Received: 15 April 2020 | Revised: 14 October 2020 | Accepted: 21 November 2020

DOI: 10.1002/sim.8845

RESEARCH ARTICLE

Statistics in Medicine WILEY

Heritability curves: A local measure of heritability in family models

Geir D. Berentsen[1] | **Francesca Azzolini**[2] | **Hans J. Skaug**[2] | **Rolv T. Lie**[3,4] | **Håkon K. Gjessing**[3,4]

[1]Department of Business and Management Science, NHH Norwegian School of Economics, Bergen, Norway

[2]Department of Mathematics, University of Bergen, Bergen, Norway

[3]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway

[4]Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway

Correspondence
Francesca Azzolini, Department of Mathematics, University of Bergen, P.O. Box 7803, Bergen 5020, Norway.
Email: francesca.azzolini.47@gmail.com

Funding information
Centres of Excellence, Grant/Award Number: 262700; Norges Forskningsråd, Grant/Award Number: 225912/F50 "Health Registries for Research"

Classical heritability models for family data split the phenotype variance into genetic and environmental components. For instance, the ACE model in twin studies assumes the phenotype variance decomposes as $a^2 + c^2 + e^2$, representing (additive) genetic effects, common (shared) environment, and residual environment, respectively. However, for some phenotypes it is biologically plausible that the genetic and environmental components may vary over the range of the phenotype. For instance, very large or small values of the phenotype may be caused by "sporadic" environmental factors, whereas the mid-range phenotype variation may be more under the control of common genetic factors. This article introduces a "local" measure of heritability, where the genetic and environmental components are allowed to depend on the value of the phenotype itself. Our starting point is a general formula for local correlation between two random variables. For estimation purposes, we use a multivariate Gaussian mixture, which is able to capture nonlinear dependence and respects certain distributional constraints. We derive an analytical expression for the associated correlation curve, and show how to decompose the correlation curve into genetic and environmental parts, for instance, $a^2(y) + c^2(y) + e^2(y)$ for the ACE model, where we estimate the components as functions of the phenotype y. Furthermore, our model allows switching, for instance, from the ACE model to the ADE model within the range of the same phenotype. When applied to birth weight (BW) data on Norwegian mother-father-child trios, we conclude from the model that low and high BW are less heritable traits than medium BW. We also demonstrate switching between the ACE and ADE model when studying body mass index in adult monozygotic and dizygotic twins.

KEYWORDS
correlation curve, heritability, multivariate Gaussian mixture, twin studies

1 | INTRODUCTION

Biometrical modeling of family trait correlations has a very long tradition, going back at least to Fisher[1] and Wright,[2,3] and being developed into an extensive modeling framework over the years,[4,5] with openly available software tools, such as OpenMx.[6] For a continuous trait Y, such as weight or height, the basic idea is that trait variability—or more precisely, the

variance of the measured trait, Var(Y)—can be decomposed into genetic and environmental components, each explaining a portion of the observed trait variance. Thus, the concept of *heritability* can, loosely, be defined as the proportion of trait variance explained by genetic components, with environmental influences assumed to explain the rest.[7] As an example, the most common twin model, known as the ACE model, decomposes the trait Y into additive genetic effects (A), common (shared) environment (C), and residual (random) environment (E). In terms of variances, we commonly define quantities a^2, c^2, and e^2 as the *proportions* of trait variances explained by the components A, C, and E, respectively . Thus, assuming that no other effects are present, we have $a^2 + c^2 + e^2 = 1$.

To separate genetic variance from environmental variance, family data are needed. Genetic correlations between family members decrease in more distant relationships, thus providing contrasts from which the genetic components can be estimated. For instance, in the classical ACE twin design, the additive genetic correlation in monozygotic twin pairs is assumed to be 1, whereas the corresponding correlation, or degree of shared genetic influence, is assumed to be 1/2 in dizygotic twin pairs. In addition, it is frequently assumed that the amount of shared environment is the same in dizygotic twins as is monozygotic twins. The quantities a and c above can also be seen as the degree to which the underlying genes A and shared environmental C are being "expressed" in the phenotype of each individual. Thus, the monozygotic twin pair phenotype correlation will be $\rho^{(\mathrm{MZ})} = a^2 + c^2$, and $\rho^{(\mathrm{DZ})} = \frac{1}{2}a^2 + c^2$ for the dizygotic twin pairs. As a consequence, the difference $\frac{1}{2}a^2$ between monozygotic and dizygotic twin pair correlations is ascribed to genes alone, providing an estimate of the heritability a^2.

The ACE model is very specific in its assumption of additive genetic effects, as well as independent, additive contributions from the environment. In the biometrical modeling literature, a wide range of variants and extensions have been developed. Using family structures of increasing complexity, numerous different effects can be identified, such as additive genetic effects, dominant genetic effects, X-chromosome effects, effects of maternal genes on the fetus during pregnancy, effects of mitochondrial genes, gene-gene interactions, gene-environment interactions, and so on.[5,8,9] Extending the family structures used for modeling is in general challenging since genetic correlations between more distant relatives quickly drop to nearly undetectable levels, and assumptions about how environmental factors are shared within larger families become harder to verify.[9] Still, with a steady increase in registry-based population studies with large sample sizes and available data on environmental covariates, such modeling has become feasible.

Common to practically all models in the field is that the degree of heritability is assumed constant across the full range of the phenotype. For instance, the estimated proportion a^2 of variance explained by additive genes is assumed to be the same whether the phenotype Y is small, close to its mean, or large. It seems clear, however that, for instance, rare but dramatic environmental influences on the phenotype may occasionally cause the phenotype to deviate strongly from its mean value, much more than would be expected under "normal" circumstances. Below, we illustrate our models of heritability using a child's birth weight (BW) as phenotype. While the BW distribution is close to a normal distribution, it has a heavier tail to the left (Figure B1); this may indicate a higher proportion of low BW children than what would be expected from many minor genetic and environmental components adding up during pregnancy. This simple observation may suggest that the degree of heritability of BW can differ in the different ranges of weight; perhaps the lowest BW values are caused by "rouge" environmental factors that act more strongly than genetic effects in the tail, or maybe they are caused by rare, recessive genes that only occasionally exert a strong negative influence on BW.

These observations motivate us to look for differences in heritability across the range of the trait value Y. The existing methods for investigating such differences are almost exclusively based on regression methods and twin-studies. In their seminal work,[10] DeFries and Fulker evaluate the degree of regression to the mean for cotwins of probands from strata in the tails of a continuous trait distribution. The idea is that if the trait is heritable, then we should observe DZ cotwins with a higher degree of regression to the mean compared with the MZ cotwins. This approach is known as DeFries-Fulker (DF) extremes analysis for twins. Later, a formal test was developed to examine whether the heritability of the trait for probands in the selected strata was equal or different to the unselected population.[11] This methodology was extended by Cherny et al[12] by considering interaction effects between the heritability of the trait and the realized value of the trait for the proband. This approach can be used to detect linear and quadratic changes in heritability as the trait value changes. These methods all have the drawback of only providing a rough description of how the heritability varies with the trait value. The DF approach requires the researcher to select a cut-off point (a low or high trait value) for choosing the strata; the result can thus be misleading if the heritability changes smoothly as the trait value vary. Conversely, if there exists a point in the trait distribution where the heritability jumps and then stabilizes again, the Cherny approach will only model this change by a linear or quadratic curve.

These drawbacks were addressed in Reference 13 using quantile regression; by using the extended DF extremes analysis[14] as the quantile regression equation, the authors obtain a heritability measure for each quantile of the trait

distribution. Consequently, their method results in a heritability measure for each value of the trait $Y = y$, corresponding to a specific quantile of the distribution.

However, in the present article we introduce an approach based on localizing traditional genetic models. Informally, this means making sense of estimating, for instance, the additive genetic effect as a function of the phenotype; that is, to define in a meaningful way $a^2(y)$ as the proportion of phenotype variance explained by additive genetic effects, conditional on $Y = y$. Such a definition may seem self-contradictory since one conditions on the variable whose variance is being decomposed. Nevertheless, it is fully possible to make sense of this concept, and we show in this article how to develop *heritability curves*, such as $a^2(y)$. This definition thus provides a "local" measure of heritability, depending on the phenotype value.

As for the ACE twin model, all standard biometrical models rely on the phenotype correlations between family individuals to estimate the variance components that determine heritability. Our starting point for developing a local measure of heritability is thus a local measure of dependence between family members; more specifically, we need a local measure of correlation. There are several local measures proposed in the literature, such as the local Gaussian correlation,[15] the dependence function,[16] and the correlation curve.[17] We base our approach on the correlation curve[17] $\rho(y)$, which can be defined as a measure of locally explained variance, and thus fits the framework of heritability as a proportion of explained variance. The correlation curve is similar to the traditional Pearson's correlation in that it takes values between minus one and one, and the square $\rho^2(y)$ is a measure of locally explained variance. In a bivariate Gaussian distribution, the correlation curve is constant (independent of y), and equal to the standard Pearson correlation. By contrast to the Pearson correlation the local correlation of a bivariate relationship depends on direction; for a bivariate random variable (Y_1, Y_2), the locally explained variance of Y_2 conditional on $Y_1 = y$ may differ from the locally explained variance of Y_1 conditional on $Y_2 = y$.

With phenotype measurements on, for instance, a mother (Y_1) and her child (Y_2), it may seem reasonable, for instance, to study the distribution of a child phenotype conditionally on the maternal phenotype. However, most biometrical models are formulated in terms of genetic and environmental factors *shared* by the two family members, thus assuming a form of exchangeability between the two. This is particularly clear in twin pairs, where conditioning one twin on the other twin is unnatural. In the model of Logan et al[13] this assignment was done randomly, while[12] explored both a random assignment and a double-entry approach. However, the population value of the correlation curve can be derived from the joint distribution of two variables. If the joint distribution is exchangeable, so that (Y_1, Y_2) has the same bivariate distribution as (Y_2, Y_1), the correlation curve is invariant to which variable we condition on, that is, whether we measure the locally explained variance of Y_1 conditional on Y_2 or vice versa. This means that the role of the mother and child in the above interpretation can be interchanged.

The correlation curve may be estimated parametrically or nonparametrically from observed values of a bivariate distribution (Y_1, Y_2) by conditioning on either $Y_1 = y$ or $Y_2 = y$. However, our approach is instead to first model the bivariate distribution as a Gaussian mixture distribution, where the mixture distribution is restricted in such a way as to be exchangeable. From the mixture distribution, the correlation curve can be derived explicitly. We estimate the distribution by maximum likelihood, and by allowing a sufficient number of components, a mixture distribution is very flexible and fits a wide range of distributional shapes. Having obtained the parameters of the mixture distribution, the correlation curve can be derived from its explicit expression by plugging in the estimated parameters.

The article is structured as follows. In Section 2, we define a standard mixed-effect model for continuous traits, and structure it for two specific family models: twin pairs and mother-father-child trios. Following a standard twin approach,[18] and models for family trios,[19,20] we derive expressions for the heritability estimates in both family structures. In Subsections 2.2 and 2.3, we explain the concept of correlation curves, and extend the traditional definition of heritability to the heritability curve, which depends on the trait value y. In Section 3, we introduce and analyze a Gaussian mixture[21] for bivariate phenotype distributions, parameterized to be exchangeable. We then study the limiting behavior of the correlation curve for large and small phenotype values under this model in Subsection 3.1. Finally, in Subsection 3.2, we discuss the estimation of the correlation curve for the twin-pairs and the mother-father-child trios models. Section 4 provides two applications of this approach. Namely, the first application is the analysis of body mass index (BMI) values for twin pairs collected in the dataset "twinData," found in the R-package "OpenMx";[6] the second one is the analysis of BW data of mother-father-child trios from the Medical Birth Registry of Norway. For both family structures we compute AIC and BIC values to select the best-fitting mixture models, and explore the resulting distributions and heritability curves. Section 5 outlines model extensions to more complex family structure, and also demonstrates switching between different heritability models within the same family structure. Proofs are provided in an appendix.

