Meta-Psychology, 2024, vol 8, MP.2018.1479
https://doi.org/10.15626/MP.2018.1479
Article type: Original Article

Open data: Not Applicable
Open materials: Yes
Open and reproducible analysis: Yes
Open reviews and editorial process: Yes
Preregistration: No

Edited by: Marcel van Assen
Reviewed by: Inga Schwabe, André Beauducel
Analysis reproduced by: Lucija Batinović
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/

How Close to the Mark Might Published Heritability Estimates Be?

Michael Maraun[1], Moritz Heene[2], and Philipp Sckopke[2]
[1]Department of Psychology, Simon Fraser University, Burnaby, B.C., Canada
[2]Department of Psychology, Ludwig Maximilians Universität, Munich, Germany

The behavioural scientist who requires an estimate of narrow heritability, h^2, will conduct a twin study, and input the resulting estimated covariance matrices into a particular mode of estimation, the latter derived under supposition of the standard biometric model (SBM). It is known that the standard biometric model can be expected to misrepresent the phenotypic (genetic) architecture of human traits. The impact of this misrepresentation on the accuracy of h^2 estimation is unknown. We aimed to shed some light on this general issue, by undertaking three simulation studies. In each, we investigated the parameter recovery performance of five modes- Falconer's coefficient and the SEM models, ACDE, ADE, ACE, and AE- when they encountered a constructed, non-SBM, architecture, under a particular informational input. In study 1, the architecture was single-locus with dominance effects and genetic-environment covariance, and the input was a set of population covariance matrices yielded under the four twin designs, monozygotic-reared together, monozygotic-reared apart, dizygotic-reared together, and dizygotic-reared apart; in study 2, the architecture was identical to that of study 1, but the informational input was monozygotic-reared together and dizygotic-reared together; and in study 3, the architecture was multi-locus with dominance effects, genetic-environment covariance, and epistatic interactions. The informational input was the same as in study 1. The results suggest that conclusions regarding the coverage of h^2 must be drawn conditional on a) the general class of generating architecture in play; b) specifics of the architecture's parametric instantiations; c) the informational input into a mode of estimation; and d) the particular mode of estimation employed. The results showed that the more complicated the generating architecture, the poorer a mode's h^2 recovery performance. Random forest analyses furthermore revealed that, depending on the genetic architecture, h^2, the dominance and locus additive parameter, and proportions of alleles were involved in complex interaction effects impacting on h^2 parameter recovery performance of a mode of estimation. Data and materials: https://osf.io/aq9sx/

Keywords: Heritability, heritability estimation, standard biometric model, quantitative genetics, structural equation modeling, random forest

How Close to the Mark Might Published Heritability Estimates Be?

The scientific aim of estimating the relative contributions of environment and genetics to the formation of particular aspects of psychological phenotypes is by no means new to the behavioral sciences. A traditional line of approach enlists the tools of quantitative genetics and involves the estimation of narrow- and broad-sense heritability (h^2 and H^2, respectively). The technical, quantitative sense of heritability relates to the explanation of the variability of scores on a phenotype. In particular, h^2 is the maximum proportion of phenotypic variance that can be accounted for by a linear function of allele counts; H^2, the proportion of the phenotypic variance associated with variability in the genotypes of individuals, thus reflecting all the genetic parts of a population's phenotypic variance, including additive, dominant, and gene interactions (epistasis).

At least within the behavioral sciences, it is h^2 that is most commonly estimated. And, because the selection and crossing experiments which are the basis for heritability estimation in the case of plants and animals are not applicable to humans (see, e.g., Vitzthum, 2003), the standard approach to estimation is to conduct a twin study, the yield of which is two or more covariance matrices. These matrices serve as input into either a classical coefficient, such as that of Nichols (1965) or Falconer (1960), or a structural equation model (SEM). Due most directly, it would seem, to a growing and highly accessible instructional literature bearing on the topic (e.g., Heath et al., 1989; Neale and Maes, 2004;

Rijsdijk, 2002), the latter approach to estimation has come to be ascendant, with many recent publications featuring estimates of h^2 generated through the fitting of latent variable models (e.g., Afifi et al., 2010; Nikolas and Burt, 2010; Polderman et al., 2015 see Van Houtem et al., 2013 for an overview).

All of these approaches to the estimation of h^2, the classical coefficients and SEM-based alike, are founded on what is known as the Standard Biometric Model (SBM), also known as the standard twin design, the roots of which stretch back to the work of Fisher (1919) and S. Wright (1921). The SBM (see, e.g., Holzinger, 1929; Jinks and Fulker, 1970; Rijsdijk, 2002; Schönemann, 1997; Vitzthum, 2003) is a latent variable model that portrays the dependency of phenotype upon environment and genetics in terms of a linear equation and set of moment restrictions. The linear equation is $Z = A + D + C + E$, in which A, D, C, and E are latent variables, to which the labels additive genetic variable (A), dominance genetic variable (D), environmental impact variable (C), and non-shared environmental influences (E) are commonly attached. The moment restrictions assert that all latent variables have expectations of zero and are pairwise uncorrelated. When the SBM is employed in the service of h^2 estimation, the dominance term is standardly omitted (Schönemann, 1997; Vitzthum, 2003).

Recent work of both an empirical and theoretical nature suggests strongly that the SBM can be expected to be a very poor description of the vast majority of extant phenotypic (genetic) architectures. In the first place, the SBM omits factors that are now known to be important features of many phenotypic architectures, among these, gene-environment correlation (Bowles, Gintis, et al., 2001; Bronfenbrenner, 1999; Kempthorne, 1978; Shalizi, 2007), maternal effects representing the impact of the uterine and perinatal environment, and genetic, multilocus, interaction (*epistasis*; Wei et al., 2014; Zuk et al., 2012). As Vitzthum (2003, p.541) has summarized, "Assumptions of the model include no dominance (if estimating h^2), no epistasis, no assortative mating, no genotype-environment interaction, and no genotype-environment covariance..." In the second place, what the SBM asserts about phenotypic architecture comes down to a linear model involving unmeasured, latent, variables. The variables are pinned down by spartan moment restrictions, raising the question of how much of an impact the omission of the above-mentioned effects may have on model error. As an aside, and on the methodological side of this issue, applied researchers may also consider, for example, Monte Carlo Markov Chain models which do not rely on the strict assumptions of the standard biometric model. Such models would enable them to simultaneously estimate also gene-environment interactions and gene-environment correlations (e.g., Eaves and Erkanli, 2003; He et al., 2016) to gain further insights into the genetic architecture of complex human traits.

The fitting of incorrect models to data can serve the scientific enterprise in manifold ways. However, because the scientific value inherent to an estimate of a fundamental quantity such as h^2 derives chiefly from its accuracy, it is by no means self-evident that the SBM's misrepresentation of phenotypic (genetic) architecture is benign in the context of its role as quantitative foundation for the estimation of h^2. At the least, it seems reasonable to inquire as to the degree to which this misrepresentation can be expected to impact upon the accuracy of h^2 estimates. The misestimation of h^2 under presumption of the SBM is discussed as a chief contributor to the problem of phantom heritability[1] (Wei et al., 2014; Zuk et al., 2012).

The central issue of concern in the present work is the accuracy of heritability estimates generated under presumption of the SBM. From a slightly different angle, we are interested in whether the estimates of narrow-sense heritability found in the pages of journal articles can be trusted as, on the whole, reflecting empirical reality. With the aim of shedding some light on this issue, we undertook three population-based simulation studies. In each, we constructed, quantitatively, a reasonable, non-SBM, architecture, on the basis of which we then deduced population (twin) covariance matrices. Each realization of the parameter vector associated with the architecture yields both a numerical realization of each of the matrices and a known numerical value of h^2. The matrices serve as input into a set of modes of estimation, each of which returns an "estimate" (or, more accurately, recovery) of h^2. By running the parameter vector over a selection of all possible of its realizations, we were able to produce, for each mode, an approximation to the model error distribution (i.e., the distribution of differences between true and recovered h^2). Side issues on which we hoped to throw some light included: (a) the *relative* performances of commonly employed modes of h^2 estimation; and (b), relatedly, the issue of whether the newer SEM-based modes outperform the classical correlation-based.

We emphasize that, because *population* covariance matrices were offered as input to each mode under consideration, our results are free of the obscuring effects of sampling error. Accordingly, our investigation must

[1] i.e., the phenomenon wherein the variants of a trait identified by genome-wide association studies explain only a fraction of the narrow-sense heritability indicated as inhering in the trait by conventional, SBM-based, estimates of h^2

be distinguished from prior investigations into statistical aspects of heritability estimation (e.g., Keller and Coventry, 2005; Keller et al., 2010). Our interest was in the h^2 *parameter recovery performance* of modes when these latter are offered information unadulterated by sampling error and encounter certain non-SBM architectures which can be expected to underlie empirical estimation contexts.

Thus, although our investigation was, indeed, a simulation study, it was not a conventional one, in that we did not sample individuals but, rather, numerical instantiations of particular non-SBM architectures. Hence, our study aims to investigate *model error* due to an incomplete knowledge of the actual data-generating model. We emphasize, furthermore, that the work described herein has little overlap with general critiques of heritability estimation (e.g., Schönemann, 1997), nor with accounts of the pervasive misinterpretation and misuse of the concept of heritability (e.g., Vitzthum, 2003). It is our belief that the nearest precedent to the work herein undertaken is the study on phantom heritability by Zuk et al. (2012).

Our study is thus partly motivated by the discussion about possible biasing effects of substantial non-additive genetic variation on h^2 estimation, either due to dominance (i.e., non-additivity of allelic effects within loci) or epistasis (i.e., interactions of effects among loci), or both (e.g., Eichler et al., 2010; Zuk et al., 2012). It must be noted, though, that current evidence suggests that the role of dominance or epistatic effects in human traits may be small, and most of twin data are in fact consistent with an additive polygenic genetic model, meaning that multiple genes influence the phenotypic expression of a trait. The meta-analysis by Polderman et al. (2015) found that 69% of investigated traits showed that the "... pattern of twin correlations is consistent with a (...) model of the absence of environmental effects shared by twin pairs and the presence of genetic effects that are entirely due to additive genetic variation". Thus, shared environmental as well as dominance or epistatic effects in human traits seem to play a smaller role than hitherto expected (e.g., Visscher et al., 2017 for an overview). Similarly, Zhu et al., 2015 on 79 bio-physiological human traits found little dominance genetic variance, with an average estimate of .03 and a maximum value of .19. On the other hand, things might be more complicated because the impact of non-additive genetic variation in twin studies can go under the radar of twin-study analyses. This is because dominant genetic variation might be masked by shared environment in twin studies (Chen et al., 2015), leading to h^2 estimates that are upward-biased to an unknown extent (e.g., Chen et al., 2015; Zuk et al., 2012). In addition to that, epistatic interactions can also positively bias h^2 estimates by generating real additive variation as marginal effects from higher-order genetic interactions (D. M. Evans et al., 2002; W. G. Hill et al., 2008; Marchini et al., 2005).

While it is relatively easy to raise objections against twin-based h^2 estimates based on these sources, it is harder to know to what extent non-additive effects impact upon h^2 parameter recovery performance in the context of other possibly biasing genetic parameters. Determining the extent of biasing effects of genetic non-additivity on the quality of estimation of h^2 while taking other biasing parameters into account is therefore one of the main aims of this article.

The organization of the paper is as follows: (a) In the first sections, we offer the reader necessary technical background to the problem; (b) following which, in the remainder of the paper, we address the central issue of h^2 parameter recovery. Therein, we present three simulation studies, each featuring an assessment of five modes—Falconer's coefficient and the ACDE, ADE, ACE, and AE structural equation models—under a particular choice of informational input and non-SBM architecture. The first study featured a single locus architecture with the non-SBM characteristics of dominance effects and genetic-environment correlation. The informational input was the set of covariance matrices yielded under the four twin designs: monozygotic-reared together, monozygotic-reared apart, dizygotic-reared together, and dizygotic-reared apart, denoted as $\Sigma_{MZ,T}(\theta)$, $\Sigma_{MZ,A}(\theta)$, $\Sigma_{DZ,T}(\theta)$, $\Sigma_{DZ,A}(\theta)$. The second study was identical to the first, except that the informational input was reduced to only the first two of these covariance matrices.

The third study featured a multi-locus, limiting process, architecture (Zuk et al., 2012), which, in addition to dominance effects and genetic-environment correlation, had the non-SBM property of epistasis (i.e., genetic locus-locus interaction). In particular, genome-wide association studies undertaken during the past ten years have yielded strong indications that human traits are, as a rule, multilocus or complex (Benyamin et al., 2014; D. M. Evans, 2011; Hsu, 2014; Wei et al., 2014). Though the detection of epistasis poses significant technical challenges (Wei et al., 2014), there exists now evidence that it is operative in the expression of many human phenotypes, including Alzheimer's disease (Hohman et al., 2013), multiple sclerosis (Gregersen et al., 2006), ankylosing spondylitis (D. M. Evans, 2011), breast cancer (Ritchie et al., 2001), and psoriasis (Strange et al., 2010). As noted by Zuk et al. (2012), "Quantitative geneticists have long known that genetic interactions can affect heritability calculations...". In

this final study, the informational input was identical to that of the first.

It was neither our aim nor was it necessary for the success of our investigation to construct architectures that were, in some sense, empirically *true*. In fact, for the vast majority of psychological phenotypes, little is known about the associated engendering architecture. Especially in light of manifold findings suggesting of the complex genetic architecture of certain human traits, all that was required was that a constructed architecture was *possible*. It should nevertheless be stressed that there has been progress in recent years in determining the genetic architecture of psychological traits with studies having used more complex extended twin family designs (Fedko et al., 2021; Schwabe et al., 2017), genome-wide association studies (W. D. Hill et al., 2019; Sniekers et al., 2017), or more sophisticated methodological tools (L. M. Evans et al., 2018; Feng et al., 2020; Grotzinger et al., 2019). That being said, our aim was simply to gain insight into how commonly employed modes of h^2 estimation—each derived under the supposition of the SBM—can be expected to perform when they encounter architectures that depart from the SBM in ways deemed *possible* by extant scientific theory and research.

Background and Theory

In this section, we offer the reader essential technical background to the problem, organized around a careful disambiguation of the foundational concepts of *genetic architecture*, *Fisherian decomposition*, *population variance components*, and the *standard biometric model*. Readers interested only in the results may directly consult the part of the article devoted to the simulation study.

Genetic Architecture and Phenotype

Let it be that:

1. Z is a quantitative or disease trait (*phenotype*)[2]
2. Each individual i belonging to a population P of humans has a score on Z
3. For all $i \in P$, Z_i depends genetically on n biallelic loci.

Let the alleles at locus 1 be $\{A_1, A_2\}$, at locus 2, $\{B_1, B_2\}$, at locus 3, $\{C_1, C_2\}$, etc. It follows, then, that at each of the n loci, three locus-specific genotypes are defined (e.g., at locus 1, $\{A_1A_1, A_1A_2, A_2A_2\}$); and overall, 3^n genotypes with respect to Z[3]. Define $g_j, j = 1, ..., n$, to be the gene content at locus j; i.e., the number of copies of the second allele $(A_2, B_2, C_2, ...)$ present in a locus j-specific genotype. For the locus 1 specific genotypes A_1A_1, A_1A_2, and A_2A_2, for example, g_1 is equal to 0, 1, and 2, respectively. Clearly, each of the 3^n genotypes is uniquely associated with a value of the n-vector $\mathbf{g}$, the jth element of which is g_j. We define the genetic architecture to be the scalar function $\Psi'(\mathbf{g})$ which maps each of the 3^n genotypes (equivalently, 3^n distinct values $\mathbf{g}^*$ of $\mathbf{g}$) into a genotypic value; i.e., $\Psi'(\mathbf{g}^*) : \mathbf{g}^* \subset \mathbf{R}^n \rightarrow \mathbf{R}$ (Lynch & Walsh, 1998). The genetic architecture is said to be single locus if $n = 1$, and multilocus otherwise.

