Behavior Genetics, Vol. 35, No. 5, September 2005 (© 2005)
DOI: 10.1007/s10519-005-5355-9

Cholesky Problems

Gregory Carey[1,2]

Received 16 Sep. 2004—Final 19 Apr. 2005

Behavioral geneticists commonly parameterize a genetic or environmental covariance matrix as the product of a lower diagonal matrix postmultiplied by its transpose—a technique commonly referred to as "fitting a Cholesky." Here, simulations demonstrate that this procedure is sometimes valid, but at other times: (1) may not produce fit statistics that are distributed as a χ^2; or (2) if the distribution of the fit statistic is χ^2, then the degrees of freedom (df) are not always the difference between the number of parameters in the general model less the number of parameters in a constrained model. It is hypothesized that the problem is related to the fact that the Cholesky parameterization requires that the covariance matrix formed by the product be either positive definite or singular. Even though a population covariance matrix may be positive definite, the combination of sampling error and the derived—as opposed to directly observed—nature of genetic and environmental matrices allow matrices that are negative (semi) definite. When this occurs, fitting a Cholesky constrains the numerical area of search and compromises the maximum likelihood theory currently used in behavioral genetics. Until the reasons for this phenomenon are understood and satisfactory solutions are developed, those who fit Cholesky matrices face the burden of demonstrating the validity of their fit statistics and the df for model comparisons. An interim remedy is proposed—fit an unconstrained model and a Cholesky model, and if the two differ, then report the difference in fit statistics and parameter estimates. Cholesky problems are a matter of degree, not of kind. Thus, some Cholesky solutions will differ trivially from the unconstrained solutions, and the importance of the problems must be assessed by how often the two lead to different substantive interpretation of the results. If followed, the proposed interim remedy will develop a body of empirical data to assess the extent to which Cholesky problems are important substantive issues *versus* statistical curiosities.

KEY WORDS: Cholesky; developmental genetics; lower diagonal matrix; matrix factorization; model fitting; quantitative genetics; statistics; twins.

INTRODUCTION

A common numerical method for modeling genetic and environmental covariance matrices for genetically informative data is to iterate on the elements of a lower diagonal matrix and then obtain the desired covariance matrix by postmultiplying the lower diagonal matrix by its transpose. This procedure is often referred to as "fitting a Cholesky." This note exposes a potential problem with this approach—namely, in some cases, likelihood ratio test statistics may not follow a χ^2 distribution and the degrees of freedom may not equal the number of free parameters in a general model less the number of free parameters in a nested, constrained model.

An Illustration of the Problem

I illustrate the problem with a set of 10,000 simulated twin data sets. (specific details about these

[1] Department of Psychology and Institute for Behavioral Genetics, University of Colorado, Boulder, CO, 80309-0345, USA.
[2] To whom correspondence should be addressed at Department of Psychology, University of Colorado, Boulder, CO 80309-0345, USA. Tel: +1-303-492-1658; Fax: +1-303-492-2967; e-mail: carey@colorado.edu.

simulations and other methods used herein are given in Methods at the end of this paper). Each data set consisted of 100 pairs of identical and 100 pairs of fraternal twins using only one phenotype with a heritability (a^2) of 0.50, no common environment (c^2), and unique environmentability (e^2) of 0.50. For each set of twin data, I computed the intraclass covariance matrix[1] and fitted the same model to the data using two different numerical parameterizations. The first parameterization iterated directly on a^2, c^2, and e, obtaining e^2 as $e*e$. Hence, there are no boundary constraints except for the one implied by iteration on e. The second parameterization iterated on the elements of the Cholesky, which in this case are simply a, c, and e, and obtained the variance components as the squares of these estimates.

Table I shows the distribution of the χ^2 goodness-of-fit statistics for the first and second parameterizations according to the percentage of simulations that exceeded the 0.20, 0.10, 0.05, and 0.01 levels of significance. (Note that the χ^2 has one degree of freedom because models were fitted to the intraclass and not the interclass covariance matrixes.) For the unconstrained parameterization, the distribution fits a χ^2 with 1 df (goodness of fit $\chi^2 = 0.32$, df = 1, $p = 0.57$). For the Cholesky parameterization, however, what should be the general model is rejected almost twice as often as the unconstrained model. The ostensible $\chi^2 =$ for the Cholesky does not fit a χ^2 distribution even when the df is a free parameter (best df = 1.46, $\chi^2 = 163.69$, df = 55, $p < 10^{-13}$)[2].