2 | DEVELOPMENT OF HERITABILITY CURVES

2.1 | Traditional models for twins and family trios

We first provide a basic description of how traditional biometrical models can be set up in some generality, and in particular for twins and family trios. While there are numerous ways of building, parametrizing, and interpreting such models, our approach is fairly standard, and in a form that supports our development of heritability curves. Let Y_{ij} be the trait value of individual j in a family i, and consider the mixed-effect model (see, eg, Reference 22)

$$Y_{ij} = \mu + \sum_{s=1}^{S} \alpha^{(s)} x_{ij}^{(s)} + A_{ij} + C_{ij} + D_{ij} + E_{ij}, \tag{1}$$

where A_{ij}, C_{ij}, D_{ij}, and E_{ij} represent additive genetic, common environmental, dominant genetic, and residual environmental random effects, respectively (see, eg, Reference 18). We assume the four components A_{ij}, C_{ij}, D_{ij}, and E_{ij} to be mutually independent, with mean 0 and variances σ_A^2, σ_C^2, σ_D^2, and σ_E^2. We include S continuous covariates $x_{ij}^{(s)}$, and the terms $\alpha^{(s)} x_{ij}^{(s)}$ (fixed effects) allow the average phenotype level to depend on the covariates. Note that this model assumes no gene-environment interaction. In traditional biometrical modeling (see, eg, Reference 9) the random effects are assumed to be normally distributed, that is, $A_{ij} \sim N(0, \sigma_A^2)$, $C_{ij} \sim N(0, \sigma_C^2)$, $D_{ij} \sim N(0, \sigma_D^2)$, and $E_{ij} \sim N(0, \sigma_E^2)$. The assumption of normality is seen as natural based on the central limit theorem if Y is the result of numerous small, independent genetic and environmental effects that add up to produce the trait value. Under the above assumptions the total variance of the trait is given by

$$\sigma^2 = \mathrm{Var}(Y_{ij}) = \sigma_A^2 + \sigma_C^2 + \sigma_D^2 + \sigma_E^2. \tag{2}$$

We define $a^2 = \sigma_A^2/\sigma^2$, $c^2 = \sigma_C^2/\sigma^2$, $d^2 = \sigma_D^2/\sigma^2$, and $e^2 = \sigma_E^2/\sigma^2$ as the proportions of the total variance that derive from each of the four genetic and environmental components. Note that

$$a^2 + c^2 + d^2 + e^2 = 1,$$

that is, the contributions from all components sum to one. Thus, in a model including A, C, and E, excluding dominant effects, one may quantify the genes-vs-environment contribution to trait variability as a^2. This proportion is often referred to as *heritability* and can be interpreted as how strongly the genetic effect A_{ij} contributes to the trait value. The heritability based on the additive genetic component is often referred to as *narrow sense heritability*. Some models may also include dominant genetic effects, and in such cases one may refer to $a^2 + d^2$ as the *broad sense heritability*.[23]

From independent observations of Y_{ij} alone, it is not possible to identify the individual variance components σ_A^2, σ_C^2, σ_D^2, and σ_E^2 in (2), only the total variance σ^2. In order to make the individual variances identifiable, one has to consider data on family members, for which the Y's are correlated due to shared genetic material and environment. We focus on two basic family structures—mother-father-child trios and twin pairs—in the following. As is well known, these family structures are quite restricted in the number of effects they allow to be estimated, and assumptions have to be made about what genetic and environmental effects to include in each model. In the following, we will focus on these two models as illustrations when developing heritability curves, although the principles are generally valid for more complex family structures as well.

2.1.1 | Twins

Perhaps the best known biometrical model is the ACE model for twins, complemented by the alternative ADE model. While the expressions for twin correlations in these models are very well known, we state them here as a starting point for the heritability curves.

Let Y_{ij} be the trait value of twin j ($j = 1, 2$) in twin-pair i. Let $\rho^{(\mathrm{MZ})}$ and $\rho^{(\mathrm{DZ})}$ be the phenotype correlations $\mathrm{cor}(Y_{i1}, Y_{i2})$ for MZ and DZ twins, respectively. Both ACE and ADE models include the additive genetic component A. For MZ-twins $\mathrm{cor}(A_{i1}, A_{i2}) = 1$, while for DZ-twins $\mathrm{cor}(A_{i1}, A_{i2}) = 1/2$. In the standard ACE model, the correlation for the common environmental effect is assumed to be $\mathrm{cor}(C_{i1}, C_{i2}) = 1$ in all twin pairs; thus, one makes the common assumption of DZ twins

sharing their environment to the same degree as the MZ twins. In the alternative ADE one assumes $\text{cor}(D_{i1}, D_{i2}) = 1$ for MZ twins and $\text{cor}(D_{i1}, D_{i2}) = 1/4$ for DZ twins. In both models, residual environmental effects are assumed to be independent.

Since the basic twin models utilize only the $\rho^{(\text{MZ})}$ and $\rho^{(\text{DZ})}$ phenotype correlations, they allow estimating two parameters. In addition, e^2 can be estimated from $e^2 = 1 - a^2 - c^2 - d^2$. The ACE model assumes $d^2 = 0$, and thus the parameters a^2, c^2, and e^2 can be identified; the ADE model assumes $c^2 = 0$, and thus the parameters a^2, d^2, and e^2 can be identified.

For the ACE model, it follows from the above that

$$\begin{aligned}\rho^{(\text{MZ})} &= a^2 + c^2,\\ \rho^{(\text{DZ})} &= \frac{1}{2}a^2 + c^2.\end{aligned}$$

For the ADE model, the equations are

$$\begin{aligned}\rho^{(\text{MZ})} &= a^2 + d^2,\\ \rho^{(\text{DZ})} &= \frac{1}{2}a^2 + \frac{1}{4}d^2.\end{aligned}$$

The simplest approach to estimating a^2, c^2, and d^2 is by moment estimators, that is, to solve this set of equations, using empirical values for $\rho^{(\text{MZ})}$ and $\rho^{(\text{DZ})}$, and use $e^2 = 1 - a^2 - c^2 - d^2$ to estimate e^2. The resulting solutions for the ACE model are the celebrated formulas of Falconer:[18]

$$\begin{aligned}a^2 &= 2(\rho^{(\text{MZ})} - \rho^{(\text{DZ})}),\\ c^2 &= 2\rho^{(\text{DZ})} - \rho^{(\text{MZ})},\\ e^2 &= 1 - \rho^{(\text{MZ})}.\end{aligned} \tag{3}$$

For the ADE model, the corresponding set of solutions are

$$\begin{aligned}a^2 &= 4\rho^{(\text{DZ})} - \rho^{(\text{MZ})},\\ d^2 &= 2(\rho^{(\text{MZ})} - 2\rho^{(\text{DZ})}),\\ e^2 &= 1 - \rho^{(\text{MZ})}.\end{aligned} \tag{4}$$

Without further assumptions, an informal choice between the ACE and ADE models is often made based on whether empirically $\rho^{(\text{MZ})} < 2\rho^{(\text{DZ})}$ or not. If this is the case, the ACE model is a natural choice; otherwise, the ADE model can be used. In 5.2 we will demonstrate that with local heritability models, one can actually switch between the ACE and ADE models within the same phenotype analysis.

2.1.2 | Mother-father-child trios

Let Y_{ij} be the observed trait value of individual j in nuclear family trio i. We let $j = 1, 2, 3$ correspond to the mother, father, and child, respectively. A phenotype correlation between mother and father may signify, for instance, assortative mating, inbreeding, or social homogamy among the parents. However, the correlation is typically low, and we will here assume it is zero.[19] There are thus only two correlations that provide information: the mother-child and father-child correlations. There are numerous ways of parametrizing correlations in nuclear families,[9,19,20,24,25] but being restricted to two correlations means that these cannot be separated. In our setting, we assume, for additive autosomal genes, that $\text{cor}(A_{i1}, A_{i3}) = \text{cor}(A_{i2}, A_{i3}) = 1/2$, and that $\text{cor}(A_{i1}A_{i2}) = 0$ for the parents. In addition, we assume that mother and child share an environmental component, but no such sharing between father and child, leading to $\text{cor}(C_{i1}, C_{i3}) = 1$ and $\text{cor}(C_{i2}, C_{i3}) = 0$. Thus,

$$\Sigma_A = \sigma_A^2 \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{bmatrix}, \quad \Sigma_C = \sigma_C^2 \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad \text{and} \quad \Sigma_E = \sigma_E^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

are the covariance matrices for the vectors (A_{i1}, A_{i2}, A_{i3}), (C_{i1}, C_{i2}, C_{i3}), and (E_{i1}, E_{i2}, E_{i3}), respectively.

A graphical representation of the above model is displayed in a path diagram in Figure B2. Under the above assumptions the vectors (Y_{i1}, Y_{i2}, Y_{i3}) are independent multivariate normal with mean

$$\left(\mu + \sum_{s=1}^{S}\alpha^{(s)}x_{i1}, \mu + \sum_{s=1}^{S}\alpha^{(s)}x_{i2}, \mu + \sum_{s=1}^{S}\alpha^{(s)}x_{i3}\right), \tag{5}$$

or simply (μ, μ, μ) if covariates are not present, and covariance matrix

$$\Sigma = \Sigma_A + \Sigma_C + \Sigma_E = (\sigma_A^2 + \sigma_C^2 + \sigma_E^2)\begin{bmatrix} 1 & 0 & \frac{1}{2}a^2 + c^2 \\ 0 & 1 & \frac{1}{2}a^2 \\ \frac{1}{2}a^2 + c^2 & \frac{1}{2}a^2 & 1 \end{bmatrix}, \tag{6}$$

where a^2, c^2, and e^2 are defined as above. Again, the unknown values can simply be estimated by the methods of moments by matching the correlation matrix (6) to its empirical counterpart, and solve for a^2, c^2, and e^2 under the constraint $a^2 + c^2 + e^2 = 1$. The solution is given by the following equations

$$\begin{aligned} a^2 &= 2\rho^{(\mathrm{FC})} \\ c^2 &= \rho^{(\mathrm{MC})} - \rho^{(\mathrm{FC})} \\ e^2 &= 1 - \rho^{(\mathrm{MC})} - \rho^{(\mathrm{FC})}, \end{aligned} \tag{7}$$

where $\rho^{(\mathrm{MC})}$ and $\rho^{(\mathrm{FC})}$ are the mother-child and father-child correlations, respectively.

We will, in the following, use these solutions, and those for the twin models, to obtain local versions of a^2, c^2, d^2, and e^2. Note that in both cases, the underlying assumption is that the covariance (correlation) matrix completely characterizes the dependence structure between traits in a family and can be decomposed as in (6).

2.2 | Correlation curves for nonlinear bivariate relationships

We now explain the concept of local correlation curves, following the approach of Bjerve and Doksum.[17] To illustrate the principle of localization, we use simulated data from a hypothetical phenotype, as seen in Figure B3A. We consider two strata (A and B) consisting of all mother-child pairs for which the mother's trait $Y_1 = y_1$ falls within two intervals (interval A and B) on the x-axis. The corresponding correlation curve is shown in Figure B3B; as a function of y_1 (horizontal axis) it is smaller in stratum A than in stratum B. This indicates that the mother-child association is stronger in stratum B compared with stratum A. In a nonparametric regression setting, this would mean that the child's trait can be predicted by the mother's trait with higher precision in stratum B than in stratum A. For both strata, an increase in the mother's trait is associated with an increase in the child's trait since the correlation curve is positive. Since the correlation curve is continuous, the location argument y_1 can be seen as the center of infinitesimal intervals from which strata such as A and B can be constructed, while the value of the correlation curve is a measure of dependence for the corresponding strata. A constant correlation curve indicates that the dependence properties are constant across these strata, while a varying correlation curve indicates strata that differ in their dependence properties.

If the joint distribution is exchangeable, so that (Y_1, Y_2) has the same bivariate distribution as (Y_2, Y_1), the correlation curve is invariant to which variable we condition on, that is, whether we measure the locally explained variance of Y_1 conditional on Y_2 or vice versa. This means that the role of the mother and child in the above interpretation can be interchanged, and the dependence structure in strata A^* and B^* in Figure B3A is similar to the dependence structure in strata A and B; the correlation curve $\rho(y)$ as a function of y thus represents a measure of the mother-child trait dependence when either the mother or the child has trait value equal to y. In the following section, we show more precisely how $\rho(y)$ is defined in terms of locally explained variance.

2.2.1 | Standard correlation curves for bivariate relationships

Let (Y_1, Y_2) be random variables from a bivariate continuous distribution, and define $\tau_1^2 = \mathrm{Var}(Y_1)$, $\tau_2^2 = \mathrm{Var}(Y_2)$, and $\rho = \mathrm{cor}(Y_1, Y_2)$. Furthermore, define $\mu(y) = \mathrm{E}(Y_1|Y_2 = y)$ and $\sigma^2(y) = \mathrm{Var}(Y_1|Y_2 = y)$ as functions of y. Assuming that

$\mu(y)$ is differentiable, define $\beta(y) = \mu'(y)$, that is, the slope of the (typically nonlinear) regression curve $\mu(y)$ when Y_1 is regressed on Y_2. Recall that in a standard linear regression context, $\mu(y)$ is a linear function of y, where the slope $\beta_{1|2} := \beta(y)$ and the conditional variance $\sigma^2_{1|2} := \sigma^2(y)$ are both constant.

By the law of total variance,

$$\mathrm{Var}(Y_1) = \mathrm{Var}(\mathrm{E}(Y_1|Y_2)) + \mathrm{E}(\mathrm{Var}(Y_1|Y_2)),$$

and it thus seems natural to define in general

$$\text{Proportion of } \mathrm{Var}(Y_1) \text{ explained by } Y_2 = \frac{\mathrm{Var}(\mathrm{E}(Y_1|Y_2))}{\mathrm{Var}(\mathrm{E}(Y_1|Y_2)) + \mathrm{E}(\mathrm{Var}(Y_1|Y_2))}.$$

In the case of linear regression, $\mathrm{Var}(\mathrm{E}(Y_1|Y_2)) = \tau_2^2\beta_{1|2}^2$ and $\mathrm{E}(\mathrm{Var}(Y_1|Y_2)) = \sigma^2_{1|2}$, and the proportion of explained variance can thus be written

$$\frac{(\tau_2\beta_{1|2})^2}{(\tau_2\beta_{1|2})^2 + \sigma^2_{1|2}} = \left(\frac{\tau_2\beta_{1|2}}{\tau_1}\right)^2 = \rho^2, \tag{8}$$

which is the usual formula for explained variance in a linear regression.