The phenotypic value is the (individual level) scalar function:

$$Z_i = \Psi'(\boldsymbol{g}_i) + \gamma(E)_i \tag{1}$$

wherein $\gamma(E)_i = Z_i - \Psi'(\boldsymbol{g}_i)$ is a residual representing the aggregate impact upon Z of all effects—both main and interaction—involving the environment[4].

Because each $i \in P$ has a set of values $\{\mathbf{g}_i, \gamma(E)\}$, the induced distribution of Z in P is determined by the joint distribution of $\{\mathbf{g}, \gamma(E)\}$.

Fisherian Decomposition of $\Psi'(\mathrm{g})$ and Population Variance Components

At the root of the quantity h^2 is the variance component σ^2_{A} (and all other genetic variance components) of H^2. This variance component, called the *additive variance*, is defined on the basis of the *Fisherian decomposition* of $\Psi'(g)$. In the single locus case, the Fisherian decomposition is

$$\Psi'(g_1) = \Psi'(g_1)_{\mathrm{lin}} + \delta(g_1) \tag{2}$$

in which $\Psi'(g_1)_{\mathrm{lin}}$ - the linear predictor of genotypic value on the basis of gene content - is called the *additive component*, and $\delta(g_1)$ - the residual, and also, because there are but three genotypic values, the quadratic fit to these values - the *dominance component*. The additive and dominance components can be expressed as

$$\Psi'(g_1)_{\mathrm{lin}} = [a + \beta g_1] \tag{3}$$

and

$$\delta(g_1) = \Psi'(g_1) - \Psi'(g_1)_{\mathrm{lin}} \tag{4}$$

[2] If Z is a disease trait, then it is treated as a [0,1] dichotomous variable given rise to by dichotomization, at τ, of an underlying quantitative trait, or "liability", Z^*; i.e., if $Z^* \geq \tau$, then $Z = 1$; else, $Z = 0$ (Zuk et al., 2012).

[3] If $n = 2$, e.g., the nine genotypes are: $\{A_1A_1B_1B_1$, $A_1A_1B_1B_2, A_1A_1B_2B_2, A_1A_2B_1B_1, A_1A_2B_1B_2$, $A_1A_2B_2B_2, A_2A_2B_1B_1, A_2A_2B_1B_2, A_2A_2B_2B_2\}$

[4] Genotypic values are taken, variously, to be either theoretical values or $E(Z|\mathbf{g})$ (Lynch & Walsh, 1998). In the latter case, the function $\Psi'(.)$ is, then, simply the conditional mean function, in which case $E\gamma(E) = 0$.

respectively, in which $\beta = \frac{\sigma_{\Psi'(g_1),g_1}}{\sigma^2_{g_1}}$ and $\alpha = \mu_{\Psi'} - \beta\mu_{g_1}$.

Because $\Psi'(g_1)_{\text{lin}}$ and $\delta(g_1)$ are orthogonal by construction, equation (2) implies that the population genetic variance can be decomposed as follows:

$$\mathrm{V}\left(\Psi'\left(g_1\right)\right) = \mathrm{V}\left(\Psi'\left(g_1\right)_{\text{lin}}\right) + \mathrm{V}\left(\delta\left(g_1\right)\right) = \sigma^2_{\mathrm{A}} + \sigma^2_{\mathrm{D}'} \quad (5)$$

wherein σ^2_{D} is called the dominance variance component. Symbolize $\mathrm{P}(\mathrm{A}_2)$ - the proportion of alleles in P of type A_2 - as φ, and let it be the case that the probability of occurrence of each of the three genotypes is governed by the Hardy-Weinberg principle (see Lynch and Walsh, 1998), in which case $\mathrm{P}(\mathrm{A}_1\ \mathrm{A}_1) = (1-\varphi)^2$, $\mathrm{P}(\mathrm{A}_1\ \mathrm{A}_2) = 2(1-\varphi)\varphi$, and $\mathrm{P}(\mathrm{A}_2\ \mathrm{A}_2) = \varphi^2$. Then

$$\sigma^2_{\mathrm{A}} = \beta^2\sigma^2_{g_1} = \frac{\sigma^2_{\Psi'(g_1),g_1}}{\sigma^2_{g_1}} = \frac{\sigma^2_{\Psi'(g_1),g_1}}{2\varphi(1-\varphi)} \quad (6)$$

and

$$\sigma^2_{\mathrm{D}} = \mathrm{V}\left(\Psi'\left(g_1\right)\right) - \sigma^2_{\mathrm{A}} \quad (7)$$

We can parameterize the genotypic values as follows: $\Psi'(g_1 = 0) = 0$, $\Psi'(g_1 = 1) = (k+1)a$, and $\Psi'(g_1 = 2) = 2a$ (see Lynch and Walsh, 1998, p. 62). Clearly, parameter a controls the linear rate of change of $\Psi'(g_1)$; parameter k, the degree of nonlinearity of $\Psi'(g_1)$. Specifically, if $k = 0$, then $\Psi'(g_1)$ is a linear function of gene content; else, it is a quadratic function, the genotypic value for $\mathrm{A}_1\ \mathrm{A}_2$ $(g_1 = 1)$ lying nonequidistant between the genotypic values for $\mathrm{A}_1\ \mathrm{A}_1$ $(g_1 = 0)$ and $\mathrm{A}_2\ \mathrm{A}_2$ $(g_1 = 2)$ (see Lynch and Walsh, 1998). Under the $\{a, k\}$ parameterization (Lynch & Walsh, 1998),

$$\sigma^2_{\mathrm{A}} = 2\varphi(1-\varphi)[a(1+k(1-2\varphi))]^2 \quad (8)$$

and

$$\sigma^2_{\mathrm{D}} = (2\varphi(1-\varphi)ak)^2 \quad (9)$$

In the multilocus case, the Fisherian decomposition is

$$\Psi'(\mathbf{g}) = \mu\Psi'(\mathbf{g}) + \sum_{j=1}^{n}\mathrm{B}\left(g_j\right) + \sum_{j=1}^{n}\delta\left(g_j\right) + \sum_{l=1}^{m}\mathrm{I}_{(\mathbf{g})l'} \quad (10)$$

in which: $\mathrm{B}\left(g_j\right) = \Psi'(\mathbf{g})_{\text{lin}j} - \mu_{\Psi'(g)}$; $\Psi'(\mathbf{g})_{\text{lin}j} = \mathrm{a}_j + \beta_j g_j$, is the linear predictor of $\Psi'(\mathbf{g})$ on the basis of g_j; $\delta\left(g_j\right) = \Psi'(\mathbf{g}) - \Psi'(\mathbf{g})_{\text{lin}j}$; and the $\mathrm{I}_{(\mathbf{g})l}$ are $m = \sum_{r=2}^{n}\begin{pmatrix} n \\ r \end{pmatrix} = (2^n - n - 1)$ *epistatic* (locus-locus) *interactions*, there being $\begin{pmatrix} n \\ r \end{pmatrix}$ interactions of the rth order, $r = 2..n$. Because the latter three components are orthogonal by construction, equation (10) implies that

$$\mathrm{V}\left(\Psi'(\mathbf{g})\right) = \sigma^2_{\mathrm{A}} + \sigma^2_{\mathrm{D}} + \sigma^2_{\mathrm{EP}} \quad (11)$$

wherein σ^2_{EP} is the variance due to epistasis.

We note, for later use, the fact that a vector of parameters $\boldsymbol{\theta}_{\text{gen}}$ can be defined which determines the magnitudes of all genetic variance components – $\sigma^2_{\mathrm{A}'}$, $\sigma^2_{\mathrm{D}'}$, etc. defined under a particular architecture. The number (and identities) of the elements of $\boldsymbol{\theta}_{\text{gen}}$ will depend upon the particular architecture in question. Thus, for example, in a single locus architecture, the 3-element vector $\begin{pmatrix} \varphi \\ a \\ k \end{pmatrix}$ is one option for $\boldsymbol{\theta}_{\text{gen}}$, in that the three parameters it contains determine the two variance components σ^2_{A} and σ^2_{D}.

The Quantities h^2, H^2, and σ^2_{Z}

By definition, narrow- and broad-sense heritability are, respectively,

$$h^2 = \frac{\sigma^2_{\mathrm{A}}}{\sigma^2_{\mathrm{Z}}} \quad (12)$$

and

$$H^2 = \frac{\mathrm{V}\left(\Psi'(\mathbf{g})\right)}{\sigma^2_{\mathrm{Z}}} \quad (13)$$

From (1), it follows that

$$\sigma^2_{\mathrm{Z}} = \mathrm{V}\left(\Psi'(\mathbf{g})\right) + \mathrm{V}(\gamma(E)) + 2\sigma_{\Psi'(\mathbf{g})\gamma(\mathrm{E})} \quad (14)$$

Symbolizing $\mathrm{V}(\gamma(\mathrm{E}))$ as $\sigma^2_{\mathrm{E}'}$, expressions (5) and (14) imply that

$$\sigma^2_{\mathrm{Z}} = \sigma^2_{\mathrm{A}} + \sigma^2_{\mathrm{D}} + \sigma_{\mathrm{E}}\left(\sigma_{\mathrm{E}} + 2\rho_{\Psi'(g_1)_{\text{lin},\gamma(\mathrm{E})}}\sigma_{\mathrm{A}} + 2\rho_{\delta(g_1),\gamma(\mathrm{E})}\sigma_{\mathrm{D}}\right) \quad (15)$$

For the multilocus case, (11) and (14) imply that

$$\sigma^2_{\mathrm{Z}} = \sigma^2_{\mathrm{A}} + \sigma^2_{\mathrm{D}} + \sigma^2_{\mathrm{EP}} + \sigma_{\mathrm{E}}\left(\sigma_{\mathrm{E}} + 2\rho_{\sum_{j=1}^{n}B(g_j),\gamma(\mathrm{E})}\sigma_{\mathrm{A}} + 2\rho_{\sum_{j=1}^{n}\delta(g_j),\gamma(\mathrm{E})}\sigma_{\mathrm{D}} + 2\rho_{\sum_{l=1}^{m}I_{(g)l,\gamma(\mathrm{E})}}\sigma_{\mathrm{EP}}\right) \quad (16)$$

The rho parameters – $\rho_{\Psi'(g_1)_{\text{lin},\gamma(\mathrm{E})}}$, $\rho_{\delta(g_1),\gamma(\mathrm{E})}$, $\rho_{\sum_{j=1}^{n}B(g_j),\gamma(\mathrm{E})}$, $\rho_{\sum_{j=1}^{n}\delta(g_j),\gamma(\mathrm{E})}$, $\rho_{\sum_{l=1}^{m}I_{(g)l,\gamma(\mathrm{E})}}$ – are, of course, genetic-environment correlations.

We observe, for later use, that a vector of parameters $\boldsymbol{\theta}_{\mathrm{V}}$ can be defined such that $\left[\boldsymbol{\theta}_{\text{gen}} : \boldsymbol{\theta}_{\mathrm{V}}\right]$ determines the magnitude of δ^2_{Z}. The number (and identities)

of the elements of $\boldsymbol{\theta}_{\mathrm{V}}$ is dependent upon the particular architecture in question. In the case of a single locus architecture, one option for $\boldsymbol{\theta}_{\mathrm{V}}$ is the vector $\begin{pmatrix} \sigma^2_{\mathrm{E}} \\ \rho^2_{\Psi^r(g_1)_{\mathrm{lin}},\gamma(\mathrm{E})} \\ \rho_{\delta(g_1),\gamma(\mathrm{E})} \end{pmatrix}$. Note, finally, that, because σ^2_{Z} is determined by $\left[\boldsymbol{\theta}_{\mathrm{gen}} : \boldsymbol{\theta}_{\mathrm{V}}\right]$, so, too, is $h^2\left(H^2\right)$.

Twin Studies and the Covariance Matrix $\Sigma_{\mathrm{R,L}}(\theta)$

From an examination of expressions (15) and (16), it is evident that neither the parameter $\sigma^2_{\mathrm{A}}\left(\sigma^2_{\mathrm{A}'}\sigma^2_{\mathrm{D}'},\sigma^2_{\mathrm{EP}}\right)$, nor, consequently, $h^2\left(H^2\right)$, can be recovered on the basis of the information present in the distribution of Z. The researcher who requires an estimate of $h^2\left(H^2\right)$ must increase the amount of information that can be brought to bear on the problem, and the way he or she does this, is by collecting data under two or more twin designs. Let $(Z,Z')_{R,L}$ be the set of score-pairs of twins, collected under twin design $\{R,L\}$, in which $R \equiv \{$ MZ (monozygotic), DZ (dizygotic) $\}$ and $L \equiv \{$T (reared together),A (reared apart)$\}$. The yield from collecting data under two or more such designs is a set of sample covariance matrices, each matrix, $\hat{\Sigma}_{R,L'}$, an estimate of the corresponding population covariance matrix $\Sigma_{R,L} = \begin{pmatrix} \sigma^2_{Z} & \sigma_{Z,Z'}(R,L) \\ \sigma_{Z,Z'}(R,L) & \sigma^2_{Z} \end{pmatrix}$. This set of sample covariance matrices is the input into a chosen mode of estimation.

Because any particular $\Sigma_{R,L}$ is uniquely associated with a particular crossing of trait Z and population P, it is numerically engendered by the (unknown) architecture which underlies Z within P. In particular, the elements of any $\Sigma_{R,L}$ are a function of three parameter vectors: $\boldsymbol{\theta}_{\mathrm{gen}}, \boldsymbol{\theta}_{\mathrm{V}}$, and a vector $\boldsymbol{\theta}_{\mathrm{C}(R,L)}$ which, jointly with $\left[\boldsymbol{\theta}_{\mathrm{gen}} : \boldsymbol{\theta}_{\mathrm{V}}\right]$, yields the off-diagonal element of $\Sigma_{R,L'}$, i.e., $\sigma_{Z,Z'}(R,L)$. The number (and identities) of the elements of $\boldsymbol{\theta}_{C(R,L)}$ will, of course, depend upon the architecture in question. In the case of a single locus architecture (see Schönemann, 1989, 1997)

$$\sigma_{Z,Z'}(R,L) = \rho_{\mathrm{A}}(R)\sigma^2_{\mathrm{A}} + \rho_{\mathrm{D}}(R)\sigma^2_{\mathrm{D}} + \rho_{\mathrm{E}}(R,L)\sigma^2_{\mathrm{E}'} \qquad (17)$$

in which: $\rho_{\mathrm{A}}(R)[\rho_{\mathrm{D}}(R)]$ is the correlation over twin pairs having genetic relationship R between the $\Psi'(\mathrm{g}_1)$ lin $[\delta(\mathrm{g}_1)]$ components of Z and Z'; and $\rho_{\mathrm{E}}(R,L)$, the correlation over twin pairs having genetic relationship R and rearing relationship L - between the $\gamma(E)$ components of Z and Z'. Consequently, one choice of $\boldsymbol{\theta}_{C(R,L)}$ is the vector

$$\begin{pmatrix} \rho_{\mathrm{A}}(R) \\ \rho_{\mathrm{D}}(R) \\ \rho_{\mathrm{E}}(R,L) \end{pmatrix}$$

The Standard Biometric Model for Twin Data

When applied to data from a twin study, the standard biometric model assumes the form

$$\begin{pmatrix} Z \\ Z' \end{pmatrix} = \begin{pmatrix} A + D + C + E \\ A' + D' + C' + E' \end{pmatrix} \qquad (18)$$

in which: (A,A') are additive genetic variables; (D,D') are dominance genetic variables; (C,C') are variables representing the total impact, upon Z and Z', respectively, of all environmental influences to which the twin pair was exposed in common; and (E,E'), variables representing the total impact of all environmental influences to which the members were uniquely exposed. The standard moment restrictions stipulate that the variables: a) all have expectations of zero; and b) are pairwise uncorrelated under all combinations of R and L, save for the following: if $R = MZ$, $\rho(A,A') = \rho(D,D') = 1$; if $R = \mathrm{DZ}$, $\rho(A,A') = 0.5$, $\rho(D,D') = 0.25$; and if $L = \mathrm{T}$, $\rho(C,C') = 1$.