The reason is easily seen if we consider for the moment only estimates of c^2 (unconstrained) and c (constrained Cholesky parameterization). Because the population value for c^2 is 0, the unconstrained estimate of c^2 had a mean of 0 with half of the estimates being lower than 0. Iteration on c, however, cannot give a negative estimate of c^2, so the value of c becomes so small that it is effectively fixed at a number very close to 0. Depending upon minimization software (and how that software is used, of course), this fact may or may be brought to the attention of the user. When these boundary conditions are encountered for the single parameter c, then likelihood theory maintains that c be treated as a fixed parameter and one degree of freedom is gained.

Table I. Percentage of Simulated Twin Data Sets that Exceed the Critical Value for a p Level: One Phenotype

p value	Unconstrained	Cholesky, ignoring boundary conditions	Cholesky, fixing parameters to bounds. Number of parameters fixed (df):	
	df = 1	df = 1	0 (df = 1)	1 (df = 2)
	$N = 10{,}000$	$N = 10{,}000$	$N = 4799$	$N = 5201$
0.20	20.4	33.1	20.2	21.6
0.10	10.5	19.4	10.7	10.6
0.05	5.1	10.5	5.1	5.4
0.01	1.1	2.5	1.0	1.0

N equals the number of simulated data sets for a particular

The two rightmost columns of Table I illustrate this principle. Here, the 10,000 Cholesky results are divided into two groups, those in which one and only one parameter converged to a bound (arbitrarily defined as an absolute value of the parameter less than 0.001) and those with no boundary condition. When there is no boundary condition, then the test statistic fits the distribution of a χ^2 with 1 df ($\chi^2 = 0.26$, df = 1, $p = 0.61$). When boundary constraints are met, the test statistic is also χ^2, but with 2 df ($\chi^2 = 0.91$, df = 1, $p = 0.34$). The fixed parameter was c in 97.4% of the cases and a in the remaining 2.6%.

The Problem Gets More Complicated

This illustration suggests a simple *ad hoc* solution—keep track of the number of parameters with final values close to 0 and adjust the degrees of freedom accordingly. Unfortunately, this strategy appears appropriate only for the analysis of a single phenotype.

[1] I use the intraclass covariance matrix for three reasons. First, the parameter estimates from fitting an intraclass model (same means and covariance matrices for twin 1 and twin 2 and a symmetric cross-twin covariance matrix) to the intraclass matrix are identical to fitting the same model to raw data. They are not identical when the interclass matrix is used. Second, it is much easier to inspect visually the differences between an observed and predicted covariance matrix using the intraclass matrix—one does not have to mentally average two estimates of the same statistic as one must do for the interclass matrix. Third, in interpreting results and diagnosing models, one does not have to account for the sampling error due to the assignment of one twin as twin 1 and the other as twin 2.

[2] See the Methods section for a full explanation of testing whether the empirical distribution of ostensible χ^2 statistics fits a χ^2 distribution. This usually involves 50 or more df for the test. Given that the distribution is reasonably approximated by a χ^2, then testing whether the best fitting distribution fit the nominal df involves a likelihood ratio test with one df (i.e., comparing the free parameter of the best fitting df to fixed parameters of the nominal df))

I performed 30,000 simulations on two phenotypes using 100 pairs of MZ and 100 pairs of DZ twins. Again, two different parameterizations of the same model were fitted to the data. Both iterated on the Cholesky for the unique environmental matrix. The first parameterization—i.e., the unconstrained solution—treated each element of the additive genetic and common environmental matrices as free, unbounded parameters. The second parameterization iterated on the elements of the Cholesky factors for the genetic and for the common environmental covariance matrices.

