &= (e_C/e_M) \equiv \beta_1 \quad \text{and} \\ (\alpha_C/a_M) &= (\varepsilon_C/e_M) \equiv \beta_2. \end{aligned} \quad (3)$$

Then, replacing a_C , e_C , α_C , and ε_C with $\beta_1 a_M$, $\beta_1 e_M$, $\beta_2 a_M$, and $\beta_2 e_M$, respectively, (2) can be rewritten as

$$P = \mu_P + \beta_1 M + \beta_2 M^2 + (a_U + \alpha_U M)A_U + (e_U + \varepsilon_U M)E_U \quad (4)$$

(Fig. 1b). That is, in this sub-model the common factors A_M and E_M on P only operate through the manifest value M to influence P .

Before introducing model (1–2), Purcell (2002, pp. 555–560) presents model (4) with $\beta_2 = 0$ for situations where one is interested in modeling the genetic and environmental contributions to P “controlling for” or “partialing out” M . A limitation of (4) is that genetic and environmental influences on P that are mediated through M are masked, potentially distorting the overall gene–environment picture. This issue directly motivated his GxM with rGM model given in (1) and (2).

Model (1–2) parameterizes the notions of rGM and GxM. The idea of rGM is that there are genetic influences on P that also impact the individual's measured environment M . This is captured jointly by parameters (a_C, α_C) and can be examined by testing whether these quantities deviate significantly from zero. The notion of GxM is that genetic effects on P are moderated by environment M . This is captured jointly by parameters (α_C, α_U) and the presence of GxM can therefore be similarly tested in the context of model (2). The parameter α_U is easier to interpret than α_C : When $\alpha_U \neq 0$, this indicates that the genes A_U that influence P but *do not* influence M are more (or less) potent in the causation of P for larger values of M than they are for smaller values of M . This is one clear form of GxM. Interpretation of parameter α_C is a bit more subtle: $\alpha_C \neq 0$ indicates that the genes A_M that influence both M and P are more (or less) potent in the causation of P for larger values of M . However, as will be seen in “Sub-models and equivalent models” the interpretation of α_C should be considered jointly with ε_C .

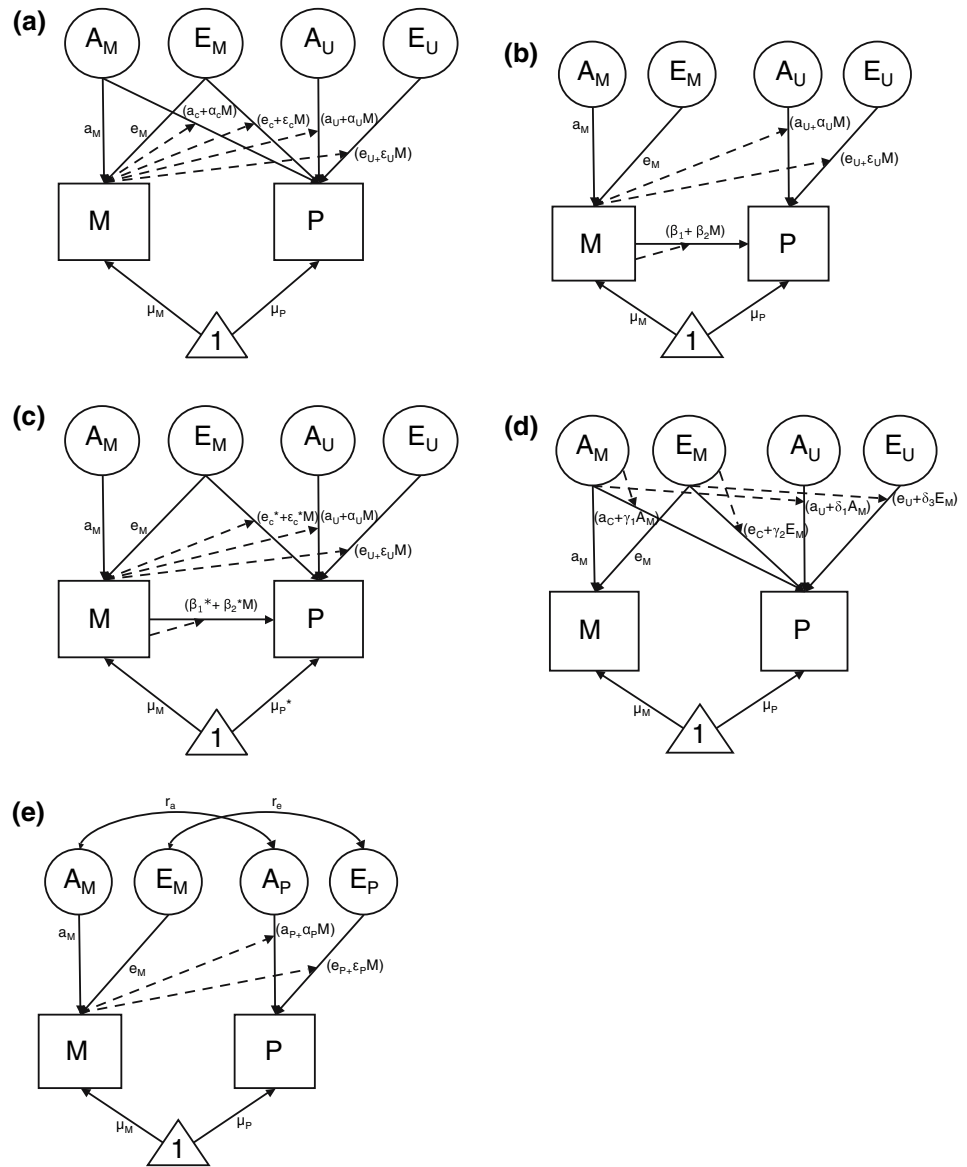
Specification and testing of GxM with rGM models

Sub-models and equivalent models

Here, we explore sub-models of, and alternative models equivalent to, Purcell's model (1–2) that do not include GxM in the sense meant by Purcell. We also show that (1–2) is a restriction of a more general model containing multiplicative terms of genetic and environmental factors. The restrictions imposed yield GxM, but there are other restrictions that do not involve GxM that are equally parsimonious and scientifically plausible.²

² Models for the simulated examples were fitted and tested in Mplus v.4.21 (Muthén and Muthén 2006). Scripts and output files are available for download from the first author's web site at <http://www.health.bsd.uchicago.edu/rathouz/GxM>.

Fig. 1 Path diagrams for several models which may or may not contain GxM. **(a)** Purcell’s (2002) model for testing and quantifying GxM in the presence of rGM. **(b)** Main effects model containing no A_M -by- M or E_M -by- M terms. **(c)** Alternative model containing main effects of M and M^2 instead of A_M -by- M terms. **(d)** Alternative model containing no multiplicative terms in M . **(e)** Correlated factors model for testing and quantifying GxM in the presence of rGM. *Note:* All latent variables have mean 0 and variance 1. Dashed arrows pointing from a variable to a path indicate moderation of the indicated path by that variable



Main effects model

We showed in “Purcell’s GxM with rGM model” that (4) is a sub-model of (2). A concern, therefore, is that a non-linear effect of M on P , parameterized by β_2 in (4), could lead to an artifactual detection of common genetic (A_M)-by-measured environment (M) interaction—i.e., $\alpha_C \neq 0$ —if the nonlinear effect β_2 is not specifically tested. Because (4) imposes constraints (3) on (2), this concern can be alleviated by testing (2) vs. (4) via a likelihood ratio test (LRT). Rejection of the null hypothesis (4) in favor of the alternative (2) would provide support for the interaction of A_M and M .

Simulated Example 1a. We simulated data on 1000 monozygotic (MZ) and 1000 dizygotic (DZ) twins under models (1) and (4), setting $\mu_M = \mu_P = 0.00$, $a_M = e_M$

$= 1.00$, $\beta_1 = 0.30$, $\beta_2 = 0.20$, $a_U = \sqrt{1.00 - 0.30^2}$, $e_U = 1.00$, and $\alpha_U = \epsilon_U = 0.00$. We fitted model (1–2) to these data and obtained estimates for GxM of $\hat{\alpha}_C = 0.18$ and $\hat{\alpha}_U = 0.01$. The LRT statistic for the GxM null hypothesis that $\alpha_C = \alpha_U = 0.00$ was $\chi^2 = 29.9$ which is highly significant on 2 degrees of freedom (df). Fitting instead model (4), we obtain $\hat{\beta}_1 = 0.30$, $\hat{\beta}_2 = 0.20$, and $\hat{\alpha}_U = 0.01$. The LRT for model (2) vs. (4) was $\chi^2 = 0.11$ on 2 df. If in (4) we test the GxM null hypothesis that $\alpha_U = 0.00$, we obtain $\chi^2 = 0.20$ on 1 df.

