HHS Public Access

Author manuscript

Behav Genet. Author manuscript; available in PMC 2022 July 01.

Published in final edited form as:

Behav Genet. 2021 July ; 51(4): 425–437. doi:10.1007/s10519-021-10055-x.

The analytic identification of variance component models common to behavior genetics

Michael D. Hunter,
School of Psychology, Georgia Institute of Technology, Atlanta, GA 30313

S. Mason Garrison,
Department of Psychology Wake, Forest University, Winston-Salem, NC 27109

S. Alexandra Burt,
Department of Psychology, Michigan State University East, Lansing, MI 48824

Joseph L. Rodgers
Department of Psychology and Human Development, Vanderbilt University, Nashville, TN 37203

Abstract

Many behavior genetics models follow the same general structure. We describe this general structure and analytically derive simple criteria for its identification. In particular, we find that variance components can be uniquely estimated whenever the relatedness matrices that define the components are linearly independent (i.e., not confounded). Thus, we emphasize determining which variance components can be identified given a set of genetic and environmental relationships, rather than the estimation procedures. We validate the identification criteria with several well-known models, and further apply them to several less common models. The first model distinguishes child-rearing environment from extended family environment. The second model adds a gene-by-common-environment interaction term in sets of twins reared apart and together. The third model separates measured-genomic relatedness from the scanner site variation in a hypothetical functional magnetic resonance imaging study. The computationally easy analytic identification criteria allow researchers to quickly address model identification issues and define novel variance components, facilitating the development of new research questions.

Keywords

behavior genetics; model identification; variance components; structural equation modeling

Introduction

Many models in the field of behavior genetics take the form

Correspondence may be addressed to Michael D. Hunter, 654 Cherry Street, Georgia Institute of Technology, Atlanta, GA 30332-0170; or email sent to michael-hunter@gatech.edu.

$$\Sigma(\theta) = \widehat{Cov(y)} = \sum_{i=1}^{m} R_i \sigma_i^2 = R_1\sigma_1^2 + R_2\sigma_2^2 + \ldots + R_m\sigma_m^2 \quad (1)$$

where $\Sigma(\theta)$ is the expected covariance matrix as a function of the vector of free parameters $\theta = \left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2\right)$, $\boldsymbol{y}$ is the vector of all observed outcome variables for all members of a family, $\widehat{Cov(y)}$is the model-implied approximation to the covariance of the observed outcome variables, R_i is a relatedness matrix, and σ_i^2 is the variance attributable to R_i. Models of this form occur in both classical twin and family designs, and in modern molecular designs. In this paper we will (1) show how several common designs in classic and modern behavior genetics fit this general pattern, (2) derive the analytic solution for model identification of any model in this form, (3) validate several example models with previously known identification, and (4) show how this analytic identification solution can aid in (a) evaluating novel research designs and (b) finding problems of identification and their solutions in these novel settings. To illustrate this last point, we use two examples: first, a twin and family model that separates child-rearing environmental effects from broader extended family effects, and second, a modern molecular model that separates additive genetic similarly from similarity that may be due to the same measurement instrument (e.g., fMRI scanner site). Throughout, we emphasize model identification – the ability to uniquely determine parameter estimates – rather than the particular parameter estimation method. By creating easily computed model identification criteria, other research can then conceive new kinds of variance components to address novel research questions.

Common Designs

Classical Twin Design

In the classical twin design for a single phenotype, the ACE model of additive genetics, common environments, and unique environments separates Equation 1 into two special forms: one for monozygotic (MZ) twins and one for dizygotic (DZ) twins. For MZ twins Equation 1 takes the form of Equation 2.

$$\begin{aligned}\Sigma_{MZ}(\theta) = &\underbrace{\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}}_{R_A MZ}\sigma_A^2 \\ &+ \underbrace{\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}}_{R_C}\sigma_C^2 + \underbrace{\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}}_{R_E}\sigma_E^2\end{aligned} \quad (2)$$

For DZ twins Equation 1 takes the form of Equation 3.

$$\Sigma_{DZ}(\theta) = \underbrace{\begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}}_{R_A DZ}\sigma_A^2 + \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}}_{R_C}\sigma_C^2 + \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{R_E}\sigma_E^2 \quad (3)$$

Equations 2 and 3 show the common form of the classical twin design for a single phenotype. There is a model-implied covariance matrix for the MZ twins and a model-implied covariance matrix for the DZ twins. The two covariance matrices have different structures, but depend on the same free parameters (σ_A^2, σ_C^2,σ_E^2). The model cannot be estimated using only MZ twin pairs or only DZ twins pairs. Rather, the two kinds of twin pairs and their model-implied covariance matrices must be combined into a multiple group model: one group for MZ twins and one group for DZ twins. The combined multiple group model for MZ and DZ twins is merely the block-diagonal form of Equation

$$\Sigma(\theta) = \begin{pmatrix} \Sigma_{MZ}(\theta) & 0 \\ 0 & \Sigma_{DZ}(\theta) \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & .5 \\ 0 & 0 & .5 & 1 \end{pmatrix}}_{R_A}\sigma_A^2 + \underbrace{\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}}_{R_C}\sigma_C^2 + \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{R_E}\sigma_E^2 \quad (4)$$

The block-diagonal multigroup construction of the ACE model is less common in behavior genetics writings, but is entirely equivalent to that of Equations 2–3 and has advantages for understanding model identification. Consequently, the remainder of this work will use the block-diagonal construction and variations on it.

The scores for a single phenotype are stored in the 4-dimensional vector $\boldsymbol{y}$ and thus have a corresponding 4 × 4 model-implied covariance matrix, $\Sigma(\boldsymbol{\theta})$. The additive genetics relatedness matrix is R_A; the common environments relatedness matrix is R_C; and the unique environments relatedness matrix is R_E. The variance components associated with each relatedness areσ_A^2, σ_C^2, andσ_E^2, respectively.

Each matrix describes the relation between pairs of twins for the defined component. The 1s on the diagonal of R_A imply that each twin is genetically identical to themselves. The off-diagonal 1s in the upper left block of R_A are for MZ twins, representing their genetic equivalency. The off-diagonal .5s in the lower right block of R_A are for DZ twins, representing that they share on average half of their segregating genes. The blockwise unit matrices of R_C indicate which pairs of twins have the same rearing environment. Finally, the identity matrix of R_E represents the unique variance contribution of each twin whatever its source. The variance σ_E^2 is generally considered a combination of several attributes, including unique environmental effects, model misspecification[1], and measurement error (cf. Plomin, 2011; Turkheimer & Waldron, 2000).

