ON NEGATIVE HERITABILITY AND NEGATIVE ESTIMATES OF HERITABILITY

DAVID STEINSALTZ, ANDY DAHL, AND KENNETH W. WACHTER

Abstract. We consider the problem of interpreting negative maximum likelihood estimates of heritability that sometimes arise from popular statistical models of additive genetic variation. These may result from random noise acting on estimates of genuinely positive heritability, but we argue that they may also arise from misspecification of the standard additive mechanism that is supposed to justify the statistical procedure. Researchers should be open to the possibility that negative heritability estimates could reflect a real physical feature of the biological process from which the data were sampled.

1. Introduction: The meaning of heritability

1.1. **Operational definitions of heritability.** As Albert Jacquard [15] pointed out decades ago, *narrow-sense heritability* — commonly denoted h^2 — has conventionally two distinct meanings:

1. The proportion of total variance attributable to additive genetic effects;
2. The slope of the linear regression of children's phenotypes on the mean parental phenotypes.

Both meanings appear in the earliest works to give a quantitative operational definition to *heritability*, in particular [21]. (For more about the history of the notion of heritability see [2].)

The correspondence between these two meanings depends on an additive model, where genetic and non-genetic effects are independent and sum together to produce the phenotype. When we have general genetic relatedness (rather than parental relations with fixed 50% relatedness) heritability is analogous to a regression coefficient relating phenotypic similarity to genotypic similarity.

We are particularly concerned here with the interpretation of negative estimates of heritability. The appearance of negative estimates for a parameter of crucial scientific interest that is *prima facie* positive is unusual, as has often been noted. Negative estimates of the heritability parameter are often dismissed as a mathematical abstraction, values in a range that arises purely formally and that may only be reported for formal purposes, as part of an ensemble of estimates that collectively are unbiased. Several recent studies [32, 30, 3, 7, 34, 11, 10, 14] have reported individual negative heritability estimates in this

DS supported by Grant ES/N011856/1 from the UK Economic and Social Research Council. AD supported by Grant 1U01HG009080-01 from the U.S. National Institute of Health. KWW supported by Grant 5P30AG012839 from the U.S. National Institute on Aging.

way, including them in averages that themselves came out positive. We illustrate one such analysis [17] using RNA-sequencing data from the GEUVADIS consortium in section 5.

In [16] a point estimate of -0.109 is obtained for heritability of horn length in Soay sheep. It is immediately dismissed with the statement that "it is impossible to have negative heritability" and the inference is drawn that the true heritability must actually be a small positive number toward the upper end of the confidence interval.

We wish to argue that negative heritability estimates need to be taken more seriously. The confusion, we contend, comes from the overlap between statistical models that operationalize the two different interpretations of heritability described above. The argument for rejecting negative estimates appears compelling just so long as the focus is only on the random-effects probability model in equation (1) that typically motivates our definition 1 of heritability. Variance is nonnegative, hence the ratio of two variances cannot be negative. The denominator represents total variance and the numerator represents one component of variance, implying a ratio in $[0, 1]$ if the two components are independent (as the model presumes).

While "variance attributable to additive genetic effects" is a basic element of the genetic model, it has no place in the statistical algorithms such as GREML derived from this model that is widely used to estimate heritability from experimental data. The GREML algorithm is actually (as we will explain in section 2.1) the realization of a multivariate normal model that is naturally constrained to have the parameter $h^2 \geq -1/(\max\{s_i^2\} - 1)$, where $(s_i)_{i=1}^n$ are the singular values of the genotype matrix. If the phenotypes were derived from summing independent additive genetic effects then the true h^2 must indeed be nonnegative, but that must be recognized as an assumption that must be scientifically warranted, as it is not compelled on formal grounds.

In contemporary analyses of genomic data, two roles played by alleles may be distinguished. Alleles exercise or tag actual causal influences on traits. Alleles also serve as markers of family and ancestry, markers of relatedness among individuals that may structure historical, behavioral, social, and environmental influences on traits. As attention expands beyond the additive genetic model to data-analytic genome-wide statistical studies, heritability and the parameter called heritability are not necessarily the same.

1.2. **The meaning of negative heritability.** Once we have accepted the GREML multivariate normal framework — which we will define precisely — we must admit the possibility that the joint distribution of phenotypes and genotypes in a given dataset may be best described by an h^2 value that is negative. The question this raises is, can such a negative heritability estimate be biologically sensible? As described in Section 2, the parameter for heritability may be identified, in a precise way, with a correlation between genotype similarity and phenotype similarity. The model invites us to select an estimate of h^2 that will best match the genetic covariance between individuals to the similarity in their traits. Even if we *want* heritability to be interpreted in the first sense, as a partition of variance, this will not, in general, be correct. All we have access to from the data is an estimate of something like heritability in sense 2. High heritability means that individuals with similar genotype tend to have similar trait values. Zero heritability means that genotypes tell us

nothing about similarities in trait values. Negative heritability, then, could be perfectly sensible as a description of the data: It means that individuals with similar genotypes are likely to have more divergent trait values than those with highly disparate genotypes.

Saying that a given set of data might be best described by a negative heritability estimate goes only part of the way toward answering the question of the biological plausibility of the concept. Suppose you were estimating the weight of water droplets by successively adding them to a small container, and estimating the slope from the sequence of weights. If the scale is sufficiently imprecise it is hardly unlikely that we could estimate a negative slope, yet common sense tells us that negative estimates should be dismissed as unrealistic, and truncated at zero (or combined with prior expectations to form a positive Bayesian estimate). Statistical theory tells us that this system should produce slope estimates that may be positive or negative, but that the probability of a negative estimate goes to zero as the number of measurements goes to infinity, and the estimate converges to its true (positive) value. The essential question is, is there a plausible mechanism that could produce genuine negative heritability, so that as the amount of data generated by the model goes to infinity, the estimate converges to a negative quantity.