It should be emphasized that the moment restrictions have, to say the least, a tenuous relationship to extant scientific understanding. We note, in particular, that: a) (A,A') and (D,D') are latent variables and are in no way identical to the additive and dominance components yielded by the Fisherian decomposition, which are orthogonal by construction. b) The genetic correlations prescribed under the SBM are expectations deduced under particular, and contentious, assumptions. c) There is little, if any, theoretical basis for the SBM stipulated correlations involving (C,C') and (E,E'), and the questionability of these stipulations (that, e.g., $\rho(C,C') = 0$ for twins reared apart) is well-documented (see, e.g., Shalizi, 2007; Vitzthum, 2003).

Modes of Estimation

For the purposes of the present work, a mode of estimation was defined to be any strategy that takes as input a set of sample covariance matrices produced in one or more twin studies and returns a recovered value of h^2. As earlier noted, we distinguish between classical, correlation-based modes, such as the coefficients due to Falconer, $2\left(\hat{\rho}_{Z,Z|MZ,T} - \hat{\rho}_{Z,Z|DZ,T}\right)$, and Nichols, $\frac{2\left(\hat{\rho}_{Z,Z|MZ,T} - \hat{\rho}_{Z,Z|DZ,T}\right)}{\hat{\rho}_{Z,Z|MZ,T}}$, and the more recent strategies based on SEM.

The Study

Overview

To any fixed set $\{P, Z, \Psi'(\mathbf{g}), R, L\}$, there can be associated a parameter vector $\boldsymbol{\theta} \equiv \begin{pmatrix} \boldsymbol{\theta}_{\text{gen}} \\ \boldsymbol{\theta}_{\text{V}} \\ \boldsymbol{\theta}_{\text{C}(R,L)} \end{pmatrix}$. We express the dependency of $\Sigma_{R,L}$ and h^2 upon $\boldsymbol{\theta}$ as $\Sigma_{R,L}(\boldsymbol{\theta})$ and $h^2(\boldsymbol{\theta})$, respectively. As $\boldsymbol{\theta}$ moves over its parameter space, yielding novel realizations $\boldsymbol{\theta}^*$, all possible instantiations of $\{P, Z, \Psi'(\mathbf{g}), R, L\}$ are spelled out. For each, there is an associated pair of numerical realizations of $\Sigma_{R,L}(\boldsymbol{\theta})$ and $h^2(\boldsymbol{\theta})$, say, $\Sigma_{R,L}(\boldsymbol{\theta}^*)$ and $h^2(\boldsymbol{\theta}^*)$. Our approach is to examine the performance of particular modes in recovering $h^2(\boldsymbol{\theta}^*)$ when the informational input is a set of $\Sigma_{R,L}(\boldsymbol{\theta}^*)$. Let there be: a set $S \equiv \{m_1, m_2, .., m_t\}$ of t modes of estimation; a set $H(\boldsymbol{\theta}) \equiv \left\{\Sigma_{R,L}(\boldsymbol{\theta})_g, g = 1..f\right\}$ of input population covariance matrices, each matrix produced under one of f twin studies undertaken; and a set $R(H(\boldsymbol{\theta})) \equiv \left\{h_1^{2'}, h_2^{2'}, .., h_t^{2'}\right\}$ of t (recovered) narrow heritability values, where the sth $h_s^{2'} = m_s(H(\boldsymbol{\theta}))$ is yielded by the sth mode with informational input $H(\boldsymbol{\theta})$. To this end, we conducted three simulation studies. The same five modes—Falconer's coefficient and the four structural equation models, ACDE, ACE, ADE, and AE—were investigated under each. Studies 1 and 2 shared the same single locus architecture but differed in the informational input $H(\boldsymbol{\theta})$ offered to the modes. Study 3 featured a more complex multilocus architecture. With the aim of improving readability, we present the details of construction and results of each simulation study separately.

Data Generation

Data simulation and analysis were undertaken using the following R packages: data.table (Dowle et al., 2017), doMC (Revolution Analytics & Weston, 2022), dplyr (Wickham et al., 2023), foreach (Calaway et al., 2015), lavaan (Rosseel, 2012), psych (Revelle, 2016), ranger (M. N. Wright & Ziegler, 2017), and mlr (Bischl et al., 2017). The code and data sets can be downloaded from https://osf.io/aq9sx/.

Structural Equation Models (SEM)

Structural equation models (ACDE, ACE, ADE, and AE) were fit by lavaan (Rosseel, 2012) with the following restrictions: a) identification was achieved by fixing the variances of all latent variables to unity; b) correlations between latent variables were fixed following the standard biometric model (see (18) and the following); c) analogous factor loadings to be estimated (depending on the model: additive, dominance, common environment, and unique environment) were fixed to be equal within twin pairs and across twin designs. The four models are graphically depicted in Figure 1. Because the input into analyses consisted of sets of population covariance matrices, the sample size was set arbitrarily large to 1,000,000 within lavaan.

Following estimation of model parameters, recovered narrow heritability was calculated as $\hat{h} = \frac{\hat{a}^2}{\hat{\sigma}_Z^2}$, where $\hat{a}^2$ is the squared recovered additive factor loading, and $\hat{\sigma}_Z^2$ is the recovered model-implied variance of Z (Z').

Analysis

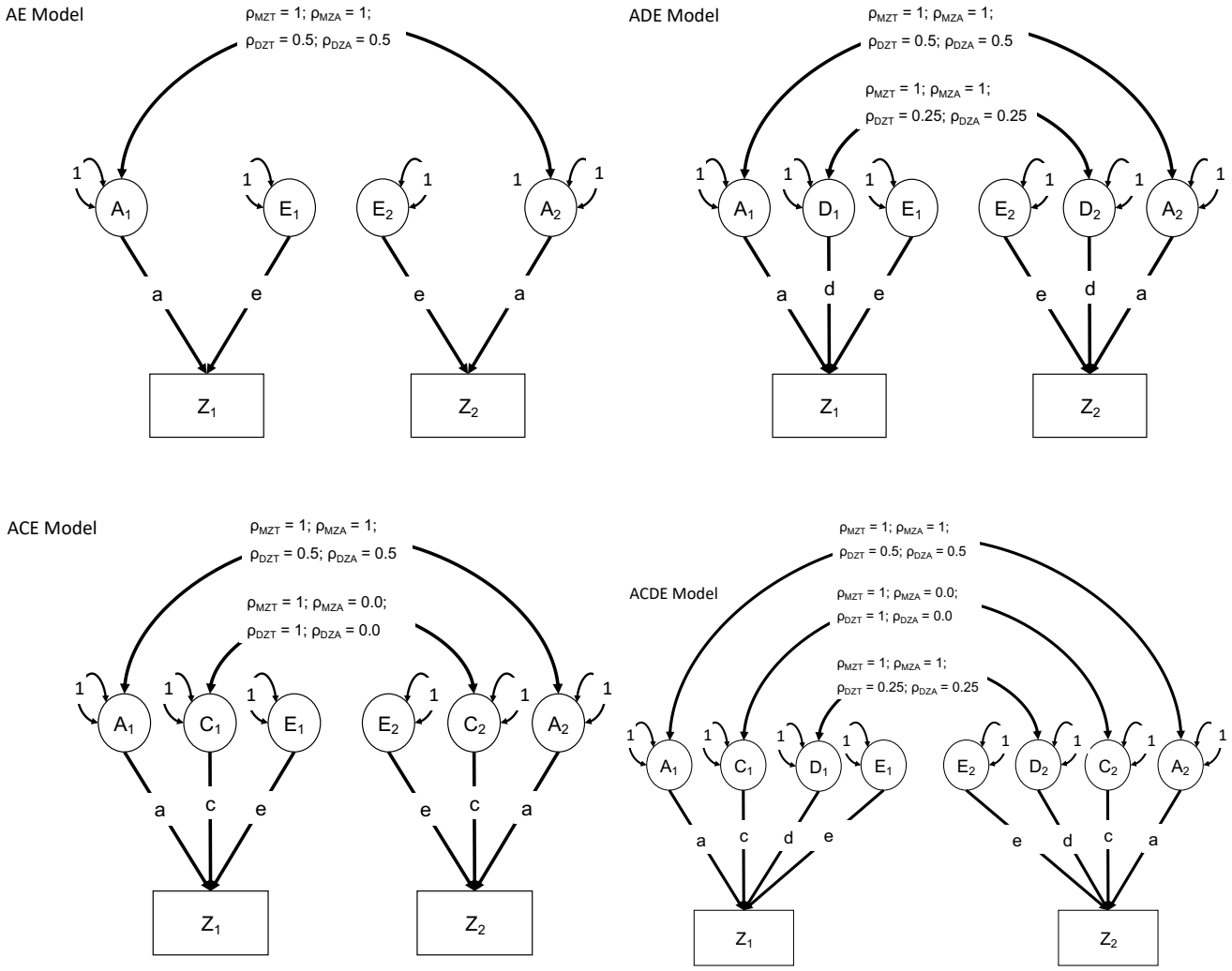
For each study and each of the 5 simulated model error distributions (one for each mode), we report the mean, standard deviation, minimum, and maximum. For the absolute model error $|h_r^{2'} - h^2(\theta)|$ distribution, we report the mean, standard deviation, and every tenth percentile point. Both our central and side issues can be adjudicated in manifold ways, and we found it informative to consider each from different angles. We employed the mean absolute model error as a criterion of overall recovery performance, indicating the expected closeness to the mark of the recovered values yielded by a given mode. The variance of absolute model error served as a criterion of the sensitivity of the parameter recovery performance of each mode to the specific features, which vary over θ^*, of the particular architecture in play. Finally, the minimum and maximum of the model error distribution served as an indication of how far off the mark recovered values delivered by a mode could potentially be. In other words, how severely could a mode misrepresent the true state of nature under the chosen genetic architectures? Just because researchers will, in general, be blind to the particulars of the engendering architectures that underlie their estimations of h^2, this feature of recovery performance seems, to us, a particularly important one. To ensure that the assessment of the parameter recovery performance of the SEM-based modes is not confounded with the issue of model fit, we undertook a sub-analysis in which the former issue is assessed based on data contributed only by those solutions for which the standardized root mean squared error of approximation (RMSEA) as the noncentrality parameter of tolerable model misfit in the population is less than .06[5] (i.e., instances in which the application of a particular model to a set of input covariance matrices resulted in a notably good data-model fit).

Finally, with the aim of providing some insight into

[5] see Hu and Bentler, 1999.

Figure 1

Graphical representations of the four structural equation models (ACDE, ACE, ADE, and AE).



the kinds of architectures - i.e., parametric instantiations of $\{P, Z, \Psi'(\mathbf{g}), R, L\}$ - under which a mode delivers its poorest h^2 parameter recovery performance, we undertook variable importance analyses, with model error as the dependent variable and the six factors $\left\{\varphi, a, k, \rho_{\Psi'(g_2)_{\text{lin}},\gamma(\text{E})}, \rho_{\delta(g_2),\gamma(\text{E})}, h^2\right\}$ as predictors. These analyses were carried out using the method of random forest regression trees (RFRT; Jones and Linder, 2015; Friedman, 2001, p. 587ff; Breiman, 2001). Each analysis was based on 1,000 trees, and variable importance (VIMP) measures calculated using the permutation accuracy importance mode (e.g., Strobl et al., 2008). The reader should keep in mind that a VIMP for a given predictor does not capture merely the additive effect (direct impact) of a given predictor but also the impacts upon the dependent variable of interaction effects involving the predictor and other predictors. A comparatively large VIMP, then, signifies a predictor possessing overall predictive ability or significance but does not imply anything about how this overall predictive efficacy is partitioned into the various effects (i.e., additive and/or interaction effects) in which the given predictor is involved.

We attempted to gain some insight into the issue of whether a relatively large VIMP is attributable to only additive (direct) effect or a blending of additive and interaction effects through computation of individual conditional expectation (ICE) plots, as suggested by Gold-

stein et al. (2015). This approach is preferable to that of the classical partial dependence plot (Friedman, 2001), which graphically represents the average partial relationship between the predicted values on the dependent variable and one or more predictors, conditional on the remaining predictors. Insofar as substantial interaction effects exist, and the partial individual response relationship is heterogeneous, this state of affairs will be reflected in a crossing of the ICE curves.

Simulation Study 1

Introduction

In this first simulation study, we assessed the parameter recovery performance of the five modes when they encounter a single locus architecture with the non-SBM properties of dominance effects and genetic-environment (either one or both of gene content-environment and dominance-environment) correlation. The informational input into each mode was $H \equiv \{\Sigma_{MZ,T}(\theta), \Sigma_{DZ,T}(\theta), \Sigma_{MZ,A}(\theta), \Sigma_{DZ,A}(\theta)\}$.

Details of construction.

We employed $\{a, k\}$ parameterization for the genotypic values and parameterized the architecture in terms of the vectors $\boldsymbol{\theta}_{\mathbf{gen}} = \begin{pmatrix} \varphi \\ a \\ k \end{pmatrix}$, $\boldsymbol{\theta}_{\mathbf{V}} = \begin{pmatrix} \rho_{\Psi'(g_1)_{\text{lin}},\gamma(\text{E})} \\ \rho_{\delta(g_1),\gamma(\text{E})} \\ h^2 \end{pmatrix}$, $\boldsymbol{\theta}_{\mathbf{C(MZ,T)}} = \begin{pmatrix} \rho_A(MZ) \\ \rho_D(MZ) \\ \rho_E(MZ,T) \end{pmatrix}$, $\boldsymbol{\theta}_{\mathbf{C(DZ,T)}} = \begin{pmatrix} \rho_A(DZ) \\ \rho_D(DZ) \\ \rho_E(DZ,T) \end{pmatrix}$, $\boldsymbol{\theta}_{\mathbf{C(MZ,A)}} = \begin{pmatrix} \rho_A(MZ) \\ \rho_D(MZ) \\ \rho_E(MZ,A) \end{pmatrix}$, and $\boldsymbol{\theta}_{\mathbf{C(DZ,A)}} = \begin{pmatrix} \rho_A(DZ) \\ \rho_D(DZ) \\ \rho_E(DZ,A) \end{pmatrix}$.