Twins Raised Apart and Together Design

In the twins raised apart and together design, MZ and DZ twins raised apart are added to the classic twin design. For brevity, we often refer to this design as the “twins reared apart design”. The special case of Equation 1 now becomes Equation 6.

$$\Sigma(\theta) = \begin{pmatrix} 1 & 1 & 0 & 0 & & & & \\ 1 & 1 & 0 & 0 & & & & \\ 0 & 0 & 1 & .5 & & & & \\ 0 & 0 & .5 & 1 & & & & \\ & & & & 1 & 1 & 0 & 0 \\ & & & & 1 & 1 & 0 & 0 \\ & & & & 0 & 0 & 1 & .5 \\ & & & & 0 & 0 & .5 & 1 \end{pmatrix} \sigma_A^2 + \begin{pmatrix} 1 & 1 & 0 & 0 & & & & \\ 1 & 1 & 0 & 0 & & & & \\ 0 & 0 & 1 & 1 & & & & \\ 0 & 0 & 1 & 1 & & & & \\ & & & & 1 & 0 & 0 & 0 \\ & & & & 0 & 1 & 0 & 0 \\ & & & & 0 & 0 & 1 & 0 \\ & & & & 0 & 0 & 0 & 1 \end{pmatrix} \sigma_C^2 + \begin{pmatrix} 1 & 0 & 0 & 0 & & & & \\ 0 & 1 & 0 & 0 & & & & \\ 0 & 0 & 1 & 0 & & & & \\ 0 & 0 & 0 & 1 & & & & \\ & & & & 1 & 0 & 0 & 0 \\ & & & & 0 & 1 & 0 & 0 \\ & & & & 0 & 0 & 1 & 0 \\ & & & & 0 & 0 & 0 & 1 \end{pmatrix} \sigma_E^2 \quad (5)$$

[1]Of course, model misspecification may also impact any component of the model (see Keller & Coventry, 2005; Coventry & Keller, 2005, for a detailed analysis)

$$= \begin{pmatrix} R_A & \\ & R_A \end{pmatrix} \sigma_A^2 + \begin{pmatrix} R_C & \\ & R_E \end{pmatrix} \sigma_C^2 + \begin{pmatrix} R_E & \\ & R_E \end{pmatrix} \sigma_E^2 \quad (6)$$

The empty off-diagonal blocks in Equation 6 signify zero entries. The first of the two diagonal blocks is identical to the classical twin design. The second diagonal blocks (i.e., the lower right) create the unique feature of the twins reared apart design: the additive genetic and unique environments blocks are the same as the classical twin design, but the common environments blocks differ. The first block for the shared environmental relatedness corresponds to twins raised together, whereas the second block corresponds to twins raised apart. One of the novel features of this design relates to its identification. Additional variance components for dominance-genetic effects and other nonlinear effects are identified in the twins reared apart design that are not identified in the classical twin design[2]. This will be shown analytically in the section on identification.

General Pedigree Design

In the general pedigree design, the block design of Equation 6 is further extended to as many families as desired. Suppose there are F families, where a family is defined as the union of all people in a data set with nonzero relatedness and relatedness may be along any known genetic or environmental pathway[3]. The covariance structure for such a pedigree model is block diagonal with F blocks each of size equal to that family's number of members. No blocks need to be the same size for different families, but all the blocks for a single family need to be the same size across all contributing variance components. The general schema is in Equation 7.

$$\Sigma(\theta) = \begin{pmatrix} R_{A1} & & \\ & \ddots & \\ & & R_{AF} \end{pmatrix} \sigma_A^2 + \begin{pmatrix} R_{C1} & & \\ & \ddots & \\ & & R_{CF} \end{pmatrix} \sigma_C^2 + \begin{pmatrix} R_{E1} & & \\ & \ddots & \\ & & R_{EF} \end{pmatrix} \sigma_E^2 \quad (7)$$

We illustrate with A, C, and E variance components, but depending on the family structure more or fewer components may be identified.

Modern Molecular Design

In the modern molecular design with nominally unrelated people, all people in the data are considered to be members of the same family. Thus, at a conceptual level the modern

[2]So-called ACE and ADE models are identified in the classical twin design, but not ACDE models. All of these are identified in the twins reared apart design.

[3]The border where one family ends and another begins may differ for different research purposes. For example, the definition of a "family" here does not exclude people related only by marriage because these people are "related" along the environmental pathway of marriage, regardless of whether a marriage variance component is in the fitted model.

molecular design is the limiting case of the pedigree design in which there is only one, large family. Equation 1 now takes the special form of Equation 8,

$$\Sigma(\theta) = \sum_{i=1}^{m} R_i \sigma_i^2 + I\sigma_E^2 \tag{8}$$

where each R_i is a relatedness matrix the size of the whole sample and the σ_i^2 are variance components associated with each relatedness matrix. The I matrix is the identity matrix, equal in structure to the previous R_E blocks corresponding to the σ_E^2 component. Of course, the last variance component of Equation 8 could be incorporated into the sum (e.g., as in Equation 1); however, conventionally the terms in the sum are conceived as additive genetic relatednesses due to distinct biological sources (e.g., due to each chromosome or to specific rare variants). Thus, the σ_i^2 terms form a genomic partitioning of the σ_A^2 component from twin and family modeling.

The degree of genetic relatedness in each R_i is measured by the single nucleotide polymorphism (SNP) correlations between all pairs of people rather than average genetic relatedness due to inheritance for known descendants (see Yang, Lee, Goddard, & Visscher, 2011, for some details). The SNP correlation between person j and person k defines the row j and column k entry of R_i:

$$R_{i\{jk\}} = \frac{1}{N} \sum_{h=1}^{N} \frac{\left(x_{hj} - 2p_h\right)\left(x_{hk} - 2p_h\right)}{2p_h(1-p_h)} \tag{9}$$

where N is the number of SNPs, x_{hj} is the number of copies of the reference allele for person j on SNP h, x_{hk} is the number of copies of the reference allele for person k on SNP h, and p_h is the frequency of the reference allele.

Each R_i is generally a SNP-measured genomic relatedness matrix. For example with $m = 1$, R_1 is the genomic relatedness matrix across all chromosomes. With $m = 22$, the R_i could be the SNP relatednesses due to each of the 22 autosomes, respectively. However, there is no mathematical reason to require each R_i originate from a measured genetic association. The subsequent section will show an example considering non-genetic components in at least one of the R_i matrices.

In the next section, we consider the identification of models given by the general form of Equation 1. Equations 4, 6, 7, and 8 are each special cases of the general model. We then consider the identification of these special cases.