We want to define a "local" variant of ρ^2, describing the proportion of explained variance when $Y_2 = y$, thus to define $\rho^2(y)$ as a function of y. To this end, (8) is a natural starting point, and the extension to a nonlinear setting would thus be to allow both $\beta(y)$ and $\sigma^2(y)$ to depend on y. This leads to the definition

$$\rho(y) = \frac{\tau_2\beta(y)}{[(\tau_2\beta(y))^2 + \sigma^2(y)]^{1/2}}, \tag{9}$$

where we recall that $\tau_2^2 = \mathrm{Var}(Y_2)$, $\beta(y) = \frac{d}{dy}\,\mathrm{E}(Y_1|Y_2 = y)$, and $\sigma^2(y) = \mathrm{Var}(Y_1|Y_2 = y)$.

Indeed, this is the formula developed by Bjerve and Doksum[17] and Doksum et al.[26] As pointed out by Bjerve et al, the correlation curve should not be confused with the conditional correlation obtained by applying the usual correlation formula to the conditional distribution of (Y_1, Y_2) given $Y_2 = y$, which would always be zero. It should also be noted that while τ_2 is kept fixed in (9), the denominator $(\tau_2\beta(y))^2 + \sigma^2(y)$ is no longer necessarily equal to $\tau_1^2 = \mathrm{Var}(Y_1)$ from the original distribution. In fact, for a *fixed* $y = y_0$, it corresponds to $\mathrm{Var}(Z_1)$ from a hypothetical bivariate distribution (Z_1, Z_2) where $\mathrm{Var}(Z_2) = \tau_2^2$ and $\mathrm{Var}(Z_1)$ is determined from having a linear regression of Z_1 on Z_2 with constant slope $\beta(y_0)$ and constant conditional variance $\mathrm{Var}(Z_1|Z_2) = \sigma^2(y_0)$.

2.2.2 | Correlation curves for symmetric bivariate relationships

In our setting, we are interested in relationships between pairs of family members, for example, a pair of twins or a child and a parent. We denote the pair's respective trait values by Y_1 and Y_2. At first glance, it may seem natural to ask about the explained variation of a child trait Y_1, conditional on its parental value Y_2. However, this is less natural for twins, who are from the same generation. Indeed, most biometrical models assume that the positive correlation between the trait values is generated by shared genes and shared environment; the sharing is symmetrical between family members, and the generational aspect is only used to compute the degree of relatedness. That is, in pairs of family members, the two members should be exchangeable, so that (Y_1, Y_2) and (Y_2, Y_1) have the same bivariate distribution. Clearly, this means that when applying (9) in a heritability setting, it would be reasonable to expect that Y_1 conditional on Y_2 should provide the same answers as Y_2 conditional on Y_1. While exchangeability is obviously not the case for general bivariate distributions, we achieve pairwise exchangeability by a corresponding restriction of our parametric models for the bivariate distributions, as described later. When including covariates, the assumption of pairwise exchangeability should apply to the residuals, that is, the covariate-adjusted traits $Y_1 - \sum_{s=1}^{S}\alpha^{(s)}x_1^{(s)}$ and $Y_2 - \sum_{s=1}^{S}\alpha^{(s)}x_2^{(s)}$.

Note that it would suffice to assume that, for all y,

$$\begin{aligned}
\tau_1^2 = \mathrm{Var}(Y_1) = \mathrm{Var}(Y_2) = \tau_2^2 &=: \sigma,\\
\mathrm{E}(Y_1|Y_2 = y) = \mathrm{E}(Y_2|Y_1 = y) &=: \mu(y),\\
\mathrm{Var}(Y_2|Y_1 = y) = \mathrm{Var}(Y_1|Y_2 = y) &=: \sigma^2(y),
\end{aligned} \tag{10}$$

since this would imply that (9) would be invariant to the direction of conditioning. However, the models presented in this article all imply full pairwise exchangeability. We do *not*, however, ask for full exchangeability of the multivariate outcome distribution; for instance, a mother-father-child trio would clearly not have the same trivariate distribution as a child-father-mother trio. Nevertheless, the pairwise exchangeability implies that all family members have the same marginal distributions. The appropriateness of the exchangeability assumptions will be addressed in the Discussion.

2.3 | Heritability curves

Assuming $\rho(y)$ to be well defined for the joint distribution of the two family members, we are interested in the degree to which the value of $\rho(y)$ can be attributed to heritability on one side, and to environment on the other. In particular, we are interested in knowing how these contributions vary with y.

Definition 1 (Heritability curve for the twin ADE model). Assume the exchangeability property (10) holds for both MZ and DZ bivariate distributions. Adopting the moment Equation (4), we define the heritability curve by

$$a^2(y) = 4\rho^{(\mathrm{DZ})}(y) - \rho^{(\mathrm{MZ})}(y), \tag{11}$$

where $\rho^{(\mathrm{MZ})}(y)$ and $\rho^{(\mathrm{DZ})}(y)$ are the correlation curves of MZ and DZ twins calculated according to (9). Similarly, (4) allows local versions of the dominance effect

$$d^2(y) = 2[(\rho^{(\mathrm{MZ})}(y) - 2\rho^{(\mathrm{DZ})}(y)] \tag{12}$$

and residual environment

$$e^2(y) = 1 - \rho^{(\mathrm{MZ})}(y) \tag{13}$$

to be defined.

Note that with Equation (11), a trait value can in principle display a nonlinear association within both MZ and DZ twins, but have constant local heritability $a^2(y)$ due to a canceling effect in $4\rho^{(\mathrm{DZ})}(y) - \rho^{(\mathrm{MZ})}(y)$.

Definition 2 (Heritability curve for the twin ACE model). Assume the exchangeability property (10) holds for both MZ and DZ bivariate distributions, and again let $\rho^{(\mathrm{MZ})}(y)$ and $\rho^{(\mathrm{DZ})}(y)$ be the correlation curves of MZ and DZ twins calculated according to (9). Based on (3), we let

$$\begin{aligned} a^2(y) &= 2(\rho^{(\mathrm{MZ})}(y) - \rho^{(\mathrm{DZ})}(y)), \\ c^2(y) &= 2\rho^{(\mathrm{DZ})}(y) - \rho^{(\mathrm{MZ})}(y), \\ e^2(y) &= 1 - \rho^{(\mathrm{MZ})}(y). \end{aligned} \tag{14}$$

define the corresponding heritability curve as well as the local common and residual environment curves in the ACE twin model.

We similarly define the heritability curve for family trios by adopting the genetic model described in Section 2.1.2 locally:

Definition 3 (Heritability curve for the mother-father-child trio ACE model). Assuming the exchangeability property (10), let $\rho^{(\mathrm{MC})}(y)$ and $\rho^{(\mathrm{FC})}(y)$ be correlation curves (9) for mother-child and father-child relationships, respectively. The heritability curves $a^2(y)$, $c^2(y)$, and $e^2(y)$ are then given by

$$a^2(y) = 2\rho^{(\mathrm{FC})}(y), \tag{15}$$

$$c^2(y) = \rho^{(\mathrm{MC})}(y) - \rho^{(\mathrm{FC})}(y), \tag{16}$$

$$e^2(y) = 1 - \rho^{(\mathrm{MC})}(y) - \rho^{(\mathrm{FC})}(y). \tag{17}$$

We next define a parametric class of multivariate densities for family data that can easily be fit by maximum likelihood, allows for nonlinear dependence, and admits an analytical expression for the correlation curve (9).

3 | CORRELATION AND HERITABILITY CURVES FOR GAUSSIAN MIXTURES

Throughout this article we denote by $\phi_d(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma})$ a d-dimensional Gaussian density, evaluated at $\boldsymbol{y}=(y_1,\ldots,y_d)$, and with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. While we can analyse data up to $d=3$, our method of analysis is based on bivariate distributions.

Consider the observed trait vector $\boldsymbol{y}=(y_1,y_2)$ for a pair of family members. We assume that it follows a m-component Gaussian mixture with density

$$\sum_{k=1}^{m} p_k \phi_2(\boldsymbol{y};\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k), \tag{18}$$

where $\sum_{k=1}^{m} p_k = 1$. The mean and covariance structure of the kth mixture component is taken to be

$$\boldsymbol{\mu}_k=(\mu_k,\mu_k), \quad \boldsymbol{\Sigma}_k=\begin{pmatrix}\sigma_k^2 & \sigma_k^2\rho_k\\ \sigma_k^2\rho_k & \sigma_k^2\end{pmatrix}, \tag{19}$$

where $\rho_k\in(-1,1)$ is the correlation parameter. The components of the mixture are ordered such that $\sigma_1\le\cdots\le\sigma_m$. If $\sigma_q=\sigma_{q+1}=\cdots=\sigma_m$ for some $q<m$, then we order the components in ascending order with respect of the means, that is, $\mu_q<\cdots<\mu_m$. Note that under the above constraints on $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$, the exchangeability condition (10) is satisfied. In addition, Y_1 and Y_2 have the same marginal distribution, with marginal density

$$g(y)=\sum_{k=1}^{m} g_k(y) \tag{20}$$

as the sum over the individual (weighted) components $g_k(y) := p_k\phi_1(y;\mu_k,\sigma_k^2)$. The (total) marginal mean, marginal variance, and correlation are given by

$$\mu=\sum_{k=1}^{m} p_k\mu_k, \quad \sigma^2=\sum_{k=1}^{m} p_k\left[\sigma_k^2+(\mu_k-\mu)^2\right] \quad \text{and} \quad \rho=\sigma^{-2}\sum_{k=1}^{m} p_k\left[\rho_k\sigma_k^2+(\mu_k-\mu)^2\right]. \tag{21}$$

We next derive local versions of μ and σ. Let δ be a latent variable with $P(\delta=k)=p_k$, $k=1,\ldots m$, showing which mixture component is realized. From Bayes' rule, it follows that the distribution of $\delta|Y_2=y$ is given as

$$p_k^*(y) := P(\delta=k|Y_2=y)=\frac{g_k(y)}{g(y)}. \tag{22}$$

In addition, by the assumed normality of each mixture component, it follows that

$$\mu_k(y) := \mathrm{E}(Y_1|Y_2=y,\delta=k)=\mu_k+\rho_k\cdot(y-\mu_k),$$

that is, $\mu_k(y)$ is a line with slope ρ_k, going through the point (μ_k,μ_k). By the law of total expectation,

$$\begin{aligned}\mu(y) &:= \mathrm{E}(Y_1|Y_2=y)=\mathrm{E}[\mathrm{E}(Y_1|Y_2=y,\delta)|Y_2=y]\\ &=\sum_{k=1}^{m} p_k^*(y)\mu_k(y).\end{aligned} \tag{23}$$

Similarly, by the law of total variance:

$$\begin{aligned}\sigma^2(y) &:= \mathrm{Var}(Y_1|Y_2=y) = \mathrm{E}[\mathrm{Var}(Y_1|Y_2=y,\delta=k)|Y_2=y] + \mathrm{Var}[\mathrm{E}(Y_1|Y_2=y,\delta=k)|Y_2=y] \\ &= \mathrm{E}(\sigma_\delta^2(1-\rho_\delta^2)|Y_2=y) + \mathrm{Var}(\mu_\delta(y)|Y_2=y) \\ &= \sum_{k=1}^{m} p_k^*(y)[\sigma_k^2(1-\rho_k^2) + [\mu_k(y)-\mu(y)]^2]. \end{aligned} \tag{24}$$

We are now ready to give the expression for $\beta(y) = \mu'(y)$, to be used in the correlation curve (9) for the mixture distribution.

Proposition 1. *Define*

$$d_k(y) := -(y-\mu_k)/\sigma_k^2.$$

Then,

$$\beta(y) = \sum_{k=1}^{m} p_k^*(y)[\rho_k + (\mu_k(y)-\mu(y))d_k(y)], \tag{25}$$

where $p_k^*(y)$ is given by (22).

Proof. See Appendix. ∎

Notice that when there is only a single mixture component ($m=1$), yielding a bivariate Gaussian distribution, the above expressions reduce to $\sigma = \sigma_1$, $\mu(y) = \mu_1$, $\sigma^2(y) = \sigma_1^2(1-\rho_1^2)$, and $\beta(y) = \rho_1$. Inserting these expressions in (9) we get a constant correlation curve, $\rho(y) = \rho_1$ for every y. Hence, if $m=1$ the heritability curve $a^2(y)$, given by (11) or (15), reduces to the ordinary heritability coefficient a^2.

3.1 | Properties of the correlation curve under a Gaussian mixture

It is of interest to investigate the asymptotic behavior of $\rho(y)$ as $y \to \pm\infty$ under the mixture (18) since this can be used to evaluate the asymptotic behavior of the heritability curve $a^2(y)$, which in general will depend on the family design. We state the result in the following theorem, which also includes the limit behavior of $\beta(y)$ and $\sigma^2(y)$.

Intuitively, a one-dimensional mixture distribution is asymptotically dominated in the tails by the component with the largest variance; if two or more components all share the largest variance, the sizes of the mean values come into play, with the component with the smallest mean value dominating when $y \to -\infty$, and the largest when $y \to +\infty$. While this in itself is fairly obvious, we here use it to develop the resulting asymptotic behavior of $\beta(y)$, $\sigma^2(y)$, and $\rho(y)$.