The term "negative heritability" appeared for the first time, so far as we are aware, in a paper [12] by J. B. S. Haldane, written around 1960, but first published posthumously in 1996. Haldane described how the maternal-effect trait of neonatal jaundice could be said to display negative heritability: Because the disease results from maternal antibodies against a fetal antigen, it will not arise in a fetus whose mother herself experienced neonatal jaundice.[1] Haldane then calculates a negative heritability from a model that is specialized to the peculiar inheritance structure of neonatal jaundice.

We will suggest one such mechanism in Section 4. As with Haldane's model (which may be understood as a special case), this mechanism has implications which may be implausible or even obviously false in a given experimental setting. It involves interactions between individuals that are not primarily genetic, and so may be dismissed as irrelevant to the study of genetic heritability. The point we want to suggest, though, is that as an abstract physical mechanism that could be producing our data it is as mathematically plausible as the linear random-effects model that undergirds GREML. This is only one example of such a mechanism, and the conclusion we advocate is that negative heritability must be acknowledged as a genuine phenomenon for genotype-phenotype data, even if it may be reasonably excluded by the context of some particular studies. Thinking about what sorts of biological settings could yield negative heritability can also prove an effective guide to understanding when negative heritability estimates may be reliably truncated or ignored.

This is very much like the advice on "interpretation of negative components of variance" propounded in a very different context by the statistician J. A. Nelder [22] in 1954. Nelder considered the problem of ANOVA testing on split-plot experiments, where error for main plots was found to be smaller than the error for subplots, producing a negative estimate for the residual subplot error. As we have done here, Nelder showed how the apparently negative "variance component" could arise either from sampling error from a positive variance

[1]We thank Jonathan Marchini for pointing out this reference to us.

component, or from a misspecification of the model, where correlations between measurements have been neglected. "In any particular situation," Nelder concludes, "it is the statistician's responsibility to decide which model is more appropriate."

2. The GREML model as linear regression

2.1. **The random-effects model.** For the remainder of this paper we follow [26] in using the letter ψ to represent heritability, to avoid the confusing implication built in to the nomenclature h^2 that this parameter cannot be negative.

Underlying GREML, as well as alternative approaches to heritability estimation such as LD-score and Haseman–Elston regression, is a basic random-effects model. Following the notation of [26], our basic object is a data set consisting of an $n \times p$ matrix Z, taken to represent the genotypes of n individuals, measured at p different loci. There is a vector $\mathbf{y}$, representing a scalar observation for each of the n individuals. The underlying observations are counts of alleles taking the values $0, 1$, or 2, but the genotype matrix is centered to have mean zero in each column and normalized to have mean square over the whole matrix equal to 1. (Often, columns are further standardized to variance 1, but we do not make this assumption). The model posits the existence of a random vector $\mathbf{u} \in \mathbb{R}^p$ of genetic influences from the individual SNPs such that

$$\mathbf{y} = Z\mathbf{u} + \varepsilon. \tag{1}$$

The vectors $\mathbf{u}$ and ε are assumed to be independent and to have zero means and i.i.d. normal components. The variances are determined by two parameters, which are to be estimated: θ represents the precision (reciprocal variance) of the non-genetic noise and ψ represents the heritability, entering the model as the ratio of genetic variance to total variance. We will also use the notation $\phi = \psi/(1-\psi)$ in some places, for concision.

The GREML model has been formulated as a random-effects model, but it is equivalent to a multivariate normal model corresponding to the covariance matrix

$$C^2 := \theta_0^{-1} \left((\psi_0/(1-\psi_0))\, A + I_n \right), \tag{2}$$

where $A = ZZ^*/p$ is the Genetic Relatedness Matrix (GRM), and θ_0 and ψ_0 are the true values of the parameters. In this section we describe how the model may also be understood as a linear regression model.

In their original paper [31], Yang and coauthors spell out an analogy between GCTA and a different form of linear regression. They regress squared trait differences between pairs of individuals on corresponding elements of the GRM, with $n(n-1)/2$ points and correlated errors (this is Haseman–Elston regression, which has recently become a popular heritability estimation method due to its speed and robustness to some degree of model misspecification [9, 5]). Instead, we draw an approximate comparison between GREML and regression with n points and independent errors.

Let $Z/\sqrt{p} = U \operatorname{diag}(s_i)\, V^*$ be the singular-value decomposition of $Z/\sqrt{p}$, and rotate the observations to diagonalize the covariance matrix, obtaining

$$\mathbf{z} := U^* \mathbf{y}.$$

Because the columns of Z have zero means, one of the singular values is zero and the corresponding column of U is proportional to a vector with all elements equal to 1. Thus in each other (orthogonal) column of U the elements sum to zero, and each column defines a contrast between weighted groupings of individuals in the sample.

The elements of $\mathbf{z}$ are independent centered normal random variables, and z_i has variance $(1 + (\psi_0/(1-\psi_0))s_i^2)/\theta_0$. It follows that $z_i^2\,\theta_0(1-\psi_0)\,/(1+\psi_0(s_i^2-1))$ are independent chi-square random variables each on one degree of freedom and

$$\log z_i^2 = -\log(\theta_0) - \log(1-\psi_0) + \log\Big(1 + \psi_0(s_i^2 - 1)\Big) + \epsilon_i^*$$

where the ϵ_i^* are distributed as the logarithms of the independent chi-square variables, long-tailed to the left, with mean ≈ -1.302, standard deviation ≈ 2.266, and skewness ≈ -1.643.

