Testing the moderation of quantitative gene by environment interactions in unrelated individuals

Rasool Tahmasbi[1*]; Luke M. Evans[1]; Eric Turkheimer[2]; Matthew C. Keller[1,3]

[1] Institute for Behavioral Genetics (IBG), University of Colorado, Boulder CO; [2] Department of Psychology, University of Virginia, Charlottesville, VA; [3] Department of Psychology and Neuroscience, University of Colorado, Boulder CO

The environment can moderate the effect of genes – a phenomenon called gene–environment (GxE) interaction. There are two broad types of GxE modeled in human behavior – qualitative GxE, where the effects of individual genetic variants differ depending on some environmental moderator, and quantitative GxE, where the additive genetic variance changes as a function of an environmental moderator. Tests of both qualitative and quantitative GxE have traditionally relied on comparing the covariances between twins and close relatives, but recently there has been interest in testing such models on unrelated individuals measured on genomewide data. However, to date, there has been no ability to test quantitative GxE effects in unrelated individuals using genomewide data because standard software cannot solve nonlinear constraints. Here, we introduce a maximum likelihood approach with parallel constrained optimization to fit such models. We use simulation to estimate the accuracy, power, and type I error rates of our method and to gauge its computational performance, and then apply this method to IQ data measured on 40,172 individuals with whole-genome SNP data from the UK Biobank. We found that the additive genetic variation of IQ tagged by SNPs increases as socioeconomic status (SES) decreases, opposite the direction found by several twin studies conducted in the U.S. on adolescents, but consistent with several studies from Europe and Australia on adults.

The effects of genes do not exist in a vacuum; they are likely to be influenced by the environmental background to various degrees. Understanding such GxE interactions has been a major focus of disease and behavioral genetic research over the past twenty years. Much of this research has investigated qualitative GxE effects using a candidate gene approach, such that the effects of specific genetic polymorphisms chosen a-priori based on biological hypotheses were modeled as a function of environmental moderators (e.g., [1]). However, concerns of high false positive rates [2], a history of poor replication [3], and the realization that individual genetic effect sizes are typically very small [4] has cast doubt on the utility of candidate gene-by-environment interaction studies. An alternative approach is to ask whether genetic effects across the genome change, on average, across an environmental moderator [5]. Qualitative GxE effects (see Supplemental Text) manifest as a non-unity genetic correlation between the same trait at different levels of an environment. Tests of such qualitative GxE effects have long been employed in samples of close relatives and twins [6], but have recently been tested among unrelated individuals using genome-wide SNP data, instantiated in the popular GCTA software using a mixed linear effects approach [7].

Maximum likelihood methods using close relatives have also been used to test quantitative GxE effects, in which genetic or environmental variance components change across the level of moderator [8]. Twin analyses of depression [1] [9] [10] [11] [12] [13], schizophrenia and bipolar disorder [14], alcohol and drug use and abuse [15] [16] [17], and others traits [18] [16] [17] have shown that the genetic and/or environmental variation underlying human behavior is often non-constant across different environments. Perhaps the best known example of this approach was Turkheimer's [19], finding that the additive genetic variance, V_A, of IQ was lower for low SES than high SES individuals, which had also been reported previously [20] [21] [22] [23] [24] [25] [26]. This study prompted multiple follow-up studies, with some replicating the original finding and others not (Table S1).

Testing for quantitative GxE effects in unrelated individuals is important because close family members share environmental and non-additive genetic factors that, in combination, can lead to serious biases in estimates of additive genetic variation [27] [28] [29]. Furthermore, much more genome-wide data is available to researchers than twin/family data, and this is especially so for rare disorders. To date, however, there has been no ability to directly test quantitative GxE effects in a unified modeling approach using genome-wide SNPs in unrelated samples. Instead, to investigate changes in the genetic variance tagged by SNPs across a moderator, samples have been binned at different levels of the moderator, with genetic variance or SNP-heritability assessed separately in each group [30] [31]. Unfortunately, such an approach loses power compared to an approach that models all the data simultaneously, assumes that variances do not change as a function of the moderator within bins, and it make it difficult to test functionally different forms of possible interactions. Furthermore, if heritability (rather than additive genetic variance) is estimated separately per bin, it is implicitly assumed that variances are equal across bins, whereas what is often of interest is whether the absolute magnitude of genetic or environmental variation changes.

In this paper, we introduce a model for testing quantitative GxE effects in unrelated samples using genome-wide SNPs and assess its accuracy by simulation. We then apply this method to a sample of 40,172 individuals in the UK Biobank to understand whether and how genetic variation underlying IQ changes as a function of SES in this population.