Simulated Example 1b. We simulated data as in Example 1a, but set $\beta_2 = 0.00$ and $\alpha_U = 0.20$. Fitting model (1–2), we obtained GxM estimates of $\hat{\alpha}_C = 0.00$ and $\hat{\alpha}_U = 0.21$, with $\chi^2 = 88.0$ on 2 df for testing $\alpha_C = \alpha_U = 0.00$. Model (4) yielded $\hat{\beta}_1 = 0.30$, $\hat{\beta}_2 = 0.00$ and $\hat{\alpha}_U = 0.21$. The LRT for model (2) vs. (4) was $\chi^2 = 0.07$ on 2 df. Testing the

Table 1 Components and parameters in structural GxM with rGM models

Model components	
A_M	Additive genetic influences on M (standardized)
E_M	Nonshared environmental effects on M (standardized)
A_U	Additive genetic influences on P independent of A_M (standardized)
E_U	Nonshared environmental influences on P independent of E_M (standardized)
A_P	Additive genetic influences on P , possibly correlated with A_M (standardized)
E_P	Nonshared environmental influences on P , possibly correlated with E_M (standardized)
Model parameters (“main” effects)	
a_M	Effect of A_M on M
e_M	Effect of E_M on M
β_1/β_1^*	Effect of M on P
β_2/β_2^*	Effect of M^2 on P
a_C	“Main” effect of A_M on P , i.e., additive genetic effects on P in common with M
e_C/e_C^*	“Main” effect of E_M on P , i.e., nonshared environmental effects on P in common with M
a_U	“Main” effect of A_U on P , i.e., additive genetic effects on P unique to P
e_U	“Main” effect of E_U on P , i.e., nonshared environmental effects on P that are unique to P
a_P	Effect of A_P on P
e_P	Effect of E_P on P
Model parameters (interaction effects)	
α_C	Interaction effect on P between M and A_M , i.e., “common-genetic-by-environment” interaction
ϵ_C/ϵ_C^*	Interaction effect on P between M and E_M
α_U	Interaction effect on P between M and A_U , i.e., “unique-genetic-by-environment” interaction
ϵ_U	Interaction effect on P between M and E_M
α_P	Interaction effect on P between M and A_P , i.e., “additive-genetic-by-environment” interaction
ϵ_P	Interaction effect on P between M and E_P
γ_j	Multiplicative effects on P of A_M and E_M , $j = 1,2,3$
δ_j	Multiplicative effects on P of A_M/E_M with A_U/E_U , $j = 1,2,3,4$
λ_j	Multiplicative effects on P of A_M/E_M with A_P/E_P , $j = 1,2,3,4$

Notes: The terms “common” and “unique” refer to overlap (or lack thereof) in causal influences on M and P . The term “nonshared” refers to the fact that the effect is not shared across family members. The term “additive” refers to the fact that genetic effects are highly polygenic, allelic variation at each locus presumably making a small independent contribution to the overall variance

GxM null hypothesis that $\alpha_U = 0.00$ in (4), we obtain $\chi^2 = 87.9$ on 1 df.

Simulated Example 1c. We then simulated data on 1000 monozygotic (MZ) and 1000 dizygotic (DZ) twins under model (1–2), instead of under (4). We set $\mu_M = \mu_P = 0.00$, $a_M = e_M = 1.00$, $a_C = e_C = 0.30$, $\alpha_C = 0.02$, $\epsilon_C = 0.00$, $a_U = \sqrt{1.00 - 0.30^2}$, $e_U = 1.00$, and $\alpha_U = \epsilon_U = 0.00$. Fitting model (1–2), we obtained GxM estimates of $\hat{\alpha}_C = 0.20$ and $\hat{\alpha}_U = 0.02$, with $\chi^2 = 33.1$ on 2 df for testing $\alpha_C = \alpha_U = 0.00$. Model (4) yielded $\hat{\beta}_1 = 0.31$, $\hat{\beta}_2 = 0.09$ and $\hat{\alpha}_U = 0.01$. The LRT for model (2) vs. (4) was $\chi^2 = 13.0$ on 2 df. Testing the GxM null hypothesis that $\alpha_U = 0.00$ in (4), we obtain $\chi^2 = 0.27$ on 1 df.

These three examples illustrate that (i) if model (4) is the true data generating mechanism for M and P , use of model (2) could lead to artifactual detection of GxM (Example 1a); (ii) it is possible to test (4) versus the more general (2) to provide support for true GxM over

nonlinear effects of M on P (Example 1c); and (iii) GxM can also be detected in model (4) by focusing on parameter α_U (Example 1b).

In summary, model (2) distinguishes between moderation of A by M in the unique versus the common part of P . By contrast, in model (4), only GxM in the unique part of P is modeled, and any GxM in the common part is partialled out through β_1 and β_2 . While use of (4) risks masking GxM, use of (2) risks mistaking nonlinear effects of M for GxM. It therefore seems important to consider both complementary models when testing GxM.

Equivalent models without A_M -by- M

A second concern with detecting A_M -by- M interaction is that model (1–2) is *equivalent* to models containing no such effects. An equivalent model is an alternative model that is “indistinguishable from the original model in terms

of goodness of fit to the data;” often these alternative models may offer substantively different interpretations of the data (MacCallum et al. 1993). To see the equivalence, note that from (1) we can write $A_M = (M - \mu_M - e_M E_M)/a_M$. Inserting this expression into (2) and re-arranging terms yields

$$P = \mu_P + (a_C + \alpha_C M)(M - \mu_M)/a_M + \{(e_C + \varepsilon_C M) - (e_M/a_M)(a_C + \alpha_C M)\}E_M + (a_U + \alpha_U M)A_U + (e_U + \varepsilon_U M)E_U,$$

i.e., (2) can be re-expressed as a model with a main and a quadratic effect of M and a nonshared environment (E_M)-by-measured environment (M) interaction, with no A_M -by- M interaction. That is, additive and multiplicative effects of M , A_M and E_M are not separately identifiable. Specifically, (2) can be rewritten as

$$P = \mu_P^* + \beta_1^* M + \beta_2^* M^2 + (e_C^* + \varepsilon_C^* M)E_M + (a_U + \alpha_U M)A_U + (e_U + \varepsilon_U M)E_U \tag{5}$$

(Fig. 1c), where parameters $(\mu_P^*, \beta_1^*, \beta_2^*, e_C^*, \varepsilon_C^*)$ in (5) replace parameters $(\mu_P, a_C, \alpha_C, e_C, \varepsilon_C)$ in (2). A one-to-one relationship between these two sets of parameters exists and is given in an online Appendix.³ One key relation is that $\alpha_C = a_M \beta_2^*$, so that when the true data generating mechanism is (5), and $a_M > 0$ and $\beta_2^* \neq 0$, the analyst using (2) would incorrectly find that $\alpha_C \neq 0$ and claim to have detected an A_M -by- M interaction. This equivalence does not preclude testing for unique genetic (A_U)-by-measured environment (M) interactions. Nonetheless, in the absence of a large A_U -by- M term, it is impossible to unambiguously detect GxM because it cannot be distinguished from nonlinear main effects of M and E_M -by- M interaction.