The mean of s_i^2 is 1, and when $\psi(s_i^2 - 1)$ are uniformly small we may approximate our equation by

$$\log z_i^2 \approx -\log(\theta_0) - \log(1-\psi_0) + \psi_0(s_i^2 - 1) + \epsilon_i^* \tag{3}$$

Here ψ_0 takes on the role of the true slope for a regression of $\log z_i^2$ on $(s_i^2 - 1)$. It can be estimated by least squares, bearing in mind that the skew of the ϵ_i^* affects standard errors of estimation.

Practitioners of GREML instead usually estimate ψ via maximum likelihood. It is shown in [26] that the maximum likelihood estimates can be expressed in terms of quantities $w_i(\psi) := (1-\psi)\,/\,(1+\psi(s_i^2-1))$ and $v_i(\psi) := w_i(\psi)z_i^2$. They satisfy

$$\begin{aligned} 0 &= \mathrm{Cov}\Big(\mathbf{w}(\hat\psi), \mathbf{v}(\hat\psi)\Big) \\ \hat\theta &= n\,/\,\sum_{i=1}^n v_i(\hat\psi) \end{aligned} \tag{4}$$

Here Cov is the empirical covariance of vector elements, an operation on vectors defined by $\mathrm{Cov}(\mathbf{x},\mathbf{y}) := n^{-1}\sum(x_i - \bar x)(y_i - \bar y)$, and Var is similarly defined. We also set $\tau_2(\psi) = \psi^{-2}\mathrm{Var}(w(\psi))$, and omit the dependence on ψ when helpful. When ψ_0 is modest and the variance of the squared singular values is small, the least-squares and maximum likelihood estimates are close to each other.

Suppose, however, that the true variances of the z_i include a phenotypic contribution that varies inversely with the singular values of Z. In the model developed in Section 4 to first order in $s_i^2 - 1$ the true slope is $(\psi_0 - \alpha_0 - \psi_0^2)/(1-\psi_0)^2$ as a function of a repulsion parameter α. When α_0 exceeds ψ_0, the true slope turns negative and estimated slopes correctly include negative values. The precision of the estimated slopes depends on the variance of the squared singular values. When the $s_i^2 - 1$ are tightly clustered, there will be large standard errors for the estimates of ψ.

2.2. **Simulations.** We illustrate in Figure 1 a close match between regression-based and maximum likelihood estimates for ψ. The plot also emphasizes the wide estimation uncertainty when the singular values s_i^2 are tightly clustered.

We draw s_i^2 from the so-called independent setting, which assumes the genotype matrix Z has independent entries. The known limiting measure for these singular values is the Marcenko–Pastur distribution, from which we draw n equally-spaced values. We directly simulate the independent whitened phenotypes z_i as independent normal variables with mean zero and variance $(1 + (\psi_0/(1-\psi_0))s_i^2)/\theta_0$. This procedure gives ψ estimates that are approximately equivalent to those obtainable from simulating a random Z with i.i.d. entries along with the associated random trait vector y. See section 5 of [26] for details.

We choose Z to have $n = 375$ rows and $p = 1,000$ columns, similar to the dimensions of the *cis*-window genotype matrices in Section 5. Nonetheless, the s_i^2 from our independent setting differ from those empirical singular value distributions. We could simulate a different model for the GRM simply by choosing a different distribution for s_i^2. We also choose $\psi^* = 0.50$ and $\theta^* = 1$.

To compute the MLE, we solve the first equation in (4), a simple univariate function of ψ, and substitute into the second equation. To compute our regression-based estimator, we perform ordinary least squares regression of $\log(z_i^2)$ on s_i^2 and an intercept.

Figure 1 shows a scatterplot of one realization of pairs $(s_i^2, \log z_i^2)$. The red line corresponds to the true value of ψ, i.e. $\psi^* = .5$, and describes the expectation of $\log(z_i^2)$ as a function of s_i^2 given $\theta^* = 1$. The solid pink line corresponds to the ordinary least squares estimate and is very similar to the dashed blue line for the MLE. Dotted green lines are shown for several more distant alternative values for ψ. It is no wonder that slopes are hard to pin down in the face of such huge scatter in observations.

3. Bias from rejecting negative heritability estimates

The common practice of truncating the maximum likelihood calculation to non-negative values introduces bias that is well-known and may be serious for samples of moderate size, both when estimates are truncated at zero and when negatives are ignored. It is thus worth looking beyond the original motivation to the actual structure of the GREML model and considering what meaning negative parameter values might turn out to have.

The problem of estimating the probability of negative heritability estimates was studied fifty years ago by [8]. We add here a few comments about how the framework described in [26] may contribute to understanding the magnitude of the negative heritability estimate problem that arises from sampling noise in settings where the true heritability is understood to be nonnegative, hence where truncation at zero (or rejection of negative estimates) is warranted and guarantees improved estimates in, say, mean squared error. We gain a rough idea of the impact of rejecting negative estimates from a normal approximation

$$\hat{\psi} - \psi_0 \approx \frac{\sqrt{2}}{\sqrt{n\,\tau_2}}\psi_0(1-\psi_0)X,$$

where X has standard normal distribution (see [26] for derivation). Rejecting estimates where $\hat{\psi} < 0$, we have the conditioning bias

$$\mathbb{E}[\hat{\psi}] - \psi_0 \approx \frac{(1-\psi_0)}{\sqrt{n\tau_2}} z_0 \frac{e^{-z_0^2/2}}{\sqrt{\pi}\Phi(z_0)}, \tag{5}$$