To model unrelated individuals, let

$$\mathrm{z_{ik}} = \frac{\mathrm{x_{ik}} - 2\mathrm{p_k}}{\sqrt{2\mathrm{p_k q_k}}}, \qquad (1)$$

be the standardized genotype for individual i and variant k, where $\mathrm{x_{ik}} \in \{0,1,2\}$ and $\mathrm{p_k}$ is the allele frequency for the kth variant. Denote by $\mathrm{V_P}$ the phenotypic variance component and $\mathrm{V_E}$ non-genetic variance component, such that $\mathrm{V_P} = \mathrm{V_A} + \mathrm{V_E}$.

We write the quantitative gene by environment interaction model for individual i as

$$y_\mathrm{i} = \beta_0 + \textstyle\sum_\mathrm{l} \beta_\mathrm{l}\, \mathrm{w_{il}} + \lambda \mathrm{M_i} + (a + a'\mathrm{M_i})\mathrm{g_i} + (e + e'\mathrm{M_i})\epsilon_\mathrm{i}, (2)$$

where β_l's are coefficients corresponding to $\mathrm{w_{il}}$ covariates, $\mathrm{M_i}$ is the standardized moderator and λ its effect, $\mathrm{g_i} = \sum_{\mathrm{k=1}}^{\mathrm{m}} \mathrm{z_{ik}}\, \alpha_\mathrm{k}$, with the coefficients α_k's representing genetic effects for each of m SNPs assumed to be drawn from a normal distribution with mean zero and variance $1/\mathrm{m}$, and ϵ_i's representing error effects drawn from a standard normal distribution and independent from α_k. The a and e coefficients represent the importance of additive genetic and environmental factors, respectively, while the a' and e' coefficients represent the degree to which the additive genetic and environmental influences change as a function of the moderator, M. In this (full) model, denoted Model 1, the additive and error variances are $\mathrm{V_A} = (a + a'\mathrm{M})^2$ and $\mathrm{V_E} = (e + e'\mathrm{M})^2$, which change as a function of moderator. Purcell [8] used a similar model for twin data.

From within this framework, we can define other models where $\mathrm{V_A}$ is constant but $\mathrm{V_E}$ changes as a function of moderator by setting $a' = 0$ (Model 2), where $\mathrm{V_A}$ changes as a function of moderator but $\mathrm{V_E}$ is constant by setting $e' = 0$ (Model 3), and where both $\mathrm{V_A}$ and $\mathrm{V_E}$ are constant by setting $a' = e' = 0$ (Model 4). Model 4 is the same as the base REML model instantiated in GCTA, such that $\mathrm{V_A}$, $\mathrm{V_E}$ and h^2 are constant, but is useful for comparison and hypothesis testing. We also define a final model, Model 5, where $\mathrm{V_A}$ and $\mathrm{V_E}$ change as a function of moderator, but h^2 is constant, by constraining $e' =$

ea'/a (note that setting $e' = a'$ does not accomplish this; see Supplement). This model is useful for testing if changes in $\mathrm{V_A}$ and $\mathrm{V_E}$ are more parsimoniously explained by a change in V_P. To the best of our knowledge, Model 5 or equivalent models, where the proportionate changes of $\mathrm{V_A}$ and $\mathrm{V_E}$ are constrained to be equal, have not been developed or tested in models designed for twin/family data (e.g., [8]). The best model for fitting the data can be determined based on formal hypothesis tests or, for models that are not nested, on the AIC/BIC fit criteria.

For Model 1,

$$\mathrm{cov}[\mathrm{Y}|\mathrm{Z},\mathrm{T}] = \frac{1}{\mathrm{m}}\mathrm{T} \circ \mathrm{ZZ}' + \mathrm{K} \circ \mathrm{I_n}, \quad (3)$$

where $Y = [y_1, \ldots, y_n]'$ is the column vector of phenotypes and I_n is the identity matrix of size n, $Z = [z_{ik}]$ is the standardized genotype matrix, $T = [(a + a'M_i)(a + a'M_j)]$, $K = [(e + e'M_i)(e + e'M_j)]$, and the operator $\circ$ is Schur product (or matrix element–wise product). (Details for computing the covariance matrices for all five models are shown in the Supplement). We assume $\mathrm{Y} = [\mathrm{y}_1, \ldots, \mathrm{y}_\mathrm{n}]'$ follows a normal distribution with mean $\beta_0 + \sum_l \beta_l \mathrm{w_{il}} + \lambda \mathrm{M_i}$ and the covariance matrix estimated for each model. If we let $A = \frac{1}{\mathrm{m}}\mathrm{ZZ}'$ be the estimated genetic relationship matrix (GRM) from whole genome SNP data, equation (3) becomes

$$\mathrm{cov}[\mathrm{Y}|\mathrm{Z},\mathrm{T}] = a^2\mathrm{A} + aa'\mathrm{A_2} + a'^2\mathrm{A_3} + \mathrm{K} \circ \mathrm{I_n} \quad (4)$$

where A_2 and A_3 are additional GRMs that are functions of both the moderator and A (see Supplement for details). As can be seen, the coefficient of the second term (aa') is a function of the first and third term, which makes equation (3) a constrained covariance matrix. REML (implemented by GCTA) can deal with multiple GRMs if their coefficients are independent of each other, but that is not the case here. If these three GRMs (A, $\mathrm{A_2}$, $\mathrm{A_3}$) are entered into GCTA, it will estimate a coefficient of the second term that is not constrained to equal aa'. Here, we maximize the log-likelihood function using parallel constrained optimization (see the definition of matrix V in [32], page 77) assuming that the phenotypes follow a multivariate normal distribution (see Methods).