Simulated Example 2. We again simulated data on 1000 MZ and 1000 DZ twins under models (1) and (5), setting $\mu_M = \mu_P = 0.00$, $a_M = e_M = 1.00$, $\beta_1^* = 0.30 = -e_C^*$, $\beta_2^* = 0.20 = -\varepsilon_C^*$, $a_U = \sqrt{1.00 - 0.30^2}$, $e_U = 1.00$, and $\alpha_U = \varepsilon_U = 0.00$. We fitted models (1) and (5) to these data and obtained the following estimates demonstrating that the fitted model is reflecting the true data generating mechanism: $\hat{\beta}_1^* = 0.31$, $\hat{\beta}_2^* = 0.21$, $\hat{e}_C^* = -0.31$, $\hat{\varepsilon}_C^* = -0.22$, $\hat{a}_U = 0.98$, $\hat{\alpha}_U = 0.01$, $\hat{e}_U = 1.00$, and $\hat{\varepsilon}_U = 0.00$. Fitting model (1–2) to these data, we obtained estimates of $\hat{\alpha}_C = 0.21$ and $\hat{\alpha}_U = 0.01$, with a LRT statistic for the 2 df GxM null hypothesis that $\alpha_C = \alpha_U = 0.00$ of $\chi^2 = 39.6$. If we test the 1 df null hypothesis that $\alpha_U = 0.00$, we obtain a LRT statistic of $\chi^2 = 0.36$. Similarly, if instead we had fitted model (4) and tested $\alpha_U = 0.00$, we would have obtained

$\chi^2 = 2.52$ on 1 df. This example illustrates how, (i) if model (5) is the true data generating mechanism, then fitting model (2) could lead to artifactual detection of GxM, and (ii) testing just $\alpha_U = 0.00$ in either model (2) or (4) yields a more robust test of GxM than tests including parameter α_C .

Expanded structural model with GxM

We now examine a more general structural model for P as a function of additive and multiplicative effects of A_M , E_M , A_U , and E_U . In this model, M is only indirectly represented by A_M and E_M , but the model contains (2) as a special case. For ease of exposition, we assume without loss of generality that $\mu_M = 0.00$.

Beginning with (2) and replacing M with $a_M A_M + e_M E_M$, we obtain a model containing all multiplicative terms of A_M , E_M , A_U , E_U , except those composed only of a product of unique effects A_U and E_U . This more general model is given by

$$P = \mu_P + a_C A_M + e_C E_M + \gamma_1 A_M^2 + \gamma_2 E_M^2 + \gamma_3 A_M E_M + a_U A_U + \delta_1 A_M A_U + \delta_2 A_U E_M + e_U E_U + \delta_3 E_M E_U + \delta_4 A_M E_U. \tag{6}$$

Here, the three γ parameters represent multiplicative or quadratic effects of common genes and common environment, and the four δ parameters represent multiplicative effects between common genes or environment and unique genes or environment. Model (2) is recovered through constraints on (6). Specifically, (2) holds if

$$\gamma_3 = \frac{e_M}{a_M} \gamma_1 + \frac{a_M}{e_M} \gamma_2, \quad \delta_2 = \frac{e_M}{a_M} \delta_1, \quad \text{and} \quad \delta_4 = \frac{a_M}{e_M} \delta_3, \tag{7}$$

in which case

$$\alpha_C = \gamma_1/a_M, \quad \varepsilon_C = \gamma_2/e_M, \quad \alpha_U = \delta_2/e_M, \quad \text{and} \quad \varepsilon_U = \delta_4/a_M.$$

Model (6) with constraints (7) conveys several important messages about (2). First, (2) is a special case of a model which contains nonlinear effects of A_M , E_M , A_U , and E_U . Second, the fit of (2) to the data can be tested by comparing it via LRT to (6) with only a 3-df test corresponding to the three constraints in (7). Third, the fact that model (2) imposes constraints (7) on (6) means that it is conceivable that in an analysis using (2), values of (α_C, α_U) significantly different than zero could arise spuriously as an artifact of effects of one or more of A_M^2 , E_M^2 , $A_M A_U$, or $E_M E_U$, none of which are gene-by-environment interactions. This is because constraints (7) link effects of A_M^2 , E_M^2 , $A_M A_U$, and $E_M E_U$ with effects of $A_M E_M$, $A_U E_M$, and $A_M E_U$.

³ See <http://www.health.bsd.uchicago.edu/rathouz/GxM/>.

This last point suggest directly testing (6) for GxM by testing whether

$$\gamma_3 = \delta_2 = \delta_4 = 0, \quad (8)$$

versus the alternative that one of them is different than zero. A significant result would suggest GxM. By contrast, if (8) holds, the resulting model is given by

$$P = \mu_P + a_C A_M + e_C E_M + \gamma_1 A_M^2 + \gamma_2 E_M^2 + a_U A_U + \delta_1 A_M A_U + e_U E_U + \delta_3 E_M E_U \quad (9)$$

(Fig. 1d) which, because of (7) and (8), contains no GxM interactions in the sense of Purcell. Note that model (9) requires the same number of parameters as (2); in many settings, it may be equally parsimonious, easy to interpret and scientifically plausible as (2).⁴

Simulated Examples 3a and 3b. We simulated data on 1000 MZ and 1000 DZ twins under models (1) and (9). For Example 3a, we set $\mu_M = \mu_P = 0.00$, $a_M = e_M = 1.00$, $a_C = 0.30$, $a_U = \sqrt{1.00 - 0.30^2}$, $e_C = 0.00$, $e_U = 1.00$, $\gamma_1 = 0.20$, $\delta_1 = 0.00$, and $\gamma_2 = \delta_3 = 0.00$. Fitting model (1–2) to these data, we obtained estimates for GxM of $\hat{\alpha}_C = 0.16$ and $\hat{\alpha}_U = 0.01$ with a LRT statistic for the 2 df GxM null hypothesis that $\alpha_C = \alpha_U = 0.00$ of $\chi^2 = 26.9$. For Example 3b, we simulated data as in Example 3a, but set $\gamma_1 = 0.00$ and $\delta_1 = 0.20$. We obtained estimates of $\hat{\alpha}_C = 0.00$ and $\hat{\alpha}_U = 0.21$ with a LRT statistic of $\chi^2 = 85.4$. These examples illustrate that, if the true data generating mechanism for P is model (9), this could lead to artifactual detection of GxM in model (2).

Summary: identifiability and misspecification

The existence of models (4), (5), and (9) point out potential problems of identifiability and misspecification in using Purcell's model (2) for testing GxM. Identifiability is a problem in many structural equations models, and is not specific to Purcell's model. Our concern here is that these models are equally as or more parsimonious than (2), yet do not contain GxM in the sense of Purcell. For example, (9) and the nonunique parts of (4) and (5) are additive in the effects of genetic and environmental factors, even while allowing those effects to operate nonlinearly. Such nonlinear effects could easily arise if the scale on which genes contained in A_M operate on P is different than the scale on which they operate on M . Those models do not

contain gene–environment interaction terms, but may equally well explain the observed associations between M and P .

Correlated factors model for GxM with rGM

Here, we question the Cholesky parameterization of the bivariate biometric model as a baseline for testing GxM, propose the correlated factors model as an alternative baseline model, and show how testing GxM is simpler in this model.

Cholesky parameterization

As noted above, model (1–2) extends the Cholesky parameterization of the classical bivariate AE model to allow for GxM interaction. As discussed by Loehlin (1996), the Cholesky is one of several equivalent bivariate biometric models, and, although the models are mathematically equivalent, the interpretation of the parameters is not. The Cholesky parameterization is most interpretable when there is a clear theoretical, causal, and/or temporal ordering of the variables M and P in the model (Johnson 2007). In the absence of such ordering, it is incorrect to interpret A_M as representing the shared genetic influences on M , as A_M almost certainly contains genetic factors that are unique to M (Loehlin 1996). For example, Johnson and Krueger (2005a) analyze data on income and BMI in a representative sample of adult twins. Their objective was to test whether “genetic variance associated with physical health decreases with increasing levels of income” (p. 581). They provide a clear theoretical argument for treating income as the moderator M and measured BMI as the phenotype P of interest in their analysis. Nonetheless, income and BMI were measured concurrently in their sample and, as pointed out by Johnson (2007), the validity of their findings hinges on the validity of the ordering of these variables in the causal model. In situations without a clear causal ordering, Loehlin suggests using either a “common factor” or a “correlated factors” model. These models, which treat M and P on equal footing, might be simpler and more defensible base models for analyses of GxM. In the classical ACE model with no GxM moderation, the results of a Cholesky analysis can be transformed to yield an equivalent common factor analysis (Loehlin 1996). This equivalence breaks down, however, when moderation by M is incorporated into the model.