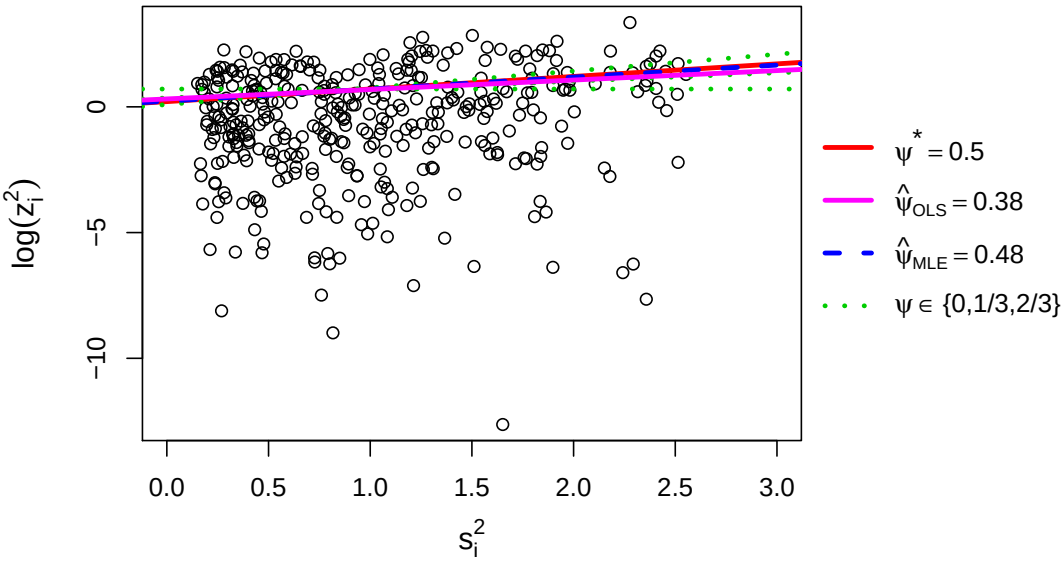

FIGURE 1. Linear-regression approximation of the estimation problem for ψ. The solid red line shows the true line, corresponding to $\psi = 0.5$, while the blue and pink lines respond to the MLE and linear regression estimator, respectively. The dotted green lines correspond to $\psi = 0, 0.33, 0.67, 0.84$.

here $z_0 = \sqrt{n\tau_2}/\sqrt{2}(1-\psi_0)$ and Φ is the standard normal c.d.f. If we instead truncate the estimates — raising all negative $\hat{\psi}$ to 0 — we get the truncation bias

$$\mathbb{E}[\hat{\psi}] - \psi_0 \approx \frac{(1-\psi_0)}{\sqrt{n\tau_2}} z_0 \frac{e^{-z_0^2/2}}{\sqrt{\pi}} - \psi_0\big(1 - \Phi(z_0)\big). \tag{6}$$

4. The phenotypic repulsion model

The notion that new species force their way into phenotypic gaps in the existing ecological community was termed by Darwin the "principle of divergence" and has been further developed by ecologists under the name "phenotypic repulsion" or "phylogenetic repulsion" [28]. Species living in close proximity — which are often closely related phylogenetically — coexist by separating from each other phenotypically. A similar kind of competitive exclusion has been proposed [27] on the individual level to explain observed pattern of developmental variation within human families. Social niche-formation within families has also been proposed by [13] — without an explicit mathematical model — as the basis for an evaluation of gene-environment interaction based on misclassified twin types. While we are not aware of mathematical models of this phenomenon, one could certainly imagine local competition for sunlight, combined with range-limited seed dispersion, yielding an effective phenotypic repulsion between related plants in a forest setting, or monozygotic twins who seek to distinguish themselves from their sibling.

We propose a model of phenotypic repulsion where individuals that are most closely related genetically strive to avoid each other phenotypically. We begin with a model like that described in Section 2.1, where individuals have phenotypes determined by normally distributed effect sizes acting on their individual genotypes. We introduce a penalty term to the probability, of the form

$$\exp\Big\{-\alpha_0\theta_0 \sum_{1\le i<j\le n} a_{ij} y_i y_j\Big\}$$

where $a_{ij} = \frac{1}{p}\sum_{k=1}^{p} Z_{ik}Z_{jk}$ is the (i,j) entry of the GRM, and $\alpha \le 1$ is a parameter with true value α_0 measuring the extent to which genetically similar individuals are pushed to have differing phenotypes. Of course, this setup could be generalized to higher-dimensional phenotypes, with $y_i y_j$ replaced by an arbitrary inner product. The penalty term is inspired by the statistical mechanics models that have been applied to geographically-structured population dynamics, such as the Contact Process [20], used to model the spread of epidemics.

Combining this specification with (2) we see that the phenotypes will now be multivariate normal with mean 0 and covariance matrix

$$\theta_0^{-1}\left[(\phi_0 A + I_n)^{-1} + \alpha_0\,(A - I_n)\right]^{-1}. \tag{7}$$

It follows that the transformed phenotypes $\mathbf{z} = U^*\mathbf{y}$ are independent normal with mean zero and variance

$$\mathrm{Var}(z_i) = \theta_0^{-1}\frac{1+\phi_0 s_i^2}{1-\alpha_0+\alpha_0 s_i^2(1-\phi_0)+\alpha_0\phi_0 s_i^4}\,.$$

Suppose the data have come from this phenotypic–repulsion model, and we analyze them using the random-effects model. While it is always possible to get $\hat{\psi} < 0$ because of random fluctuations, we would like to show that the heritability implied by this model is "really" negative, in the sense that the distribution of $\hat{\psi}$ converges to a strictly negative value as the number of subjects goes to ∞. This will follow from Proposition 4.1 (below) with

$$f(t) = \frac{1+\phi_0 t}{1-\alpha_0+\alpha_0(1-\phi_0)t+\alpha_0\phi_0 t^2},$$

as long as $\phi_0 < \alpha_0$, since

$$f'(t) = \frac{\phi_0 - \alpha_0(1+\phi_0 t)^2}{(1-\alpha_0+\alpha_0(1-\phi_0)t+\alpha_0\phi_0 t^2)^2},$$

which is less than 0 for all $t \ge 0$.