We ran a comprehensive set of simulations (Method section) to investigate the performance of the proposed method. Phenotypes were simulated from each of the 5 models with 6 different sets of parameters (Table S2) and different sample sizes. The results are shown in Tables S3-S6. The biases of the estimated parameters were not statistically significant from zero. The simulation results for type I error are presented in Supplementary Figures S1-S9 and Tables S7-S12, and show no inflation of type-I error rates. Figure 1 presents the statistical power for testing $a' = 0$ in Model 3, and shows 80% power for detecting a 5% increase in V_A for every standard deviation increase in the moderator (when a=.63 and a'=.04) once sample sizes are above 8000. The power for a given parameter differs across models (see Supplementary Tables S7-S12 and Supplementary Figures S10-S17), and is lower in models attempting to estimate more parameters due to correlations between the estimates. For example, 80% power for detecting a' in Model 3 requires a sample size of 1000, but due to the correlation between estimates of a' and e', requires a sample of size 6000 in Model 1.

Figure 2 shows results from a sensitivity analysis, where the data are simulated from Model 1 and the parameters are estimated from Models 1-5, to show the effects of model misspecification on parameter estimates. Estimates are unbiased when the correct model is used, but a' is overestimated and a underestimated when e' is incorrectly dropped, and e' is overestimated and e underestimated when a' is incorrectly dropped. When both a' and e' are incorrectly dropped, such as would occur using the traditional approach and not allowing for moderation of V_A or V_E, estimates for a are unbiased but estimates for e are overestimated, leading to underestimation of h^2. Figures S18-S20 show similar results where data are simulated from models 2, 3, 4, and 5 respectively (see also Tables S18-S21). Overall, our results indicate that estimates are unbiased when the correct model is chosen but can be biased to various degrees under model misspecification.

It is important to note that environmental effect may be correlated with the genetic effect on the trait (r_{GE}) rather than modifying the genetic effects on the trait (GxE). r_{GE} implies that certain alleles are over- or under-represented depending on the value of the moderator, and can appear as quantitative GxE in certain ways of modeling GxE, e.g., by stratifying the sample by the moderator. Entering the moderator in the means model as a main effect, as is done here, will effectively remove from the covariance model any genetic effects that are shared between trait and moderator [8].

The use of unrelated samples with genome-wide SNP data allow investigations of quantitative GxE hypotheses in larger sample sizes and on more phenotypes than are available in twin and family datasets while avoiding potential biases that exist when close relatives are modeled. To demonstrate our approach, we investigate the moderation of variance components of IQ as a function of a measure of SES (the reverse-scaled Townsend Deprivation Index) in the UK Biobank, given that this has been a hypothesis of great interest (Table S1). The estimated parameters along with their 95% confidence intervals (CI) and p-values for all five models are presented in Table 1. Across all models, the estimated parameters a and e are similar, showing consistency of the estimated $\mathrm{V_A}$ and $\mathrm{V_E}$ at the mean level of SES, and estimates of a' and e' are negative, showing that estimated $\mathrm{V_A}$ and $\mathrm{V_E}$ decrease as a function of SES in the range of SES investigated. Constraining the heritability to be the same across the moderator by setting $e' = ea'/a$ (Model 5 vs. Model 1) led to a non-significant decrease in fit (p = .145), suggesting that overall V_P changes as a function of SES and that $\mathrm{V_A}$ and $\mathrm{V_E}$ change roughly proportionately. Consistent with this, Model 5 had the lowest AIC and BIC values, making it the most parsimonious model (see Figure 3).

Our results are in the opposite direction of those reported by several studies conducted in the US [19] [21] [22]. However, our results are more consistent with several findings from Western Europe and Australia, where $\mathrm{V_A}$, on average, decreases slightly as a function of SES [20]. However, even studies from Western Europe and Australia have tended to find virtually no change in overall V_P (due to a counter-balancing effect of $\mathrm{V_E}$ increases as a function of SES), whereas we found a significant decrease in V_P across SES. While it is possible that moderation of unmodeled non-additive genetic effects in twin studies could lead to discrepancies between the current results and those based on twins, this cannot explain different patterns of changes in V_P. Thus, the source of discrepancies across this studies and previous ones based on twins may have to do with differences in measures of IQ, of SES, or in differences in study populations. Almost all the US twin studies are conducted in adolescent and early childhood, while this study and [20] are on adults (see Table S1).