Under this choice of $\boldsymbol{\theta}_{\text{V}}$: i) σ^2_{Z} is determined as $\frac{\sigma^2_{\text{A}}}{h^2}$; and ii) σ^2_{E}

as the largest root of the quadratic equation

$$0 = \sigma^2_{\text{A}} + \sigma^2_{\text{D}} - \sigma^2_{\text{Z}} + \sigma_{\text{E}}\left(\sigma_{\text{E}} + \rho_{\Psi'(g_1)_{\text{lin}},\gamma(\text{E})}\sigma_{\text{A}} + \rho_{\delta(g_1),\gamma(\text{E})}\sigma_{\text{D}}\right)$$

under the restriction that

$$h^2 < \frac{\sigma^2_{\text{A}}}{\left(\sigma^2_{\text{A}} + \sigma^2_{\text{D}}\right)}$$

(which ensures that σ^2_{E} is positive) and

$$-\sqrt{1 - \rho_{\Psi'(g_1)_{\text{lin}},\gamma(\text{E})}} \leq \rho_{\delta(g_1),\gamma(\text{E})} \leq \sqrt{\rho_{\Psi'(g_1)_{\text{lin}},\gamma(\text{E})}}$$

(which ensures that the covariance matrix of $\{\Psi'(g_1)\text{lin}, \delta(g_1), \gamma(E)\}$ is gramian).

Each element of $H(\boldsymbol{\theta}) \equiv \{\Sigma_{\text{MZ,T}}(\boldsymbol{\theta}), \Sigma_{\text{DZ,T}}(\boldsymbol{\theta}), \Sigma_{\text{MZ,A}}(\boldsymbol{\theta}), \Sigma_{\text{DZ,A}}(\boldsymbol{\theta})\}$ takes a nine-element parameter vector $\boldsymbol{\theta}$ as its argument. As $\boldsymbol{\theta}$ moves over its parameter space, yielding novel realizations $\boldsymbol{\theta}^*$, all possible instantiations of $\{P, Z, \Psi'(\mathbf{g}), R, L\}$ are spelled out. For each realization $\boldsymbol{\theta}^*$, realizations $H(\boldsymbol{\theta}^*) \equiv \{\Sigma_{\text{MZ,T}}(\boldsymbol{\theta}^*), \Sigma_{\text{DZ,T}}(\boldsymbol{\theta}^*), \Sigma_{\text{MZ,A}}(\boldsymbol{\theta}^*), \Sigma_{\text{DZ,A}}(\boldsymbol{\theta}^*)\}$ and $h^2(\boldsymbol{\theta}^*)$ are yielded, and, from $H(\boldsymbol{\theta}^*)$, a realization $R(H(\boldsymbol{\theta}^*))$. Though any admissible realization $\boldsymbol{\theta}^*$ is *possible* in nature, we selected a sample of $3^5 \times 5 = 1,215$ of these, generated by crossing the sets of theoretically guided selections for $\left\{\varphi, a, k, \rho_{\Psi'(g_2)_{\text{in}},\gamma(\text{E}'}\rho_{\delta(g_2),\gamma(\text{E}'}h^2\right\}$ described below. Selection was, in fact, motivated by imagining the phenotype in question to be psychopathy as quantified by the PCL-R (Hare, 1991).

φ: Let the A_2A_2 genotype represent that of "true" psychopaths. Estimates of the proportion of true psychopaths, $P(A_2A_2)$ are generally contained within the interval [.01,.25], see Hare (1991, 1996). This implies that h^2 is contained in the interval [.1,.5], and we selected the values [.1,.3,.5].

a: Under the a, k parameterization, a is equal to $\frac{(\mu_{Z|g_1=2} - \mu_{Z|g_1=0})}{2}$ (Lynch & Walsh, 1998). Because the PCL-R assigns scores on the interval [0,63], sensible values for a are [10,20,30]. The three sensible values were derived by means of application of the given formula to three pairs of conditional means: [62,2], [50,10], [40,20]. So, these values, (62-2)/2 = 30, (50-10)/2 = 40, and (40-20)/2 = 10 are not based on any empirical distribution of PCL values. Only on the interval of possible PCL-R scores, and candidate pairs of conditional means, the first for the $g_1 = 2$ (psychopathy) group, and the second, the group genetically most antithetical to the psychopathy group; i.e., the $g_1 = 0$ group.

k: There do not exist, to our knowledge, many studies in which k has directly been estimated. However, recent studies (e.g., Herzig et al., 2018; Nolte et al., 2017) suggest that the contributions from dominance genetic components are relatively small compared to the additive genetic components. These findings notwithstanding, Chen et al. (2015) showed in a twin-based analysis that while the variance of each of 19 human complex traits was largely due to additive components, dominance components played a larger role than had been suggested by other studies featuring fewer complex traits. In light of these findings, one should therefore rather aim for a broader range of dominance deviation effects, although it should be noted that recent evidence suggests the majority of dominance effects to be rather small compared to the additive effects. We therefore chose the *possible* values [0,.4,.8].

$\rho_{\Psi'(g_2)_{\text{lin}},\gamma(\text{E})}$: As one can see from the relation $V(\gamma(E)|g_1) = \sigma^2_E\left(1 - \rho^2_{\Psi'(g_2)_{\text{lin}},\gamma(\text{E})}\right)$, this parameter con-

trols the *degree of variability* of the impacts of environment on Z, conditional on subpopulations of individuals with identical gene content. There appears to be no scientific reason why the degree of *variability* in impacts should differ much between the general population and subpopulations with identical gene content. Thus, we chose a set of values on the lower end of the spectrum; [0,.3,.6].

$\rho_{\delta(g_2),\gamma(E)}$: For this parameter, we offer precisely the same argument, and selected the set of values, [0,.1,.3].

h^2: h^2 is the parameter the recovery of which is under investigation. As we would like to gain insight into the recovery capabilities of the modes under a wide range of possibilities, we selected the following values: [.1,.3,.5,.7,.9].

The twin correlations – $\rho_A(MZ)$, $\rho_D(MZ)$, $\rho_A(DZ)$, $\rho_D(DZ)$, $\rho_E(MZ,T)$, $\rho_E(DZ,T)$, $\rho_E(MZ,A)$, $\rho_E(DZ,A)$ – were held constant over realizations θ^*. The magnitudes of the four environment correlations can be expected to vary over $P, Z, \Psi'(\mathbf{g}), R, L$, and is very much an empirical issue. Genetic theory suggests that the first four of these correlations should be in the vicinity of 1, 1, .5, and .25, respectively. Because correlations of a magnitude of unity are empirically implausible, we selected the following set of *reasonable* values: .97, .96, .52, .21, .70, .50, .30, .05. To provide a justification for these values: Of course, on the estimation side of the coin, one should stick to those values that theory nominates as most reasonable. Just because the architecture to be estimated will be unknown, there is nothing else to do. However, for any given architecture, nature will fix the values of parameters; and nature is not constrained by best present theory. It is entirely reasonable to construct target architectures for which, e.g., a twin correlation is set a touch less than unity. In nature, states of affairs are dichotomous; either a given architecture has a correlation of unity, or it does not. Theory does not govern which is the case. Estimation will always be in the face of individual, particular, unknown architectures, and the task is to assess the performance of the estimators when, merely, they encounter reasonable possibilities. What is believed, theoretically, is already enshrined within the estimation schemes; to insist that our models of reality accord with these assumptions is akin to insisting that the Copernican theory be by checked through comparison of a circle to a circle.

Of the $3^5 \times 5 = 1,215$ sampled realizations of θ^*, 1,080 instantiations of the architecture that satisfied the restrictions earlier described. The set of realizations $H(\theta^*)$ based on this set of 1,080 θ^* served, then, as input into the modes. Table A1 of Appendix A shows the range, over the 1,080 θ^* , of each of the induced components σ_A^2, σ_D^2 , and σ_E^2. For the structural equation models, the number of input data points, parameters to be estimated, and degrees of freedom, were as follows[6] : ACDE(12,4,8), ADE(12,3,9), ACE(12,3,9), and AE(12,2,10).

Results

Parameter Recovery Performance of Modes

If the non-SBM architecture of study 1 happened to be underlying an empirical setting in which h^2 was being recovered, and the informational input was $\{\Sigma_{MZ,T}(\boldsymbol{\theta}), \Sigma_{DZ,T}(\boldsymbol{\theta}), \Sigma_{MZ,A}(\boldsymbol{\theta}), \Sigma_{DZ,A}(\boldsymbol{\theta})\}$, how successfully would each of the modes be in recovering the true h^2? Our central and side issues can be addressed with reference to Tables 1 and 2, which present a quantitative summary of the model error and absolute model error distributions, respectively. Boxplots of the five model error distributions appear in Figure 2.

How close to the mark can a narrow-sense heritability recovered value be expected to be? From Table 2, the answer is that it depends upon the mode employed. In particular, the ordering of the modes with respect to the criterion of mean absolute model error (from best to worst recovery performance) was as follows: ACDE (.02), ADE (.04), ACE (.08), AE (.08), and Falconer (.10).

How sensitive is h^2 parameter recovery performance to the specifics of $P, Z, \Psi'(\mathbf{g}), R, L$? The answer is that it depends upon the mode employed. The ordering of the modes with respect to the criterion of variance of absolute model error (from least to most sensitive) was as follows: ACE (.012), AE (.012), ACDE (.022), ADE (.022), and Falconer (.142).

How far off the mark (how badly misrepresenting of the true state of nature) can a recovered value be? The answer, once again, is that it depends upon the mode employed. None of the modes were much at risk of underestimating h^2, minimum model error ranging from -0.03 (ACDE, ADE, ACE, and AE) to -0.09 (Falconer). However, all of them were susceptible to yielding substantial upward-biased recovered values, the ordering with respect to maximum model error being ACDE (0.06), ADE (0.14), ACE (0.45), AE (0.47), and Falconer (0.66).

Relative Parameter Recovery Performance of Modes: On all of the relevant criteria, the ordering of the modes –

[6] Each matrix (twin design) contributes 3 input data points; matrix Φ, the correlation matrix of latent variables, contributes no parameters to be estimated (the variances are set to unity, and off-diagonal elements are as prescribed by the standard biometric model); the number of loadings to be estimated is simply equal to the number of letters describing the model.

Table 1

Summary of Model Error, Study 1

	$h_r^{2'} - h^2(\theta)$							
	n	$n_{\text{inadm. sol.}}$	M	SD	Median	Mad	min	max
ACDE	805	102	.02	.02	.02	.03	-.03	.06
ADE	805	3	.04	.02	.04	.02	-.03	.14
ACE	805	19	.08	.10	.04	.04	-.03	.45
AE	805	0	.08	.10	.05	.04	-.03	.47
Falconer	805	-	.10	.15	.03	.04	-.09	.66

Note. Model Error: $h_r^{2'} - h^2(\theta)$; n: Total number admissible scenarios (those satisfying inequalities); $n_{\text{inadm. sol.}}$: number of inadmissible solutions; Mad: Median absolute deviation.

Table 2

Summary of Absolute Model Error, Study 1

	$\lvert h_r^{2'} - h^2(\theta)\rvert$										
			Percentile Points								
	M	SD	.10	.20	.30	.40	.50	.60	.70	.80	.90
ACDE	.02	.02	.00	.01	.01	.02	.02	.03	.04	.04	.05
ADE	.04	.02	.01	.02	.03	.04	.04	.05	.05	.06	.06
ACE	.08	.01	.01	.02	.03	.04	.05	.06	.14	.20	.20
AE	.08	.01	.01	.02	.03	.04	.05	.06	.07	.14	.20
Falconer	.10	.14	.02	.02	.03	.04	.05	.09	.19	.27	.27

Note. Absolute Model Error: $\lvert h_r^{2'} - h^2(\theta)\rvert$

from best to poorest performing – was ACDE, ADE, ACE, AE, and Falconer.

Do the newer SEM-based modes outperform the classical, correlation-based modes? In this particular simulation setting, the answer is an unqualified yes.

Parameter Recovery Performance of SEM-based Modes, for Subset of Solutions with RMSEA < 0.06: In the case of the ACDE mode, RMSEA did not exceed 0.06 for any of the admissible solutions. Accordingly, Table 3 provides a summary of the parameter recovery performance of ADE, ACE, and AE modes, conditioning on the subset of admissible solutions for which RMSEA < 0.06.

The high degree of similarity between the entries of Tables 1 and 3 indicates that the conclusions drawn about parameter recovery performance are in no way confounded by the issue of model fit.

Impact of Features of Architecture on Parameter Recovery Performance: Results of the variable importance analysis are presented in Table 4.

As is apparent from the Table 4 entries, the general conclusion is that only k, h^2, and φ had notable impacts upon h^2 parameter recovery performance, with the ordering of these factors, in respect to overall strength of impact, as listed above. The modes, however, did vary with respect to their sensitivity to these factors, with the strongest impact being that of k upon Falconer's. In keeping with its position as delivering the best overall parameter recovery performance, ACDE was the least affected by the factors. As can be seen from R^2 values of about 0.94, most of the variance in model error could be explained. As indicated by heterogeneous ICE curves[7],

Table 3

Summary of Model Error for solutions with RMSEA < .06, Study 1

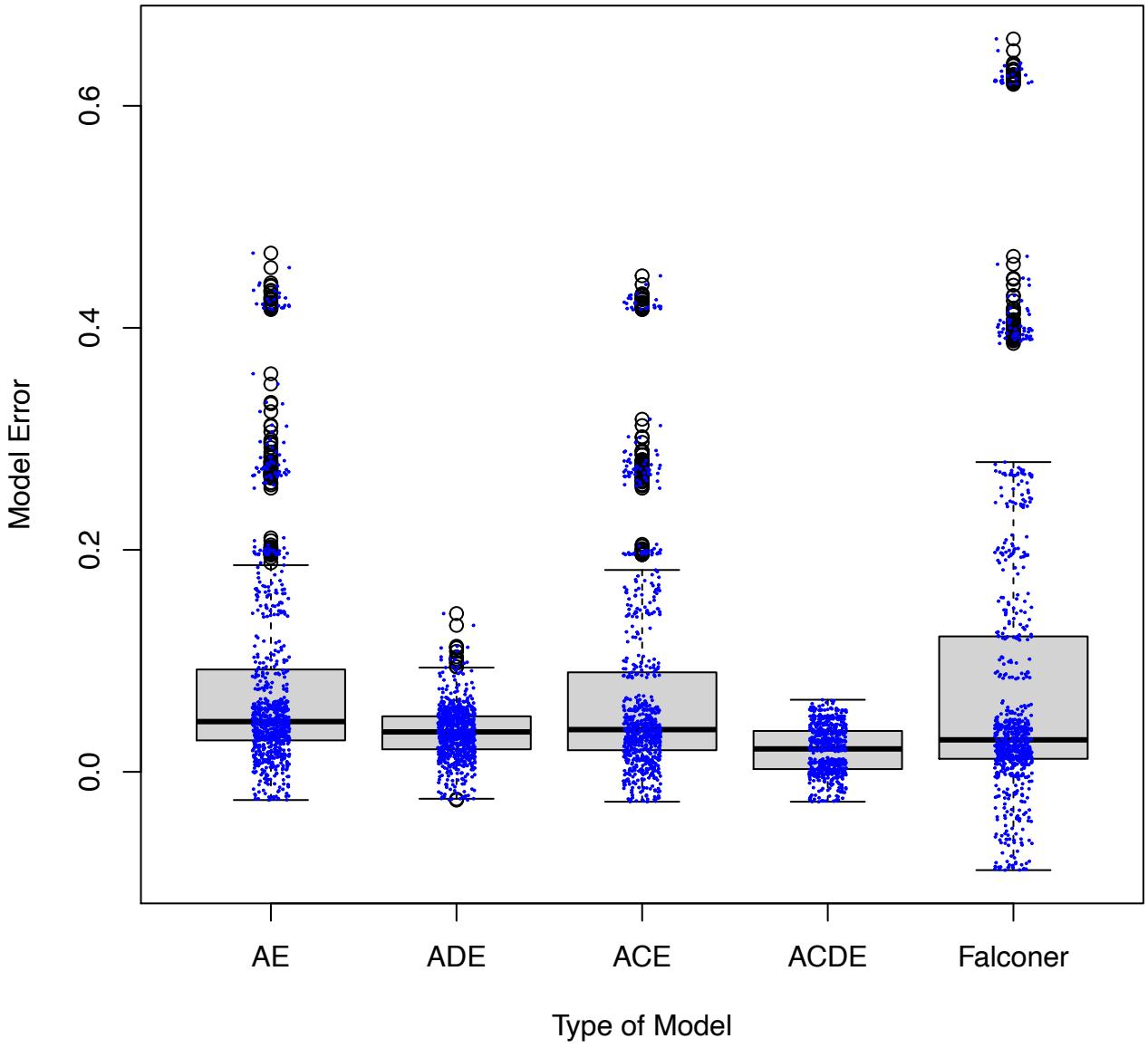
	n	M	SD	Median	Mad	min	max
ADE	796	.04	.02	.04	.02	-.03	.14
ACE	762	.06	.08	.04	.03	-.03	.32
AE	790	.08	.09	.04	.03	-.03	.42

Note. n: Number of solutions satisfying the condition RMSEA $< .06$; Mad: Median absolute deviation

[7] For the sake of stringency, these multiple graphs can be found in the electronic supplementary material on https://osf.io/aq9sx/, under a CC-BY4.0 license.