Correlated factors model with GxM

Let A_P and E_P be the genetic and environmental influences on P . In a classical correlated factors (CF) model (Loehlin

⁴ Currently available structural equations modeling software such as Mplus (Muthén and Muthén 2006) or Mx (Neale et al. 2003) does not permit estimation of models (6) or (9) owing to nonlinearity of the A and E terms and difficulties of high-dimensional numerical integration. Development of specialized software to fit these models is an area for future development.

1996), A_P is correlated with A_M and E_P is correlated with E_M . This formulation suggests that an alternative to (2) be generated by extending the CF model to allow for GxM, viz,

$$\left. \begin{aligned} P &= \mu_P + (a_P + \alpha_P M)A_P + (e_P + \varepsilon_P M)E_P \\ r_a &= \text{corr}(A_M, A_P) \text{ and } r_e = \text{corr}(E_M, E_P). \end{aligned} \right\} \quad (10)$$

Figure 1e presents a path diagram of this model. Here, A_P and E_P are standard normal random variables independent of each other, a_P and α_P represent additive genetic effects on P , e_P , and ε_P represent nonshared environmental effects on P , and r_a and r_e represent the correlation in genetic and in environmental influences between M and P . When $\alpha_P = \varepsilon_P = 0.00$, (10) is the correlated factors parameterization of the bivariate biometric model. Additionally, (10) is a special case of model (2). To see this, suppose in (2) that

$$\left. \begin{aligned} (\alpha_C/a_C) &= (\alpha_U/a_U) \equiv \gamma_a \quad \text{and} \\ (\varepsilon_C/e_C) &= (\varepsilon_U/e_U) \equiv \gamma_e, \end{aligned} \right\} \quad (11)$$

and define standard normal random variables

$$A_P = \frac{a_C A_M + a_U A_U}{\sqrt{a_C^2 + a_U^2}} \quad \text{and} \quad E_P = \frac{e_C E_M + e_U E_U}{\sqrt{e_C^2 + e_U^2}}.$$

A_P and E_P are just Z-scores of the sums $a_C A_M + a_U A_U$ and $e_C E_M + e_U E_U$. Replacing α_C , α_U , ε_C , ε_U with $\gamma_a a_C$, $\gamma_a a_U$, $\gamma_e e_C$, and $\gamma_e e_U$, respectively, (2) becomes

$$P = \mu_P + (1 + \gamma_a M)(a_C A_M + a_U A_U) + (1 + \gamma_e M)(e_C E_M + e_U E_U),$$

from which (10) is recovered by setting $a_P = \sqrt{a_C^2 + a_U^2}$, $\alpha_P = \gamma_a \sqrt{a_C^2 + a_U^2}$, $r_a = a_C / \sqrt{a_C^2 + a_U^2}$, $e_P = \sqrt{e_C^2 + e_U^2}$, $\varepsilon_P = \gamma_e \sqrt{e_C^2 + e_U^2}$, and $r_e = e_C / \sqrt{e_C^2 + e_U^2}$.

We offer (10) as a more straightforward model in which to test and quantify GxM than (2) when M and P do not have a clear causal ordering. Model (10) allows for multiplicative effects with M with two fewer parameters than (2), and GxM can be tested in (10) with a single parameter α_P , which could increase the power to detect GxM. Model (10) does not, however, permit differentiation of GxM effects into unique and common parts. Model (2) may therefore be more appropriate where a causal ordering is justified, as when M and P are two longitudinal measurements. It is also possible to examine empirical support for (2) as an alternative to (10) via a LRT, imposing constraints (11).

Simulated Example 4a. To investigate the performance of CF model (10), we simulated data on 1000 MZ and 1000 DZ twins under model (1–2), setting $\mu_M = \mu_P = 0.00$, $a_M = e_M = 1.00$, $a_C = 0.50$, $a_U = \sqrt{1.00 - 0.50^2}$, $e_C = 0.10$, $e_U = \sqrt{1.00 - 0.10^2}$, $\alpha_C = 0.05$, $\alpha_U = 0.15$, and $\varepsilon_C = \varepsilon_U = 0.00$. This specification comes close to satisfying

constraints (11) so that model (10) holds approximately. Fitting model (1–2) to these data, we obtain estimates for GxM of $\hat{\alpha}_C = 0.05$ and $\hat{\alpha}_U = 0.16$ with a LRT statistic of $\chi^2 = 59.4$ for the 2 df GxM null hypothesis that $\alpha_C = \alpha_U = 0.00$. Alternatively, we might start out by testing the 2 df null hypothesis that (10) holds versus the more general (2); this yields $\chi^2 = 4.50$, so we do not reject the CF model in favor of the Cholesky model. In (10), we obtain a GxM estimate of $\hat{\alpha}_P = 0.16$ with a 1 df LRT test statistic of $\chi^2 = 56.5$, which, given the lower df, is a stronger result for GxM than that obtained with model (2).

Simulated Example 4b. We again simulated data as in Example 4a, but set $\alpha_C = -0.05$ and $\alpha_U = 0.15$. This specification does not conform to CF model (10). Under model (1)–(2), we obtain $\hat{\alpha}_C = -0.05$ and $\hat{\alpha}_U = 0.16$ with LRT statistics of $\chi^2 = 42.8$ for $\alpha_C = \alpha_U = 0.00$ and $\chi^2 = 41.8$ for $\alpha_U = 0.00$. In (10), we obtain $\hat{\alpha}_P = 0.08$ with a 1 df LRT test statistic of $\chi^2 = 10.6$, a substantially weaker result for detecting GxM than those obtained with model (2). Testing model (10) versus the more general (2), however, yields $\chi^2 = 49.0$, so, it is unlikely we would have proceeded to test GxM under model (10); rather ultimate GxM detection would be based on the test that $\alpha_U = 0.00$.

Simulated Example 4c. We again simulated data as in Example 4b, but set $\alpha_C = -0.15$ and $\alpha_U = 0.05$. As in that example, this specification does not conform to CF model (10), but here, α_C (and not α_U) is the larger source of GxM. Under model (1–2), we obtain $\hat{\alpha}_C = -0.16$ and $\hat{\alpha}_U = 0.06$ with a LRT statistic of $\chi^2 = 24.7$ for $\alpha_C = \alpha_U = 0.00$. In (10), we obtain $\hat{\alpha}_P = -0.11$ with a 1 df LRT test statistic of $\chi^2 = 22.5$, an equally strong result for detecting GxM than that obtained with (2). Testing (10) vs. (2), yields $\chi^2 = 77.1$. Thus, whether one first tests (10) vs. (2) and, having selected (2) proceeds to test GxM in (2), or rather tests GxM directly in (10), GxM would be correctly detected.

These three examples suggest that, when CF model (10) holds approximately, a test of (10) versus the Cholesky model (2) will lead us to choose (10). Under (10) the test for GxM is more powerful than that in (2). Alternatively, even when (10) does not hold, GxM sometimes can still be successfully tested in (10) or, when (10) is rejected in favor of (2), in the more flexible model (2)

Variance decomposition in BG models

Misrepresentation of variance components in Purcell’s model

In order to quantify GxM in the context of (2), Purcell (2002, p. 564) and others (e.g., Johnson and Krueger

2005a; Johnson 2007) have described the total (unstandardized) variance in P due to genetic factors as given by

$$(a_C + \alpha_C M)^2 + (a_U + \alpha_U M)^2, \quad (12)$$

and, with parameter estimates for a_C , α_C , a_U , and α_U , have plotted this quantity as a function of M .⁵ Expression (12) presumably arises from the assumption that A_M and A_U are independent standard normal variables in model (2). This expression is incorrect, however, because it presumes that for fixed M both A_M and A_U are freely varying quantities. Whereas this is true for A_U , it is not for A_M ; when M is fixed the variance of A_M is conditional on that fixed value of M . To formalize this point, note that M and A_M are jointly bivariate normal random variables with $\text{var}(M) = a_M^2 + e_M^2$, $\text{var}(A_M) = 1$ and $\text{corr}(M, A_M) = a_M / \sqrt{a_M^2 + e_M^2}$. From standard bi-variate normal theory, given M ,

$$E(A_M|M) = \frac{a_M}{a_M^2 + e_M^2}(M - \mu_M) \neq 0 \quad \text{and}$$

$$\text{var}(A_M|M) = \frac{e_M^2}{a_M^2 + e_M^2} < 1,$$

which differ from the values $E(A_M) = 0$ and $\text{var}(A_M) = 1$ which hold marginally over M .