There are two limitations regarding the modeling approach for quantitative GxE we introduced. First, because codes were written in R, the computational speed is not optimal (see Table 1), although we have partially resolved this problem by finding a better starting point from moment matching methods (Haseman-Elston regression). Second, we have not yet developed methods to estimate quantitative GxE for categorical outcomes, such as occurs in case-control studies. Both issues are potentially addressable with further refinement of the code and model in the future.

We have demonstrated a general approach for estimating quantitative GxE in unrelated samples using constrained optimization. We showed by simulation that the bias of the estimated parameters is negligible, that type-I errors are appropriately controlled, and that estimates can be biased under model misspecification. In particular, if quantitative GxE effects occur, we showed that traditional approaches that do not model GxE underestimate heritability. We applied our method to whole-genome SNP data from the UK Biobank, and found that phenotypic variance of IQ decreases as a function of SES, but that heritability of SES remains roughly constant.

Materials and Methods

Data. UK Biobank recruited 500,000 people aged between 40-79 years in 2006-2010 from across the UK. Prospective participants were invited to visit an assessment center, at which they completed an automated questionnaire and were interviewed about lifestyle, medical history and nutritional habits; basic variables such weight, height, blood pressure etc. were measured; and blood and urine samples were taken, and DNA was extracted from blood. Genotyping was done using two closely related arrays, with each having ~800,000 SNP markers. Samples were analyzed in batches of approximately 4700 individuals.

Data quality control. Participants were tested for fluid intelligence at up to three separate occasions; when more than one score was available for an individual, we selected the first score. Fluid intelligence score is a simple unweighted sum of the number of correct answers given to the 13 fluid intelligence questions. Participants who did not answer all of the questions within the allotted 2-minute limit were scored as zero for each unanswered question. The mean for standardized fluid intelligence score was .046 in males and -.042 in females ($p < 0.001$). Participant age (mean=58.2, SD=7.99) was computed from the appropriate fluid intelligence collection date minus the birthday. The standardized fluid intelligence score decreased slightly as age increased (beta = -.0095, $p \sim 0$, adjusted $R^2 = 0.006$). Townsend deprivation index (TDI) was calculated immediately prior to participant joining UK Biobank based on the area in which their postcode was located. The mean for standardized TDI was 0.0028 and -0.0025 in males and females, respectively ($p = 0.60$). In this paper, we used reverse-coded TDI as a measure for SES.

After merging fluid intelligence scores with non-missing TDI, sex, age at recruitment, place born and genotype measurement batch, 41,908 Caucasian individuals remained with genotype information. After dropping individuals with SNP missingness >

.03 and dropping a minimal number of individuals in pairs with genomic relatedness > 0.05, the final sample size was 40,172. We used the first 15 principal components as covariates (see Table S25). In addition to the UKB standard genotypic quality control, we dropped SNPs with missingness > .05 and with Hardy-Weinberg equilibrium threshold $p < 10^{-6}$, leaving 345,767 SNPs.

Simulation procedure. We simulated populations with sizes $N \in \{500,1000,2000,4000,8000\}$ for different sets of parameters $\theta = (a, a', e, e')$. These values for θ are shown in Table S2. For θ_6 with $(a, a') = (0.633,0.038)$, the genotypic variance was $V_A = a^2 = .4$ for $M = 0$ and was $(a + a' M)^2 = .45$ for $M = 1$, i.e., increasing one standard unit of the moderator led to an increase of 0.05 units of V_A. Similarly, for $(e, e') = (0.774, 0.032)$, the non-genotypic variance was $V_E = e^2 = .6$ for $M = 0$ and was $(e + e' M)^2 = .65$ for $M = 1$.

Genotypes were simulated from UK Biobank array data and phenotypes were simulated using models 1–5 for different sets of 1000 causal variants (CVs) in each replication. For each set of the parameters, we simulated $r = 200$ replications (for N=8000, $r = 100$). To gauge the performance of the proposed method, we estimated the parameters for each replication and computed the bias and variance of each estimate as

$$\begin{aligned} \mathrm{bias}(\hat{\theta}) &= \frac{1}{r}\sum_r \left(\hat{\theta}_r - \theta\right), \\ \mathrm{var}(\hat{\theta}) &= \frac{1}{r}\sum_r \left(\hat{\theta}_r - \mathbb{E}(\hat{\theta})\right)^2, \end{aligned}$$

for $\theta \in \{a, a', e, e'\}$, where $\mathbb{E}(\hat{\theta}) = 1/r \sum_r \hat{\theta}_r$ (Tables 2 and S3-S6).