Figure 2

Box plots with jittered data points of the empirical distribution of model error $h_r^{2'} - h^2(\theta)$*, Study 1*



the predictor impacts of k, h^2, and φ on h^2 model error were characterized by interaction effects, particularly pronounced for population h^2.

Simulation Study 2

Introduction

The second study was identical to the first, except that the input into each mode was $H \equiv \{\Sigma_{MZ,T}(\boldsymbol{\theta}), \Sigma_{DZ,T}(\boldsymbol{\theta})\}$. The interest here was in the recovery performance of the modes under this reduced

Table 4

Variable Importance Analysis, Study 1

	Input Parameter						
	a	k	φ	$P_{\Psi'(g_2)/\text{lin},\gamma(\text{E})}$	$P_{\delta(g_2),\gamma(\text{E})}$	h^2	R^2
ACDE	0.0000	0.0008	0.0001	0.0000	0.0000	0.0002	.96
ADE	0.0002	0.0007	0.0001	0.0000	0.0000	0.0002	.91
ACE	-0.0002	0.0141	0.0022	-0.0002	-0.0002	0.0033	.95
AE	-0.0001	0.0132	0.0023	-0.0002	-0.0002	0.0031	.95
Falconer	-0.0005	0.0338	0.0051	-0.0005	-0.0006	0.0073	.95

Note. a: additive parameter; k: Dominance parameter; φ: $P(A_2)$ or proportion of alleles of type A_2; $P_{\Psi'(g_2)/\text{lin},\gamma(\text{E})}$: correlation between additive component and environmental impact; $P_{\delta(g_2),\gamma(\text{E})}$: correlation between dominance component and environmental impact; h^2: narrow heritability; R^2: Multiple R^2

informational input.

Details of Construction

For the structural equation models, the number of input data points, parameters to be estimated, and degrees of freedom were as follows: ACDE(6,4,2), ADE(6,3,3), ACE(6,3,3), and AE(6,2,4). A possible confusion concerning the reported positive number of degrees of freedom of the ACDE model might arise here. According to Ozaki, Toyada, Iwama, Kubo, and Ando (2011), the number of degrees of freedom should be -1 because the covariance matrix of the observed variables contains three observations (one variance and two covariances), but the ACDE model contains four free parameters, hence $df = 3 - 4 = -1$. However, one must keep in mind that *each* covariance matrix of MZ and DZ twins consists of $k(k+1)/2$ observations (where k is the number of observed variables). Consequently, we have $2(2(2+1)/2 = 6$ observations, not three. Therefore, the correct number of degrees of freedom in this case is $6 - 4 = 2$, not -1. An interesting wrinkle was that, the positive degrees of freedom notwithstanding, none of the ACDE models were identified in study 2; though, of course, certain *functions* of model parameters may have been. Appendix B provides a formal explanation that the ACDE model is identified in study 1 and 3 but not in study 2. The non-identifiability of the ACDE model in study 2 notwithstanding, we kept the ACDE model to see how well it would perform under this non-identified state of affairs.

Results

Parameter Recovery Performance of Modes

Tables 5 and 6 present a quantitative summary of the model error and absolute model error distributions when the modes faced the non-SBM architecture of study 2, and the informational input was $\{\Sigma_{MZ,T}(\boldsymbol{\theta}), \Sigma_{DZ,T}(\boldsymbol{\theta})\}$. Boxplots of the five model error distributions appear in Figure 3. In such an empirical setting, our central and side issues can be answered as follows.

How close to the mark can a narrow-sense heritability recovered value be expected to be? From Table 6, the answer is that it depends on the mode employed. In particular, the ordering of the modes in terms of the criterion of mean absolute model error (from best to worst recovery performance) was as follows: ADE (0.05), ACE (0.08), AE (0.09) Falconer (0.10), and ACDE (0.29).

How sensitive is h^2 parameter recovery performance to the specifics of $P, Z, \Psi'(\mathbf{g}), R, L$? The answer is that it depends on the mode employed, the ordering of the modes with respect to the criterion of variance of absolute model error (from least to most sensitive), as follows: ADE (0.042), ACE (0.082), AE (0.092), Falconer (0.102), and ACDE (0.292).

How far off the mark (how badly misrepresenting of the true state of nature) can a recovered value be? In the direction of downward-bias, the ordering of the modes (from least to most severe downward-bias) was ADE and AE (-0.02), ACE and Falconer (-0.09), and ACDE (-0.88); in the direction of upward-bias (from least to most severe upward-bias), ACDE (0.04), ADE (0.31), ACE and AE (0.49), and Falconer (0.66). Evidently, then, there exists an interaction between mode and direction of model error, with ACDE susceptible to miscalculate h^2 as much smaller than its true value, and Falconer, AE, ACE, and AE, susceptible to misrepresent it as much larger.

Relative Parameter Recovery Performance of Modes: In the present case, it was much less straightforward to derive an overall parameter recovery performance ordering of the modes. On the criteria of expected closeness to the mark and sensitivity to the particulars of the architecture in play, the modes were ordered (from best to worst) as follows: ADE, ACE, AE, Falconer, and ACDE.

Table 5

Summary of Model Error, Study 2

	$h_r^{2'} - h^2(\theta)$							
	n	$n_{\text{inadm. sol.}}$	M	SD	Median	Mad	min	max
ACDE	805	0	-.29	.20	-.27	.25	-.88	.04
ADE	805	4	.05	.04	.05	.02	-.02	.31
ACE	805	3	.07	.11	.03	.04	-.09	.49
AE	805	0	.09	.10	.05	.04	-.02	.49
Falconer	805	-	.09	.15	.03	.04	-.09	.66

Note. Model Error: $h_r^{2'} - h^2(\theta)$; n: Total number admissible scenarios (those satisfying inequalities); $n_{\text{inadm. sol.}}$: number of inadmissible solutions; Mad: Median absolute deviation.

Table 6

Summary of Absolute Model Error, Study 2

	$\lvert h_r^{2'} - h^2(\theta)\rvert$										
	Percentile Points										
	M	SD	.10	.20	.30	.40	.50	.60	.70	.80	.90
ACDE	.29	.20	.04	.09	.17	.20	.27	.33	.40	.47	.56
ADE	.05	.04	.02	.03	.03	.04	.05	.05	.06	.07	.09
ACE	.08	.10	.01	.02	.02	.03	.04	.05	.08	.15	.21
AE	.09	.10	.02	.03	.04	.04	.05	.06	.09	.15	.22
Falconer	.10	.14	.01	.02	.02	.03	.04	.05	.09	.19	.27

Note. Absolute Model Error: $\lvert h_r^{2'} - h^2(\theta)\rvert$

However, as noted previously, ACE, AE, and Falconer were susceptible to delivering large upward-biased values.

Do the newer SEM-based modes outperform the classical, correlation-based modes? In this particular setting, the answer depends on the criterion chosen. On most, Falconer's occupies a mid-pack positioning. Certainly, in this setting, it cannot be concluded unequivocally that an SEM-based mode will outperform the classical Falconer's coefficient.

Parameter Recovery Performance of SEM-based Modes, for Subset of Solutions with RMSEA < 0.06: There were admissible solutions with an RMSEA that exceeded 0.06 only in the case of the ACE and AE modes. Accordingly, Table 7 provides a summary of the parameter recovery performance of these two modes, conditioning on the subset of solutions for which RMSEA < 0.06.

Comparing Tables 5 and 7, makes clear that the conclusions we draw about parameter recovery performance are in no way confounded by the issue of model fit.

Impact of features of architecture on parameter recovery performance. Results of the variable importance analysis are presented in Table 8.

As is apparent from the Table 8 entries, the general conclusion is that only k and h^2 impacted upon h^2 parameter recovery performance. The ordering of these factors in respect strength of impact, was, for ACDE and ACE, h^2 followed k, and, for all other modes, k followed by h^2. With R^2 values of about .93 most of the variance in model error could be explained. Once again, for all of the modes, heterogenous ICE curves for h^2 and k suggested that the impacts, especially of population h^2, were attributable to interaction effects.

Simulation Study 3

Introduction

In this third study, we assessed the parameter recovery performance of the modes when they encounter a multilocus architecture with the non-SBM properties of dominance effects, genetic-environment (either one or both of gene content-environment and dominance-environment) correlation, and epistasis, and the informational input is $H \equiv \{\Sigma_{MZ,T}(\boldsymbol{\theta}), \Sigma_{DZ,T}(\boldsymbol{\theta}), \Sigma_{MZ,A}(\boldsymbol{\theta}), \Sigma_{DZ,A}(\boldsymbol{\theta})\}$.

Figure 3

Box plots with jittered data points of the empirical distribution of model error $h_r^{2'} - h^2(\theta)$, Study 2

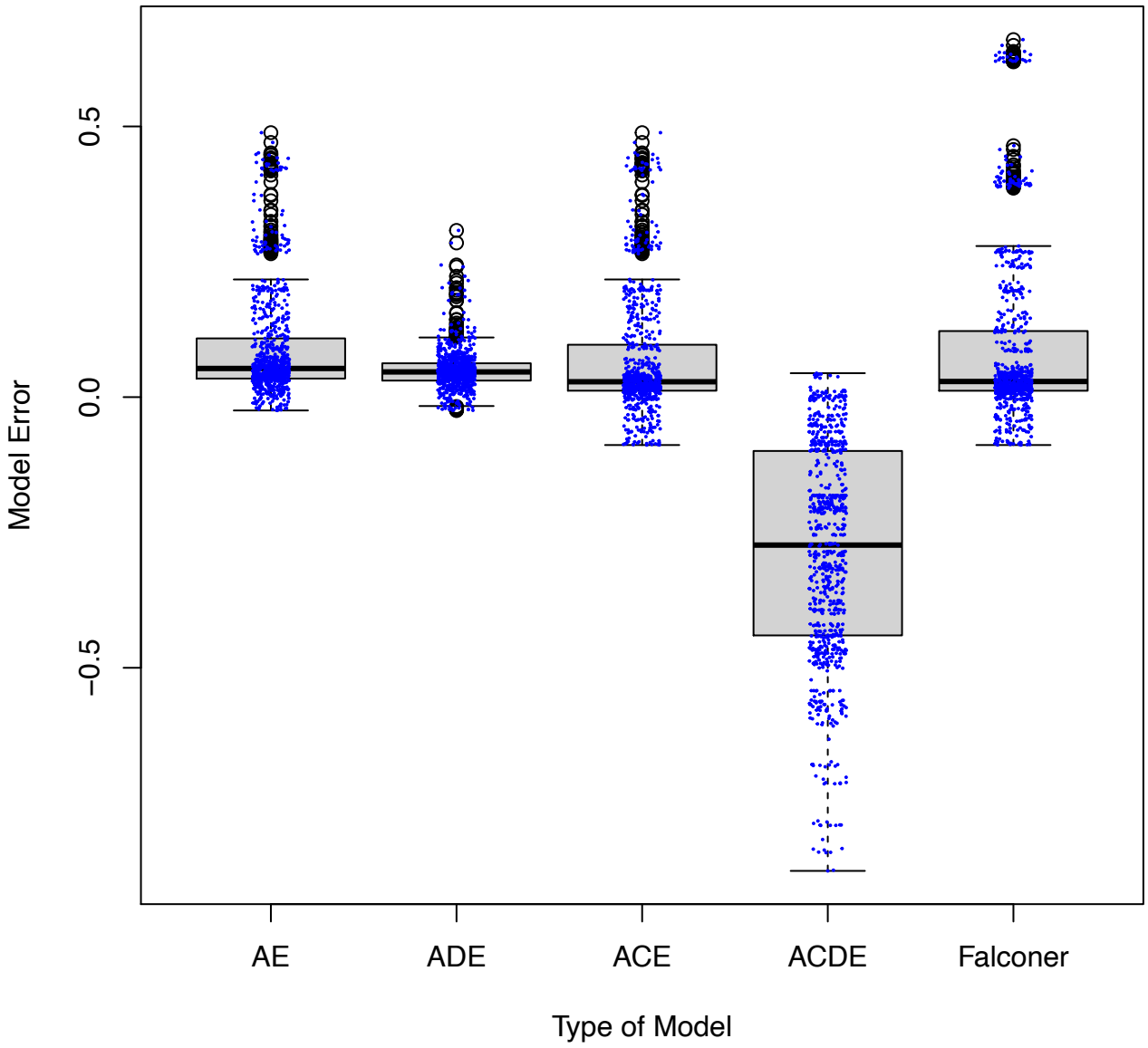


Table 7

Summary of Model Error for solutions with RMSEA < .06, Study 2

	n	*M*	*S D*	Median	Mad	min	max
ACE	778	.06	.09	.03	.03	-.09	.41
AE	783	.08	.08	.05	.04	-.02	.49

Note. Model Error: $h_r^{2'} - h^2(\theta)$; *n*: Number of solutions satisfying the condition RMSEA < .06; Mad: Median absolute deviation.