In other words, to the extent that we say that heritability is defined by the linear model, heritability can be negative if genotypes and phenotypes interact through the environment in a manner like the phenotypic repulsion model. This will be true even if the phenotypic interactions are limited to small family groups. We prove that this is the case — that the heritability to which the estimates converge with increasing population size is negative — in the following Proposition, which is proved in section **??**.

Proposition 4.1. *Suppose we have a family of $n \times n$ GRMs A_n for $n \to \infty$, with eigenvalues $s_{n,i}^2$. We suppose that the distributions of eigenvalues converge to a nontrivial distribution $d\sigma(s^2)$.*

Let $U^{(n)}$ be the corresponding eigenvector matrix. For each n we have a multivariate normal random vector $\mathbf{y}^{(n)}$ with covariance matrix $U^{(n)} \operatorname{diag}\big(f(s_{n,i}^2)\big) U^{(n)*}$, where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a strictly decreasing, continuously differentiable function. We assume that the singular values $s_{n,i}$ are bounded above by $s_{\max}$, and that

$$C_1 := \inf_{0 \le t \le s_{\max}^2} (-f'(t)) > 0. \tag{8}$$

(We maintain the normalization assumption that $\sum_i s_{n,i}^2 = n$.)

Let $\hat{\psi}_n$ be the MLE for an observation $\mathbf{y}^{(n)}$, calculated from the random-effects model with GRM A_n. Then $\hat{\psi}$ is bounded above in probability by a strictly negative quantity $-\delta$, depending on C_1 and the distribution σ, as $n \to \infty$. That is, the probability of $\hat{\psi}_n > -\delta$ goes to 0 as $n \to \infty$.

5. Transcriptome-wide Heritability Estimation in GEUVADIS

We end on a brief example to illustrate the practical significance of negative heritability estimates. Although negative estimates of heritability for a single, fixed trait are rarely published, it is common to include negative estimates when profiling heritability across a large number of roughly exchangeable traits [32, 30, 3, 7, 34, 11, 10, 14]. Characterizing such -omic-wide heritability is common in functional genomics, where high-throughput measurements of some genomic property are made at thousands of genomic positions. The most common measurement is (RNA) gene expression, but other prominent examples include methylation levels, chromatin accessibility, expression response to stimuli, or protein expression.

We analyzed an RNA-sequencing dataset from the consortium on Genetic European Variation in Health and Disease (GEUVADIS) [17] [2]. We aligned the raw transcript reads from the European individuals to the reference hg19 transcriptome with the RSEM software package that implements an Expectation-Maximization (EM) algorithm [19]. We removed perfectly correlated genes and genes with low expression mean or variance.

For each i in 375 people and j in $4,154$ genes, we define the phenotype $y_i^{(j)}$ as $\log\big(1 + n_{ij}\big)$ where n_{ij} is the number of observed RNA reads for gene j measured in person i. Our preprocessing resulted in a sample for $4,154$ genes measured on 375 people, each of which we centered and scaled to mean zero, variance one.

Separately for each gene $y^{(j)}$, we estimate its *cis*-heritability, that is, the heritability in expression levels explained by SNPs near to the gene. We do so by fitting our standard model (1) with a genotype matrix $Z^{(j)}$ whose columns correspond to SNPs located up to 1 megabase upstream or downstream of gene j's transcription start site. Restricting to SNPs near a gene is a common way to enrich for functionally causal SNPs. We discard

[2]We thank David Siegel for help processing the GEUVADIS data.

rare SNPs, which we define as SNPs with minor allele frequencies below 2.5%. Finally, we remove genes with less than $1,000$ corresponding SNPs, which excludes 35 genes.

The column dimensions (p) of the separate genotype matrices range from $1,000$ to $20,523$ with a mean of $3,027$ and median of $2,754$. We fit each $\hat{\psi}$ with the maximum likelihood routine from [6], yielding $4,119$ values reflecting systematic variation across genes in their *cis*-heritability, within the limits imposed by sampling error.

The distribution of the resulting transcriptome-wide (*cis*-)heritability estimates is shown in Figure 2 in the form of a smoothed histogram. Clearly, a substantial number of the estimates are negative. The mode is close to zero. Removing negative heritability estimates increases the transcriptome-wide average heritability from 6.2% to 9.0%, and truncating at zero increases it from 6.2% to 6.6%.

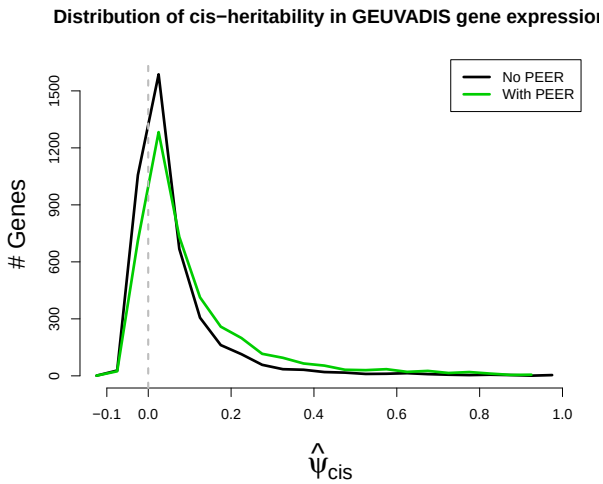

FIGURE 2. Transcriptome-wide density of gene expression *cis*-heritability estimates in the GEUVADIS data.