We consider the following two cases: Recall the ordering $\sigma_1^2 \le \cdots \le \sigma_m^2$, and define $q = \min\{l : \sigma_l^2 = \sigma_m^2\}$. We define Case I as $q=m$. For the alternative, Case II, where $q<m$, our conventions is that the mean values are then ordered such that $\mu_q < \mu_m$. To simplify the notation, define the constant K as follows:

$$\begin{aligned} &\text{Case I } (q=m), \quad y \to \pm\infty, \quad K := m, \\ &\text{Case II } (q<m), \quad y \to -\infty, \quad K := q, \\ &\text{Case II } (q<m), \quad y \to +\infty, \quad K := m. \end{aligned}$$

Theorem 1. *The asymptotic behavior of $\beta(y)$, $\sigma^2(y)$, and $\rho(y)$, given by (25), (24), and (9), are*

$$\begin{aligned} &\lim_y \beta(y) = \rho_K, \\ &\lim_y \sigma^2(y) = \sigma_K^2(1-\rho_K^2), \\ &\lim_y \rho(y) = \tilde{\rho}_K := \frac{\sigma\rho_K}{\left[\sigma^2\rho_K^2 + \sigma_K^2(1-\rho_K^2)\right]^{1/2}}. \end{aligned} \tag{26}$$

The global variance σ^2 is defined as in (21).

Proof. See Appendix. ∎

Theorem 1 shows that $\beta(y)$, $\sigma^2(y)$, and $\rho(y)$ all stabilize to finite limits as $y \to \pm\infty$, and their behavior is determined by the variance and correlation of mixture component K, in addition to the global variance σ^2. In Case I we have that the asymptotic correlation is the same in both tails, as exemplified in Figure B4A where $K = 3$ and $\tilde{\rho}_3 \approx 0.5$. Identical correlations in both tails may seem unmotivated for family data. Still, within the data range the correlation curve will be determined by all of the mixture components, in accordance with (9), which allows for different behavior in the tails.

Case II, on the other hand, allows for different asymptotic correlation in the left and right tail, with the differences being the use of ρ_n vs ρ_m in (26).

Theorem 1 is further shown in Figure B4 showing the limiting behavior of $\beta(y)$, $\sigma^2(y)$, and $\rho(y)$ for a three-component mixture under Case I. Note that the limiting correlation satisfies $\tilde{\rho}_3 < \min(\rho_1, \rho_2, \rho_3)$ for the parameter values used in the figure. This is counter-intuitive because the posterior probability $p_3^*(y)$ approaches 1 in the tails (upper left panel), but still the limiting correlation is not simply ρ_3. The peak in correlation around $\mu_2 = 2$ is reasonable as the second component has the highest ρ.

3.1.1 | The case of equal σ_k's

It is worth studying the special case that $\sigma_1 = \sigma_2 = \cdots = \sigma_m$, with their common value denoted by σ_0. This is Case II of Theorem 1 with $q = 1$. From (21) we get $\sigma^2 = \sigma_0^2 + \sigma_\mu^2$, where

$$\sigma_\mu^2 = \sum_{k=1}^{m} p_k(\mu_k - \mu)^2, \tag{27}$$

which is the variance due to differences in locations of mixture components. Recall the convention that the mixture components are ordered such that $\mu_1 < \mu_2 < \cdots < \mu_m$. We are now ready to state the following corollary to Theorem 1.

Corollary 1. *When $\sigma_1 = \cdots = \sigma_m$ the asymptotic behavior of $\rho(y)$, given by (9), is*

$$\lim_{y\to-\infty} \rho(y) = \rho_1 \sqrt{\frac{1+\gamma}{1+\gamma\rho_1^2}} \quad \text{and} \quad \lim_{y\to\infty} \rho(y) = \rho_m \sqrt{\frac{1+\gamma}{1+\gamma\rho_m^2}}, \tag{28}$$

where $\gamma = \sigma_\mu^2/\sigma_0^2$ is the ratio of between and within-component variance in the Gaussian mixture.

The limiting correlations always exceed (in absolute value) ρ_1 and ρ_m, respectively. When $\gamma \to \infty$, that is, the mixture components gets increasingly spread out, both limits approach 1 in absolute value.

3.2 | Estimation

In this section, we explain how to fit Gaussian mixtures to family data. On one hand, they are fully parametric distributions, which can be exploited in estimation and inference. On the other hand, allowing the number of mixture components m to grow, mixtures become increasingly flexible, which allows us to view them also as nonparametric tools. In particular, Gaussian mixtures seem well suited to model small perturbations from Gaussianity.

3.2.1 | Mixtures for family trios

First, let $\boldsymbol{y} = (y_1, y_2, y_3)$ denote the trait vector for the mother-father-child trio, which is assumed to have the following mixture density:

$$\sum_{k=1}^{m} p_k \phi_3(\boldsymbol{y}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k).$$

Here $\boldsymbol{\mu_k}$, $\boldsymbol{\Sigma_k}$ are structured in the following way:

$$\boldsymbol{\mu}_k = (\mu_k, \mu_k, \mu_k), \quad \boldsymbol{\Sigma}_k = \begin{pmatrix} \sigma_k^2 & \sigma_k^2\rho_k^{(\mathrm{MF})} & \sigma_k^2\rho_k^{(\mathrm{MC})} \\ \sigma_k^2\rho_k^{(\mathrm{MF})} & \sigma_k^2 & \sigma^2\rho_k^{(\mathrm{FC})} \\ \sigma_k^2\rho_k^{(\mathrm{MC})} & \sigma_k^2\rho_k^{(\mathrm{FC})} & \sigma_k^2 \end{pmatrix}, \tag{29}$$

where we use superscripts on the ρ's to denote relationship. Integrating the above joint density with respect to any one of the three family members (y_1, y_2, or y_3) will result in the bivariate Gaussian mixture (18) from which we defined the correlation curve. The reason for performing joint estimation, rather than pairwise, is to optimally utilize the information contained in mother-father-child trios. Note that the three marginals are identical by construction, although the joint distribution is not exchangeable unless $\rho_k^{(\mathrm{MF})} = \rho_k^{(\mathrm{MC})} = \rho_k^{(\mathrm{FC})}$ for $k = 1, \ldots, m$.

Given n such trios, the parameters $(\mu_k, \sigma_k, \rho_k, p_k)$ can be estimated by maximizing the following log-likelihood:

$$\log L = \sum_{i=1}^{n} \log \left[\sum_{k=1}^{m} p_k \phi_3(\boldsymbol{y}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right]. \tag{30}$$

Once the parameters are estimated, the heritability curve $a^2(y)$ can be obtained via the correlation curves as described in Definition 3.

3.2.2 | Mixtures for twins

For twins, consider first a dizygotic pair with trait vector $\boldsymbol{y} = (y_1, y_2)$. The likelihood contribution from $n^{(\mathrm{MZ})}$ such pairs is:

$$\log L^{(\mathrm{MZ})} = \sum_{i=1}^{n^{(\mathrm{MZ})}} \log \sum_{k=1}^{m} p_k \phi_2(\boldsymbol{y}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{31}$$

where $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are structured as in (19). The likelihood contribution of $n^{(\mathrm{DZ})}$ dizygotic twin pairs, $\log L^{(\mathrm{DZ})}$, is defined analogously using the same number m of mixture components. The only parameters that differ between the MZ and DZ cases are the correlation parameters ρ_k in (19). The fact that p_k, μ_k, and σ_k are shared across the MZ and DZ mixtures, calls for using a combined log-likelihood $\log L = \log L^{(\mathrm{MZ})} + \log L^{(\mathrm{DZ})}$. Once the parameters are estimated, the heritability curve $a^2(y)$ can be obtained via the correlation curves as described in Definition 1.

3.2.3 | Covariates

The same framework can be used when including covariates. Assume S continuous covariates $x_{ij}^{(s)}$, $s = 1, \ldots, S$, are measured for each individual j in family i. The likelihood can then be parameterized in terms of the covariate-adjusted traits. For example, for mother-father-child trios, the likelihood becomes

$$\log L = \sum_{i=1}^{n} \log \left[\sum_{k=1}^{m} p_k \phi_3 \left(\boldsymbol{y}_i - \sum_{s=1}^{S} \alpha^{(s)} \boldsymbol{x}_i^{(s)}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \right) \right], \tag{32}$$

where $\boldsymbol{x}_i^{(s)} = (x_{i1}^{(s)}, x_{i2}^{(s)}, x_{i3}^{(s)})$. The scalar parameters $\alpha^{(1)}, \ldots, \alpha^{(S)}$ are estimated jointly with parameters of the Gaussian mixture, from the likelihood.

Note in particular that this allows, for instance, a generational effect on the average phenotype. The generational effect can be accommodated by choosing $\boldsymbol{x}_i^{(1)} = c(1, 1, 0)$, that is, $\boldsymbol{x}^{(1)}$ serves as a dummy variable for the first generation, leaving the second generation as reference. Similarly, two dummy variables can provide separate intercepts for mothers, fathers, and children, respectively, by choosing $\boldsymbol{x}_i^{(1)} = c(1, 0, 0)$ and $\boldsymbol{x}_i^{(2)} = c(0, 1, 0)$ as dummies for mothers and fathers, respectively, leaving children as reference. Sex differences can be included by a dummy covariate $\boldsymbol{x}_i^{(1)} = c(1, 0, x_{i3}^{(1)})$, where

$x_{i3}^{(1)} = 1$ if the child in family i is a girl, and zero otherwise. Categorical covariates with more categories can be broken down into dummy variables in a standard fashion.

The resulting correlation curves, computed as described in Section 3 and Equation (9), then correspond to the joint distribution of the covariate-adjusted traits. This is entirely in line with standard modeling approaches in biometrical genetics; covariates can explain a part of the trait variability, and the remaining variance is decomposed, for instance, as in Equation (6).

There is, however, one very important distinction between local heritability and the traditional models. In traditional models, the variance decomposition is independent of covariate values. Certainly, adding covariates in the adjustment may impact the actual variance component estimates, but since there are no interactions between fixed and random effects, the decomposition of the adjusted trait variances is the same for all individuals, independent of their covariate or trait values. Local heritability curves, however, depend—by definition—on the trait value y. When presenting a local heritability curve, it is thus natural to compute it for, for instance, average covariate values. Consequently, an useful approach would be to center all covariates by subtracting their mean, that is, to replace the original covariates by their centered counterparts $\tilde{x}_{ij}^{(s)} = x_{ij}^{(s)} - \frac{1}{3n}\sum_{i,j} x_{ij}^{(s)}$ in the likelihood (32). The estimated correlation and heritability curves will then correspond to a family trio consisting entirely of "average" individuals. Alternatively, they could be presented for specific subgroups, and so on, although they would only differ by a horizontal shift of the curve. This discussion is not unique to local heritability curves; similar issues are frequently met when presenting standardized mortality rates, standardized survival curves, and so on in other settings.

3.2.4 | Software

We maximize both of the log-likelihoods (30) and (31) using the R-package TMB.[27] In TMB the (negative) log-likelihood is implemented as a C++ function, which is compiled and linked into the R session, where the standard function minimizer `nlminb` is employed. In addition, TMB calculates the gradient and Hessian (first- and second-order derivatives) of the log-likelihood by Automatic Differentiation.[27] Such derivative information can substantially speed up the minimizer and make it more robust. Finally, TMB uses derivatives to calculate the approximate standard deviation of any interest quantity, as a function of the parameters, using the delta method. This feature of TMB will be used to estimate pointwise confidence intervals of correlation and heritability curves.

For the purpose of selecting the number of mixture components, m, we calculate both of the criteria $\text{AIC} = -2\log(L) + 2Q$ and $\text{BIC} = -2\log(L) + \log(n)Q$ for each candidate model, where Q is the number of parameters and $\log(L)$ is obtained either from (30) or (31). Contributing to Q is the total number of p_k's, μ_k's, σ_k's, and ρ_k's, but due to the constraint $\sum_{k=1}^{m} p_k = 1$ there are only $m-1$ free p_k's. Hence, for the trio likelihood (30) we have $Q = 6m-1$, while for the twin likelihood (31), with different ρ_k for MZ and DZ twins, we have $Q = 5m-1$. When covariates are accounted for, via the likelihood (32), the number S of α parameters must be added to Q. It is clear that for $\log(n) > 2$, BIC will be more conservative than AIC, in the sense of favoring smaller values of m. In our experience, given a dataset with a fixed number of observations, the correlation curve tends to be more unstable (fluctuating) for larger values of m. For this reason we will use BIC as our model selection criterion, but we will still report AIC as a comparison.

4 | APPLICATIONS

4.1 | BMI of twins

We use the "twinData" dataset found in the R-package "OpenMx".[6] As our response, we take BMI measurements (around age 18) for $n^{(MZ)} = 534$ monozygotic and $n^{(\text{DZ})} = 328$ dizygotic female-female twin pairs. Table 1 compares models in the range $1 \le m \le 5$, and it is seen that the pure bivariate Gaussian model ($m = 1$) fits considerably worse than any of the mixture models ($m > 1$). The lowest AIC and BIC values occur for $m = 5$ and $m = 2$, respectively, but it is seen that AIC is almost indecisive between models with $m > 1$. Due to its heavier penalization, $\log(n^{(\text{MZ})} + n^{(\text{DZ})}) = \log(862) = 6.8$, of the number of parameters, BIC more clearly favors $m = 2$. According to our decision to base model selection on BIC, we choose the model with $m = 2$.