Table 8

Variable Importance Analysis, Study 2

	Input Parameter						
	a	k	φ	$P_{\Psi'(g_2)/\text{lin},\gamma(\text{E})}$	$P_{\delta(g_2),\gamma(\text{E})}$	h^2	R^2
ACDE	0.0023	0.0049	0.0019	-0.0014	-0.0015	0.0585	.91
ADE	0.0005	0.0012	0.0004	0.0000	0.0000	0.0004	.90
ACE	-0.0001	0.0187	0.0026	-0.0003	-0.0003	0.0036	.96
AE	0.0001	0.0132	0.0027	-0.0002	-0.0002	0.0030	.95
Falconer	-0.0008	0.0516	0.0074	0.0000	-0.0004	0.0107	.95

Note. a: additive parameter; k: Dominance parameter; φ: $P(A_2)$ or proportion of alleles of type A_2; $P_{\Psi'(g_2)/\text{lin},\gamma(\text{E})}$: correlation between additive component and environmental impact; $P_{\delta(g_2),\gamma(\text{E})}$: correlation between dominance component and environmental impact; h^2: narrow heritability; R^2: Multiple R^2

Details of Construction

The number of loci, n, was set to three, and—following Zuk et al. (2012), who provide argument as to its "ecological validity"—the genotypic function $\Psi'(g)$ was chosen to be $\max(\Psi'(g_j), j = 1..3)$. This non-additive *limiting process* architecture (see Zuk et al., 2012) induces epistasis in an empirically realistic fashion. We employed $\{a, k\}$ parameterization, once again, and parameterized the architecture in terms of the following vectors: $\boldsymbol{\theta}_{\text{gen}} = \{\boldsymbol{\varphi}, \boldsymbol{a}, \boldsymbol{k}\}$, wherein the jth element of each of the 3-The vectors φ, a, and k is φ_j, a_j, and k_j respectively; $\boldsymbol{\theta}_V = \begin{pmatrix} \sum_{j=1}^3 \rho_{B_j,\gamma(E)} \\ \sum_{j=1}^3 \rho_{\delta_j,\gamma(E)} \\ \sum_{j=1}^3 \rho_{l_j,\gamma(E)}/h^2 \end{pmatrix}$; $\boldsymbol{\theta}_{C(MZ,T)} = \begin{pmatrix} \rho_A(MZ) \\ \rho_D(MZ) \\ \rho_{EP}(MZ) \\ \rho_E(MZ,T) \end{pmatrix}$; $\boldsymbol{\theta}_{C(MZ,A)} = \begin{pmatrix} \rho_A(MZ) \\ \rho_D(MZ) \\ \rho_{EP}(MZ) \\ \rho_E(MZ,A) \end{pmatrix}$, and $\boldsymbol{\theta}_{C(DZ,A)} = \begin{pmatrix} \rho_A(DZ) \\ \rho_D(DZ) \\ \rho_{EP}(DZ) \\ \rho_E(DZ,A) \end{pmatrix}$. This particular choice of $\boldsymbol{\theta}_V$ implies that: i) σ_z^2 is determined as σ_A^2/h^2; and ii) σ_E^2 is the largest root of the quadratic equation,

$$0 = \sigma_A^2 + \sigma_D^2 + \sigma_{EP}^2 + \sigma_E$$

$$\left(\sigma_E + \sum_{j=1}^3 \rho_{B_j,\gamma(E)}\sigma_A + \sum_{j=1}^3 \rho_{\delta_j,\gamma(E)}\sigma_D + \sum_{j=1}^3 \rho_{l_j,\gamma(E)}\sigma_{EP}\right),$$

under restriction that $h^2 < \frac{\sigma_A^2}{(\sigma_A^2+\sigma_D^2+\sigma_{EP}^2)}$ and $|\Sigma| \geq 0$, wherein $\Sigma = \begin{pmatrix} \sigma_A^2 & 0 & 0 & \rho_{\sum B_j,\gamma(E)} \\ 0 & \sigma_B^2 & 0 & \rho_{\sum \delta_j,\gamma(E)} \\ 0 & 0 & \sigma_{EP}^2 & \rho_{\sum l_j,\gamma(E)}/\sigma_E^2 \end{pmatrix}$.

Realizations of $\boldsymbol{\theta}^*$ were produced by crossing the following sets of parameter selections:

- φ: for $j = 1..3$, $[.1, .3, .5]$;
- a: for $j = 1..3$, $[10, 20, 30]$;
- k: for $j = 1..3$, $[0, .2, .4]$;
- $\rho_{\sum_{j=1}^3 B_j,\gamma(E)}$: $[0, .3, .6]$;
- $\rho_{\sum_{j=1}^3 \delta_j,\gamma(E)}$: $[0, .1, .3]$;
- $\rho_{\sum_{j=1}^3 l_j,\gamma(E)}$: $[0, .2, .5]$;
- h^2: $[.1, .3, .5, .7, .9]$;
- $\rho_{A(MZ)}, \rho_{D(MZ)}, \rho_{EP(MZ)}, \rho_{A(DZ)}, \rho_{D(DZ)}, \rho_{EP(DZ)}, \rho_{E(MZ,T)}, \rho_{E(DZ,T)}, \rho_{E(MZ,A)}, \rho_{E(DZ,A)}$: $[.97, .96, .95, .52, .21, .5, .7, .5, .3, .05]$.

Of the $3^9 \times 3^3 \times 5 = 2,657,205$ sampled realizations of $\boldsymbol{\theta}^*$, $2,174,537$ yielded instantiations of the architecture which satisfied the restrictions earlier described. The set of realizations $H(\boldsymbol{\theta}^*)$ based on this set of $2,174,537$ $\boldsymbol{\theta}^*$ served, then, as input into the modes.

Results

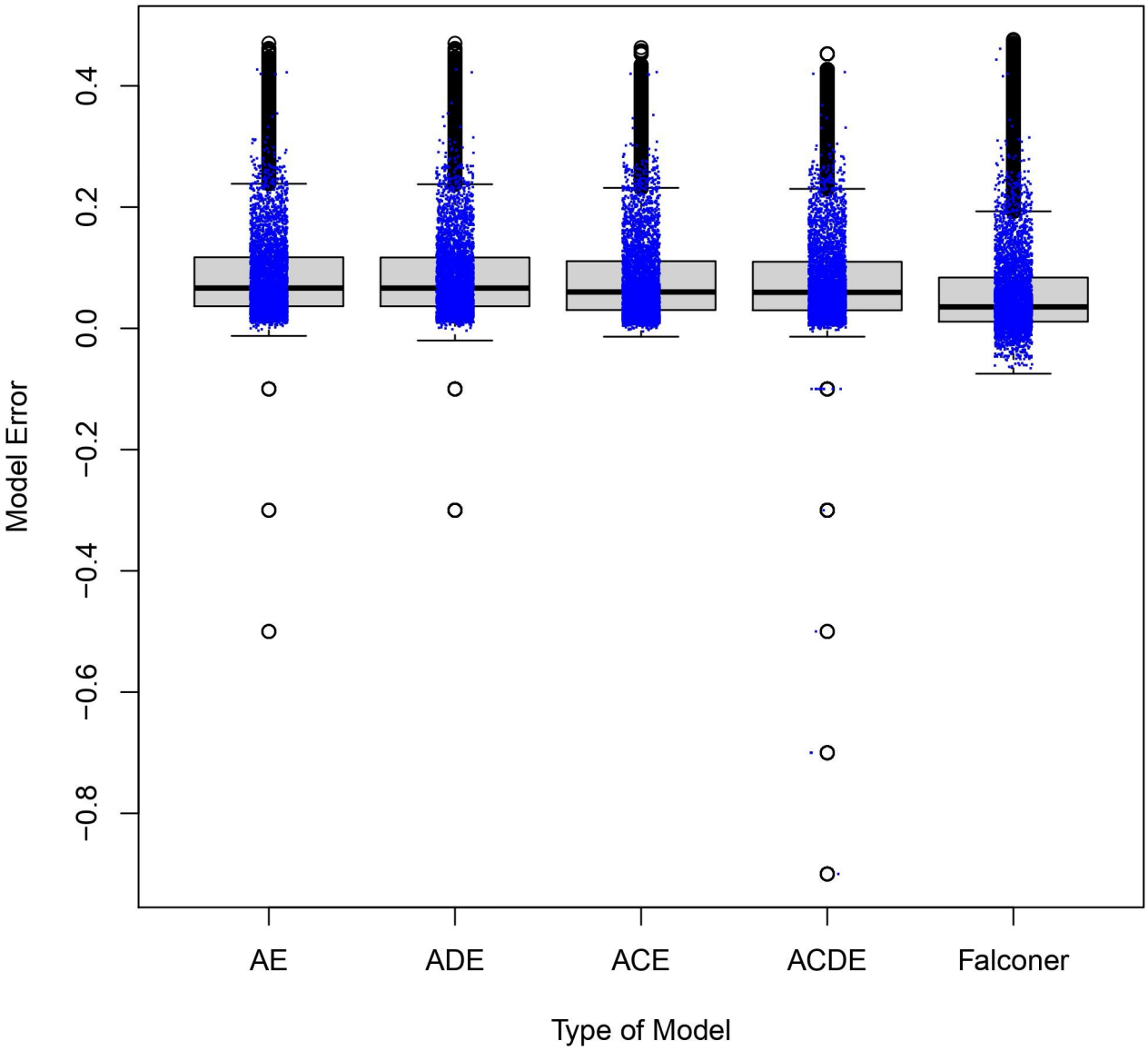
Parameter recovery performance of modes

Tables 9 and 10 present a quantitative summary of the model error and absolute model error distributions, when the modes were faced with the multilocus non-SBM architecture of study 3 and informational input was $\{\Sigma_{\text{MZ,T}}(\boldsymbol{\theta}), \Sigma_{\text{DZ,T}}(\boldsymbol{\theta}), \Sigma_{\text{MZ,A}}(\boldsymbol{\theta}), \Sigma_{\text{DZ,A}}(\boldsymbol{\theta})\}$. Boxplots of the five model error distributions appear in Figure 4. In such an empirical setting, our central and side issues can be answered as follows.

How close to the mark can a narrow heritability recovered value be expected to be? Table 10 shows that, with respect to the criterion of mean absolute model error, all the SEM-based modes delivered precisely the same parameter recovery performance (.08). Falconer's coefficient performed marginally better (.06).

Figure 4

Box plots with jittered data points of the empirical distribution of model error $h_r^{2'} - h^2(\theta)$*, Study 3*



Note. Individual data points (blue dots) are based on a random selection of 5000 cases.

Table 9

Summary of Model Error, Study 3

	$h_r^{2'} - h^2(\theta)$							
	n	$n_{\text{inadm. sol.}}$	M	SD	Median	Mad	min	max
ACDE	1,963,413	3,312	.08	.06	.06	.05	-.90	.45
ADE	2,174,537	98	.08	.06	.07	.05	-.30	.47
ACE	2,174,537	4,280	.08	.06	.06	.05	-.01	.46
AE	2,174,537	275	.08	.06	.07	.05	-.30	.47
Falconer	2,174,537	-	.05	.06	.04	.04	-.07	.48

Note. Model Error: $h_r^{2'} - h^2(\theta)$; n: Total number admissible scenarios (those satisfying inequalities); $n_{\text{inadm. sol.}}$: number of inadmissible solutions; Mad: Median absolute deviation.

Table 10

Summary of Absolute Model Error, Study 3

	$\lvert h_r^{2'} - h^2(\theta)\rvert$										
	Percentile Points										
	M	SD	.10	.20	.30	.40	.50	.60	.70	.80	.90
ACDE	.08	.06	.02	.03	.04	.05	.06	.08	.10	.13	.17
ADE	.08	.06	.02	.03	.04	.05	.07	.08	.10	.13	.18
ACE	.08	.06	.02	.03	.04	.05	.06	.08	.10	.13	.17
AE	.08	.06	.02	.03	.04	.05	.07	.08	.10	.13	.18
Falconer	.06	.06	.01	.01	.02	.03	.04	.05	.07	.10	.15

Note. Absolute Model Error = $\lvert h_r^{2'} - h^2(\theta)\rvert$

Table 11

Summary of Model Error for solutions with RMSEA < .06, Study 3

	n	M	SD	Median	Mad	min	max
ADE	1,963,403	.08	.06	.07	.05	-.30	.46
AE	1,963,320	.08	.06	.07	.05	-.01	.46

Note. n: Number of solutions satisfying the condition RMSEA < .06; Mad: Median absolute deviation.

How sensitive is h^2 parameter recovery performance to the specifics of $P, Z, \Psi'(\mathbf{g}), R, L$? The answer is that, the five modes were equally sensitive, the variance of each of the five absolute model error distributions equal to $.06^2$.

How far off the mark (how badly misrepresenting of the true state of nature) can a recovered value be? In the direction of downward-bias, the ordering of the modes (from least to most severe downward-bias) was ACE (-.01), Falconer (-.07), ADE (-.30), AE (-.30), and ACDE (-.90). In the direction of upward-bias, the modes were tightly packed within the .45 to .48 interval. Evidently, all the modes were susceptible to misrepresent the magnitude of heritability deriving from a multilocus architecture of the present type as much larger than it actually is; all but Falconer and ACE were susceptible, additionally, to misrepresent it as much smaller than it actually is.

Relative parameter recovery performance of modes. Save for the fact that Falconer's and ACE were not susceptible to yielding large down-ward biases, there was not much to choose between the modes.

Do the newer SEM-based modes outperform the classical, correlation-based? When the underlying architecture is a multilocus, non-SBM, architecture of the sort considered, herein, the answer is, clearly not.

Parameter recovery performance of SEM-based modes, for subset of solutions with RMSEA < .06. Only in the case of the AE and ADE modes, were there admissible solutions for which RMSEA exceeded .06.

Accordingly, Table 11 provides a summary of the parameter recovery performance of AE and ADE, conditioning on the subset of solutions for which RMSEA < .06.

The high degree of similarity between the entries of Tables 9 and 11, indicates that the conclusions we draw about parameter recovery performance are in no way confounded by the issue of model fit.

Impact of features of architecture on parameter recovery performance. Results of the variable importance analysis are presented in Table 12.

As is apparent from the Table 12 entries, the general conclusion is that only h^2 and a had notable impacts upon h^2 parameter recovery performance. Most of the variance in model error could be explained, as indicated by R^2 values of about .95. For all the modes, ICE curves were predominantly homogeneous, largely following the same course, which indicates that these impacts were attributable mainly to additive effects[8].

Discussion

The results of our investigation suggest that conclusions regarding our focal issue must be drawn conditional on a number of different factors: a) the general class of generating architecture in play; b) specifics of the architecture's parametric instantiations; c) the informational input into a mode; and d) the particular mode employed. In Study 1, the modes confronted 1,080 instantiations of a single locus architecture which had the non-SBM properties of dominance effects and either one or both of gene content-environment, and dominance-environment, correlation. The informational input was a set of population covariance matrices under the four twin designs: monozygotic-reared together, monozygotic-reared apart, dizygotic-reared together, and dizygotic-reared apart: $\{\Sigma_{MZ,T}(\theta), \Sigma_{DZ,T}(\theta), \Sigma_{MZ,A}(\theta), \Sigma_{DZ,A}(\theta)\}$. In this setting: a) on all recovery performance criteria, the ordering of the modes was (from best to poorest performing) ACDE, ADE, ACE, AE, and Falconer's; b) the overall h^2 recovery performance delivered by the modes (as quantified by mean absolute model error) ranged from a best of .02 (ACDE) to a worst of .10 (Falconer's); c) the sensitivity of the modes to the architecture's parametric specifics (quantified by the variance of absolute model error) ranged from a best of .022 (ACDE) to a worst of .142 (Falconer's); and c) the susceptibility of the modes to yielding downward-biases ranged from a best of -.03 (ACDE, ADE, and ACE) to a worst of -.09 (Falconer), and to upward-biases, from a best of .06 (ACDE) to a worst of .66 (Falconer's). Random forest analyses on the impact of genetic parameters on the model error of h^2 revealed that only k (genetic dominance), h^2 (population narrow heritability), and φ (proportion of alleles of type A2) had notable impacts upon h^2 parameter recovery performance. Average explained variance in model error across models was .94.