We repeated the analysis with adjustments based on the technique of Probabilistic Estimation of Expression Residuals (PEER) [24]. We projected out the top 10 PEER factors from the expression phenotypes and the SNPs in Z. This practice, or close variants based on gene expression principal components [1] or surrogate variables [18], is nearly universal in functional genomics [25]. The common aim of these approaches is to approximate latent confounding variation, like experimental batch effects, which can be captured by dimensionality reduction under certain assumptions. The confounder estimates are treated as known covariates and residualized from the phenotype and genotype data.

Correcting for 10 PEER factors increases many of the $\hat{\psi}$ values and reduces the incidence of negative $\hat{\psi}$ as shown in the green curve in Figure 2. However, it is clear that many negative estimates remain. Negative estimates are bound to be part of the picture whenever ψ_0 is small and estimated with low precision, both conditions which will likely hold in most

functional genomic analyses for at least the near future. On the question of whether some negative estimates may be meaningful reflections of non-genetic phenotypic structure, it is well to keep an open mind.

6. Proof of Proposition 4.1

Proof. We follow the general principle, enunciated by [29], that the MLE for the misspecified model will converge to the closest fit in the Kullback–Leibler sense. In other words, the parameter estimate converges in probability to the location of the maximum *expected value* of the log-likelihood function. The arguments of [29] do not apply directly here, because we are not sampling i.i.d. random variables; however, the score function may be written

$$\frac{1}{\bar{v}(\psi)} \cdot G_n(\psi; \mathbf{x}) := \frac{1}{n(1-\psi)\bar{v}(\psi)} \sum_{i=1}^{n} a_i(\psi) x_i, \tag{9}$$

for $-1/(s_{\max}^2 - 1) < \psi \leq 1$, where (x_i) are i.i.d. χ_1^2 random variables and

$$\begin{aligned} a_i(\psi) &\left(\frac{1}{1-\psi+\psi s_{n,i}^2} - n^{-1} \sum_{j=1}^{n} \frac{1}{1-\psi+\psi s_{n,j}^2} \right) \\ &:= \frac{f(s_{n,i}^2)}{1-\psi+\psi s_{n,i}^2} \left(\frac{s_{n,i}^2}{1-\psi+\psi s_{n,i}^2} - n^{-1} \sum_{j=1}^{n} \frac{s_{n,j}^2}{1-\psi+\psi s_{n,j}^2} \right). \end{aligned} \tag{10}$$

We note that the maximum likelihood occurs either at a zero of G_n, or at $\psi = 1$ if G_n is everywhere positive. (It goes to $+\infty$ at the left boundary.) We restrict to $\psi \in [\beta_1, \beta_2]$ for some fixed $\beta_1 < \beta_2$ strictly in the interval $(-1/(s_{\max}^2 - 1), 1)$. The coefficients $a_i(\psi)$ are uniformly bounded and uniformly Lipschitz, so, by a variant of the central result of [33], $G_n(\psi; \mathbf{x})$ converges uniformly in ψ to the function that is the limit of the expected values

$$G(\psi) = \lim_{n\to\infty} G_n(\psi; 1) = \frac{1}{1-\psi} \operatorname{Cov}_\sigma \left(\frac{f(S^2)}{1-\psi+\psi S^2}, \frac{S^2}{1-\psi+\psi S^2} \right).$$

The covariance is understood here to be with respect to S^2 having distribution σ. (This result does not satisfy exactly the conditions of [33], so we provide a proof of the result, stated as Lemma 6.1.) We have assumed that the limiting distribution σ for the squared singular values is nontrivial, hence with nonzero variance. We may assume without loss of generality that

$$C_2 := \inf_n \left(n^{-1} \sum_{i=1}^{n} s_{n,i}^4 - 1 \right) > 0$$

We need to show that G is negative for all ψ above the bound given in (11).

Near $\psi = 0$ the function $G_n(\psi; 1)$ is well-behaved, and takes on the value $\operatorname{Cov}(f(s_{n,i}^2), s_{n,i}^2)$ at $\psi = 0$. Since $f(t) + C_1 t$ is a decreasing function of t, for $t \in [0, s_{\max}^2]$, we have by Harris's

inequality [4, Theorem 2.15]

$$\begin{aligned} G(0) &= \operatorname{Cov}\left(f(S^2) + C_1 S^2, S^2\right) - C_1 \operatorname{Var}\left(S^2\right) \\ &\le -C_1 C_2 \\ &< 0. \end{aligned}$$

We also have

$$\begin{aligned} (1-\psi)G'(\psi) = &- \operatorname{Cov}\left(\frac{f(S^2)}{1-\psi+\psi S^2}, \frac{S^2(S^2-1)}{(1-\psi+\psi S^2)^2}\right) \\ &- \operatorname{Cov}\left(\frac{(S^2-1)f(S^2)}{(1-\psi+\psi S^2)^2}, \frac{S^2}{1-\psi+\psi S^2}\right) \\ &+ (1-\psi)^{-1} \operatorname{Cov}\left(\frac{f(S^2)}{1-\psi+\psi S^2}, \frac{S^2}{1-\psi+\psi S^2}\right). \end{aligned}$$

Since f is decreasing, we have for $-\frac{1}{2s_{\max}^2-1} < \psi < 0$ the bound

$$\left|G'(\psi)\right| \le 20 s_{\max}^4 f(0).$$

Thus $G(\psi) < 0$ for all

$$\psi \in \left(-\frac{C_1 C_2}{20 s_{\max}^4 f(0)}, 0\right). \tag{11}$$

It follows that $G_n(\psi; \mathbf{x})$ is negative for all ψ between 0 and the bound given in (11), with probability tending to 1 as $n \to \infty$.