Identification

A model for the observed covariance matrix is identified when there is a unique set of parameters that defines the model-implied covariance matrix (Bekker & Wansbeek, 2001). The free parameters, $\boldsymbol{\theta}$, are mapped to the model-implied covariance matrix, $\Sigma(\boldsymbol{\theta})$. Identification requires that this mapping between the free parameters and the model-implied

covariance matrix be unique. A unique mapping has $\Sigma(\boldsymbol{\theta}_a) = \Sigma(\boldsymbol{\theta}_b)$ if and only if $\boldsymbol{\theta}_a = \boldsymbol{\theta}_b$. Colloquially, a model is identified when free parameters cannot trade off one another to produce the same model-implied covariance matrix. A typical example of model non-identification from psychometrics occurs in a one factor model when all factor loadings are freely estimated along with the factor variance. The one factor model is typically identified by constraining either one factor loading or the factor variance to some fixed value. When both the factor variance and all the factor loadings are freely estimated, an increase in the factor variance can be compensated for by a proportionate decrease in all the factor loadings to produce an identical model-implied covariance matrix. Thus, more than one set of parameter values produces the same model-implied covariance and consequently the same fit to the data. Thus, the factor variances and the factor loadings are not mutually identified.

The same identification issue occurs in biometrics between the “path coefficients” (i.e., factor loadings) and the “variance components” (i.e., factor variances); these model specifications either estimate the factor loadings or the factor variances, respectively, because both are not simultaneously identified. However, the emphasis of the present work is on which variance components are identified for a given set of genetic and environmental relationships, rather than the method of specifying the components which is already well-handled.

Within model identification, there are two types: global and local. Global identification requires that the mapping between the free parameters and the model-implied covariance matrix is unique over the entire parameter space, whereas local identification only requires this mapping be unique in the neighborhood of the current free parameter values (Bollen & Bauldry, 2010). That is, local identification requires that if $\Sigma(\boldsymbol{\theta}_a) = \Sigma(\boldsymbol{\theta}_b)$ and $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$ are sufficiently close to one another then $\boldsymbol{\theta}_a = \boldsymbol{\theta}_b$. The closeness criterion means that local identification depends on the numerical values of the free parameters. A model may be locally identified for some free parameter values, but not others[4]. In the present work, we are only concerned with the easier problem of resolving local model identification. A sufficient condition for the *local* identification of a model for the covariance matrix is that the first derivative (Jacobian) of the function that maps free parameters to the non-redundant elements of the model-implied covariance matrix has full column rank (Bekker & Wansbeek, 2001; Bekker, 1986; Bekker & ten Berge, 1997). We will next show how this applies to the model of Equation 1.

The General Case

The vector of free parameters, σ_i^2, along with the fixed relatedness matrices, R_i, completely specify the covariance matrix of a single phenotype across a family of arbitrary size and structure as in Equation 1.

[4]A classic example of local but not global identification is a one-factor model with free factor loadings, but a fixed factor mean and a fixed factor variance. Such a model is locally identified everywhere except where all the factor loadings are zero. So, the model is locally identified where the factor loadings are all ones, but not where the factor loadings are all zeros. Therefore, the model is not globally identified.

According to the methods described by Bekker and colleagues (Bekker, 1986; Bekker & ten Berge, 1997), such a model is (locally) identified if and only if the matrix of first partial derivatives of the model-implied covariance matrix with respect to the free parameters has rank equal to the number of free parameters. Thus, the model is identified when the columns of the matrix of first partial derivatives are linearly independent. Conceptually, this means that each free parameter leads to a unique, linearly separable change in the model-implied covariance matrix. Mathematically, we derive from Equation 1 the half-vectorization[5] of the model-implied covariance matrix.

$$vech(\Sigma(\theta)) = vech\left(R_1\sigma_1^2 + R_2\sigma_2^2 + \ldots + R_m\sigma_m^2\right) \tag{10}$$

$$= vech(R_1)\sigma_1^2 + vech(R_2)\sigma_2^2 + \ldots + vech(R_m)\sigma_m^2 \tag{11}$$

$$= \underbrace{\left(vech(R_1) \quad vech(R_2)\ldots vech(R_m)\right)}_{X}\underbrace{\begin{pmatrix}\sigma_1^2\\ \sigma_2^2\\ \vdots\\ \sigma_m^2\end{pmatrix}}_{\theta} \tag{12}$$

$$= X\theta \tag{13}$$

where X is the half vectorization of each relatedness matrix concatenated into separate columns, and $\boldsymbol{\theta}$ is the vector of all free parameters. Then taking the first partial derivatives (Jacobian) of $vech(\Sigma(\boldsymbol{\theta}))$ with respect to $\boldsymbol{\theta}$ yields

$$\frac{\partial}{\partial\theta}vech(\Sigma(\theta)) = \frac{\partial}{\partial\theta}(X\theta) = X \tag{14}$$

Equation 14 implies that the general behavior genetics model in Equation 1 is identified if and only if X has rank equal to the number of free parameters. Thus, identification occurs if and only if the relatedness matrices are linearly independent.

Importantly, the identification criterion derived above does not depend on the phenotype, on the observed covariance data, or even on the numerical values of the free parameters in $\boldsymbol{\theta}$. Instead, the identification criterion only depends on the structural relations in the data. Thus, a researcher can determine the identification of a model *before* collecting any data, provided that the structural features of the data are know. All that is needed is the linear independence of the structural features.

[5]The half vectorization of a matrix concatenates all the unique elements of a symmetric matrix into a single column vector. That is, the half vectorization concatenates the lower triangle of a symmetric matrix into a single column vector. It is often notated by *vech*(), and is distinct from the full vectorization notated by *vec*() which concatenates *all* elements of any matrix into a single column vector.

For any given family structure, the linear independence of the relatedness matrices is exceedingly simple to determine. Multiple methods exist for determining the linear independence of a set of vectors. Introductory linear algebra and statistics books contain many of them (e.g., Leon, 2006; Lay, 2003; Cohen, Cohen, West, & Aiken, 2003; Scharf, 1991). A matrix X as structured above will have linearly independent columns if any of the following criteria are met:

1. The determinant of the Gram matrix, $X^{\mathrm{T}}X$, is nonzero: det ($X^{\mathrm{T}}X$) ≠ 0 (Scharf, 1991, p. 26, Theorem 2.1).
2. The rank of X as computed by its QR decomposition is equal to the number of columns (Lay, 2003, p. 402–416).
3. The rank of the null space of X is equal to the number of rows of X minus the number of columns of X (Leon, 2006, p. 164, rank-nullity theorem).
4. The multiple R^2 coefficient of determination for all possible multiple regressions between the columns of X is always less than 1 (Cohen et al., 2003, p. 631–632). Computationally, $R_i^2 = 1 - \frac{1}{\mathrm{diag}\left(\mathrm{Cor}(X)^{-1}\right)_i} < 1$ for all columns i, where diag( )$_i$ is the i^{th} diagonal element of a matrix and Cor(X)$^{-1}$ is the inverse of the correlation matrix of X.