Figure 1 presents the distribution of the observed χ^2 goodness of fit statistics for the unconstrained solution (panel a) and the Cholesky solution (panel b). Fitting models to intraclass matrices here gives 3 degrees of freedom for the goodness of fit statistics, so the "nominal df" for both distributions should be 3. Also shown in the Figure is the best fitting degrees of freedom to the observed statistics for each solution.

The unconstrained solution clearly fits the expected distribution. The χ^2 goodness of fit for the observed χ^2 statistics assuming 3 df is 74.70, df = 79, $p = 0.62$. The best fitting degrees of freedom is 3.004, within rounding error of 3. As a result, the curves for the nominal and best fitting degrees of freedom overlap so much that they appear as one curve in the Figure.

The Cholesky solution clearly does not agree with the nominal 3 degrees of freedom ($\chi^2 = 300.86$, df = 98, $p<10^{-8}$). Furthermore, the distribution of the observed statistic is not a χ^2 even with its best fitting degrees of freedom of 4.62 ($\chi^2 = 12,\ 558.17$, df = 97, $p<10^{-16}$).

These results agree with those for a single phenotype in showing that the ostensible χ^2 from a Cholesky solution is not, in fact, a χ^2. Will setting a parameter to a bound and adding a degree of freedom rectify the situation? The answer is "No."

Table II gives the results of selecting those cases in which one and only one parameter hit a boundary constraint and then subdividing them by whether that parameter was the lower right-hand element of the genetic Cholesky or the lower right-hand element of the common environmental Cholesky. (In the present simulations, these situations account for over 99.7% of the cases in which only one parameter hit a bound.) The table gives the percent of simulations that exceeds a given p level for the nominal df (3) and for adding a degree of freedom so that the distribution has 4 df.

Note first that the perturbation away from a χ^2 distribution depends on *which* parameter encounters the boundary condition. Let Λ_A and Λ_C respectively denote the Choleskys that generate the additive genetic matrix (**A**) and the common environment matrix (**C**). If we compare the two columns with 3 df, 28.2% of the statistics exceed a p value of 0.20 when $\Lambda_A(2,2)$ hits the boundary constraint; but 40.4% of the statistics exceed that p value when $\Lambda_C(2,2)$ is bounded.

Furthermore, adding a degree of freedom no longer rectifies the situation. With $\Lambda_A(2,2)$ set to its bound, the resulting distribution of the fit statistic departs significantly from a χ^2 with 4 df ($\chi^2 = 43.53$, df = 1, $p<10^{-10}$). Neither does adding a degree of freedom work when $\Lambda_C(2,2)$ encounters a bound ($\chi^2 = 612.26$, df = 1, $p < 10^{-16}$).

Finally, the distribution of fit statistics appears to follow a χ^2—it is just that the degrees of freedom are no longer integers. The best fitting degrees of

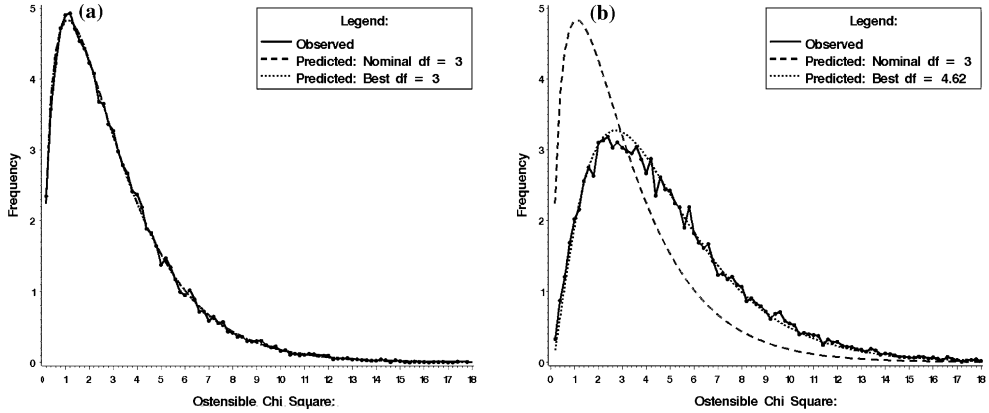

Fig. 1. Distribution of ostensible χ^2 statistics for an unconstrained solution (Panel a) and a Cholesky solution (Panel b) with theoretical distributions for the nominal degrees of freedom and for the best fitting degrees of freedom: two phenotypes.