We note now that for $\psi \in [0,1]$ $f(t)/(1-\psi+\psi t)$ is a decreasing function of t, and $t/(1-\psi+\psi t)$ is increasing, so (again by Harris's Inequality) $G(\psi) < 0$, which completes the proof. □

Lemma 6.1. *Let $a : [0, s_{\max}] \times [\beta_1, \beta_2] \to \mathbb{R}$ be a function that is bounded and Lipschitz. Let $s_{n,1}, \ldots, s_{n,n}$ be a triangular array of uniformly bounded real numbers, with a limiting measure σ, so that $n^{-1}\sum_{i=1}^n \delta_{s_{n,i}}$ converges weakly to σ, where δ_x represents unit point mass at x. Let $x_{n,1}, \ldots, x_{n,n}$ be independent random variables with $\mathbb{E}[x_{n,i}] = 1$ and $\operatorname{Var}(x_{n,i}) \le V$ for a fixed number V. Define*

$$G_n(\psi) := n^{-1} \sum_{i=1}^n x_{n,i} a\big(s_{n,i}, \psi\big).$$

Then G_n converges weakly in the uniform topology to

$$G(\cdot) := \int_0^{s_{\max}} a(s, \cdot) d\sigma(s).$$

That is,

$$\sup_{\psi \in [\beta_1, \beta_2]} \left|G(\psi) - G_n(\psi)\right| \xrightarrow{n\to\infty}_P 0.$$

Proof. We have

$$\sup_{\psi\in[\beta_1,\beta_2]}\big|G(\psi)-G_n(\psi)\big| \le \sup_{\psi\in[\beta_1,\beta_2]}\Big|n^{-1}\sum_{i=1}^n\big(x_{n,i}-1\big)\,a\big(s_{n,i},\psi\big)\Big| + \sup_{\psi\in[\beta_1,\beta_2]}\Big|n^{-1}\sum_{i=1}^n a\big(s_{n,i},\psi\big)-\int_0^{s_{\max}} a(s,\psi)d\sigma(s)\Big|. \tag{12}$$

The Lipschitz property of a implies that the second term may be bounded, for any fixed positive integer k, by

$$\frac{2\alpha}{k}(\beta_2-\beta_1)+\max_{1\le j\le k}\Big|n^{-1}\sum_{i=1}^n a\big(s_{n,i},\psi_j\big)-\int_0^{s_{\max}} a\left(s,\psi_j\right)d\sigma(s)\Big|, \tag{13}$$

where

$$\psi_j := \left(1-\frac{j}{k}\right)\beta_1+\frac{j}{k}\beta_2 \text{ for } 1\le j\le k,$$
$$\alpha := \max\{\sup|a|,\mathrm{Lip}(a)\}$$

Because of the assumed weak convergence of the distribution of the $s_{n,i}$, this converges to $\frac{2\alpha}{k}(\beta_2-\beta_1)$ as $n\to\infty$ for each fixed k; hence the second term converges to 0 as $n\to\infty$.

To deal with the first term we use the standard method of *chaining* (*cf.* [23, chapter 3]): We define finite skeletons of $[\beta_1,\beta_2]$, subsets $D_0\subset D_1\subset\cdots$ with $|D_j|=2^j$, defined by

$$D_j := \left\{\beta_1+\frac{2\ell+1}{2^{j+1}}(\beta_2-\beta_1)\,:\,\ell=0,\cdots,2^j-1\right\}.$$

Note that each $\psi\in D_j$ for $j\ge 1$ has a unique nearest neighbour in D_{j-1}, which we will denote by ψ', and $|\psi-\psi'|=(\beta_2-\beta_1)2^{-j-1}$. Define

$$R_0 := n^{-1/2}\left|\sum_{i=1}^n\big(x_{n,i}-1\big)\,a\big(s_{n,i},\psi_0\big)\right|,$$
$$R_j := \max_{\psi\in D_j} n^{-1/2}\left|\sum_{i=1}^n\big(x_{n,i}-1\big)\,a\big(s_{n,i},\psi\big)-\sum_{i=1}^n\big(x_{n,i}-1\big)\,a\big(s_{n,i},\psi'\big)\right|.$$

We have

$$\mathbb{E}\left[R_0^2\right]\le n^{-1}V\sum_{i=1}^n a\left(s_{n,i},\psi_0\right)^2\le\alpha^2V.$$

For any collection of random variables $\xi_1,\ldots,\xi_m$ we know that

$$\mathbb{E}\left[\max\xi_k^2\right]\le m\max\mathbb{E}\left[\xi_k^2\right],$$

so

$$\mathbb{E}\left[R_j^2\right]\le 2^j\left(2^{-j-1}(\beta_2-\beta_1)\alpha\right)^2V=2^{-j-2}(\beta_2-\beta_1)^2\alpha^2V.$$

By the Cauchy–Schwarz inquality we have

$$\mathbb{E}\left[(R_0+\cdots+R_k)^2\right] \le \Big(\sum_{j=0}^{k}\mathbb{E}\left[R_j^2\right]^{1/2}\Big)^2 \le 2\alpha^2(\beta_2-\beta_1)^2 V.$$

So finally, since

$$\sup_{\psi\in[\beta_1,\beta_2]}\Big|n^{-1}\sum_{i=1}^{n}\big(x_{n,i}-1\big)\,a\big(s_{n,i},\psi\big)\Big| \le n^{-1/2}\sum_{j=0}^{\infty}R_j,$$

we have

$$\mathbb{E}\left[\sup_{\psi\in[\beta_1,\beta_2]}\Big|n^{-1}\sum_{i=1}^{n}\big(x_{n,i}-1\big)\,a\big(s_{n,i},\psi\big)\Big|^2\right] \le \frac{2\alpha^2 V(\beta_2-\beta_1)^2}{n}. \tag{14}$$

By Markov's inequality, combining this with (13) completes the proof. □

References

[1] O Alter et al. "Singular value decomposition for genome-wide expression data processing and modeling." In: *Proceedings of the National Academy of Sciences* 97.18 (Aug. 2000), pp. 10101–10106.