There are numerous other criteria, but the above span a reasonable computational and conceptual set of options. We next illustrate the linear independence method of local model identification using the example models described above.

Classical Twin Design

Behavior geneticists have relied on the classical twin design for its ability to separate genetic and family effects. Critically, both MZ and DZ twins are required to identify the A, C, and E components. There are various ways to determine this, but we use Equation 14 here. For MZ twins only, Equation 4 implies that X of Equation 14 is

$$X_{MZ} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \tag{15}$$

Because the first two columns are equal, and thus linearly dependent, the model is not identified. For DZ twins alone, the analogous X is

$$X_{DZ} = \begin{pmatrix} 1 & 1 & 1 \\ .5 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \tag{16}$$

which does not contain duplicated columns, but two times the first column minus the third column is equal to the middle column, so this model is also not identified. However, combining MZ and DZ twins as in Equation 4 yields

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \\ .5 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \xrightarrow[\text{Drop rows of zeros}]{} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ .5 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} X_{MZ} \\ X_{DZ} \end{pmatrix} \tag{17}$$

which is identified, as is well-known and previously established. Now we consider the ACDE model.

$$\Sigma(\theta) = R_A\sigma_A^2 + R_A^2\sigma_D^2 + R_C\sigma_C^2 + R_E\sigma_E^2 \tag{18}$$

where R_A^2 indicates that R_A is multiplied by itself elementwise, rather than matrix multiplied[6]. The X matrix in this case (omitting rows of all zeros) is

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ .5 & .25 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \tag{19}$$

which has rank 3 so it is not identified.

Twins Raised Apart Design

It follows from the classical twin design that the ACE model is also identified in the twins raised apart design. The ACDE model, however, is now identified in the twins raised apart design because its design matrix has rank 4. The D component in the ACDE model is a kind of nonlinear gene-by-gene interaction. Not only is this gene-by-gene interaction identified, but a gene-by-common-environment (i.e., A*C) interaction is also identified. Hunter, McKee, and Turkheimer (2020) developed a mathematical model that suggests an A*C interaction may account for some of the reduced heritability found in modern molecular designs compared to that for twins and families (see e.g., Plomin, 2014; Sauce & Matzel, 2018; Turkheimer, 2011; Vinkhuyzen, Wray, Yang, Goddard, & Visscher, 2013, for further information and reviews). For the present purpose, it is immaterial whether an A*C interaction really leads to missing heritability. We want to know whether a model with an A*C interaction is identified.

[6]Note that the dominance-genetic design matrix, R_A^2, could also be notated as R_D but this would hide the mathematical relation between additive-genetics and dominance-genetics: namely, that dominance-genetics are a kind of nonlinear gene-by-gene interaction.

For the twins raised apart design, we consider a novel model with additive genetics, common environments, a gene-by-common-environment interaction, a gene-by-gene dominance interaction, and unique environments as given in Equation 20.

$$\Sigma(\theta) = R_A\sigma_A^2 + R_A R_C \sigma_{A*C}^2 + R_C\sigma_C^2 + R_A^2\sigma_D^2 + R_E\sigma_E^2 \tag{20}$$

We call the model of Equation 20 an A*CDE model because it includes the product of additive genetics and common environments as an interaction term. The novel part of Equation 20 is the $R_A R_C \sigma_{A*C}^2$ term which adds a variance component due to a gene-by-common-environment interaction. The relatedness matrix for the A*C variance component, $R_A R_C$, is the elementwise product of the relatedness matrix for the additive genetics and common environments relatedness matrix. We determine the identification of the model in Equation 20 by creating the design matrix, X_{A*C}, and checking its rank. After dropping all-zero rows, X_{A*C}, is given by Equation 21.

$$X_{A*C} = \begin{array}{c} \\ MZT \\ DZT \\ MZA \\ DZA \end{array} \begin{array}{c} \begin{array}{ccccc} A & A*C & C & D & E \end{array} \\ \begin{pmatrix} 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 0.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 0.50 & 0.50 & 1.00 & 0.25 & 0.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 0.00 & 0.00 & 1.00 & 0.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 0.50 & 0.00 & 0.00 & 0.25 & 0.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \end{pmatrix} \end{array} \tag{21}$$

The columns are labeled with their corresponding variance components. The rows are labeled with the blocks that make-up the four groups of the twins raised apart design: *MZT* is for MZ twins raised together; *DZT* is for DZ twins raised together; *MZA* is for MZ twins raised apart; and *DZA* is for DZ twins raised apart. The design matrix, X_{A*C}, has rank 5 so the model of Equation 20 is identified.

General Pedigree Design

For the purpose of model identification, the off-diagonal blocks in Equation 7 can be omitted without influencing the rank of the design matrix. Therefore, for a general pedigree model the design matrix for F families becomes

$$X_{pedigree} = \begin{pmatrix} vech(R_{A1}) & vech(R_{C1}) & \dots & vech(R_{E1}) \\ vech(R_{A2}) & vech(R_{C2}) & \dots & vech(R_{E2}) \\ \vdots & \vdots & \vdots & \vdots \\ vech(R_{AF}) & vech(R_{CF}) & \dots & vech(R_{EF}) \end{pmatrix} \quad (22)$$

where arbitrarily many relatedness matrices can be included. Depending on the structure of the relatedness matrices, a large number of novel variance components may be identified. The strength of more complex family designs is that they afford the identification of novel variance components. For example, Rodgers, Bard, Johnson, D'Onofrio, and Miller (2008) used the intergenerational complex family structure of the National Longitudinal Survey of Youth to conceive of the Mother-Daughter Aunt-Niece (MDAN) design which alters the form of the equal environments assumption and affords the identification of novel variance components. The prospect of easy analytic model identification opens new possibilities for researchers to develop relatedness matrices that allow variance components to be conceived that have not previously been possible.

Modern Molecular Design

The modern molecular design generalizes the AE model to measured genotypes. Depending on the structure of the genomic relatedness matrices and on the structure of other known relations in the data, a large number of additional variance components may be identified. The additional identified variance components in modern molecular designs are almost entirely unexplored. The conceptually limited set of variance components addressed in molecular designs fails to fully utilize many aspects of the data collected. Example data characteristics include geographic site of data collection as a member of a consortia, data collector, residential area of the participant, and laboratory where assays are conducted. Each of these characteristics may have critical methodological or theoretical import. We therefore urge researchers to use the ability to analytically determine model identification to conceptualize additional variance components that afford novel hypotheses. We illustrate one such novel variance component.