To investigate statistical power and type-I error rates, we simulated r data sets with sizes $N \in \{500,1000,2000,4000,8000\}$ from models defined under the alternative and null hypotheses respectively, and then computed the maximum value of the log-likelihood for the alternative, $\ell(\Theta_1)$ and the null, $\ell(\Theta_0)$. The test statistics is

$$\chi^2 = -2(\ell(\Theta_0) - \ell(\Theta_1))$$

was compared to the critical value $Q_{1-\alpha}$, which is obtained from the central chi-square distribution with df degrees of freedom and $\alpha = 0.05$, where df is the difference between the number of free parameters of models alternative and null. Power was computed as

$$\text{Power} = \frac{1}{r}\sum_{i=1}^{r} 1[\chi_i^2 > Q_{1-\alpha}]\,,$$

where $r = 200$ replications. Similarly, we computed the type I error by simulating r data sets from the null distribution, and calculated the proportion of rejected test,

$$\text{Type I Error} = \frac{1}{r}\sum_{i=1}^{r} 1[\chi_i^2 > Q_{1-\alpha}]\,.$$

Code availability. The R codes are freely available at https://github.com/rtahmasbi/GxE.

References

[1] E. Strachan, G. Duncan, E. Horn and E. Turkheimer, "Neighborhood deprivation and depression in adult twins: genetics and gene× environment interaction," *Psychological medicine,* vol. 47, pp. 627-638, 2017.

[2] L. E. a. K. M. C. Duncan, "A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry," *American Journal of Psychiatry,* vol. 168, no. 10, pp. 1041-1049, 2011.

[3] R. a. S. N. a. H. A. a. M. Y. a. A. K. a. B. T. a. B. M. a. C.-W. S. a. E. B. a. F. H. a. o. Culverhouse, "Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression," *Molecular psychiatry,* pp. 1-10, 2017.

[4] P. M. a. W. N. R. a. Z. Q. a. S. P. a. M. M. I. a. B. M. A. a. Y. J. Visscher, "10 Years of GWAS Discovery: Biology, Function, and Translation," *The American Journal of Human Genetics,* vol. 101, no. 1, pp. 5-22, 2017.

[5] S. H. a. Y. J. a. G. M. E. a. V. P. M. a. W. N. R. Lee, "Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood," *Bioinformatics,* vol. 28, no. 19, pp. 2540-2542, 2012.

[6] M. C. a. C. L. R. Neale, "Data summary," in *Methodology for Genetic Studies of Twins and Families*, Springer, 1992, pp. 35-53.

[7] J. Yang, B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery and others, "Common SNPs explain a large proportion of the heritability for human height," *Nature genetics,* vol. 42, pp. 565-569, 2010.

[8] S. Purcell, "Variance components models for gene-environment interaction in twin analysis," *Twin research,* vol. 5, pp. 554-571, 2002.

[9] D. Molenaar, C. M. Middeldorp, G. Willemsen, L. Ligthart, M. G. Nivard and D. I. Boomsma, "Evidence for Gender-Dependent Genotype by Environment Interaction in Adult Depression," *Behavior genetics,* vol. 46, pp. 59-71, 2016.

[10] K. Keskitalo-Vuokko, T. Korhonen and J. Kaprio, "Gene-Environment Interactions Between Depressive Symptoms and Smoking Quantity," *Twin Research and Human Genetics,* vol. 19, pp. 322-329, 2016.

[11] L. Mandelli and A. Serretti, "Gene environment interaction studies in depression and suicidal behavior: an update," *Neuroscience & Biobehavioral Reviews,* vol. 37, pp. 2375-2397, 2013.

[12] T. Klengel and E. B. Binder, "Gene-Environment Interactions in Major Depressive Disorder," *The Canadian Journal of Psychiatry,* vol. 58, pp. 76-83, 2013.

[13] S. Sharma, A. Powers, B. Bradley and K. J. Ressler, "Gene× environment determinants of stress-and anxiety-related disorders," *Annual review of psychology,* vol. 67, pp. 239-261, 2016.

[14] R. Uher, "Gene-environment interactions in severe mental illness," *Frontiers in psychiatry,* vol. 5, p. 48, 2014.

[15] P. B. Barr, J. E. Salvatore, H. Maes, F. Aliev, A. Latvala, R. Viken, R. J. Rose, J. Kaprio and D. M. Dick, "Education and alcohol use: A study of gene-environment interaction in young adulthood," *Social Science & Medicine,* vol. 162, pp. 158-167, 2016.

[16] J. M. Vink, "Genetics of addiction: future focus on gene× environment interaction?," *Journal of studies on alcohol and drugs,* vol. 77, pp. 684-687, 2016.

[17] M. D. Li, R. Cheng, J. Z. Ma and G. E. Swan, "A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins," *Addiction,* vol. 98, pp. 23-31, 2003.

[18] A. Caspi and T. E. Moffitt, "Gene-environment interactions in psychiatry: joining forces with neuroscience," *Nature reviews. Neuroscience,* vol. 7, p. 583, 2006.

[19] E. Turkheimer, A. Haley, M. Waldron, B. Donofrio and I. I. Gottesman, "Socioeconomic status modifies heritability of IQ in young children," *Psychological science,* vol. 14, pp. 623-628, 2003.