Table II. Percentage of simulations exceeding a critical level when the lower right-hand element of the genetic Cholesky or the common environmental Cholesky hit a bound: two phenotypes using only simulations where one and only one parameter hit a bound

	Parameter with the boundary constraint			
	Genetic parameter, $N = 2271$		Common environment parameter, $N = 14{,}631$	
p level	df = 3	df = 4	df = 3	df = 4
0.20	28.2	16.5	40.4	25.7
0.10	14.8	8.5	23.6	13.9
0.05	8.2	4.1	13.8	7.5
0.01	1.9	0.7	3.4	1.5

Both nominal df (3) the effect of adding one df are shown.

freedom when $\Lambda_A(2,2)$ is fixed at its bound is 3.66. This gives a good fit to the data ($\chi^2 = 62.15$, df = 59, $p = 0.36$). With $\Lambda_C(2,2)$ at a bound, the best fitting df is 4.54, although the fit is not as good ($\chi^2 = 120.98$, df = 90, $p = 0.02$).

Simulations using three, four and five phenotypes gave similar results with the one important difference noted below. In all cases, the distribution of the ostensible fit statistic for the Cholesky solution departed significantly from a χ^2, even using the best fitting degrees of freedom. Also, adding a degree of freedom for each parameter that encountered a bound in the Cholesky matrices did not result in a χ^2 distribution. Finally, the best fitting degrees of freedom when parameters were bounded were fractional.

What was that "one important difference?" It was this—as more phenotypes were simulated, Cholesky problems got worse.

A Conjecture about the Problem

Before presenting and discussing further results, it is important to step back and reflect on the nature of the problem. I cannot provide a formal analytical proof of why the Cholesky parameterizations behave as they do in these simulations. Such a proof requires analytic equations for the derivatives of the likelihood with respect to the free parameters and would be very time consuming to develop. Instead I offer a conjecture.

Texts in matrix algebra demonstrate that any symmetric matrix, say **S**, can be factorized as

$$\mathbf{S} = \mathbf{QDQ}^t, \tag{1}$$

where **Q** is a lower diagonal (or if one prefers, lower triangular) matrix with 1s on the diagonal and **D** is a diagonal matrix. (This is sometimes referred to Gauss or Gauss-Jordan factorization). Matrix **S** may be negative semi definite (i.e., at least one negative eigenvalue), singular (i.e., at least one eigenvalue equal to 0), or positive definite (i.e., all eigenvalues greater than 0), and this status of **S** depends on the elements of **D**. If **D** has one or more 0s on its diagonal, then **S** is singular. Otherwise, if **D** has one or more negative elements on the diagonal, then **S** is negative semi definite. Otherwise (and now all diagonal elements of **D** will be greater than 0), **S** is positive definite. If all of the elements of **D** are positive, then let $\mathbf{W} = \mathbf{QD}^{1/2}$ so that

$$\mathbf{S} = \mathbf{WW}^t. \tag{2}$$

Matrix **W** is defined as the Cholesky matrix and computation of **W** from an observed **S** is called Cholesky factorization.

Note the discrepancy between the way a Cholesky matrix is defined in matrix algebra and the way in which it is defined in behavioral genetics. In matrix algebra, a Cholesky must be nonsingular and the product of it and its transpose must be a positive definite matrix. When we in behavioral genetics "fit a Cholesky," we really fit a lower diagonal matrix that it sometimes deliberately set to be singular by fixing a diagonal element to 0. If we wish to be consistent with mathematical definitions, we should really speak of "lower diagonal matrices" instead of Choleskys. (Interestingly, the initial papers in behavioral genetics that fitted "Choleskys"—e.g., Cantor (1983); Fulker (1978); Martin *et al.*, (1984)—did not call them Choleskys.) To avoid confusion, I rewrite Eq. (2) as

$$\mathbf{S} = \mathbf{XX}^t,$$

where **X** is a lower diagonal matrix that may or may not contain a 0 as a diagonal element. I will now speak of a lower diagonal or LD matrix instead of a Cholesky.