Additionally, statements such as “the variance of P due to additive genetic factors,” when attached to (12), carry an implicit presumption that genetic factors A_M are independent of environmental factors E_M . This is also incorrect; again due to conditioning on M , A_M , and E_M are not independent as they are marginally over (i.e., averaging over) M . Rather, once M is fixed, E_M is a *deterministic function* of A_M . One can see this by writing $E_M = (M - \mu_M - a_M A_M) / e_M$. Thus, conditional on M , A_M no longer represents independent additive genetic effects on P , but rather the trade-off between genetic (A_M) and environmental (E_M) contributions to a given level of M .

For these reasons, (12) does not represent total additive genetic variance in phenotype P as a function of moderator M and, therefore, does not properly characterize the extent of GxM. Note, however, that, because A_U and E_U are independent of M , $(a_U + \alpha_U M)^2$ and $(e_U + \varepsilon_U M)^2$ do still represent the additive genetic and environmental contributions to the variance of P given M that are unique to P , and there is no loss of coherence in interpreting these quantities conditionally on M . In particular, $(a_U + \alpha_U M)^2$ does indeed capture the GxM that arises from genetic influences unique to P (i.e., with no influence on M).

A similar problem arises in obtaining rGM from (2). Purcell (2002, p. 564) quantifies rGM by

$$r_A = \frac{a_C + \alpha_C M}{\{(a_C + \alpha_C M)^2 + (a_U + \alpha_U M)^2\}^{1/2}}, \quad (13)$$

i.e., the signed square-root of the proportion of total genetic variance in P that is common to M . Johnson (2007, p. 428) correctly points out that r_A in (13) is not so much rGM, which would be the correlation between M and the genetic effects on P , as it is the correlation between the genetic effects on P and the genetic effects on M . Nevertheless, whether one redefines rGM or renames r_A , the interpretation assigned to it is incorrect. Again, the error is that one is attempting to condition on M while at the same time allowing A_M to freely vary as if M were not fixed, which is not logical. Therefore, the expression $(a_C + \alpha_C M)^2$ appearing in (13) does not represent genetic variance in P that is common to M . Given M , not only is $(a_C + \alpha_C M)^2$ not the variance of $(a_C + \alpha_C M)A_M$, but it no longer represents purely genetic variance. Rather it is the variance in P due to the tradeoff between the genetic and environmental contributions to a given level of M .

A method for variance decomposition

If expressions (12) and (13) are not mathematically coherent, how then can one use the parameters of model (1–2) to quantify and summarize GxM and rGM? One straightforward answer to this question is to focus on the key structural parameters; specifically, (a_C, α_C) jointly capture rGM, while (α_C, α_U) capture GxM. However, there are strong and well-founded traditions in BG of interpreting models in terms of variance components and derivative quantities such as the proportion of common genetic variance between two traits. This approach has served BG research well because in traditional ACE and related models, the terms of interest are uncorrelated. Owing to the fact that model (2) contains multiplicative terms, however, it is not expressible in terms of uncorrelated effects on P due separately to factors A_M , E_M , A_U , and E_U , even though these quantities are mutually independent.

Here, we propose an approach that does allow statistically meaningful variance decompositions to be obtained in structural models such as (2) for phenotype P . Our proposal involves re-expressing the model as a sum of uncorrelated additive terms. These terms represent average and partial effects of genetic and environmental factors A_M , E_M , A_U , and E_U on P . Because the terms are uncorrelated, their variances are additive, and can therefore be used as measures of effect size of these factors. Unlike the ACE model, however, the decomposition of P into such terms is neither unique nor determined by the data.

⁵ This formula is not explicitly stated but is unambiguously implied by Purcell’s formula for rGE given below in (13).

The AE model revisited and expanded

Consider a general structural model for P as a function only of additive genetic effects A and independent environmental effects E .⁶ That is, P is a function of A and E , generally denoted $P \equiv g(A, E)$, which may depend on A and E in a possibly nonlinear way. The classical AE model is one such example, but is restricted to linear combinations of A and E .

When interest lies in isolating the effects of genes A on P , one might choose to write the model for P in one of two ways:

$$P = E(P|A) + \{P - E(P|A)\} \quad (14)$$

or, alternatively,

$$P = E(P|E) + \{P - E(P|E)\}. \quad (15)$$

Note importantly that in both of these representations, the two additive components on the right-hand side are *per force* uncorrelated, the implication of which is that

$$\text{var}(P) = \text{var}\{E(P|A)\} + \text{var}\{P - E(P|A)\},$$

i.e., the variances are additive. Therefore, representations (14) and (15) of the model for P yield valid decompositions of the variance of P .

The interpretation of decomposition (14) is as follows: The first term represents the *average effect* of A on P in the sense that it quantifies how the mean of P varies as a function of A ignoring (i.e., averaging over) E . It is what we would obtain if A were measured and we regressed P on A . The variance of this term is that part of the variance of P that is accounted for by the effect of A on P , i.e., that is explained by variability in A in the population, ignoring E .

The second term in (14) represents the *partial* or *residual effect* of E , *removing* or *controlling for* effects due to A in the sense that it quantifies how P varies with E for fixed values of A . The variance of this partial effect is that part of the variance in P that is accounted for by the partial effect given A of E on P , i.e., that is explained by variability in E holding A fixed.

Under an additive AE structural model for P , namely $P = aA + eE$, it can be shown that $E(P|A) = \{P - E(P|E)\} = aA$, and $E(P|E) = \{P - E(P|A)\} = eE$, owing to the facts that A and E are (i) independent and (ii) figure additively into the model for P . Therefore, $a^2 = \text{var}(aA)$ represents both the average and the partial contribution of additive genetic effects to the variance of P , with a similar interpretation for e^2 .

Suppose now for sake of illustration that the structural model for P is given by

$$P = aA + eE + bAE, \quad (16)$$

i.e., that independent factors A and E combine multiplicatively to yield the realized value of P . Using decomposition (14), we obtain

$$P = E(P|A) + \{P - E(P|A)\} = (aA) + (eE + bAE).$$

In this decomposition, the first term has $\text{var}(aA) = a^2$ and the second has

$$\text{var}(eE + bAE) = e^2 + b^2. \quad (17)$$

The first variance component represents the part of the variance of P that is due to additive genes alone, ignoring variability in environmental factors, whereas the second one is that part due to variability in environment factors controlling for additive genetic effects. We make two remarks about this interpretation.

First, if P had been subjected to decomposition (15) rather than (14), different values and different interpretations would have obtained for the variance components due to A and E . This is analogous to what occurs with R^2 and partial R^2 in the multiple linear regression model. In that model, if two predictors X_1 and X_2 are correlated, then the total R^2 can be expressed as the variance explained by X_1 , plus the variance explained by X_2 , controlling for X_1 . Reversing the order of the predictors yields a different interpretation.

Second, under decomposition (14), it is valid to view the variance of the partial effect of E controlling A both conditionally on—and marginally over— A . That is, given A , the conditional variance is

$$\text{var}(eE + bAE|A) = (e + bA)^2, \quad (18)$$

and this allows the analyst to interpret the relative contribution of E to P conditional on various levels of A . The unconditional (marginal) variance (17) of effects due to E controlling A is simply the average over A of (18). Expression (18) quantifies how the importance of the partial effect of E controlling A varies with levels of A .

Variance decomposition in Purcell's model

As just shown in the simple AE model, the construction of partial effects and corresponding variance quantities provides a statistically coherent and defensible way to decompose the total variance in P and assign it to different factors contributing to P . We now apply this technique to the more complex GxM with rGM model (1–2). We propose two alternative variance decompositions, although there are others that one could construct. The first is based on the analytic aim of Purcell's model, which is to evaluate interaction effects between genetic factors and a measured

⁶ Recall that we omit shared environmental effects C only for simplicity of exposition; these models can be expanded to include C .

moderator M . The second ignores this aim, instead presenting variance due to all genetic influences followed by that due to partial effects of environment, controlling for and as a function of genetic effects. In the presentation below, the formulae for functions g_j and h_j and their corresponding variances are not critical for grasping the key ideas; in applications, such functions would simply be computed and, possibly, plotted using estimated parameters in model (1–2). Note that, because (10) is a special case of (2), the proposed decompositions apply equally well to the correlated factors model (10) as they do to (2).