[20] E. M. Tucker-Drob and T. C. Bates, "Large cross-national differences in gene × socioeconomic status interaction on intelligence," *Psychological science,* vol. 27, pp. 138-149, 2016.

[21] R. M. Kirkpatrick, M. McGue and W. G. Iacono, "Replication of a gene-environment interaction via multimodel inference: additive-genetic variance in Adolescents ́ General Cognitive Ability Increases with Family-of-Origin Socioeconomic Status," *Behavior genetics,* vol. 45, pp. 200-214, 2015.

[22] T. C. Bates, G. J. Lewis and A. Weiss, "Childhood socioeconomic status amplifies genetic effects on adult intelligence," *Psychological Science,* pp. 1-6, 2013.

[23] E. M. Tucker-Drob and K. P. Harden, "Intellectual interest mediates Gene × Socioeconomic Status interaction on adolescent academic achievement," *Child development,* vol. 83, pp. 743-757, 2012.

[24] M. Rhemtulla and E. M. Tucker-Drob, "Gene-by-socioeconomic status interaction on school readiness," *Behavior genetics,* vol. 42, pp. 549-558, 2012.

[25] E. M. Tucker-Drob, M. Rhemtulla, K. P. Harden, E. Turkheimer and D. Fask, "Emergence of a gene × socioeconomic status interaction on infant mental ability between 10 months and 2 years," *Psychological Science,* vol. 22, pp. 125-133, 2011.

[26] K. P. Harden, E. Turkheimer and J. C. Loehlin, "Genotype by environment interaction in adolescents cognitive aptitude," *Behavior genetics,* vol. 37, pp. 273-283, 2007.

[27] L. J. a. L. K. A. a. Y. P. A. a. M. N. G. a. o. Eaves, "Model-fitting approaches to the analysis of human behaviour," vol. 41, no. 3, pp. 249-320, 1978.

[28] M. C. a. C. W. L. Keller, "Quantifying and addressing parameter indeterminacy in the classical twin design," *Twin Research and Human Genetics,* vol. 8, no. 3, pp. 201-213, 2005.

[29] O. a. H. E. a. S. S. R. a. L. E. S. Zuk, "The mystery of missing heritability: Genetic interactions create phantom heritability," *Proceedings of the National Academy of Sciences,* vol. 109, no. 4, pp. 1193-1198, 2012.

[30] M. R. a. E. G. a. M. G. a. L.-J. L. R. a. T. M. A. a. Z. Z. a. N. I. M. a. v. V.-O. J. V. a. S. H. a. E. T. a. o. Robinson, "Genotype-covariate interaction effects and the heritability of adult body mass index.," *Nature genetics,* vol. 49, no. 8, pp. 1174-1181, 2017.

[31] T. a. C. C.-Y. a. N. B. M. a. S. M. R. a. S. J. W. Ge, "Phenome-wide heritability analysis of the UK Biobank," *PLoS genetics,* vol. 13, no. 4, p. e1006711, 2017.

[32] J. Yang, S. H. Lee, M. E. Goddard and P. M. Visscher, "GCTA: a tool for genome-wide complex trait analysis," *The American Journal of Human Genetics,* vol. 88, pp. 76-82, 2011.

Acknowledgements

This work was supported by the National Institutes of Mental Health grant R01MH100141 to Dr. Keller. We thank Jian Yang and Peter Visscher for helpful comments on the project.

Author contributions

R.T. and M.K. designed the study. E.T. discussed the results and commented on the final manuscript. All authors wrote the manuscript.

Additional information

Supplementary information is available for this paper.

Competing interests

The authors declare no competing interests.

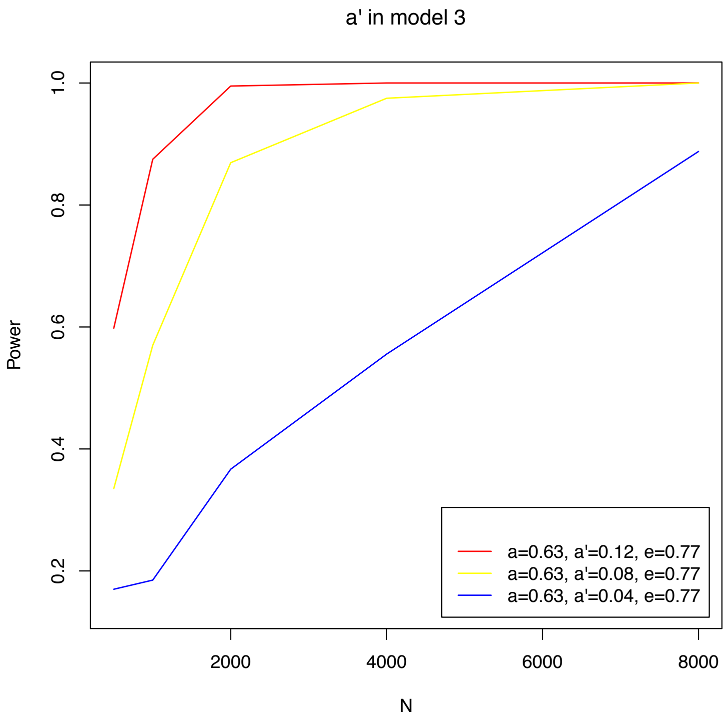


Figure 1. Power plot for testing $a' = 0$ in model 3 with different set of parameters.