Consider a sample of six people with measured genomes. The size of the sample is small only for illustrative purposes; a real data example would be cumbersome to fully describe here. For these six people, four different functional magnetic resonance imaging (fMRI) scanners were used. In fMRI studies, scanner and site differences can lead to differences in the measured phenotype across scanners, and an additional variance component for scanner is one method of reducing these effects (Zhou et al., 2012). We would like to add a variance component, σ^2_S, that accounts for variability due to the scanner.

The model for this scanner example takes the form of Equation 23, a special case of Equation 8 with three components: one for measured genetic similarity, one for the scanner, and one for the residual/environment.

$$\Sigma(\theta) = \begin{pmatrix} 0.94 & 0.04 & -0.04 & -0.01 & -0.01 & 0.03 \\ 0.04 & 1.02 & 0.04 & 0.03 & -0.01 & 0.03 \\ -0.04 & 0.04 & 0.98 & -0.01 & -0.03 & 0.07 \\ -0.01 & 0.03 & -0.01 & 0.97 & 0.03 & -0.04 \\ -0.01 & -0.01 & -0.03 & 0.03 & 1.05 & 0.07 \\ 0.03 & 0.03 & 0.07 & -0.04 & 0.07 & 0.99 \end{pmatrix} \sigma^2_A + \begin{pmatrix} 1 & & & & & \\ & 1 & 1 & & & \\ & 1 & 1 & & & \\ & & & 1 & 1 & \\ & & & 1 & 1 & \\ & & & & & 1 \end{pmatrix} \sigma^2_S + \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{pmatrix} \sigma^2_E \tag{23}$$

If the standard model in Equation 8 is a generalization of the AE model to measured genomes, then the model in Equation 23 which adds an effect for scanner could be called an ASE model.

The first 6 × 6 matrix is the genomic-relatedness matrix. The diagonals indicate “self-relatedness” (i.e., how related each person is to themselves), but are often taken as a measure of inbreeding. The off-diagonals of the genomic-relatedness matrix can be thought of as the correlations between the genotypes of each pair of people.

The second 6 × 6 matrix is for the scanner. Participant 1 has scanner 1; participants 2 and 3 both have scanner 2; participants 4 and 5 both have scanner 3; and participant 6 has scanner 4. The structure illustrated here is block-diagonal for readability, but does not need to be. The key feature to estimate a random effect for scanner is to have 1s on the diagonal, and 1s on any off-diagonal element for a pair of participants with the same scanner. Other kinds of random effects could be similarly defined for values other than 1s and 0s. The critical aspect is to have the relatedness matrix indicate the degree of similarity between participants for the desired random effect.

As before, the last matrix is for residual variance, measurement error, or any other feature of the data that inflates the diagonals of the observed covariance relative to the expected covariance matrix from the other components.

The model in Equation 23 is identified because the corresponding matrix X_S has rank

$$X_S = \begin{pmatrix} 0.94 & 1 & 1 \\ 0.04 & 0 & 0 \\ -0.04 & 0 & 0 \\ -0.01 & 0 & 0 \\ -0.01 & 0 & 0 \\ 0.03 & 0 & 0 \\ 1.02 & 1 & 1 \\ 0.04 & 1 & 0 \\ 0.03 & 0 & 0 \\ -0.01 & 0 & 0 \\ 0.03 & 0 & 0 \\ 0.98 & 1 & 1 \\ -0.01 & 0 & 0 \\ -0.03 & 0 & 0 \\ 0.07 & 0 & 0 \\ 0.97 & 1 & 1 \\ 0.03 & 1 & 0 \\ -0.04 & 0 & 0 \\ 1.05 & 1 & 1 \\ 0.07 & 0 & 0 \\ 0.99 & 1 & 1 \end{pmatrix} \quad (24)$$

Using methods similar to those outlined above, new variance components can be also be considered for modeling in behavior genetics. The more general designs of modern measured-genome methods afford many opportunities for such novel variance components.

Computational Examples

Appendix A defines several R (R Development Core Team, 2020) functions written by the first author for (1) determining whether a model following the structure of Equation 1 is identified, and (2) fitting such a model to covariance data. These functions are also available as an R package on CRAN (Garrison, Hunter, Burt, & Trattner, 2020) and on GitHub (https://github.com/R-Computing-Lab/BGMisc).

Consider the classical twin model with only MZ twins. We check its identification with the following code.

```
identifyComponentModel(
A=matrix(1, nrow=2, ncol=2),
C=matrix(1, nrow=2, ncol=2),
E=diag(1, nrow=2))
```

This returns the following output.

```
Component model is not identified.
Non-identified parameters are A, C
$identified
[1] FALSE
nidp
[1] “A” “C”
```

This means that the “A” and “C” components are not simultaneously identified in this model. However, consider the classical twin design with both MZ and DZ twins. We check its identification with the following code.

```
identifyComponentModel(
A=list(matrix(1, nrow=2, ncol=2),
matrix(c(1, .5, .5, 1),
nrow=2, ncol=2)),
C=list(matrix(1, nrow=2, ncol=2),
matrix(1, nrow=2, ncol=2)),
E=list(diag(1, nrow=2), diag(1, nrow=2)))
```

The above code returns the following output.

```
Component model is identified.
$identified
[1] TRUE
$nidp
character(0)
```

Thus, as is well-known, the classical twin model is identified. Again, the benefit here is not to show that the classical twin model is identified. Rather, the benefit is to provide a way to determine if *any* model following the broad structure of Equation 1 is identified.

One model with less established identification is the AC’RE model (Garrison et al., 2018). In the AC’RE model, the structures for the additive genetics (A) and unique environments (E) are left unchanged, but the common environments (C) are partitioned into a rearing environment (R) and an extended community or family environment (C’). Its aim was to produce results comparable to the twins-raised apart design, with less rare kinship groups, such as cousins. Accordingly, a reasonable question then is: given a particular family structure, are all the components of the AC’RE model identified?

In the classical twin design, the relatedness matrix for the C’ component and the R component would be identical, and thus the AC’RE model would not be identified. However, in the twins reared apart and together design the relatedness matrices for C’ and R are now

distinct. We can construct the relatedness matrices for the AC'RE model by extending the model in Equation 6 with the following R code:

```
arel <- rep(list(matrix(1, 2, 2),
matrix(c(1, .5, .5, 1), 2, 2)), times=2))
cprel <- rep(list(matrix(1, 2, 2)), times=4))
rrel <- list(matrix(1, 2, 2), matrix(1, 2, 2),
diag(1, 2), diag(1, 2)))
erel <- rep(list(diag(1, 8)), times=4)
```

Then we can determine this model's identification in the twins reared together and apart design with the helper function used previously.

```
identifyComponentModel(
A=arel,
Cp=cprel,
R=rrel,
E=erel)
```

The output from this function indicates that the AC'RE model is identified for the twins reared apart and together design.

```
Component model is identified.
$identified
[1] TRUE
$nidp
character(0)
```

Similar to the identification problem, a model in the form of Equation 1 is simple to fit. Using twin and sibling data from the National Longitudinal Survey of Youth (NLSY) available in the NlsyLinks package (Beasley et al., 2016) along with kinship links created by Rodgers et al. (2016), we fit such a model using ordinary least squares and compare our estimates to standard maximum likelihood estimates from the same data.