For the first decomposition, we write

$$\begin{aligned}
 P &= E(P|E_M, A_M) + \{E(P|A_U, E_M, A_M) - E(P|E_M, A_M)\} \\
 &+ \{E(P|E_U, E_M, A_M) - E(P|E_M, A_M)\} \\
 &= g_1(A_M, E_M) + g_2(A_U; E_M, A_M) + g_3(E_U; E_M, A_M),
 \end{aligned}
 \tag{19}$$

where we note that last two terms are valid because, given (A_M, E_M) , partial effects due to A_U and E_U are uncorrelated with each other. These terms can be written as:

$$\begin{aligned}
 g_1(A_M, E_M) &= \mu_P + a_C A_M + \alpha_C a_M A_M^2 + e_C E_M + \varepsilon_C e_M E_M^2 \\
 &+ (\alpha_C e_M + \varepsilon_C a_M) A_M E_M
 \end{aligned}
 \tag{20}$$

$$\begin{aligned}
 g_2(A_U; E_M, A_M) &= a_U A_U + \alpha_U a_M A_M A_U + \alpha_U e_M A_U E_M \\
 &= (a_U + \alpha_U M) A_U
 \end{aligned}
 \tag{21}$$

$$\begin{aligned}
 g_3(E_U; E_M, A_M) &= e_U E_U + \varepsilon_U e_M E_M E_U + \varepsilon_U a_M A_M E_U \\
 &= (e_U + \varepsilon_U M) E_U,
 \end{aligned}
 \tag{22}$$

from which one can check that the sum is equal to P and that each g_j is uncorrelated with the others.

The interpretation of these three terms is as follows: $g_1(A_M, E_M)$ is the (average) effect on P of genes and environment in common with M . Because M is a function of A_M and E_M , g_1 also automatically includes effects of M as well. We combine the effects of A_M and E_M here into a single effect g_1 because, as shown in “Sub-models and equivalent models,” effects of A_M , E_M , and M are not separately identifiable. Regarding the other two terms, $g_2(A_U; E_M, A_M)$ is the partial effect of unique genes, controlling for and as a function of common effects (A_M, E_M) and consequently M as well; finally, $g_3(E_U; E_M, A_M)$ is the partial effect of unique environmental factors, controlling for and as a function of common effects (A_M, E_M) and M . Table 2 contains variance expressions for the g_j 's, both conditionally on and marginally over M . These variances can be used as measures of effect size of (A_M, E_M) , of A_U and of E_U .

An alternative decomposition is given by

$$\begin{aligned}
 P &= E(P|A_M, A_U) + \{P - E(P|A_M, A_U)\} \\
 &= h_1(A_M, A_U) + h_2(E_M, E_U; A_M, A_U),
 \end{aligned}
 \tag{23}$$

where

$$\begin{aligned}
 h_1(A_M, A_U) &= \mu_P + a_C A_M + \alpha_C a_M A_M^2 + \varepsilon_C e_M \\
 &+ a_U A_U + \alpha_U a_M A_M A_U
 \end{aligned}
 \tag{24}$$

$$\begin{aligned}
 h_2(E_M, E_U; A_M, A_U) &= e_C E_M + \varepsilon_C e_M (E_M^2 - 1) \\
 &+ (\alpha_C e_M + \varepsilon_C a_M) A_M E_M \\
 &+ \alpha_U e_M A_U E_M + e_U E_U \\
 &+ \varepsilon_U e_M E_M E_U + \varepsilon_U a_M A_M E_U.
 \end{aligned}
 \tag{25}$$

Conditional and marginal variance expressions are given in Table 2. In (23), $h_1(A_M, A_U)$ is the average effect of all genetic factors on P , while $h_2(E_M, E_U|A_M, A_U)$ is the partial effect of all environmental factors on P , controlling for and as a function of genetic factors. Note that both A_M and A_U play a role in the conditional variance of h_2 . Note also that h_1 could be further decomposed into an effect due to A_M and an effect of A_U controlling for and as a function of A_M . Finally, decomposition (23) admits a complementary representation as average environmental effects plus residual effects due to genetics.

This latter variance decomposition provides a valid quantification of rGM as the correlation between the average genetic effect on M and the average genetic effect on P . That is, $r_A = \text{corr}\{E(M|A_M, A_U), E(P|A_M, A_U)\}$. As M and A_U are independent, r_A simplifies to $r_A = \text{corr}\{E(M|A_M), E(P|A_M, A_U)\}$, yielding

$$\begin{aligned}
 r_A &= \text{corr}\{a_M A_M, h_1(A_M, A_U)\} \\
 &= \frac{a_C}{\{a_C^2 + 2\alpha_C^2 a_M^2 + a_U^2 + \alpha_U^2 a_M^2\}^{1/2}},
 \end{aligned}
 \tag{26}$$

to replace (13). Note that (26) does not and should not depend on M because in order to compute it, we are required to average over M .

Hypothetical examples

We illustrate the variance decompositions in subsections “Misrepresentation of variance components in Purcell’s model” and “A method for variance decomposition” through two hypothetical examples.⁷ In the first, suppose that model (1–2) has been fitted to data and the following parameter estimates have been obtained: $a_M = e_M = 1.00$,

⁷ R (R Development Core Team 2005) scripts and output files for obtaining the results in this section are available for download from the first author’s web site at <http://www.health.bsd.uchicago.edu/rathouz/GxM/>.

Table 2 Conditional and marginal variance expressions for decomposition of GxM with rGM model according to common effects and unique effects (g_j 's) and to genetic and environmental effects (h_j 's)

Effect	Conditional variance	Marginal variance
$g_1(A_M, E_M)$	–	$a_C^2 + 2\alpha_C^2 a_M^2 + e_C^2 + 2\varepsilon_C^2 e_M^2 + (\alpha_C e_M + \varepsilon_C a_M)^2$
$g_2(A_U; E_M, A_M)$	$(a_U + \alpha_U M)^2$	$a_U^2 + \alpha_U^2 (a_M^2 + e_M^2)$
$g_3(E_U; E_M, A_M)$	$(e_U + \varepsilon_U M)^2$	$e_U^2 + \varepsilon_U^2 (e_M^2 + a_M^2)$
$h_1(A_M, A_U)$	–	$a_C^2 + 2\alpha_C^2 a_M^2 + a_U^2 + \alpha_U^2 a_M^2$
$h_2(E_M, E_U; A_M, A_U)$	$e_C^2 + 2\varepsilon_C^2 e_M^2 + (\alpha_C e_M + \varepsilon_C a_M)^2 A_M^2 + \alpha_U^2 e_M^2 A_U^2 + e_U^2 + \varepsilon_U^2 e_M^2 + \varepsilon_U^2 a_M^2 A_M^2$	$e_C^2 + 2\varepsilon_C^2 e_M^2 + (\alpha_C e_M + \varepsilon_C a_M)^2 + \alpha_U^2 e_M^2 + e_U^2 + \varepsilon_U^2 e_M^2 + \varepsilon_U^2 a_M^2$

$a_C = -0.40$, $a_U = \sqrt{1.00 - 0.40^2} = 0.92$, $\alpha_C = 0.05$, $\alpha_U = 0.15$, $e_C = 0.10$, $e_U = \sqrt{1.00 - 0.10^2} = 0.99$, and $\varepsilon_C = \varepsilon_U = 0.00$. This specification encodes a model wherein genes and environment contribute equally to M and approximately equally to P , the larger proportion of genetic contributions to P are unique to P (versus in common with M), and almost all of environmental contributions to P are unique to P .