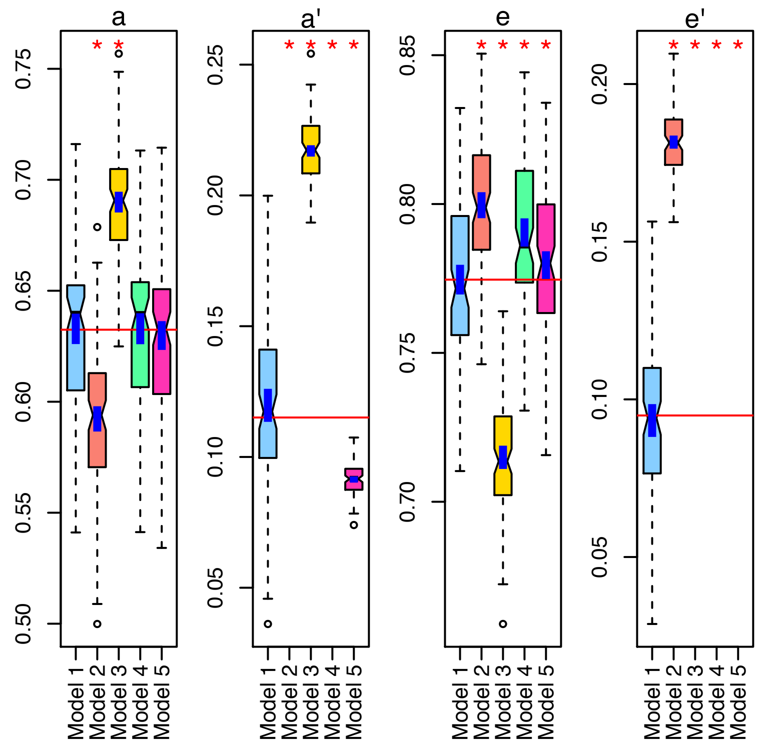
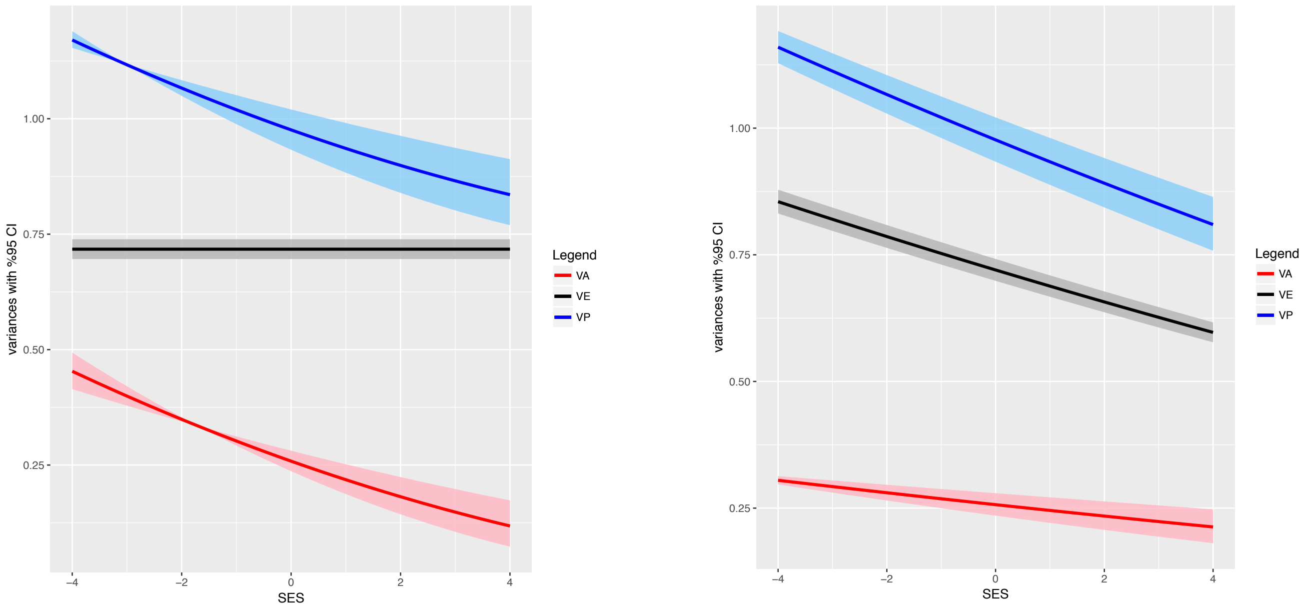


Figure 2. Data were simulated from model 1, and estimated with 5 different model. The vertical red lines are the true values. A red star appears above each boxplot if the estimated parameter is significantly different from its true value. The vertical blue liens are 95% confidence intervals. Model 1 is the only model that can estimate all the parameters, accurately.