First, we create a cleaned and structured data set for standardized math scores.

```
library(NlsyLinks)
# Start with the built-in data.frame in NlsyLinks
dsLinks <- Links79PairExpanded
# Use only Gen2 Siblings (NLSY79-C)
dsLinks <- dsLinks[dsLinks$RelationshipPath=='Gen2Siblings', ]
oName_S1 <- "MathStandardized_S1" #Stands for Outcome1
oName_S2 <- "MathStandardized_S2" #Stands for Outcome2
```

```
dsGroupSummary <- RGroupSummary(dsLinks, oName_S1, oName_S2)
dsClean <- CleanSemAceDataset(dsDirty=dsLinks, dsGroupSummary, oName_S1,
oName_S2)
# Drop small number of ambiguous relationships
dsClean <- dsClean[dsClean$R != .375, ]
```

Next, we create covariance matrices for each of the three relatedness types: cousins, DZ twins and full siblings, and MZ twins.

```
cousinCov <- cor(dsClean[dsClean$R==.25, c('O1', 'O2')], use='pair')
sibCov <- cor(dsClean[dsClean$R==.5, c('O1', 'O2')], use='pair')
mzCov <- cor(dsClean[dsClean$R==1, c('O1', 'O2')], use='pair')
```

Finally, we fit the ACE model

```
# Fit model using maximum likelihood and NlsyLinks function
ace <- AceLavaanGroup(dsClean)
# Fit model using OLS regression
ols <- fitComponentModel(
covmat=list(mzCov, sibCov, cousinCov),
A=list(
matrix(1, nrow=2, ncol=2),
matrix(c(1, .5, .5, 1), nrow=2, ncol=2),
matrix(c(1, .25, .25, 1), nrow=2, ncol=2)),
C=list(
matrix(1, nrow=2, ncol=2),
matrix(1, nrow=2, ncol=2),
matrix(1, nrow=2, ncol=2)),
E=list(
diag(1, nrow=2),
diag(1, nrow=2),
diag(1, nrow=2))
)
ace
#[1] "Results of ACE estimation: [show]"
# ASquared CSquared ESquared CaseCount
# 0.6137716 0.2138675 0.1723609 8205.0000000
coef(ols)/sum(coef(ols))
# compmA compmC compmE
#0.81239650 0.10949326 0.07811024
```

The estimated variances based on ordinary least squares and limited information pairwise covariances broadly correspond to the full-information maximum likelihood estimates from the same data.

The provided R functions will help other researchers readily determine whether a candidate model of interest is identified given the structural patterns of relatedness found in their data. The functions also provide preliminary estimates of the variances associated with each component.

At this point, consideration of other available software to identify models is merited. The OpenMx package (Neale et al., 2016) has a function called mxCheckIdentification() that allows users to check whether a given OpenMx model is locally identified. The function uses the same underlying mathematical methods as those discussed here, and will produce identical identification results to the proposed methods for any model that meets the applicable criteria for both packages.

We feel there are five advantages of the proposed software over the mxCheckIdentification() function in OpenMx. First, the mxCheckIdentification() function uses a numerical Jacobian for its version of Equation 14. Because the mxCheckIdentification() function is more general than the methods proposed here, the numerical Jacobian is necessary, but it has much higher computational cost than the more specialized solution we proposed. Second, the mxCheckIdentification() function is not implemented for modern molecular designs (i.e., its genomic-relatedness-matrix restricted maximum likelihood features; Kirkpatrick, Pritikin, Hunter, & Neale, 2021). So, researchers interested in identifying novel components of modern molecular designs would not be able to use mxCheckIdentification(). Third, there is far less code needed to specify a variance component model and check its identification using the current methods than is needed for the same process in OpenMx. Although OpenMx is a powerful tool and can be used to identify a broad range of models, the particularly simple solution for variance component models means a specialized solution is beneficial. Moreover, we provide R code to exemplify our methods, but python, C/C++, or FORTRAN routines would be much easier to write for this specialized problem than for the general model identification problem solved by OpenMx.

The fourth and fifth advantages of the proposed model identification methods relate to formulating the identification problem as one of ordinary least squares regression. Fourth, the proposed methods yield initial estimates of all variance components using ordinary least squares (OLS). These OLS estimates could then be used as starting values for maximum likelihood estimation. The OLS estimates are only possible because the variance component model has a particularly simple structure. More complicated models are not amenable to the same OLS solution. Fifth, the OLS regression method readily yields estimates of multicollinearity among the variance components: the tolerance and the variance inflation factor. The multicollinearity regression diagnostics can suggest that although the variance components are identified, they may be poorly estimated because their relatedness matrices are collinear. These diagnostics alert researchers to estimation problems beyond mere identification. The mxCheckIdentification() function cannot provide this information on multicollinearity because the identification of most models yields no such information. It is

only because variance component models have such simple identification criteria that identification also provides information on multicollinearity.

Discussion

In this brief note, we have discussed a general modeling form that encompasses many models in behavior genetics. In particular, we limited ourselves to the subset of behavior genetics models characterized as variance component models. We then showed how several common models fit this form. We next derived an analytic method of model identification for this general model, and showed several special cases of its use to both common and uncommon designs. We considered ACE and ACDE models in the context of twins, A*CDE and AC'RE models in the context of twins raise apart and together, the mother-daughter-aunt-niece model in the context of pedigrees, and AE and ASE models in the context of modern molecular designs. The most important contribution of this work on model identification is that it allows researchers to quickly, easily, and accurately test whether a model is identified based purely on the structure of the relatedness matrices, independent of the observed phenotypes. Thus, novel relatedness patterns and variance components can be explored in the burgeoning data sources that involve large and complex family structures including modern molecular studies of nominally unrelated people.