Table 2 shows the parameter estimates. The first mixture component is dominating with $p_1 = 0.81$. For MZ twins there is high correlation (ρ_k) within in each of the two components, while for DZ twins ρ_2 is close to zero. The (global)

m	No. of parameters	ΔAIC	ΔBIC
1	4	259.4	227.6
2	9	8.0	0
3	14	2.8	18.5
4	19	6.5	46.0
5	24	0	63.3

Note: AIC and BIC values are relative to the best fitting models (respectively, $m = 5$ and $m = 2$).

TABLE 1 Model comparison for the twin BMI data, where m is the number of mixture components and $5m - 1$ is the number of parameters in the model

	$k = 1$		$k = 2$		
Parameters	**Estimate**	**Std error**	**Estimate**	**Std error**	**Global**
μ_k	21.20	0.03	22.20	0.15	21.39
σ_k	0.63	0.02	1.26	0.07	0.88
$\rho_k^{(\mathrm{MZ})}$	0.75	0.03	0.70	0.06	0.78
$\rho_k^{(\mathrm{DZ})}$	0.28	0.07	−0.04	0.16	0.30
p_k	0.81	0.04	0.19	0.04	

Note: The mixture components are ordered according to the value of σ_k. The global quantities, μ, σ, $\rho^{(\mathrm{MZ})}$, and $\rho^{(\mathrm{DZ})}$ are calculated from (21).

TABLE 2 Parameter estimates and standard deviations for the chosen Gaussian mixture ($m = 2$) for the twin data

correlations for the mixtures as a whole, matches exactly the empirical Pearson correlations, which are 0.78 (MZ) and 0.30 (DZ), respectively.

Figure B5 displays the estimated correlation curve for both MZ and DZ twins, using the parameter values from Table 2. Also shown are 95% confidence intervals calculated using the delta method. Both correlation curves are fairly flat within the center 90% data range (represented by the two vertical green bars), while they both drop for low and high BMI.

Overall, the estimated MZ correlation is more than twice the estimated DZ correlation. Based on the comment in Section 2.1.1, $\rho^{(\mathrm{MZ})} > 2\rho^{(\mathrm{DZ})}$ suggests fitting an ADE model. This yields (Figure B6) an estimated heritability curve $a^2(y)$ that does not differ significantly (except maybe around $y = 22.3$) from the classical heritability coefficient (4).

Note that in both this and the following application, there are regions where the heritability and environment curve violate their natural boundaries by exceeding one or dropping below zero. This issue will be discussed in Section 5.2 below.

The TMB (R and C++) code used to produce the parameter estimates in Table 2 plots in Figure B6 is available from https://github.com/skaug/Supplementary.

4.2 | BW of family trios

To illustrate the family trio analyses, we used BWs of $n = 81,144$ complete mother-father-child trios. The data originally derived from the Medical Birth Registry of Norway, where the BW variables were added some random noise and rounded off to guarantee anonymity. The same data with some additional restrictions on parity, plurality, and so on were previously described and analyzed elsewhere.[19] The data were restricted to all births (mother, father, and child) taking place within the years 1967 to 1998.

We did not have information about the gender of the child, and hence did not account for gender via the covariate adjusted likelihood (32). Instead, we performed a standardization of the data prior to applying the unadjusted likelihood (30). We assumed a 50% sex ratio in the offspring, and introduced the quantity $D \triangleq \frac{1}{2}(\bar{y}_M - \bar{y}_F)$, where $\bar{y}_M$ is the mean of the BWs of mothers, and $\bar{y}_F$ is the mean of the BWs of fathers. We hence added D to the father's weight and subtracted it from the mother's weight; in this way, the average among mothers and fathers is the same, and close (25 g deviation) to the average in the offspring. Technically, this standardization could have been implemented using the formalism of the covariate adjusted likelihood (32), with dummy coding of mother and father, $\boldsymbol{x}_i^{(1)} = (1, 0, 0)$ and $\boldsymbol{x}_i^{(2)} = (0, 1, 0)$, respectively. Our standardization is then achieved by fixing the parameter values $\alpha^{(1)} = -D$ and $\alpha^{(2)} = D$ in the likelihood (32). In the end, the standardization is of little consequence to the end result.

Figure B1 summarizes the marginal and bivariate properties of the data. The marginal distributions are close to a Gaussian shape, but the left tail of the child BWs is slightly heavier than the right tail. As suggested in the Introduction, this may be indicative of strong but rare factors dominating in producing the lowest BWs, which is what we will confirm in our analyses of local heritability below.

The scatter plots are roughly symmetric around the identity line, which is consistent with the exchangeability assumption made in Section 2.2. It should be noted, however, that the left hand tail of the marginal distributions is somewhat heavier in the children than in the parents; this is likely because parents are selected by the fact that they have children; it is known that individuals born with low BW have somewhat reduced fertility later in life. We have, however, not taken this into consideration in our model.

From the nonparametric regression (blue curve), it is clear that there is no association between mother and father, which is reflected in the low Pearson correlation of 0.0209. For the two relationships involving the child, the nonparametric regression curve indicates a nonlinear relationship, particularly for mother-child. For BWs less than 3000 g there seems to be a low association, while for larger BWs the association is increasing.

The Gaussian mixture (18) was fit by maximum likelihood for $m = 1, \ldots, 7$. We computed both AIC and BIC values for this model. According to the BIC criterion, the best fitting mixture has $m = 4$ components (see Table 3). Parameters estimates for this model are given in Table 4. Figure B7 shows the underlying mother-child pairs, overlaid by the five mixture components.

The mother-child distribution is pear-shaped relative to a bivariate normal distribution, with more spread around the identity line ($y_1 = y_2$) for small BWs. The mixture model adapts to this shape by assigning negative ρ_k's to its two components ($k = 3, 4$) with the smallest μ_k. The remaining two components ($k = 1, 2$), which together constitute 87% of the probability mass, form a bivariate distribution that is hard to distinguish visually from a Gaussian distribution. The estimates of global correlation for the mixture in Table 4, closely match the corresponding empirical Pearson correlations given in Figure B1 for MC, FC, and MF pairs. It is seen to fit the empirical marginals fairly well, and to possess a heavier left hand tail.

TABLE 3 Model comparison for family trios, where m is the number of mixture components

m	No. parameters	Δ AIC	Δ BIC
1	5	14 848	14 749
2	11	1148	904.4
3	17	480.4	292.5
4	23	132.1	0
5	29	109.7	33.5
6	35	36.3	16.0
7	41	0	35.5

Note: The total number of (free) parameters is $6m - 1$, counting all p_k, μ_k, σ_k, $\rho_k^{(\mathrm{MC})}$, $\rho_k^{(\mathrm{FC})}$, and $\rho_k^{(\mathrm{MF})}$. AIC and BIC values are relative to the lowest one.

TABLE 4 Parameter estimates and standard deviations for the Gaussian mixture ($m = 4$) fit to the mother-father-child trios

	$k=1$		$k=2$		$k=3$		$k=4$		
Parameter	**Estimate**	**Std error**	**Estimate**	**Std error**	**Estimate**	**Std error**	**Estimate**	**Std error**	**Global**
μ_k	3516	4.247	3687	25.73	3093	11.42	2243	26.55	3493
σ_k	440.5	3.587	572.9	7.963	690.5	7.016	1116	34.70	555.0
$\rho_k^{(\mathrm{MC})}$	0.240	0.008	0.143	0.016	−0.189	0.016	−0.826	0.022	0.123
$\rho_k^{(\mathrm{FC})}$	0.134	0.008	0.054	0.017	−0.254	0.015	−0.845	0.019	0.201
$\rho_k^{(\mathrm{MF})}$	−0.011	0.008	−0.084	0.020	−0.289	0.013	0.750	0.032	0.068
p_k	0.636	0.027	0.231	0.030	0.127	0.007	0.007	0.001	

Note: The mixture components are ordered according to the value of σ_k. The global quantities, μ, σ, $\rho^{(\mathrm{MC})}$, $\rho^{(\mathrm{FC})}$, and $\rho^{(\mathrm{MF})}$ are calculated from (21).

Figure B8 shows the two estimated correlation curves $\rho^{(FC)}(y)$ and $\rho^{(MC)}(y)$, which are the components going into $a^2(y)$, $c^2(y)$, and $e^2(y)$, given, respectively, by (15) to (17). Overall, the Pearson correlation and the correlation curves for MF exceed those for FC. Both curves exceed their respective Pearson correlations in the center of the data, while they decrease for both low and high BWs. The FC curve has its maximum somewhat to the left of the maximum of the MC curve. As a robustness check, we also computed the local Gaussian correlations[15] between mother and child as displayed in Figure B9. These exhibit the same behavior as the correlation curve; large values in the center of the data which are decreasing toward both tails.

Figure B10 shows heritability and environment curves. The overall conclusion is that variation in BW is mostly attributable to environment, which was also seen in previous publications,[9,19,20] and is reflected in the classical measures of heritability $a^2 = 0.254$ and environment $c^2 + e^2 = 0.746$, and the variation in the corresponding curves. Recall that, under the assumed model (7) the heritability curve $a^2(y)$ is completely determined by the FC correlation curve $\rho^{(FC)}(y)$. Since the FC correlation curve exceeds the Pearson FC correlation in the center of the data, the heritability curve also exceeds the classical heritability measure in the same region.

5 | EXTENSIONS

5.1 | Larger family structures

The two examples presented here, the twin models and the trio model, are particularly convenient to work with since the moment equations can be solved directly, providing explicit expressions for the genetic and environmental variances. Once a Gaussian mixture is fitted to describe a bivariate twin distribution or a trivariate mother-father-child distribution, the correlation curves can be derived from each estimated pairwise relationship distribution, by using Equation (9) with $\beta(y)$ and $\sigma^2(y)$ calculated as described in Section 3. Providing a full extension to other family structures goes beyond this article's scope, but we will briefly sketch how this can be performed. As a start, consider only independent families with two individuals in each family. The MZ and DZ twin pairs would be an example of this. Alternatively it could be, for instance, pairs of full siblings, pairs of half sibs, pairs of cousins, and pairs of cousins where the parents are half sibs. Each such pair contributes an equation for the pairwise correlation. For instance, if only including the additive genetic component, we would have $\rho_1 = \frac{1}{2}a^2$, $\rho_2 = \frac{1}{4}a^2$, $\rho_3 = \frac{1}{8}a^2$, and $\rho_4 = \frac{1}{16}a^2$, respectively, for the pairwise correlations, that is,

$$\boldsymbol{\rho} = \boldsymbol{A}a^2, \tag{33}$$

with

$$\boldsymbol{\rho} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} 1/2 \\ 1/4 \\ 1/8 \\ 1/16 \end{pmatrix}.$$

Then, similarly to the twin example, Gaussian mixtures could be fitted to all pairwise relations, only varying the correlations from relation type to relation type. Or mixtures could be fitted separately for all relation types. Either way, the maximum likelihood mixture fitting would result in an estimated vector $\hat{\boldsymbol{\rho}}(y)$ of local correlation curves, with a corresponding asymptotic variance-covariance matrix $\boldsymbol{S}(y)$. In particular, if the fitting was performed independently on the various relation types, with independent families, $\boldsymbol{S}(y)$ would be diagonal; otherwise there might be off-diagonal elements in $\boldsymbol{S}(y)$. The maximum likelihood mixture fitting would provide a local version $\hat{\boldsymbol{\rho}}(y)$ of $\boldsymbol{\rho}$ on the left-hand side of Equation (33), which would be asymptotically normally distributed. The equation can then be solved using weighted least squares, or generalized least squares, to obtain

$$\hat{a}^2(y) = (\boldsymbol{A}^T\boldsymbol{S}(y)^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{S}(y)^{-1}\hat{\boldsymbol{\rho}}(y)$$

as an estimate of the local heritability. The estimate has the corresponding asymptotic variance

$$\text{Var}(\hat{a}^2(y)) = (\boldsymbol{A}^T\boldsymbol{S}(y)^{-1}\boldsymbol{A})^{-1}.$$

Clearly, Equation (33) can be directly extended to more parameters such as shared environment, maternal genes, and so on, replacing a^2 with a suitable parameter vector. Further examples of using weighted least squares can be found elsewhere.[9]

Typically, the equation system will be overdetermined, that is, with more equations than parameters. In the twin example, for the parameter vector $(a^2(y), d^2(y))$ in the ADE model, the coefficient matrix $\boldsymbol{A}$ is invertible, and the weighted least squares solution reduces to the previous solutions (11) and (12).

For family structures involving three family members, such as the mother-father-child trios presented here, a similar approach can be used to combine results from different family types. With larger family structures, the mixture fitting will become more complex. However, a simplified approach is likely to work well, particularly with large data samples; large family structures can be broken down into all pairwise relations, and the approach described above can be applied.[28] Since a single individual may appear in several pairwise relations, bootstrapping, or similar procedures may be applied to obtain an estimate of $\boldsymbol{S}(y)$, which again allows solving by generalized least squares. Approaches for extended family structures will be further explored in a separate article.