One might summarize such a fitted model via one of several variance decompositions. One decomposition for these hypothetical data is presented in the left panel of Fig. 2. This decomposition has been used a number of times in the BG literature (e.g., Johnson and Krueger 2005a; Johnson 2007). It displays variance components due to genetic factors and nonshared environmental factors as functions of M , using (12) and an analogous formula for total variance due to nonshared environment. Using these formulae, at $M = 0.00$, each of genes and environment have an effect variance on P of 1.00, accounting for 50% each of the variance of P in this example. Effect variance of A_M at $M = 0.00$ is 0.16 (8.0% of the total), while that of E_M is 0.01 (0.5%). We would conclude from this hypothetical analysis that common genes and environmental factors account for a relatively small portion of the overall variance of P , while unique genes and environmental factors predominate. Furthermore, the expression of unique genes is strongly moderated by M (GxM), with lower M values suppressing the expression of those genes. It remains to be seen if these conclusions are justified, however, because, for the reasons given in “Misrepresentation of variance components in Purcell’s model” the plot and these effect variances are mathematically incorrect.

A mathematically correct summary of the fitted model for these hypothetical data is obtained via decomposition (19–22) with variance formulae in Table 2. Under that decomposition, A_M and E_M together yield an average effect variance of 0.18 (i.e., $\text{var}(g_1) = 0.18$, Table 2, column 3), accounting for 8.6% of the variance of P . The remaining two variance components are due to uncorrelated partial effects of unique genes A_U (g_2) and unique environment E_U (g_3), each controlling (A_M, E_M). These partial effects yield

marginal variance components of 0.89 and 0.99 (Table 2, column 3), accounting for 43 and 48% of the total variance, respectively. These last two effect variances can also be presented as functions of M (Table 2, column 2), and we have done so in the center panel of Fig. 2. Note that similar conclusions are reached in this hypothetical model whether the mathematically correct (middle panel) or incorrect (left panel) effect variances are used. This is because, in this example, the environmental and genetic effects on P are almost fully accounted for by unique rather than by common factors.

An alternative model summary is given by decomposition (23–25). Under that decomposition, A_M and A_U together yield an average genetic effect variance of 1.03, accounting for 50% of the variance of P , leaving 50% to environmental partial effect variance. Here, genes influence the sensitivity of P to E_M and E_U , so the partial effect variances for environmental factors can also be presented as a function of genetic factors (A_M, A_U); see the right panel of Fig. 2. The three curves are very similar because they depend only very weakly on A_M , and moderation of environmental effects E_M and E_U by A_U is also very weak, as evidenced by the nearly flat variance curve. This decomposition can also be used to obtain valid quantification of rGM according to (26); this correlation is -0.39 . The conclusions one would reach from this decomposition with regard to GxM are complementary to those based on the decomposition presented in the middle panel of Fig. 2, and equally valid based on the data alone. Whether the decompositions presented in the second or third panels, or whether other decompositions are more pertinent will ultimately depend on the research question motivating the analysis.

Use of the mathematically correct effect variances is more important under other circumstances. For example, consider a second hypothetical fitted model in which common genetic effects on P are more important than unique genetic effects. All parameter values are the same as in the first example except that, $a_C = \sqrt{1.00 - 0.40^2} = 0.92$, $a_U = 0.40$, $\alpha_C = 0.15$, and $\alpha_U = -0.05$. The incorrect variance decomposition (12) is exactly the same as that for

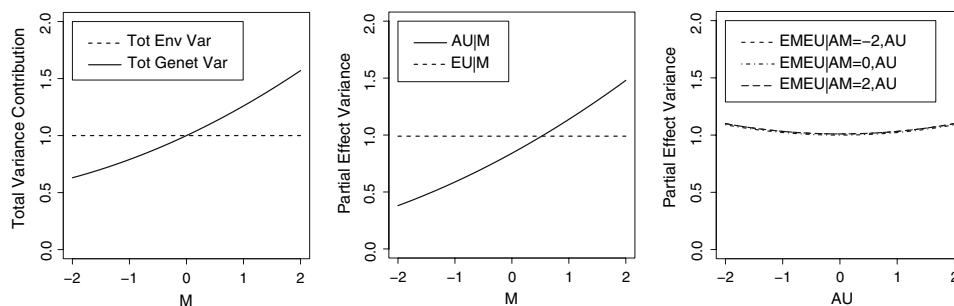


Fig. 2 Variance decompositions for first hypothetical gene–environment interaction model for P moderated by M , presented in “Hypothetical examples.” *Left*: Incorrect variance decomposition according to formula (12). *Center*: Correct variance decomposition for unique effects g_2 and g_3 controlling M . *Right*: Correct variance decomposition for environmental effects h_2 controlling genetic

effects. *Note*: For center panel, effect variance due jointly to A_M and E_M combined is 0.18 and is not shown in the plot because it is not a function of M . For right panel, effect variance due to all genetic effects A_M and A_U is 1.03 and is not shown in the plot because it is not a function of A_M or A_U

the first example and is presented in the left panel of Fig. 3. Under the variance decomposition generated by (19–22), A_M and E_M together have an average effect variance on P of 0.92, accounting for 44% of the variance. The remaining two variance components due to uncorrelated partial effects of unique genes A_U and unique environment E_U controlling (A_M , E_M) are 0.17 and 0.99, accounting for 8.0% and 48% of the total variance, respectively. Presented as a function of M , these last two components yield the center panel of Fig. 3. This figure shows a very weak effect of unique genes and a GxM effect in the opposite direction of that in the left panel. The very different results presented in the left panel are due to the fact that, in this hypothetical model, genetic effects on P are largely in common with M and are therefore inseparable from effects of M and E_M on P . The plot on the left incorrectly assumes that they are separable and yields potentially misleading results. In the third panel of Fig. 3 using (23–25), a similar picture for partial environmental effect variance is obtained as that in the first example because the latent genetic and environmental effects for P are similar in the two models, even though the sharing of those effects with M differs. This decomposition yields rGM of 0.90 according to (26).

Purcell’s variance decomposition (12) is never justified mathematically. The examples given above under “Hypothetical examples” show, however, that when the genetic components A_M common to M and P make relatively small contributions to P , and when GxM primarily arises between unique genetic components A_U and M , Purcell’s decomposition yields conclusions that are comparable to those of decomposition (19–22). One difference is that (12) attempts incorrectly to capture the total variance as a function of M , while (19–22) describes only the unique variance components as functions of M and presents the common components as constants with respect to M . When the unique components dominate the

common components, the two decompositions will yield very similar model interpretations. This was the case, for example, in the analysis presented by Johnson and Krueger (2005a) and in the primary analysis driving Johnson’s (2007) elegant paper based on Purcell’s model (Johnson 2007, Figs. 4 and 5).

By contrast, when GxM arises primarily between common genetic components A_M and M , Purcell’s decomposition (12) could be quite misleading. This is illustrated by the difference between the left and center panels of Fig. 3. This is a problem in a hypothetical model presented by Johnson (2007) based on Purcell’s decomposition. Figure 6 in Johnson (2007) yields inaccurate conclusions because it relies heavily on the incorrect expression (12) and on being able to separate multiplicative effects of M and E_M from those of M and A_M .

Choice of variance decomposition

In contrast to Purcell (2002), our proposal for obtaining statistically coherent variance decompositions in models such as (2) for phenotype P involves re-expressing the structural model as a sum of uncorrelated additive terms. In so doing, we have attempted to mimic a key feature of the ACE model, i.e., that it is specified as a sum of uncorrelated terms. Unlike the ACE model, however, the decomposition of P for a given fitted model is neither unique nor determined by the data. Many complementary decompositions exist and all are equally adequate representations of the data. Of these decompositions, which one should be used in a given analysis? There is no single answer to this question. An empirical option is to examine separately the total effects due to a variety of sources and to present the one with the largest variance first, as being the most important. Other sources of variance could then be presented as being due to partial effects.