[2] A Earl Bell. "Heritability in retrospect." In: *The Journal of Heredity* 68.5 (1976), pp. 297–300.

[3] Gaurav Bhatia et al. "Haplotypes of common SNPs can explain missing heritability of complex diseases". In: *BioRxiv* (July 2015), pp. 1–25.

[4] Stéphane Boucheron et al. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[5] Guo-Bo Chen. "Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman–Elston regression". In: *Frontiers in Genetics* 5 (2014), p. 107.

[6] Andrew Dahl et al. "A multiple-phenotype imputation method for genetic studies". In: *Nature Genetics* 48.4 (Feb. 2016), pp. 466–472.

[7] Hilary K Finucane et al. "Partitioning heritability by functional annotation using genome-wide association summary statistics". In: *Nature Genetics* (Sept. 2015).

[8] JL Gill and EL Jensen. "Probability of obtaining negative estimates of heritability". In: *Biometrics* (1968), pp. 517–526.

[9] David Golan et al. "Measuring missing heritability: inferring the contribution of common variants." In: *Proceedings of the National Academy of Sciences of the United States of America* 111.49 (Dec. 2014), E5272–81.

[10] Alexander Gusev et al. "Integrative approaches for large-scale transcriptome-wide association studies". In: *Nature Genetics* 48.3 (Feb. 2016), pp. 245–252.

[11] Melissa Gymrek et al. "Abundant contribution of short tandem repeats to gene expression variation in humans." In: *Nature Genetics* 48.1 (Jan. 2016), pp. 22–29.

[12] J. B. S. Haldane. "The negative heritability of neonatal jaundice". In: *Annals of Human Genetics* 60 (1996), pp. 3–5.

[13] "Heritability and the equal environments assumption: Evidence from multiple samples of misclassified twins". In: *Behavior Genetics* 43.5 (2013), pp. 415–426.

[14] Ryan D Hernandez et al. "Singleton Variants Dominate the Genetic Architecture of Human Gene Expression". In: *BioRxiv* (2017), p. 219238.

[15] Albert Jacquard. "Heritability: one word, three concepts". In: *Biometrics* (1983), pp. 465–477.

[16] S. E. Johnston et al. "Horn type and horn length genes map to the same chromosomal region in Soay sheep". In: *Heredity* 104 (2010), pp. 196–205.

[17] Tuuli Lappalainen et al. "Transcriptome and genome sequencing uncovers functional variation in humans." In: *Nature* 501.7468 (Sept. 2013), pp. 506–511.

[18] Jeffrey T Leek and John D Storey. "Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis". In: *PLoS Genetics* 3.9 (2007), e161.

[19] Bo Li and Colin N Dewey. "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome". In: *BMC Bioinformatics* 12.1 (2011), p. 323.

[20] Thomas Liggett. *Stochastic Interacting Systems : Contact, Voter, and Exclusion Processes*. New York: Springer Verlag, 1999.

[21] Jay L Lush. "Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics". In: *Proceedings of the American Society of Animal Nutrition* 1940.1 (1940), pp. 293–301.

[22] JA Nelder. "The interpretation of negative components of variance". In: *Biometrika* 41.3/4 (1954), pp. 544–548.

[23] David Pollard. *Empirical Processes: Theory and Applications*. Vol. 2. CBMS-NSF Regional Conference Series in Probability and Statistics. Hayward, California: Institute of Mathematical, 1990.

[24] Oliver Stegle et al. "A Bayesian Framework to Account for Complex Non-Genetic Factors in Gene Expression Levels Greatly Increases Power in eQTL Studies". In: *PLoS Computational Biology* 6.5 (2010), e1000770.

[25] Oliver Stegle et al. "Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses". In: *Nature Protocols* 7.3 (2012), pp. 500–507.

[26] David Steinsaltz et al. "Statistical properties of simple random-effects models for genetic heritability". In: *Electronic Journal of Statistics* 12.1 (2018), pp. 321–358.

[27] Frank J Sulloway. "Why siblings are like Darwin's finches: Birth order, sibling competition, and adaptive divergence within the family". In: *The Evolution of Personality and Individual Differences*. Ed. by David M. Buss and Patricia H. Hawley. Oxford University Press New York, 2011, pp. 87–119.

[28] Campbell O Webb et al. "Phylogenies and community ecology". In: *Annual review of ecology and systematics* 33.1 (2002), pp. 475–505.

[29] Halbert White. "Maximum likelihood estimation of misspecified models". In: *Econometrica: Journal of the Econometric Society* (1982), pp. 1–25.

[30] Fred A Wright et al. "Heritability and genomics of gene expression in peripheral blood". In: *Nature Genetics* 46.5 (May 2014), pp. 430–437.

[31] Jian Yang et al. "Common SNPs explain a large proportion of the heritability for human height". In: *Nature Genetics* 42.7 (2010), pp. 565–569.

[32] Jian Yang et al. "Ubiquitous polygenicity of human complex traits: genome-wide analysis of 49 traits in Koreans." In: *PLoS Genetics* 9.3 (2013), e1003355.

[33] Ke-Hai Yuan. "A theorem on uniform convergence of stochastic functions with applications". In: *Journal of Multivariate Analysis* 62.1 (1997), pp. 100–109.

[34] Zhihong Zhu et al. "Dominance genetic variation contributes little to the missing heritability for human complex traits." In: *American Journal of Human Genetics* 96.3 (Mar. 2015), pp. 377–385.