5.2 | Model switching

As seen in Figure B6, estimated heritability and environment can fall outside the natural boundaries of 0 and 1. This is a well-known problem with the moment equations, and has frequently been addressed in the literature. If the model is correctly specified in the first place, it is usually caused by sample variation when the sample size is small, and will disappear with larger sample sizes. Since the local heritability curves are more flexible than the standard model with a fixed overall heritability, the problem may occur more frequently with the local heritability models. We will not fully explore this issue here, but note that in Section 5.1, generalized least squares estimation can be performed with additional inequality restrictions,[29] which will ensure that parameters remain within range.

However, an interesting approach, available in local models, is that one can switch from one model to another over the range of the phenotype. In Figure B6, it is seen that in the middle range of the phenotype, around a BMI of 22, the dominant component drops below zero, and the heritability climbs above one, indicating that in this region the twin ACE model would be more appropriate than the twin ADE model. It is conceivable that structurally different genetic models may apply in different ranges of the phenotype. Consequently, we also fitted an ACE model to the correlation curves derived in Section 4, and replaced the ADE model by the ACE model in the middle range of the BMI scale. Note that there is no difficulty in letting the local heritability curves switch from an ADE model to an ACE model locally. In particular, we see that when $\rho^{(\mathrm{MZ})} = 2\rho^{(\mathrm{DZ})}$, both (3) and (4) provide the same estimates for a^2 and e^2, and both c^2 and d^2 are estimated as zero. As seen in Figure B11, the resulting heritability curve is continuous across the y values at which the model switches, and stays below one in the middle range.

Taken at face value, Figure B11 suggests that a common environment is more active in determining BMI at the central part of the BMI range, while dominant genes are more active for the smallest and largest BMIs. Clearly, given the restricted sample size of the BMI data, this conclusion should be treated with caution, and it also does not resolve the negative heritability estimates at the extreme ends of the scale. It illustrates, however, how local heritability models can extract more and possibly relevant information about heritability over the full range of the phenotype.

6 | DISCUSSION AND CONCLUSION

We have provided closed-form expressions for the correlation curve for exchangeable bivariate Gaussian mixtures. To our knowledge, this result is new and should be useful generally in situations where exchangeability can be assumed. Since differences in mean values may be accounted for using a linear predictor like (5), it is only exchangeability of the residuals, or the weaker condition (10), that is required. In the context of our family data, the exchangeability assumption is quite reasonable for twin data. In nuclear families, it is less obvious that parents and children have the exact same marginal distribution even when using covariates to adjust for systematic generational differences. With our generational BW data, we observe that the left hand tail in the parental distribution is smaller than that of the children distribution. As discussed in Subsection 4.2, this may well be a selection phenomenon; somebody born with a very low BW is less likely to become a parent, and are thus possibly underrepresented in our data file. For instance, increased mortality among the smallest newborns is thought to lead to a selection pressure on the BW distribution over generations.[30]

A restriction of our model is that we have applied it only in situations with simple family structures where moment estimators of the heritability are explicit. In larger family structures, several pairwise relationships may provide information about the same heritability parameters. For instance, family trios with sibling data add the sibling correlation as a source of information.[20] In Section 5, we indicated how generalized least squares estimation may provide a way of combining them into a common estimate of heritability curves.[9]

Similar to the traditional moment-based estimator of heritability, the heritability curve can potentially violate its natural boundary by exceeding one or dropping below zero. In our twin BMI example, we chose the ADE model for the estimation since for the estimated overall correlations, $\rho^{(\mathrm{MZ})} > 2\rho^{(\mathrm{DZ})}$. However, as seen in Figure B6, there are values for y (the BMI) where the estimated $d^2(y)$ drops below zero. One interesting aspect of the local heritability curves is that they not only allow, for instance, the A, D, and E components of the ADE twin model to change in size over the BMI range. They also allow switching to another model altogether—in this case the ACE model—whenever appropriate, as seen in Figure B11.

The choice of Gaussian mixtures was made due to their flexibility, in the spirit of nonparametric estimation, in addition to the fact that data are fairly close to being normally distributed. Our approach is pragmatic in the sense that we have not attempted to interpret individual mixture components as subpopulations. One reason for this is the negative estimates for some of the ρ_k seen in both Table 2 and 4, which would be hard to interpret biologically.

On the other hand, Gaussian mixtures are fully parametric models, which allows us to use the standard parametric toolbox. For instance, covariates can easily enter the mean, as in (5), and it would also be straight forward to formulate model in which the σ_k were affected by family level covariates. A further benefit of having a parametric model is that we can select model complexity (m) based on standard AIC or BIC criteria.

The parametric structure is also the basis for the results about the tail behavior of the correlation curve in Theorem 1. While the center of the distribution may have sufficient data to allow stable nonparametric estimation of the heritability, the estimates in the tails are more dependent on the model structure. This is both a strength and a weakness of the mixture model. The heritability curves converge to constant values in the tails, which makes the estimates more stable; on the other hand, those estimates depend on the dominant mixture components in the tails, and the number and placement of mixture components may not always be clear cut.

There are also well-known problems with Gaussian mixtures. Among these are local maxima on the likelihood surface,[31] which can be explored by using different initial values for the numerical optimization. We avoided the classical "label switching" problem by constraining the parameters of the mixture (σ's and μ's), but have nevertheless observed some sensitivity of the parameter estimates in Table 4. Although we cannot guarantee that we have found the global optimum of the likelihood surface for the two datasets in Section 4, the choice of model complexity (m) seems to be robust to the choice of initial values. Similarly, the shape of the correlation curves (and consequently heritability and environment curves) are quite stable. A related problem is that of singularity of the Fisher information matrix which can occur for mixture models.[32] This could potentially affect the validity of AIC and BIC criteria, as well as the standard deviations based on the observed Fisher information that have been used throughout this article. Such standard deviations are produced automatically by TMB, and are very convenient in an exploratory phase, but we recommend that they are validated by simulation (parametric bootstrap).

ACKNOWLEDGEMENTS

This research was supported by Research Council of Norway grant 225912/F50 "Health Registries for Research" and the Centres of Excellence funding scheme (Grant 262700).

DATA AVAILABILITY STATEMENT

Due to Norwegian ethical and legal restrictions, Norwegian data used in this study are available upon request to the Medical Birth Registry of Norway, the Norwegian Institute of Public Health. URL: https://www.fhi.no/hn/helseregistre-og-registre/mfr. Requests for data access can be directed to Datatilgang@fhi.no.

ORCID


Francesca Azzolini https://orcid.org/0000-0002-4780-8029

REFERENCES

1. Fisher RA. XV.–The correlation between relatives on the supposition of Mendelian inheritance. *Earth Environ Sci Trans R Soc Edinb*. 1919;52(2):399-433. https://doi.org/10.1017/S0080456800012163.

2. Wright S. The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proc Natl Acad Sci*. 1920;6(6):320-332.
3. Wright S. Correlation and causation. *J Agric Res*. 1921;20(7):557-585.
4. Bulmer M. *The Mathematical Theory of Quantitative Genetics*. Oxford, England: Clarendon Press; 1985.
5. Neale M. Twin analysis. In: Elston R, Olson J, Palmer L, eds. *Biostatistical Genetics and Genetic Epidemiology*. Wiley Reference Series in Biostatistics. West Sussex, UK: Wiley; 2002:206-217.
6. Neale M, Hunter M, Pritikin J, et al. OpenMx 2.0: extended structural equation and statistical modeling. *Psychometrika*. 2016;81(2):535-549.
7. Hopper J. Heritability. In: Elston R, Olson J, Palmer L, eds. *Biostatistical Genetics and Genetic Epidemiology*. Wiley Reference Series in Biostatistics. West Sussex, UK: Wiley; 2002:371-372.
8. Hopper J, VP. Genetic correlations and covariances. In: Elston R, Olson J, Palmer L, eds. *Biostatistical Genetics and Genetic Epidemiology*. Wiley Reference Series in Biostatistics. West Sussex, UK: Wiley; 2002:327-331.
9. Gjessing H, Lie R. Biometrical modelling in genetics: are complex traits too complex? *Stat Methods Med Res*. 2008;17(1):75-96. https://doi.org/10.1177/0962280207081241.
10. DeFries J, Fulker D. Multiple regression analysis of twin data. *Behav Genet*. 1985;15(5):467-473.
11. DeFries J, Fulker D. Multiple regression analysis of twin data: Etiology of deviant scores versus individual differences. *Acta Geneticae Medicae et Gemellologiae Twin Res*. 1988;37(3-4):205-216.
12. Cherny S, Cardon L, Fulker D, DeFries J. Differential heritability across levels of cognitive ability. *Behav Genet*. 1992;22(2):153-162.
13. Logan J, Petrill S, Hart S, et al. Heritability across the distribution: an application of quantile regression. *Behav Genet*. 2012;42(2):256-267.
14. LaBuda M, DeFries J, Fulker D, Rao D. Multiple regression analysis of twin data obtained from selected samples. *Genet Epidemiol*. 1986;3(6):425-433.
15. Tjøstheim D, Hufthammer K. Local Gaussian correlation: a new measure of dependence. *J Econ*. 2013;172(1):33-48. https://doi.org/10.1016/j.jeconom.2012.08.001.
16. Holland P, Wang Y. Dependence function for continuous bivariate densities. *Commun Stat Theory Methods*. 1987;16(3):863-876. https://doi.org/10.1080/03610928708829408.
17. Bjerve S, Doksum K. Correlation curves: measures of association as functions of covariate values. *Ann Stat*. 1993;21(2):890-902. https://doi.org/10.1214/aos/1176349156.
18. Falconer D. *Introduction to Quantitative Genetics*. Edinburgh; London: Oliver and Boyd; 1960.
19. Magnus P, Gjessing H, Skrondal A, Skjærven R. Paternal contribution to birth weight. *J Epidemiol Community Health*. 2001;55(12):873-877. https://doi.org/10.1136/jech.55.12.873.
20. Lunde A, Melve K, Gjessing H, Skjærven R, Irgens L. Genetic and environmental influences on birth weight, birth length, head circumference, and gestational age by use of population-based parent-offspring data. *Am J Epidemiol*. 2007;165(7):734-741. https://doi.org/10.1093/aje/kwk107.
21. McLachlan G, Peel D. *Finite Mixture Models*. New York, NY: Wiley; 2000.
22. McCulloch C, Neuhaus J. *Generalized Linear Mixed Models*. Hoboken, NJ: Wiley Online Library; 2001.
23. Khoury M, Beaty T, Cohen B. *Fundamentals of Genetic Epidemiology*. Oxford, UK: Oxford University Press; 1993.
24. Pawitan Y, Reilly M, Nilsson E, Cnattingius S, Lichtenstein P. Estimation of genetic and environmental factors for binary traits using family data. *Stat Med*. 2004;23:449-465.
25. Rabe-Hesketh S, Skrondal A, Gjessing H. Biometrical modeling of twin and family data using standard mixed model software. *Biometrics*. 2008;64(1):280-288.
26. Doksum K, Blyth S, Bradlow E, Meng X, Zhao H. Correlation curves as local measures of variance explained by regression. *J Am Stat Assoc*. 1994;89(426):571-582. https://doi.org/10.1080/01621459.1994.10476782.
27. Kristensen K, Nielsen A, Berg C, Skaug H, Bell B. TMB: automatic differentiation and Laplace approximation. *J Stat Softw*. 2016;70(1):1-21.
28. Varin C, Reid N, Firth D. An overview of composite likelihood methods. *Statistica Sinica*. 2011;5-42.
29. Werner H. On inequality constrained generalized least-squares estimation. *Linear Algebra Appl*. 1990;127:379-392. https://doi.org/10.1016/0024-3795(90)90351-C.
30. Cavalli-Sforza L, Bodmer W. *The Genetics of Human Populations*. North Chelmsford, MA: Courier Corporation; 1999:612-614.
31. Baudry J, Celeux G. EM for mixtures. *Stat Comput*. 2015;25(4):713-726.
32. Drton M, Plummer MA. Bayesian information criterion for singular models. *J Royal Stat Soc Ser B (Stat Methodol)*. 2017;79(2):323-380.
33. Bender C, Orszag S. *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*. Berlin, Germany: Springer Science & Business Media; 2013.