One further implication of the model identification described previously deserves special comment. It follows immediately from Equation 13 that any arbitrary variance components model is solvable as an ordinary least squares regression predicting the variances and covariances from the half vectorized relatedness matrices. The variance components then become the estimated regression weights[7]. This estimation method has been known for the classical twin model for quite some time, but here it is generalized to (1) the case of covariances rather than correlations, and more importantly (2) the case of arbitrarily many relatedness matrices, each of any structure desired. The method could be conceived as a version of DeFries-Fulker (DeFries & Fulker, 1985) analysis applied to summary statistics and extended to arbitrary family designs. Alternatively, such a method similarly extends "multiple abstract variance analysis" (MAVA; Cattell, 1960; Loehlin, 1965) and the detailed classical formulas of Falconer and MacKay (1995). At minimum, the variances estimated from this regression equation could be used as starting values for a maximum likelihood estimation. A further use occurs when only summary statistics are available or permissible to share.

A direct consequence of the regression method for fitting variance component models is using regression-based multicollinearity diagnostics to gain more detailed information about the identification of the model. A model may be identified but the relatedness matrices are highly correlated which leads to poor precision for the variance estimates. The tolerance and the variance inflation factor (Cohen et al., 2003, p. 423) are standard regression metrics for multicollinearity that could easily shed light on the ability to distinguish variance components. Even when the model is identified, some of its variance components might be

[7]If one applies this regression method it is important to *not* omit the all zero rows unless the half vectorized covariance also omits these rows.

correlated strongly enough to cause estimation problems no matter what model estimation method is used.

Finally, the identification of a model should not be confused with its statistical power. Model identification is necessary for power but not sufficient. That is, the power for a non-identified variance component is necessarily zero, but not all identified variance components have equal power. Once a novel variance component is identified, researchers should apply the methods of van der Sluis, Dolan, Neale, and Posthuma (2007) and Wu and Neale (2012) to determine what design factors influence the power to recover that component. With a computationally easy and analytic solution to model identification for the variance component models that frequently arise in behavior genetics, researchers are more free to invent their own novel components.

Acknowledgments

Author Note

This work was partially supported by the NIH grant 1 R01 HD087395.

Appendix

Helpful R Functions for Determining Identification

The below code defines several useful functions for identification and fitting of model following the structure of Equation 1. These R (R Development Core Team, 2020) functions provide some basic error checking and some minimal roxygen2 (Wickham, Danenberg, & Eugster, 2015) documentation.

```
require(Matrix)
##' Determine if a variance components model is identified
##'
##' @param … Comma-separated relatedness component matrices.
##' @param silent logical. Whether to print messages about identification.
##'
##' @details
##' Returns of list of length 2. The first element is a single logical value:
##' TRUE if the model is identified, FALSE otherwise. The second list element
##' is the vector of non-identified parameters. For instance, a model might
##' have 5 components with 3 of them identified and 2 of them not. The second
##' list element will give the names of the components that are not
##' simultaneously identified.
identifyComponentModel <- function(…, silent=FALSE){
dots <- list(…)
nam <- names(dots)
if(is.null(nam)){
nam <- paste0('Comp', 1:length(dots))
}
```

```
compl <- lapply(dots, comp2vech, include.zeros=TRUE)
compm <- do.call(cbind, compl)
rank <- qr(compm)$rank
if(rank != length(dots)){
if(!silent) cat("Component␣model␣is␣not␣identified.\n")
jacOC <- Null(t(compm))
nidp <- nam[apply(jacOC, 1, function(x){sum(x^2)}) > 1e-17]
if(!silent) {
cat("Non-identified␣parameters␣are␣",
paste(nidp, collapse=",␣"), "\n")
}
return(list(identified=FALSE, nidp=nidp))
} else{
if(!silent) cat("Component␣model␣is␣identified.\n")
return(list(identified=TRUE, nidp=character(0)))
}
}
##' Fit the estimated variance components of a model to covariance data
##'
##' @param covmat the covariance matrix of the raw data, possibly blockwise.
##' @param … Comma-separated relatedness component matrices.
##'
##' @details
##' Returns a regression (linear model fitted with \code{lm}).
##' The coefficients of the regression are the estimated variance components.
fitComponentModel <- function(covmat, …){
dots <- list(…)
compl <- lapply(dots, comp2vech, include.zeros=TRUE)
compm <- do.call(cbind, compl)
rank <- qr(compm)$rank
y <- comp2vech(covmat, include.zeros=TRUE)
if(rank != length(dots)){
msg <- paste("Variance␣components␣are␣not␣all␣identified.",
"Try␣identifyComponentModel().")
stop(msg)
}
if(rank > length(y)){
msg <- paste0("Trying␣to␣estimate␣",
rank, "␣variance␣components␣when␣at␣most␣", length(y),
"␣are␣possible␣with␣the␣data␣given.\n")
warning(msg)
}
lm(y ~ 0 + compm)
}
```

```
##' Create the half-vectorization of a matrix
##'
##' @param x a matrix, the half-vectorization of which is desired
##'
##' @details
##' Returns the vector of the lower triangle of a matrix, including the
diagonal.
##' The upper triangle is ignored with no checking that the provided matrix
##' is symmetric.
vech <- function(x){
x[lower.tri(x, diag=TRUE)]
}
##' Turn a variance component relatedness matrix into its half-vectorization
##'
##' @param x relatedness component matrix
##' @param include.zeros logical. Whether to include all-zero rows.
##'
##' @details
##' This is a wrapper around the \code{vech} function for producing the
##' half-vectorization of a matrix. The extension here is to allow for
##' blockwise matrices.
comp2vech <- function(x, include.zeros=FALSE){
if(is.matrix(x)){
return(vech(x))
}else if(is.list(x)){
if(include.zeros){
return(vech(as.matrix(Matrix::bdiag(x))))
} else {
return(do.call(c, lapply(x, vech)))
}
} else {
msg <- paste("Can't make component into a half vectorization:",
"x is neither a list nor a matrix.")
stop(msg)
}
}
##' Compute the null space of a matrix
##'
##' @param M a matrix of which the null space is desired
##'
##' @details
##' The method uses the QR factorization to determine a basis for the null
##' space of a matrix. This is sometimes also called the orthogonal
##' complement of a matrix. As implemented, this function is identical
```

```
##' to the function of the same name in the MASS package.
Null <- function (M) {
tmp <- qr(M)
set <- if (tmp$rank == 0L) {
seq_len(ncol(M))
} else {
-seq_len(tmp$rank)
}
return(qr.Q(tmp, complete = TRUE)[, set, drop = FALSE])
}
```

References

Beasley WH, Rodgers JL, Bard D, Hunter MD, Garrison SM, & Meredith K. (2016). Nlsylinks: Utilities and kinship information for research with the nlsy [Computer software manual] Retrieved from https://CRAN.R-project.org/package=NlsyLinks (R package version 2.0.6)

Bekker PA (1986). A note on the identification of restricted factor loading matrices. Psychometrika, 51(4), 607–611. doi: 10.1007/bf02295600

Bekker PA, & ten Berge JM (1997). Generic global indentification in factor analysis. Linear Algebra and its Applications, 264, 255–263. doi: 10.1016/s0024-3795(96)00363-1

Bekker PA, & Wansbeek T. (2001). Identification in parametric models. In Baltagi BH (Ed.), A companion to theoretical econometrics (pp. 144–161). Blackwell Publishing Ltd. doi: 10.1002/9780470996249.ch8

Bollen KA, & Bauldry S. (2010). Model identification and computer algebra. Sociological Methods & Research, 39(2), 127–156. doi: 10.1177/0049124110366238 [PubMed: 21769158]

Cattell RB (1960). The multiple abstract variance analysis equations and solutions: For nature-nurture research on continuous variables. Psychological Review, 67(6), 353–372. doi: 10.1037/h0043487 [PubMed: 13691636]

Cohen J, Cohen P, West SG, & Aiken LS (2003). Applied multiple regression/correlation analysis for the behavorial sciences (3rd ed.). Routledge.