In Study 2, the modes confronted the same 1,080 instantiations of the single locus architecture featured in study 1, but with the reduced informational input $\{\Sigma_{MZ,T}(\theta), \Sigma_{DZ,T}(\theta)\}$. In this empirical context: a) the ordering of the modes was not identical over recovery performance criteria; b) the overall h^2 recovery performance of the modes ranged from a best of .05 (ADE) to a worst of .29 (ACDE); c) the sensitivity of the modes to the architecture's parametric specifics ranged from a best of .042 (ADE) to a worst of .202 (ACDE); c) the susceptibility of the modes to yielding downward-biases ranged from a best of -.02 (ADE and AE) to a worst of -.88 (ACDE), and to upward-biases, from a best of .04 (ACDE) to a worst of .66 (Falconer's); and d) there existed an interaction between mode and direction of bias, with ACDE susceptible to reporting h2 as much smaller than its true value, and Falconer, AE, ACE and AE, susceptible to reporting it as much larger. Random forest analyses on the impact of genetic parameters on the model error of h2 showed an effect of k (genetic dominance) and h2 (population narrow heritability) on h^2 parameter recovery performance with an average explained variance in model error across models of .95.

In Study 3, the informational input was $\{\Sigma_{MZ,T}(\theta), \Sigma_{DZ,T}(\theta), \Sigma_{DZ,A}(\theta)\}$ and the modes confronted a multilocus architecture with the non-SBM properties of dominance effects, genetic-environment correlation, and epistasis. In this case: a) there was not much to choose between the modes; b) the SEM-based modes had precisely the same overall h^2 recovery performance (.08 mean absolute model error), with Falconer's, marginally better at .06; c) the modes had identical sensitivity to the architecture's parametric specifics, the variance of absolute model error being .062 for each and every one of them; d) the susceptibility of the modes to yielding downward-bias ranged from a best of -.01 (ACE) to a worst of -.90 (ACDE); and d) in respect their susceptibility to yielding upward-biases, they were tightly packed within the .45 to .48 interval. Random forest analyses revealed that only h^2 (population heritability) and $a(j)$ (locus j additive parameter) had notable impacts upon h^2 parameter recovery performance, with an average explained variance in model error across models of .95. This finding is of particular importance because it answers one of the

[8] For the sake of stringency, these multiple graphs can be found in the electronic supplementary material on https://osf.io/aq9sx/, under a CC-BY4.0 license.

main questions of this article regarding the impact of dominance genetic variation within a multilocus architecture. Given these results, non-additive genetic effects as induced by the dominance parameter k had no notable biasing influence on the recovery of h^2, at least not under the multilocus scenario of this study. Hence, dominance genetic variation did not contribute to the missing heritability as it did not substantially bias h^2 values (see also Zhu et al., 2015 for similar conclusions).

We acknowledge, of course, that we do not know the degree to which our constructed architectures approximate to those extant in nature. We would remind our readers, however, that those who employ, in their researches, the various modes to estimate h^2, do so under the very same blanket of uncertainty over the actual form of the genetic architecture in play. We do not claim that the results we present are the final word on the performance of each of the modes in estimating h^2; only that they provide an indication as to the trustworthiness of the estimates yielded by each mode when encountering several paradigm cases.

We note, also, with due circumspection, given the small scale of the design, what appear to be the following trends:

1. It seems that the more complicated the generating architecture, the less trustworthy will be the h2 recovered by a given mode. Even though, in Study 3, the poorest average model error yielded by a mode was only .08, all of the modes were susceptible to producing narrow-sense heritability recovered values widely off the mark. This may be seen as somewhat worrisome, for the reason that a given scientist will estimate the heritability inherent to a particular, unknown, architecture, and the resulting publication will present to the scientific community this single take on the magnitude of heritability. The relevant question is, "how far off the mark might it be"; not, "how far off the mark is the average of the population of estimates from which it was drawn." This estimate will have the impact upon scientific thought; not some unknown average of which it is a part;

2. For all of the modes, save Falconer's (which does not employ the additional information available in Studies 1 and 3) and AE, the less rich is the informational input, the poorer is the h^2 parameter recovery performance.

In light of the popularity of estimating h^2, and other quantitative genetic parameters, through the fitting of latent variable models, an interesting sub-finding is that, although, in each of the studies, the SEM-based modes were confronted with architectures which they misspecified, precious few solutions were signalled – by the RMSEA > .06 criterion– as poorly fitting. Even when faced with the most complicated of our architectures- the multilocus architecture of Study 3- instances of misfit were registered only for the AE and ADE models. This suggests, once again, the inadvisability of depending upon empirical fit indices to shed light upon underlying causal structure, in general, and phenotypic (genetic) architecture, in particular (c.f. Cliff, 1983; Long, 1981; MacCallum et al., 1993; Tomarken and Waller, 2003).

Table 12

Variable Importance Analysis, Study 3

	Input Parameter													
	a(1)	**a(2)**	**a(3)**	**k(1)**	**k(2)**	**k(3)**	**q(1)**	**q(2)**	**q(3)**	$\rho^3 \sum_{j=1}^{B_j,y}(E)$	$\rho^3 \sum_{j=1}^{\delta_j,y}(E)$	$\rho^m \sum_{j=1}^{I_j,y}(E)$	**h**^{2}	**R**^{2}
ACDE	0.0026	0.0026	0.0026	0.0001	0.0001	0.0001	0.0009	0.0009	0.0009	0.0000	0.0000	0.0000	0.0031	.92
ADE	0.0026	0.0026	0.0026	0.0001	0.0001	0.0001	0.0008	0.0008	0.0008	0.0000	0.0000	0.0000	0.0029	.96
ACE	0.0027	0.0027	0.0027	0.0001	0.0001	0.0001	0.0008	0.0008	0.0008	0.0000	0.0000	0.0000	0.0031	.96
AE	0.0026	0.0026	0.0026	0.0001	0.0001	0.0001	0.0009	0.0009	0.0009	0.0000	0.0000	0.0000	0.0029	.96
Falconer	0.0029	0.0029	0.0029	0.0002	0.0002	0.0002	0.0011	0.0011	0.0011	0.0000	0.0000	0.0000	0.0020	.95

Author Contact

Corresponding Author: Prof. Dr. Moritz Heene Ludwig Maximilian University, Munich, Germany Department of Psychology Leopoldstr. 13 D-80802 Muenchen, Germany heene@psy.lmu.de Phone: +49-(0)89-2180-72514 Fax: +49-(0)89-2180-99-6888

Conflict of Interest and Funding

The corresponding author, Moritz Heene, was a member of the editorial board of *Meta-Psychology*. There was no specific funding.

Author Contributions

Conceptualization: Michael D. Maraun, Moritz Heene.
Data Curation: Moritz Heene and Philipp Sckopke.
Formal Analysis: Michael D. Maraun, Philipp Sckopke, and Moritz Heene.
Investigation: Michael D. Maraun, Philipp Sckopke and Moritz Heene.
Methodology: Michael D. Maraun.
Project Administration: Michael D. Maraun, Moritz Heene and Philipp Sckopke.
Resources: Moritz Heene and Philipp Sckopke.
Software: Philipp Sckopke and Moritz Heene.
Supervision: Michael D. Maraun, Moritz Heene and Philipp Sckopke.
Validation: Philipp Sckopke and Moritz Heene.
Visualization: Moritz Heene.
Writing - Original Draft Preparation: Michael D. Maraun, Moritz Heene, and Philipp Sckopke.
Writing - Review & Editing: Michael D. Maraun, Moritz Heene, and Philipp Sckopke.

Notice that authors are listed according to their contribution to the work, from the most to the least.

Open Science Practices

This article earned the Open Materials badge for making the materials openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, is published in the online supplement.

References

Afifi, T. O., Asmundson, G. J., Taylor, S., & Jang, K. L. (2010). The role of genes and environment on trauma exposure and posttraumatic stress disorder symptoms: A review of twin studies. *Clinical psychology review*, *30*(1), 101–112.

Benyamin, B., Pourcain, B., Davis, O. S., Davies, G., Hansell, N. K., Brion, M.-J., Kirkpatrick, R. M., Cents, R. A., Franić, S., & Miller, M. B. (2014). Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Molecular psychiatry*, *19*(2), 253–258.

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Jones, Z., Casalicchio, G., Gallo, M., Bossek, J., Studerus, E., Judt, L., Kuehn, T., Kerschke, P., Fendt, F., Probst, P., Sun, X., Thomas, J., Vieira, B., Beggel, L., ... Coors, S. (2017). Mlr: Machine Learning in R. Retrieved May 15, 2017, from https://cran.r-project.org/web/packages/mlr/index.html

Bowles, S., Gintis, H., et al. (2001). The inheritance of economic status: Education, class and genetics [Publisher: Oxford University Press New York]. *International encyclopedia of the social and behavioral sciences: Genetics, behavior and society*, *6*, 4132–141.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bronfenbrenner, U. (1999). Nature with nurture: A reinterpretation of the evidence [Publisher: Oxford University Press New York]. *Race and IQ*, 153–183.

Calaway, R., Analytics, R., & Weston, S. (2015). Foreach: Provides Foreach Looping Construct for R. Retrieved May 15, 2017, from https://cran.r-project.org/web/packages/foreach/index.html

Chen, X., Kuja-Halkola, R., Rahman, I., Arpegård, J., Viktorin, A., Karlsson, R., Hägg, S., Svensson, P., Pedersen, N. L., & Magnusson, P. K. E. (2015). Dominant Genetic Variation and Missing Heritability for Human Complex Traits: Insights from Twin versus Genome-wide Common SNP Models. *The American Journal of Human Genetics*, *97*(5), 708–714. https://doi.org/10.1016/j.ajhg.2015.10.004

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate behavioral research*, *18*(1), 115–126.

Dowle, M., Srinivasan, A., Gorecki, J., Short, T., Lianoglou, S., & Antonyan, E. (2017). Data.table: Extension of 'data.frame'. Retrieved

May 15, 2017, from https://cran.r-project.org/web/packages/data.table/index.html

Eaves, L., & Erkanli, A. (2003). Markov Chain Monte Carlo Approaches to Analysis of Genetic and Environmental Components of Human Developmental Change and G × E Interaction. *Behavior Genetics*, *33*(3), 279–299. https://doi.org/10.1023/A:1023446524917

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics*, *11*, 446. http://dx.doi.org/10.1038/nrg2809

Evans, D. M. (2011). Gene-Gene Interaction and Epistasis. In *Analysis of Complex Disease Association Studies* (pp. 197–213). Elsevier.

Evans, D. M., Gillespie, N. A., & Martin, N. G. (2002). Biometrical genetics. *Biological psychology*, *61*(1-2), 33–51.

Evans, L. M., Tahmasbi, R., Vrieze, S. I., Abecasis, G. R., Das, S., Gazal, S., Bjelland, D. W., De Candia, T. R., Goddard, M. E., Neale, B. M., Yang, J., Visscher, P. M., & Keller, M. C. (2018). Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nature Genetics*, *50*(5), 737–745. https://doi.org/10.1038/s41588-018-0108-x

Falconer, D. S. (1960). *Introduction to quantitative genetics*. Pearson Education Limited.

Fedko, I. O., Hottenga, J.-J., Helmer, Q., Mbarek, H., Huider, F., Amin, N., Beulens, J. W., Bremmer, M. A., Elders, P. J., Galesloot, T. E., Kiemeney, L. A., Van Loo, H. M., Picavet, H. S. J., Rutters, F., Van Der Spek, A., Van De Wiel, A. M., Van Duijn, C., De Geus, E. J. C., Feskens, E. J. M., … Bot, M. (2021). Measurement and genetic architecture of lifetime depression in the Netherlands as assessed by LIDAS (Lifetime Depression Assessment Self-report). *Psychological Medicine*, *51*(8), 1345–1354. https://doi.org/10.1017/S0033291720000100

Feng, J., Chen, C., Cai, Y., Ye, Z., Feng, K., Liu, J., Zhang, L., Yang, Q., Li, A., Sheng, J., Zhu, B., Yu, Z., Chen, C., Dong, Q., & Xue, G. (2020). Partitioning heritability analyses unveil the genetic architecture of human brain multidimensional functional connectivity patterns. *Human Brain Mapping*, *41*(12), 3305–3317. https://doi.org/https://doi.org/10.1002/hbm.25018

Fisher, R. A. (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, *52*(2), 399–433. https://doi.org/10.1017/S0080456800012163

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5). https://doi.org/10.1214/aos/1013203451

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, *24*(1), 44–65.

Gregersen, J. W., Kranc, K. R., Ke, X., Svendsen, P., Madsen, L. S., Thomsen, A. R., Cardon, L. R., Bell, J. I., & Fugger, L. (2006). Functional epistasis on a common MHC haplotype associated with multiple sclerosis. *Nature*, *443*(7111), 574–577. https://doi.org/10.1038/nature05133

Grotzinger, A. D., Rhemtulla, M., De Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., Ip, H. F., Marioni, R. E., McIntosh, A. M., Deary, I. J., Koellinger, P. D., Harden, K. P., Nivard, M. G., & Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour*, *3*(5), 513–525. https://doi.org/10.1038/s41562-019-0566-x

Hare, R. D. (1991). *The Hare psychopathy checklist-revised: Manual*. Multi-Health Systems, Incorporated.

Hare, R. D. (1996). Psychopathy: A Clinical Construct Whose Time Has Come. *Criminal Justice and Behavior*, *23*(1), 25–54. https://doi.org/10.1177/0093854896023001004

He, L., Sillanpää, M. J., Silventoinen, K., Kaprio, J., & Pitkäniemi, J. (2016). Estimating Modifying Effect of Age on Genetic and Environmental Variance Components in Twin Models. *Genetics*, *202*(4), 1313–1328. https://doi.org/10.1534/genetics.115.183905

Heath, A. C., Neale, M. C., Hewitt, J. K., Eaves, L. J., & Fulker, D. W. (1989). Testing structural equation models for twin data using LISREL. *Behavior Genetics*, *19*(1), 9–35. https://doi.org/10.1007/BF01065881

Herzig, A. F., Nutile, T., Ruggiero, D., Ciullo, M., Perdry, H., & Leutenegger, A.-L. (2018). Detecting the dominance component of heritability in isolated and outbred human populations [Number: 1 Publisher: Nature Publishing Group]. *Scientific Reports*, *8*(1), 18048. https://doi.org/10.1038/s41598-018-36050-7

Hill, W. D., Harris, S. E., & Deary, I. J. (2019). What genome-wide association studies reveal about the association between intelligence and mental health. *Current Opinion in Psychology*, *27*, 25–30. https://doi.org/10.1016/j.copsyc.2018.07.007

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits [Publisher: Public Library of Science]. *PLOS Genetics*, *4*(2), e1000008. https://doi.org/10.1371/journal.pgen.1000008

Hohman, T. J., Koran, M. E., Thornton-Wells, T., & for the Alzheimer's Neuroimaging Initiative. (2013). Epistatic Genetic Effects among Alzheimer's Candidate Genes (M. Toft, Ed.). *PLoS ONE*, *8*(11), e80839. https://doi.org/10.1371/journal.pone.0080839

Holzinger, K. J. (1929). The relative effect of nature and nurture influences on twin differences. *Journal of Educational Psychology*, *20*(4), 241–248. https://doi.org/10.1037/h0072484

Hsu, S. D. (2014). On the genetic architecture of intelligence and other quantitative traits. *arXiv preprint arXiv:1408.3421*. https://arxiv.org/pdf/1408.3421.pdf

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jinks, J. L., & Fulker, D. W. (1970). Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of the human behavior. *Psychological Bulletin*, *73*(5), 311–349. https://doi.org/10.1037/h0029135

Keller, M. C., & Coventry, W. L. (2005). Quantifying and Addressing Parameter Indeterminacy in the Classical Twin Design. *Twin Research and Human Genetics*, *8*(3), 201–213. https://doi.org/10.1375/twin.8.3.201

Keller, M. C., Medland, S. E., & Duncan, L. E. (2010). Are Extended Twin Family Designs Worth the Trouble? A Comparison of the Bias, Precision, and Accuracy of Parameters Estimated in Four Twin Family Models. *Behavior Genetics*, *40*(3), 377–393. https://doi.org/10.1007/s10519-009-9320-x

Kempthorne, O. (1978). A Biometrics Invited Paper: Logical, Epistemological and Statistical Aspects of Nature-Nurture Data Interpretation. *Biometrics*, *34*(1), 1–23. https://doi.org/10.2307/2529584

Long, J. S. (1981). Estimation and hypothesis testing in linear models containing measurement error: A review of Jöreskog's model for the analysis of covariance structures. In P. V. Marsden (Ed.), *Linear models in social research* (pp. 209–256). Sage.