How to cite this article: Berentsen GD, Azzolini F, Skaug HJ, Lie RT, Gjessing HK. Heritability curves: A local measure of heritability in family models. *Statistics in Medicine*. 2021;40:1357–1382. https://doi.org/10.1002/sim.8845

APPENDIX A. PROOFS

Proof of Proposition 1. Let us recall some definitions introduced in Section 3:

$$\begin{aligned} g_k(y) &:= p_k \phi_1(y; \mu_k, \sigma_k^2) \\ g(y) &:= \sum_{k=1}^{m} g_k(y) \\ p_k^*(y) &:= \frac{g_k(y)}{g(y)} \\ d_k(y) &:= -(y - \mu_k)/\sigma_k^2. \end{aligned}$$

First, note that

$$\frac{g_k'(y)}{g_k(y)} = d_k(y).$$

Furthermore, define

$$d(y) := \sum_{i=1}^{m} p_i^*(y) d_i(y),$$

that is, the weighted average of the $d_i(y)$'s. Then

$$\frac{g'(y)}{g(y)} = \frac{\sum_{i=1}^{m} d_i(y) g_i(y)}{g(y)} = d(y).$$

For any fraction $s(y) = a(y)/b(y)$ of differentiable functions, note that the chain rule can be written as $\frac{s'(y)}{s(y)} = \frac{a'(y)}{a(y)} - \frac{b'(y)}{b(y)}$. Thus,

$$\frac{p_k^{*\prime}(y)}{p_k^*(y)} = \frac{g_k'(y)}{g_k(y)} - \frac{g'(y)}{g(y)} = d_k(y) - d(y).$$

Recall from (23) that $\mu(y) = \mathrm{E}[Y_1 | Y_2 = y] = \sum_{i=1}^{m} p_i^*(y) \mu_i(y)$ is the conditional expectation,

$$\begin{aligned} \beta(y) = \mu'(y) &= \sum_{i=1}^{m} (p_i^*(y) \mu_i'(y) + p_i^{*'}(y) \mu_i(y)) \\ &= \sum_{i=1}^{m} p_i^*(y) (\rho_i + \mu_i(y)(d_i(y) - d(y))) \\ &= \sum_{i=1}^{m} p_i^*(y) (\rho_i + (\mu_i(y) - \mu(y))(d_i(y) - d(y))) \\ &= \sum_{i=1}^{m} p_i^*(y) (\rho_i + (\mu_i(y) - \mu(y)) d_i(y)), \end{aligned}$$

where we make use of $\sum_{i=1}^{m} p_i^*(y)(d_i(y) - d(y)) = 0$ and $\sum_{i=1}^{m} p_i^*(y)(\mu_i(y) - \mu(y)) = 0$. ∎

A1 Proof of Theorem 1—asymptotic behavior of $\beta(y)$, $\sigma^2(y)$, and $\rho(y)$

For two functions $a(y)$ and $b(y)$, as $y \to \infty$ (or $-\infty$), we use the standard notation that $a(y) \sim b(y)$ means $\lim_{y\to\infty} a(y)/b(y) = 1$, and $a(y) \ll b(y)$ means $\lim_{y\to\infty} a(y)/b(y) = 0$. Our proofs below follow mostly from standard theory on asymptotic behavior of real functions.[33]

Asymptotic behavior of mixture components

For one mixture component $g_k(y)$, the asymptotic behavior when $y \to \pm\infty$ is

$$g_k(y) \sim C_k \exp\left(\frac{\mu_k}{\sigma_k^2}y - \frac{1}{2\sigma_k^2}y^2\right),$$

for a constant C_k. Comparing two components $g_k(y)$ and $g_l(y)$ with $\sigma_k^2 < \sigma_l^2$, we clearly have

$$g_k(y) \ll g_l(y) \quad \text{as} \quad y \to \pm\infty \tag{A1}$$

since the y^2-term dominates the asymptotics. If $\sigma_k^2 = \sigma_l^2$, assume that $\mu_k < \mu_l$. Then

$$g_k(y) \ll g_l(y) \quad \text{as} \quad y \to +\infty, \tag{A2}$$

and

$$g_l(y) \ll g_k(y) \quad \text{as} \quad y \to -\infty. \tag{A3}$$

Let $a_k(y)$ be nonzero polynomial functions in y for $k = 1, \dots, m$. Since polynomials are asymptotically dominated by exponentials of polynomials, the products $g_k(y)a_k(y)$ are asymptotically ordered in the same way as in (A1) to (A3) above.

Asymptotic behavior of mixtures

Recall the definition of K in Theorem 1. The results above apply directly to the sum $\sum_{k=1}^m g_k(y)a_k(y)$, which will asymptotically follow the dominant term with $k = K$. That is,

$$\sum_{k=1}^m g_k(y)a_k(y) \sim g_K(y)a_K(y).$$

In particular, for the full density we get

$$g(y) = \sum_{i=1}^m g_i(y) \sim g_K(y).$$

Similarly, if $k \neq K$,

$$p_k^*(y)a_k(y) = \frac{g_k(y)a_k(y)}{g(y)} \to 0, \tag{A4}$$

and

$$p_K^*(y)a_K(y) \sim a_K(y).$$

Conditional mean $\mu(y)$

Applying the above results to μ, we obtain

$$\mu(y) = \sum_{k=1}^m p_k^*(y)\mu_k(y) \sim \mu_K(y) \sim \rho_K \cdot y.$$

Furthermore, letting $a_k(y) = \rho_k + (\mu_k(y) - \mu(y))d_k(y)$, we get

$$\beta(y) = \sum_{k=1}^m p_k^*(y)a_k(y) \sim a_K(y).$$

However, by (A4),

$$(\mu_K(y) - \mu(y))d_K(y) = \sum_{k=1}^m p_k^*(y)(\mu_K(y) - \mu_k(y))d_K(y) \to 0$$

since the Kth term vanishes. It follows that

$$\beta(y) \sim a_K(y) \to \rho_K.$$

Conditional variance $\sigma^2(y)$

For the conditional variance,

$$\sigma^2(y) = \sum_{k=1}^{m} p_k^*(y) \left[\sigma_k^2(1-\rho_k^2) + [\mu_k(y) - \mu(y)]^2\right]$$
$$\sim \sigma_K^2(1-\rho_K^2).$$

Correlation curve $\rho(y)$

Finally, the result for the correlation curve $\rho(y)$ follows directly from the results for $\sigma^2(y)$ and $\beta(y)$.

APPENDIX B. FIGURES

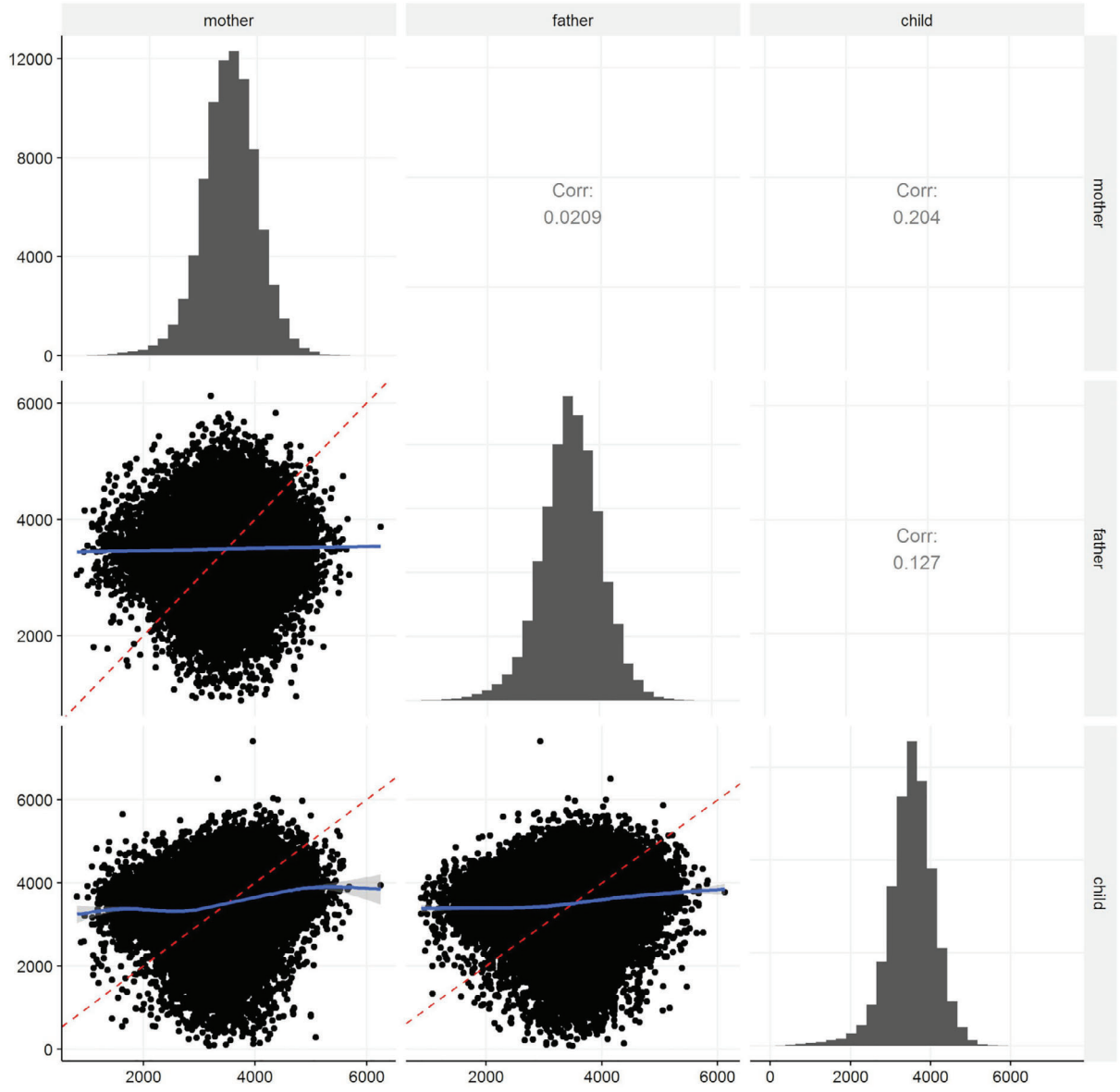


FIGURE B1 Birth weights (gram) for 81, 144 mother-father-child trios from the Norwegian Birth Registry. Diagonal: histograms of marginal birth weights. Lower triangle: pairwise scatter plots with estimated nonparametric regression line (blue) and identity line (dashed red), where $y = x$. Upper triangle: pairwise empirical correlation [Colour figure can be viewed at wileyonlinelibrary.com]

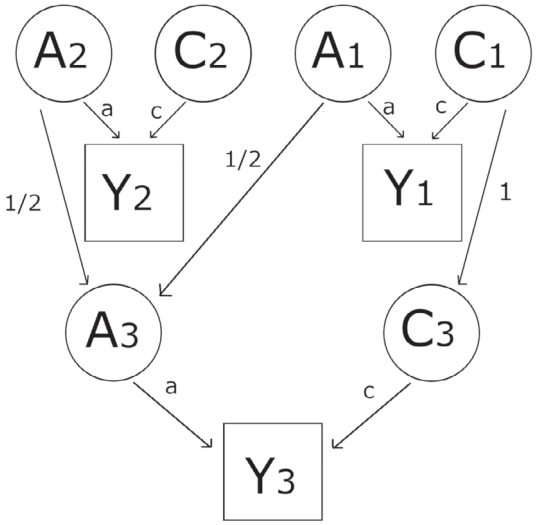


FIGURE B2 Path diagram representing the birth weight of mother Y_1, father Y_2, and child Y_3 (represented as squares). The traits are determined by the unobserved genotype values (A) and environmental values (C) (shown as circles), as well as the independent residual environmental values (E) (not shown)

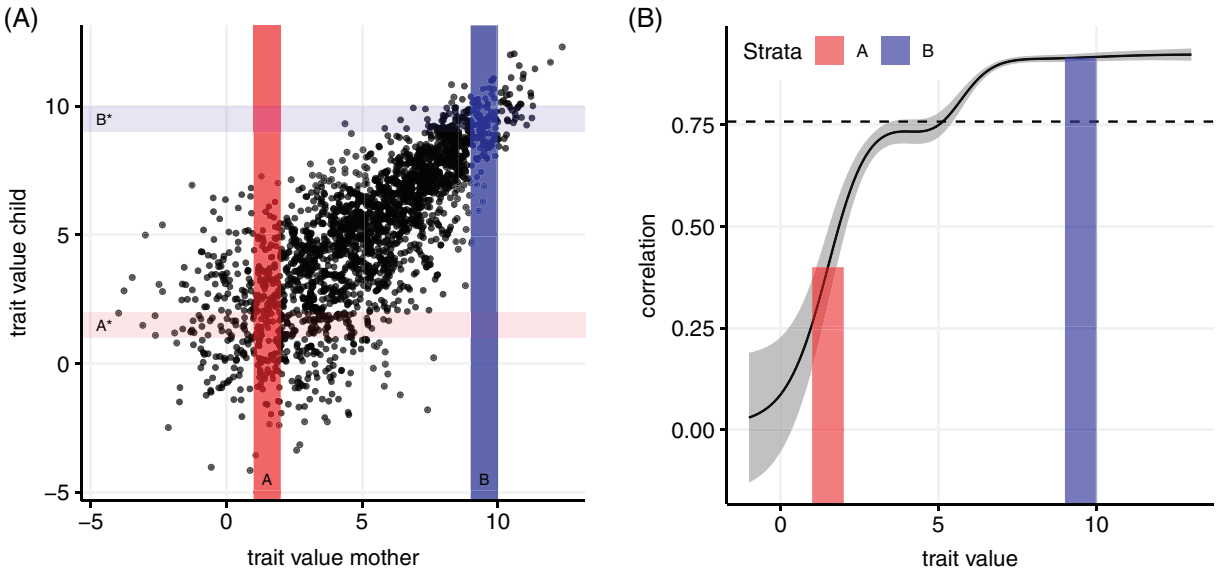
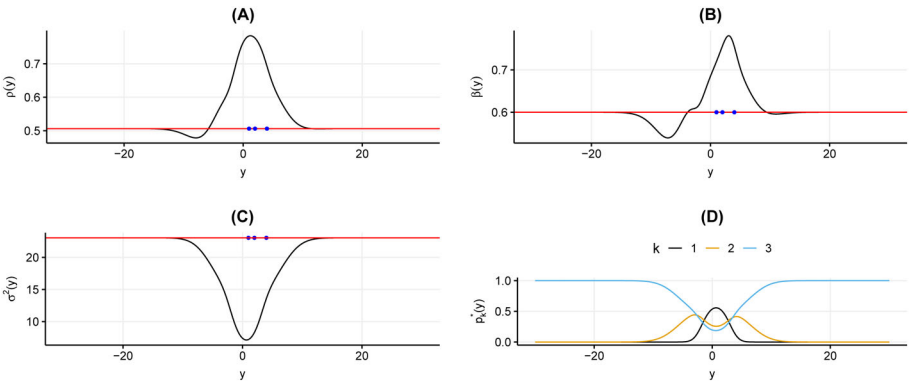