Table 1: Estimated parameters for the proposed 5 models and Turkeimer model. Green lines are CI and number in brackets are p-value. The bold numbers are the minimum AIC/BIC. Computational time and memory used in gigabyte are also reported.

Parameter	Model					
	1	**2**	**3**	**4**	**5**	**Turkeimer**
$\hat{\lambda}$	-0.110	-0.110	-0.110	-0.110	-0.110	0.360
$\hat{a}$	0.507	0.506	0.508	0.507	0.507	0.572
	(0.48,0.53)	(0.48,0.53)	(0.49,0.53)	(0.49,0.53)	(0.49,0.53)	–
$\widehat{a'}$	-0.026 [0.032]	–	-0.041 [1.4e-10]	–	-0.011	0.141
	(-0.05,0.00)	–	(-0.054,-0.028)	–	(-0.02,-0.00)	–
$\hat{e}$	0.848	0.849	0.847	0.849	0.849	–
	(0.84,0.86)	(0.84,0.86)	(0.83,0.86)	(0.84,0.86)	(0.84,0.86)	–
$\widehat{e'}$	-0.010 [0.108]	-0.024 [3e-10]	–	–	–	–
	(-0.03,0.00)	(-0.03,-0.01)	–	–	–	–
Log-like	-19234.13	-19235.85	-19234.9	-19254.82	-19234.69	-2873.2
AIC	38476.27	38477.71	38475.8	38513.64	**38475.38**	6777.3
BIC	38510.67	38503.51	38501.60	38530.84	**38501.19**	6812.5
Time	21:30:54	9:39:27	8:50:18	12:27:43	13:13:52	–
Memory (GB)	167.4	155.4	155.4	155.4	167.4	–



(a)

(b)

Figure 3. a) $\mathrm{V_A} = (\mathrm{a} + \mathrm{a'}\,\mathrm{SES})^2$, $\mathrm{V_E} = e^2$ and $\mathrm{V_P} = \mathrm{V_A} + \mathrm{V_E}$ as a function of SES with a %95 CI for model 3. b) $\mathrm{V_A} = (\mathrm{a} + \mathrm{a'}\,\mathrm{SES})^2$, $\mathrm{V_E} = (\mathrm{e} + \mathrm{e}a'/a\,\mathrm{SES})^2$ and $\mathrm{V_P} = \mathrm{V_A} + \mathrm{V_E}$ as a function of SES with a %95 CI for model 5.

Table 2. Simulation results for model 1 with different set of parameters and sample sizes. Columns are bias and standard deviation of the estimated parameters.

True Parameters				N	**a**		**a′**		**e**		**e′**	
a	a'	e	e'		bias	sd	bias	sd	bias	sd	bias	sd
0.633	0.115	0.775	0.095	2000	-0.009	0.113	0.003	0.073	-0.009	0.088	0.003	0.061
				4000	-0.005	0.059	-0.002	0.051	-0.001	0.045	0.004	0.040
				8000	0.000	0.035	0.004	0.033	0.000	0.025	-0.002	0.027
0.633	0.077	0.775	0.095	2000	-0.015	0.110	-0.001	0.075	-0.011	0.090	-0.004	0.058
				4000	-0.003	0.058	-0.001	0.057	-0.005	0.046	0.003	0.046
				8000	-0.001	0.030	-0.001	0.034	-0.003	0.022	0.001	0.027
0.633	0.077	0.775	0.063	2000	-0.004	0.105	0.000	0.073	-0.014	0.084	0.002	0.059
				4000	-0.003	0.052	0.003	0.060	-0.002	0.040	0.001	0.047
				8000	0.000	0.033	0.008	0.034	-0.002	0.024	-0.006	0.027
0.633	0.077	0.775	0.032	2000	-0.001	0.115	-0.006	0.080	-0.019	0.090	0.004	0.066
				4000	0.001	0.054	-0.010	0.056	-0.007	0.042	0.007	0.047
				8000	0.003	0.031	-0.002	0.037	-0.004	0.022	0.002	0.029
0.633	0.038	0.775	0.063	2000	-0.012	0.125	-0.004	0.075	-0.011	0.089	0.004	0.058
				4000	-0.003	0.062	0.004	0.061	-0.006	0.048	-0.003	0.049
				8000	-0.001	0.035	0.003	0.038	-0.003	0.026	-0.002	0.030
0.633	0.038	0.775	0.032	2000	-0.022	0.117	-0.004	0.077	0.000	0.081	0.004	0.062
				4000	-0.006	0.058	-0.005	0.065	-0.002	0.047	0.005	0.053
				8000	0.000	0.034	0.004	0.035	-0.002	0.024	-0.003	0.027