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*(1), 185–199. https://doi.org/10.1037/0033-2909.114.1.185

Marchini, J., Donnelly, P., & Cardon, L. R. (2005). Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature genetics*, *37*(4), 413.

Neale, M., & Maes, H. (2004). *Methodology for genetic studies of twins and families*. Virginia Commonwealth University, Department of Psychiatry.

Nichols, R. C. (1965). The national merit twin study. *Methods and goals in human behavior genetic*, 231–244. https://genepi.qimr.edu.au/contents/p/staff/1965_Nichols_Vand_Meth&Goals_231.pdf

Nikolas, M. A., & Burt, S. A. (2010). Genetic and environmental influences on ADHD symptom dimensions of inattention and hyperactivity: A meta-analysis. *Journal of Abnormal Psychology*, *119*(1), 1–17. https://doi.org/10.1037/a0018010

Nolte, I. M., Van Der Most, P. J., Alizadeh, B. Z., De Bakker, P. I., Boezen, H. M., Bruinenberg, M., Franke, L., Van Der Harst, P., Navis, G., Postma, D. S., Rots, M. G., Stolk, R. P., Swertz, M. A., Wolffenbuttel, B. H., Wijmenga, C., & Snieder, H. (2017). Missing heritability: Is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort Study. *European Journal of Human Genetics*, *25*(7), 877–885. https://doi.org/10.1038/ejhg.2017.50

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709. https://doi.org/10.1038/ng.3285

Revelle, W. (2016). *psych: Procedures for Personality and Psychological Research* [Retrieved from https://CRAN.R-project.org/package=psych.

Version 1.6.9]. Northwestern University. Evanston, Illinois, USA.

Revolution Analytics & Weston, S. (2022). doMC: Foreach Parallel Adaptor for 'parallel'. https://CRAN.R-project.org/package=doMC
R package version 1.3.8

Rijsdijk, F. V. (2002). Analytic approaches to twin data using structural equation models. *Briefings in Bioinformatics*, *3*(2), 119–133. https://doi.org/10.1093/bib/3.2.119

Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Parl, F. F., & Moore, J. H. (2001). Multifactor-Dimensionality Reduction Reveals High-Order Interactions among Estrogen-Metabolism Genes in Sporadic Breast Cancer. *The American Journal of Human Genetics*, *69*(1), 138–147. https://doi.org/10.1086/321276

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Schönemann, P. H. (1989). New questions about old heritability estimates. *Bulletin of the Psychonomic Society*, *27*(2), 175–178. https://doi.org/10.3758/BF03329932

Schönemann, P. H. (1997). On models and muddles of heritability. *Genetica*, *99*(2), 97–108. https://doi.org/10.1023/A:1018358504373

Schwabe, I., Janss, L., & Van Den Berg, S. M. (2017). Can We Validate the Results of Twin Studies? A Census-Based Study on the Heritability of Educational Achievement. *Frontiers in Genetics*, *8*, 160. https://doi.org/10.3389/fgene.2017.00160

Shalizi, C. (2007). Yet More on the Heritability and Malleability of IQ [Retrieved September 27, 2007]. http://bactra.org/weblog/520.html

Sniekers, S., Stringer, S., Watanabe, K., Jansen, P. R., Coleman, J. R. I., Krapohl, E., Taskesen, E., Hammerschlag, A. R., Okbay, A., Zabaneh, D., Amin, N., Breen, G., Cesarini, D., Chabris, C. F., Iacono, W. G., Ikram, M. A., Johannesson, M., Koellinger, P., Lee, J. J., ... Posthuma, D. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics*, *49*(7), 1107–1112. https://doi.org/10.1038/ng.3869

Strange, A., Capon, F., Donnelly, P., & Trembath, R. (2010). A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nature Genetics*, *42*, 985–990.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC bioinformatics*, *9*(1), 307. https://doi.org/10.1186/1471-2105-9-307

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology*, *112*(4), 578–598. https://doi.org/10.1037/0021-843X.112.4.578

Van Houtem, C., Laine, M., Boomsma, D., Ligthart, L., Van Wijk, A., & De Jongh, A. (2013). A review and meta-analysis of the heritability of specific phobia subtypes and corresponding fears. *Journal of Anxiety Disorders*, *27*(4), 379–388. https://doi.org/10.1016/j.janxdis.2013.04.007

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*, *101*(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Vitzthum, V. J. (2003). A Number No Greater than the Sum of Its Parts: The Use and Abuse of Heritability. *Human Biology*, *75*(4), 539–558. https://doi.org/10.1353/hub.2003.0064

Wei, W.-H., Hemani, G., & Haley, C. S. (2014). Detecting epistasis in human complex traits. *Nature Reviews Genetics*, *15*(11), 722–733. https://doi.org/10.1038/nrg3747

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). dplyr: A Grammar of Data Manipulation. https://CRAN.R-project.org/package=dplyr
R package version 1.1.2

Wright, M. N., & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, *77*(1), 1–17. https://doi.org/10.18637/jss.v077.i01

Wright, S. (1921). SYSTEMS OF MATING. I. THE BIOMETRIC RELATIONS BETWEEN PARENT AND OFFSPRING. *Genetics*, *6*(2), 111–123. https://doi.org/10.1093/genetics/6.2.111

Zhu, Z., Bakshi, A., Vinkhuyzen, A. A., Hemani, G., Lee, S. H., Nolte, I. M., van Vliet-Ostaptchouk, J. V., Snieder, H., Esko, T., & Milani, L. (2015). Dominance genetic variation contributes little to the missing heritability for human complex traits. *The American Journal of Human Genetics*, *96*(3), 377–385.

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*, *109*(4), 1193–1198. https://doi.org/10.1073/pnas.1119675109

Appendix A

Table A1

Range over architectures of proportion of variance due to additive, dominance, and environment component for the conditions satisfying the inequalities, Scenarios 1 and 2.

	M	Min	Max	Range
Additive	.45	.10	.90	.80
Dominance	.08	.00	.44	.44
Environment	.04	.00	.24	.24

Table A2

Range over architectures of proportion of variance due to additive, dominance, and environment component for the conditions satisfying the inequalities, Scenario 3

	Min	Max	Range
Additive	.10	.90	.80
Dominance	.00	.32	.32
Environment	.00	1	1
Epistasis	.00	.71	.71

Appendix B

On the Identifiability of the ACDE Model

The following answers the question under which of the scenarios realized in our study the four model parameters $\{a, b, c, d\}$ of the ACDE model are identified. The informational input of the three scenarios realized in the study to estimate the model parameters $\{a, b, c, d\}$ are:

- Scenario 1; single-locus with dominance effects and genetic-environment covariance): The informational input is $\{\Sigma_{\text{MZ,T}}, \Sigma_{\text{DZ,T}}, \Sigma_{\text{MZ,A}}, \Sigma_{\text{DZ,A}}\}$.
- Scenario 2; identical architecture as scenario 1 but the informational input is $\{\Sigma_{\text{MZ,T}}, \Sigma_{\text{DZ,T}}\}$.
- Scenario 3 (multi-locus with dominance effects, genetic-environment covariance, and epistatic interactions). The informational input is $\{\Sigma_{\text{MZ,T}}$ $\Sigma_{\text{DZ,T}}, \Sigma_{\text{MZ,A}}, \Sigma_{\text{DZ,A}}\}$.

1. Let $R\{$MZ (monozygotic), DZ (dizygotic) $\}$ and $L\{$ T (reared together), A (reared apart)$\}$;
2. The input into the SEM is $\{\Sigma_{\text{MZ,T}}(\boldsymbol{\theta}), \Sigma_{\text{DZ,T}}(\boldsymbol{\theta}), \Sigma_{\text{MZ,A}}(\boldsymbol{\theta}), \Sigma_{\text{DZ,A}}(\boldsymbol{\theta})\}$;
3. The four model implied covariance matrices are;

$$\Sigma_{MZ,T}(\boldsymbol{\theta}) = \begin{bmatrix} a^2 + c^2 + d^2 + e^2 & a^2 + c^2 + d^2 \\ a^2 + c^2 + d^2 & a^2 + c^2 + d^2 + e^2 \end{bmatrix}$$

$$\Sigma_{\text{MZ,A}}(\boldsymbol{\theta}) = \begin{bmatrix} a^2 + c^2 + d^2 + e^2 & a^2 + c^2 \\ a^2 + c^2 & a^2 + c^2 + d^2 + e^2 \end{bmatrix}$$

$$\Sigma_{\text{DZ,T}}(\boldsymbol{\theta}) = \begin{bmatrix} a^2 + c^2 + d^2 + e^2 & 0.5a^2 + 0.25c^2 + d^2 \\ 0.5a^2 + 0.25c^2 + d^2 & a^2 + c^2 + d^2 + e^2 \end{bmatrix}$$

$$\Sigma_{\text{DZ,A}}(\boldsymbol{\theta}) = \begin{bmatrix} a^2 + c^2 + d^2 + e^2 & 0.5a^2 + 0.25c^2 \\ 0.5a^2 + 0.25c^2 & a^2 + c^2 + d^2 + e^2 \end{bmatrix}$$

4. Question; can the four model parameters $\{a, b, c, d\}$ be expressed as functions of the $4 \times 3 = 12$ input parameters; i.e., with respect the input (2) and the multi-population model (3), are the model parameters identified.
5. Answer: Yes.
6. Proof.

Let the total variance be $T = a^2 + c^2 + d^2 + e^2$.

6i. $\frac{d^2}{T} = \rho_{MZ,T} - \rho_{MZ,A} = \frac{\sigma_{MZ,T[1,2]}}{\sqrt{\sigma_{MZ,T[1,1]}\sigma_{MZ,T[2,2]}}} - \frac{\sigma_{MZ,T[1,2]}}{\sqrt{\sigma_{MZ,A[1,1]}\sigma_{MZ,A[2,2]}}}$

$$\begin{aligned} & a^2 + c^2 + d^2 - \left(a^2 + c^2\right) \\ &= a^2 + c^2 + d^2 - 1\left(a^2 + c^2\right) \\ &= a^2 + c^2 + d^2 - a^2 - c^2 \\ &= \left(a^2 - a^2\right) + \left(c^2 - c^2\right) + d^2 \\ &= d^2 \end{aligned}$$

6ii. $\frac{c^2}{T} = 2\left(\rho_{MZ,A} - 2\rho_{MZ,A}\right) = 2\left(\frac{\sigma_{MZ,A[1,2]}}{\sqrt{\sigma_{MZ,A[1,1]}\sigma_{MZ,A[2,2]}}} - 2\frac{\sigma_{DZ,A[1,2]}}{\sqrt{\sigma_{DZ,A[1,1]}\sigma_{DZ,A[2,2]}}}\right)$

Hence,

$$\begin{aligned} & a^2 + c^2 - 2\left(.5a^2 + .25c^2\right) \\ &= a^2 + c^2 - a^2 - .5c^2 \\ &= \left(a^2 - a^2\right) - \left(c^2 - .5c^2\right) \\ &= .5c^2 \end{aligned}$$

and therefore $c^2 = 2 \cdot .5c^2$.

6iii. $\frac{a^2}{T} = 4\rho_{DZ,A} - \rho_{MZ,A} = 4\left(\frac{\sigma_{DZ,A[1,2]}}{\sqrt{\sigma_{DZ,A[1,1]}\sigma_{DZ,A[2,2]}}} - \frac{\sigma_{MZ,A[1,2]}}{\sqrt{\sigma_{MZ,A[1,1]}\sigma_{MZ,A[2,2]}}}\right)$

note, by the way, that this is h^2.
Hence,

$$\begin{aligned} & 4\left(.5a^2 + .25c^2\right) - \left(a^2 + c^2\right) \\ &= 2a^2 + b^2 - a^2 - b^2 \\ &= a^2 \end{aligned}$$

6iv . $\frac{e^2}{T} = 1 - \frac{a^2}{T} + \frac{c^2}{T} + \frac{d^2}{T}$
6v.T is obviously overidentified, as, under the model, it is equal to 8 different input parameters (i.e., the $2 \times 4 = 8$ distinct variances).

$\therefore$ up to an indeterminacy in the signs of $\{a, b, c, d\}$ [standard for all linear factor structures], the 4 model parameters are identified; ergo, the ACDE model is identified for the input $\{\Sigma_{\text{MZ,T}}(\boldsymbol{\theta}), \Sigma_{\text{DZ,T}}(\boldsymbol{\theta}), \Sigma_{\text{MZ,A}}(\boldsymbol{\theta}), \Sigma_{\text{DZ,A}}(\boldsymbol{\theta})\}$. Hence, the ACDE model is identified under scenarios 1 and 3 . It is, however, not identified under scenario 2 as we can see in the following.

Question: Can the four model parameters $\{a, b, c, d\}$ be expressed as functions of the $2 \times 3 = 6$ input parameters under scenario 2; i.e., with respect the input (2)?

7. Answer: No.

8. Proof.

Let the total variance be $T = a^2 + c^2 + d^2 + e^2$.
The implied covariance matrices of study 2 are

8i. $\Sigma_{MZ,T}(\boldsymbol{\theta}) = \begin{bmatrix} T & a^2 + c^2 + d^2 \\ a^2 + c^2 + d^2 & \text{T} \end{bmatrix}$

8ii. $\Sigma_{\text{DZ,T}}(\boldsymbol{\theta}) = \begin{bmatrix} \text{T} & .5a^2 + .25c^2 + d^2 \\ .5a^2 + .25c^2 + d^2 & \text{T} \end{bmatrix}$

Now,
8iii. $e^2 = \sigma^2 - \sigma_{MZ,T[1,2]}$,
whereby σ^2 is the common environmental variance.

Regarding the identifiability of c^2 and d^2 we get:
8iv. $\frac{c^2}{T} = \rho_{MZ,T} - 2\rho_{DZ,T} = \sigma_{MZ,T[1,2]} - 2\sigma_{DZ,T[1,2]}$
Hence, in terms of model parameters

$$\begin{aligned} & \text{8v. } a^2 + c^2 + d^2 - 2\left(.5a^2 + .25c^2 + d^2\right) \\ &= a^2 + c^2 + d^2 - \left(a^2 + .5c^2 + 2d^2\right) \\ &= a^2 + c^2 + d^2 - a^2 - .5c^2 - 2d^2 \\ &= \left(a^2 - a^2\right) + \left(c^2 - .5c^2\right) + \left(d^2 - 2d^2\right) \\ &= .5c^2 - d^2 \end{aligned}$$

8vi. T is obviously underidentified, as, under the model, there is no unique expression for c^2. Hence, the ACDE model is not identified with $\{\Sigma_{\text{MZ,T}}(\boldsymbol{\theta}), \Sigma_{\text{DZ,T}}(\boldsymbol{\theta})\}$.