FIGURE B3 Illustration of the concept of a correlation curve and the role of exchangeability using simulated data. Strata A and B include all mother-child pairs for which the mother's trait value falls in the intervals $[1, 2]$ and $[9, 10]$, respectively. Strata A^* and B^* include all mother-child pairs for which the child's trait value falls in the same intervals [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE B4 Illustration of the asymptotic tail behavior (red line) of the correlation curve and its builing blocks under a $m = 3$ component mixture model: A, $\rho(y)$, B, $\beta(y)$, C, $\sigma^2(y)$, and D, $p_k^*(y)$. The mixture model has parameters $(\sigma_1, \sigma_2, \sigma_3) = (2, 4, 6)$, $(\mu_1, \mu_2, \mu_3) = (1, 2, 4)$, $(\rho_1, \rho_2, \rho_3) = (0.7, 0.8, 0.6)$ and $(p_1, p_2, p_3) = (0.3, 0.3, 0.4)$ [Colour figure can be viewed at wileyonlinelibrary.com]



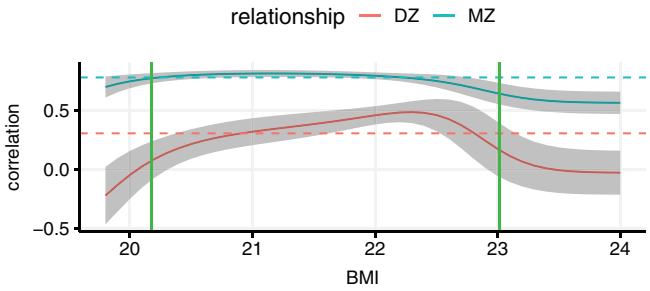


FIGURE B5 Estimated monozygotic (MZ) and dizygotic (DZ) twins correlation curves for the BMI data, with pointwise 95% confidence intervals (in gray). The dashed lines display the (overall) Pearson correlation within MZ and DZ twin pairs, respectively. The vertical green lines represent the 0.05 and 0.95 quantiles of the data [Colour figure can be viewed at wileyonlinelibrary.com]

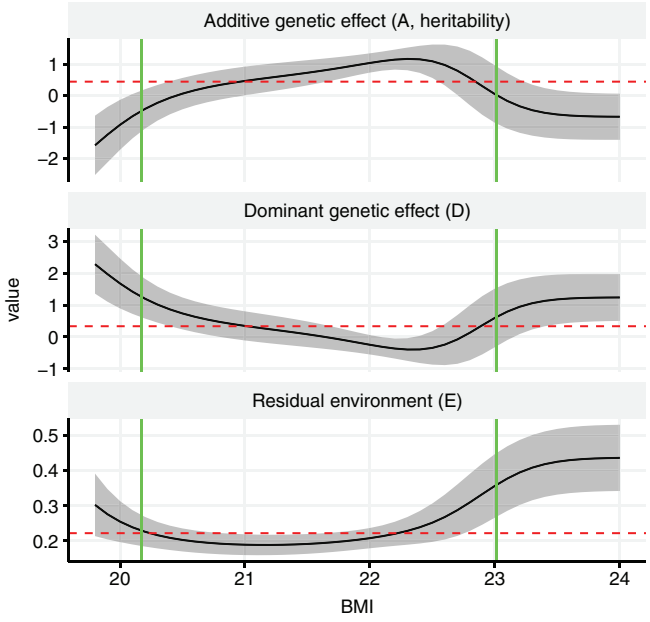


FIGURE B6 Estimated heritability curve $a^2(y)$, dominant genetic component $d^2(y)$, and environment $e^2(y)$ for the body mass index (BMI) (y) data under the ADE model (Definition 1), with pointwise 95% confidence intervals (in gray). The red dashed lines display the classical estimates of heritability, dominant component, and environment, given by (4). The vertical green lines represent the 0.05 and 0.95 quantile in data [Colour figure can be viewed at wileyonlinelibrary.com]

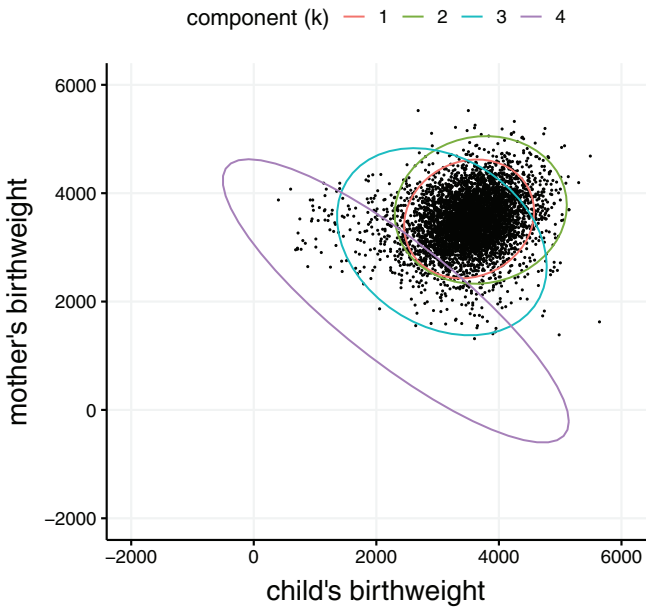


FIGURE B7 Birth weight (gram) of a random subset of 5000 mother-child pairs taken from Figure B1. Also shown are 95% level curves (ellipses) for each of the $m = 4$ mixture components in Table 4, that is, each ellipse include 95% of the probability mass for that bivariate normal component [Colour figure can be viewed at wileyonlinelibrary.com]

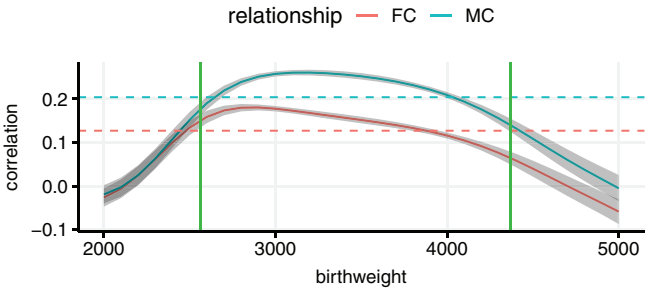


FIGURE B8 Estimated mother-child (MC) and father-child (FC) correlation curves for the Norwegian Birth Registry data, with pointwise 95% confidence intervals (in gray). The dashed lines display the (overall) Pearson correlation within MC and FC pairs, respectively. The vertical green lines represent the 0.05 and 0.95 quantiles of the data [Colour figure can be viewed at wileyonlinelibrary.com]

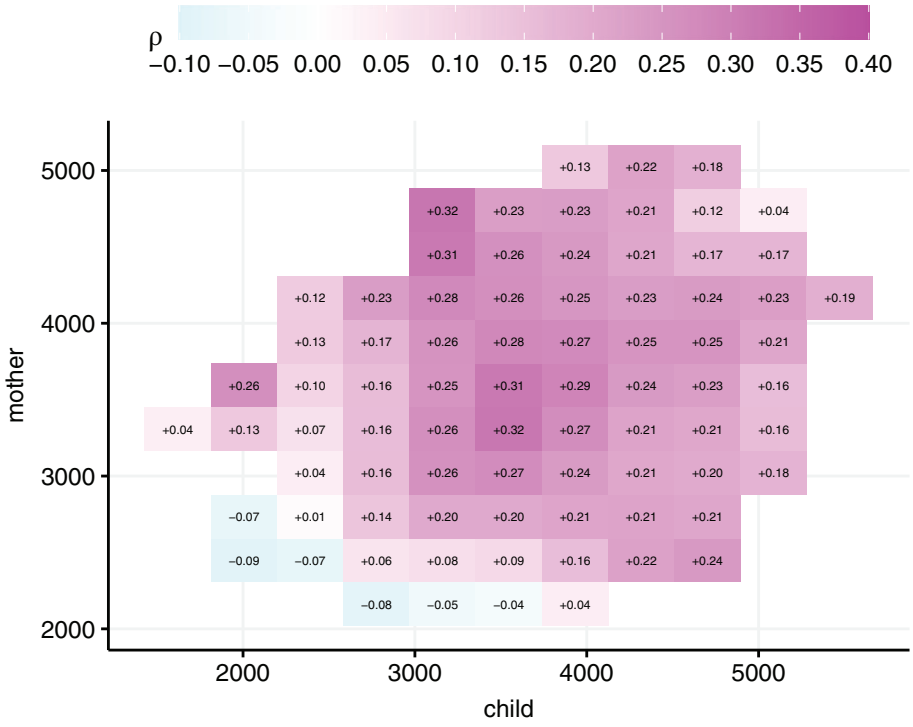


FIGURE B9 Estimated local Gaussian correlation between mother and child. Note that this correlation measure has two location arguments (y_1 and y_2) [Colour figure can be viewed at wileyonlinelibrary.com]

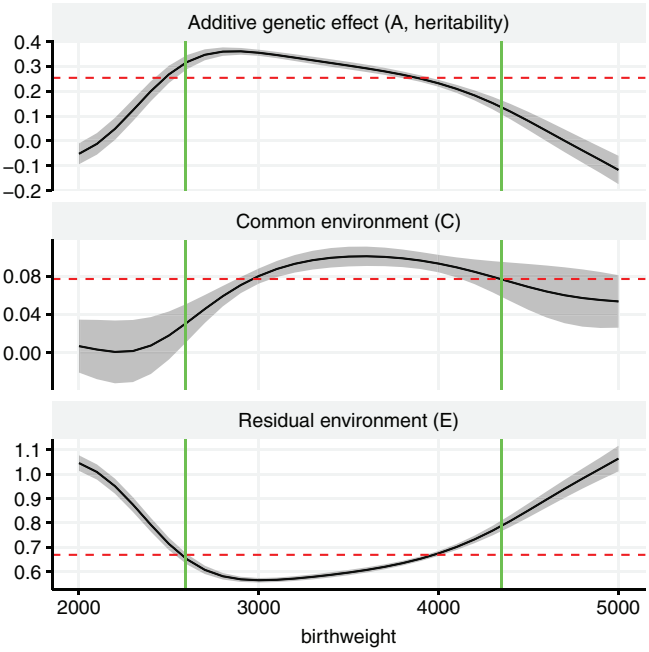


FIGURE B10 Estimated heritability curve $a^2(y)$, common environment $c^2(y)$, and residual environment $e^2(y)$ for the Norwegian Birth Registry data under the ACE model (Definition 3), with pointwise 95% confidence intervals (in gray). The red dashed lines display the classical estimates of heritability and environment, that is, empirical versions of (7). The vertical green lines represent the 0.05 and 0.95 quantiles of the data [Colour figure can be viewed at wileyonlinelibrary.com]

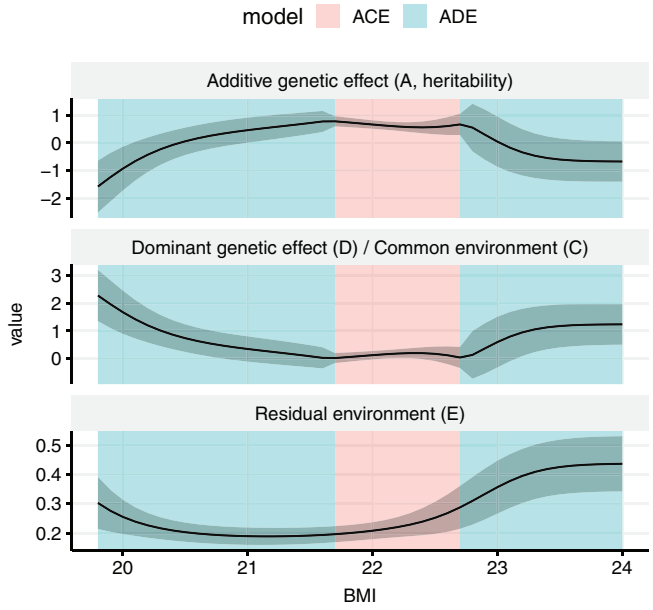


FIGURE B11 Illustration of model switching between ACE and ADE models for the body mass index (BMI) data. The quantities displayed are the same as in Figure B6 [Colour figure can be viewed at wileyonlinelibrary.com]