Coventry WL, & Keller MC (2005). Estimating the extent of parameter bias in the classical twin design: A comparison of parameter estimates from extended twin-family and classical twin designs. Twin Research and Human Genetics, 8(3), 214–223. doi: 10.1375/twin.8.3.214 [PubMed: 15989749]

DeFries JC, & Fulker DW (1985). Multiple regression analysis of twin data. Behavior genetics, 15(5), 467–473. doi: 10.1007/bf01066239 [PubMed: 4074272]

Falconer DS, & MacKay TFC (1995). Introduction to quantitative genetics. Pearson Education Limited.

Garrison SM, Hunter MD, Burt SA, & Trattner JD (2020). BGmisc: Behavior genetic modeling functions [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=BGmisc (R package version 0.1)

Garrison SM, O'Keefe P, Hunter MD, Beasley WH, Bard DE, & Rodgers JL (2018). AC'RE model: Estimating rearing effects without twins raised apart. In Behavior genetics (Vol. 48, pp. 472–472).

Hunter MD, McKee KL, & Turkheimer E. (2020). Simulated nonlinear genetic and environmental dynamics of complex traits. Unpublished manuscript.

Keller MC, & Coventry WL (2005). Quantifying and addressing parameter indeterminacy in the classical twin design. Twin Research and Human Genetics, 8(3), 201–213. doi: 10.1375/twin.8.3.201 [PubMed: 15989748]

Kirkpatrick RM, Pritikin JN, Hunter MD, & Neale NC (2021). Combining structural-equation modeling with genomic-relatedness-matrix restricted maximum likelihood in OpenMx. Behavior Genetics. doi: 10.1007/s10519-020-10037-5

Lay DC (2003). Linear algebra and its applications (Third ed.). Boston, MA: Addison Wesley.

Leon SJ (2006). Linear algebra with applications (Seventh ed.). Upper Saddle River, NJ: Prentice Hall.

Loehlin JC (1965). Some methodological problems in cattell's multiple abstract variance analysis. Psychological Review, 72(2), 156–161. doi: 10.1037/h0021706 [PubMed: 14282673]

Neale MC, Hunter MD, Pritikin JN, Zahery M, Brick TR, Kirkpatrick RM, … Boker SM (2016). OpenMx 2.0: Extended structural equation and statistical modeling. Psychometrika, 80(2), 535–549. doi: 10.1007/s11336-014-9435-8

Plomin R. (2011). Commentary: Why are children in the same family so different? non-shared environment three decades later. International Journal of Epidemiology, 40(3), 582–592. doi: 10.1093/ije/dyq144 [PubMed: 21807643]

Plomin R. (2014). Genotype-environment correlation in the era of DNA. Behavior Genetics, 44(6), 629–638. doi: 10.1007/s10519-014-9673-7 [PubMed: 25195166]

R Development Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org (ISBN 3–900051-07–0)

Rodgers JL, Bard DE, Johnson A, D'Onofrio B, & Miller WB (2008). The cross-generational mother–daughter–aunt–niece design: Establishing validity of the MDAN design with NLSY fertility variables. Behavior Genetics, 38(6), 567–578. doi: 10.1007/s10519-008-9225-0 [PubMed: 18825497]

Rodgers JL, Beasley WH, Bard DE, Meredith KM, Hunter MD, Johnson AB, … Rowe DC (2016). The NLSY kinship links: Using the NLSY79 and NLSY-children data to conduct genetically-informed and family-oriented research. Behavior Genetics. doi: 10.1007/s10519-016-9785-3

Sauce B, & Matzel LD (2018). The paradox of intelligence: Heritability and malleability coexist in hidden gene-environment interplay. Psychological Bulletin, 144(1), 26–47. doi: 10.1037/bul0000131 [PubMed: 29083200]

Scharf LL (1991). Statistical signal processing: Detection, estimation, and time series analysis. Reading, MA: Addison Wesley.

Turkheimer E. (2011). Still missing. Research in Human Development, 8(3–4), 227–241. doi: 10.1080/15427609.2011.625321

Turkheimer E, & Waldron M. (2000). Nonshared environment: A theoretical, methodological, and quantitative review. Psychological Bulletin, 126(1), 78–108. doi: 10.1037/0033-2909.126.1.78 [PubMed: 10668351]

van der Sluis S, Dolan CV, Neale MC, & Posthuma D. (2007). Power calculations using exact data simulation: A useful tool for genetic study designs. Behavior Genetics, 38(2), 202–211. doi: 10.1007/s10519-007-9184-x [PubMed: 18080738]

Vinkhuyzen AA, Wray NR, Yang J, Goddard ME, & Visscher PM (2013). Estimation and partition of heritability in human populations using whole-genome analysis methods. Annual Review of Genetics, 47(1), 75–95. doi: 10.1146/annurev-genet-111212-133258

Wickham H, Danenberg P, & Eugster M. (2015). roxygen2: In-source documentation for r [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=roxygen2 (R package version 5.0.1)

Wu H, & Neale MC (2012). On the likelihood ratio tests in bivariate ACDE models. Psychometrika, 78(3), 441–463. doi: 10.1007/s11336-012-9304-2 [PubMed: 25106394]

Yang J, Lee SH, Goddard ME, & Visscher PM (2011). GCTA: A tool for genome-wide complex trait analysis. The American Journal of Human Genetics, 88(1), 76–82. doi: 10.1016/j.ajhg.2010.11.011 [PubMed: 21167468]

Zhou B, Konstorum A, Duong T, Tieu KH, Wells WM, Brown GG, … Shahbaba B. (2012). A hierarchical modeling approach to data analysis and study design in a multi-site experimental fmri study. Psychometrika, 78(2), 260–278. doi: 10.1007/s11336-012-9298-9 [PubMed: 25107616]