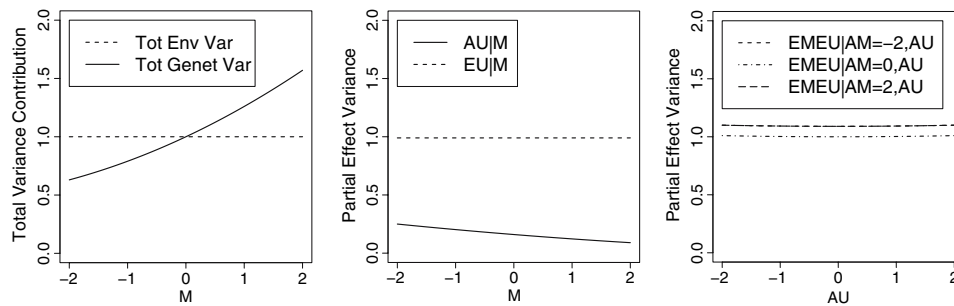


Fig. 3 Variance decompositions for second hypothetical gene–environment interaction model for P moderated by M , presented in “Hypothetical examples.” *Left*: Incorrect variance decomposition according to formula (12). *Center*: Correct variance decomposition for unique effects g_2 and g_3 controlling M . *Right*: Correct variance decomposition for environmental effects h_2 controlling genetic

effects. *Note*: For center panel, effect variance due jointly to A_M and E_M combined is 0.92 and is not shown in the plot because it is not a function of M . For right panel, effect variance due to all genetic effects A_M and A_U is 1.05 and is not shown in the plot because it is not a function of A_M or A_U

A second option is to choose decompositions based on substantive theory in the context of the research question driving the analysis. This latter approach is similar to the analyst choosing the order of predictors in a multiple regression model to reflect the scientific questions of interest, and then reporting as effect measures successive partial R^2 values for each regressor, controlling for all other regressors already entered in the model. The choice of decomposition depends on which factors should be controlled when examining each effect, and which factors are of theoretical interest as moderators of other effects. For example, decomposition (19–22) would be appropriate if unadjusted effects of moderator M and its components are of interest, and/or if M is of interest as an effect modifier of genetic and environmental factors. By contrast, decomposition (23–25) is relevant if total average genetic effects are of interest and/or if interest is on whether genetic factors moderate environmental effects.

Discussion

We revisited Purcell’s (2002) model for testing and quantifying GxM in the presence of rGM. We pointed out several problems with identifiability, specification, and testing in that model, showing how apparent GxM effects may arise spuriously due to alternative data generating mechanisms that are similar to but not the same as Purcell’s model. These misleading circumstances involve quadratic or multiplicative effects of solely genetic or solely environmental influences, but do not contain GxM. The issue of spurious moderator effects arising in the presence of untested quadratic effects has been discussed for structural equation models in general (MacCallum and Mar 1995; Lubinski and Humphreys 1990).

We have shown how to test some of these alternative models and also have proposed an alternative GxM model

based on an extension of the CF model. The availability of main effects model (4) and CF model (10) and the ability to test them each versus Purcell’s Cholesky model (2) expands the tools available to model and test GxM. Using the main effects model instead of Purcell’s model will help prevent spurious detection of GxM due to nonlinear effects of M on P and help focus attention on GxM parameter α_U for moderation of unique genetic effects versus α_C , which captures moderation of common genetic effects. The CF model provides for a more powerful test and more parsimonious quantification of GxM than does the Cholesky model and is not subject to the same problems of spurious GxM detection due to nonlinear effects of M on P . We hope that the ability to fit quadratic effects models such as (9) and analogous variations of the CF model in the future will further expand the toolbox for flexible GxM testing and modeling.

We also have discussed the calculation of effect sizes via components of variance in Purcell’s and related structural models. We pointed out flaws in the current definitions and proposed alternative calculations based on the decomposition of the model into uncorrelated terms, yielding average and partial effects of latent factors A_M , E_M , A_U , and E_U . These decompositions provide additional and complementary tools for modeling GxM. As shown via example, the proposed decompositions are constructed to automatically avoid spurious GxM detection due to nonlinear effects of common genetic or common environmental factors. Additionally, the proposed decomposition method provides more than one complementary set of effects which can be used alone or in combination depending on which factors are to be controlled, which are of interest, and/or which are putative moderators in a given analysis.

Acknowledgments The authors thank Wendy Johnson and Robert F. Krueger for insightful comments on the first draft of this article, which dramatically improved subsequent drafts.

References

- Burt SA, McGue M, DeMarte JA, Krueger RF, Iacono WG (2006) Timing of menarche and the origins of conduct disorder. *Arch Gen Psychiatry* 63:890–896
- Eaves LJ, Last K, Martin NH, Jinks JL (1977) A progressive approach to non-additivity and genotype-environment covariance in the analysis of human differences. *Br J Math Statist Psychol* 30:1–42
- Eaves L, Silberg J, Erkanli A (2003) Resolving multiple epigenetic pathways to adolescent depression. *J Child Psychol Psychiatry* 44:1006–1014
- Jang K, Dick D, Wolf H, Livesley W, Paris J (2005) Psychosocial adversity and emotional instability: an application of gene-environment interaction models. *Eur J Personality* 19:259–272
- Jinks JL, Fulker DW (1970) Comparison of the biometrical genetical, mava, and classical approaches to the analysis of human behavior. *Psychol Bull* 73:311–349
- Johnson W (2007) Genetic and environmental influences on behavior: capturing all the interplay. *Psychol Rev* 114:423–440
- Johnson W, Krueger RF (2005a) Genetic effects on physical health: lower at higher income levels. *Beh Genet* 35:579–590
- Johnson W, Krueger RF (2005b) Higher perceived life control decreases genetic variance in physical health: evidence from a national twin study. *J Personal Soc Psychol* 88:165–173
- Johnson W, Krueger R (2005c) Predictors of physical health: toward an integrated model of genetic and environmental antecedents. *J Gerontol Ser B* 60:42–52
- Johnson W, Krueger R (2006) How money buys happiness: genetic and environmental processes linking finances and life satisfaction. *J Personal Soc Psychol* 90:680–691
- Kendler K, Aggen S, Prescott C, Jacobson K, Neale M (2004) Level of family dysfunction and genetic influences on smoking in women. *Psychol Med* 34:1263–1269
- Kremen W, Jacobson K, Xian H, Eisen S, Waterman B, Toomey R, Neale M, Tsuang M, Lyons M (2005) Heritability of word recognition in middle-aged men varies as a function of parental education. *Beh Genet* 4:417–433
- Lahey BB, Waldman ID (2003) A developmental propensity model of the origins of conduct problems during childhood and adolescence. In: Lahey BB, Moffitt TE, Caspi A (eds) *Causes of conduct disorder and juvenile delinquency*. Guilford Press, New York, pp 76–117
- Loehlin JC (1996) The Cholesky approach: a cautionary note. *Beh Genet* 26:65–69
- Lubinski D, Humphreys LG (1990) Assessing spurious “moderator effects”: illustrated substantively with the hypothesized (“synergistic”) relation between spatial and mathematical ability. *Psychol Bull* 107:385–393
- MacCallum RC, Mar CM (1995) Distinguishing between moderator and quadratic effects in multiple regression. *Psychol Bull* 118:405–421
- MacCallum RC, Wegener DT, Uchino BN, Fabrigar LR (1993) The problem of equivalent models in applications of covariance structure analysis. *Psychol Bull* 114:185–199
- Moffitt TE (2005) The new look of behavioral genetics in developmental psychopathology: gene-environment interplay in antisocial behaviors. *Psychol Bull* 131:533–554
- Muthén LK, Muthén BO (2006) *Mplus user’s guide*, 4th edn. Muthén & Muthén, Los Angeles, CA
- Neale MC, Boker SM, Xie G, Maes HH (2003) *Mx: Statistical modeling*, 6th edn. Department of Psychiatry, VCU Box 900126, Richmond, VA 23298
- Plomin R, DeFries JC, Loehlin JC (1977) Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol Bull* 84:309–322
- Purcell S (2002) Variance component models for gene-environment interaction in twin analysis. *Twin Res* 5:554–571
- Rice F, Gordon H, Shelton K, Thapar A (2006) Family conflict interacts with genetic liability in predicting childhood and adolescent depression. *J Am Acad Child Adolesc Psychiat* 45:841–848
- R Development Core Team (2005) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, website: <http://www.R-project.org>
- Rutter M (2006) *Genes and behavior: nature-nurture interplay explained*. Blackwell, Malden MA
- Timberlake D, Rhee S, Haberstick B, Hopfer C, Ehringer M, Lessem M, Smolen A, Hewitt J (2006) The moderating effects of religiosity on the genetic and environmental determinants of smoking initiations. *Nicotine Tobacco Res* 8:123–133