The central substantive point is that when we iterate on the elements of the lower diagonal matrix **X**, then matrix **S** *cannot be negative semi definite*. It must be either singular (i.e., at least one diagonal element of **X** equals 0) or it must be positive definite (no diagonal element of **X** equals 0). I submit that this is the source of the fundamental problem outlined above. A variance and a variance-covariance matrix computed from raw data must be positive definite (provided, of course, that there is variability to begin with, more observations than variables, no missing values, and no variable is a perfect linear combination of other variables). But the genetic and envi-

ronmental covariance matrices in behavioral genetics are not computed directly from raw data—they are derived as mathematical functions of, say, a block in an observed MZ covariance matrix with one in a DZ matrix. Just as an estimate of a variance component in the analysis of variance can be negative, so too can sampling error give a negative estimate of, say, common environmental variance. The LD parameterization will not allow for the possibility of a negative variance or, in the multivariate case, a negative semi definite matrix.

In an unconstrained solution, the likelihood is a function of two factors—sampling error and model error. Here, the mathematical space for a numeric search is over the whole set of real numbers. The mathematical space in an LD search, however, appears to be constrained to those combinations of real numbers that generate positive definite or singular matrices. In constrained maximum likelihood estimation, the degrees of freedom are a function of both the number of free parameters and also the constraints that are active at the solution. To interpret the results from an LD solution, one must know the extent to which constraints are active at the solution; otherwise, the degrees of freedom are unknown.

To say the same thing in different terms, the LD solution may be a conditional likelihood. That is, it provides a likelihood given that sampling error produces unconstrained parameter matrices that are positive definite or singular. When such conditions are met, then the likelihood is valid. When these conditions are not met, however, the likelihood will be perturbed.

If this line of thinking is correct, then there should be a strong relationship between the eigenvalues of unconstrained matrices and the fit of the LD solution. When all unconstrained matrices are positive definite or singular, then the LD parameterization should give equivalent results to the unconstrained model. As unconstrained matrices become negative semi definite, however, the LD solution should differ from the unconstrained solution.

I examined this by computing the eigenvalues of the additive genetic matrix (**A**) and the common environmental matrix (**C**) from the unconstrained solution and then examining the properties of the ostensible χ^2 for the LD solution as a function of these eigenvalues. The first property examined was simply the difference between the ostensible χ^2 from the LD and the χ^2 from the unconstrained solution. I then fitted a model with one parameter—the degrees of freedom—to the ostensible χ^2 from the LD solution. Finally, I regressed the difference between the ostensible χ^2 from the LD and the χ^2 from the unconstrained model on the eigenvalues of the matrices in the unconstrained solution. Tables III and IV present the results from simulations on, respectively, two and three phenotypes.

Table III. Mean difference and range of differences in χ^2 values between the LD and the unconstrained solutions as a function of the number of positive eigenvalues in the additive genetic (**A**) and common environmental (**C**) matrices in the unconstrained solution: two phenotypes

Number of eigenvalues >0 in matrix							
A	**C**	*N*	Mean	Range	Best df	*p*	R^2
2	2	1,817	0.00	0.00	3.00	0.69	0.01
2	1	15,451	1.60	21.79	4.56	0.02	0.58
2	0	4,465	3.01	20.50	6.03	0.64	0.74
1	2	2,426	0.79	9.93	3.69	0.34	0.54
1	1	5,817	1.93	17.72	5.00	0.68	0.61
0	2	20	1.24	4.31	3.54	0.91	0.77

N = number of simulations in that category. Best df and *p* = best fitting degrees of freedom to the χ^2 distribution from the LD model and the *p* level for this fit. R^2 = squared multiple correlation regressing the difference in χ^2 on the four eigenvalues from the unconstrained **A** and **C** matrices.

Table IV. Mean difference and range of differences in χ^2 values between the LD and the unconstrained solutions as a function of the number of positive eigenvalues in the additive genetic (**A**) and common environmental (**C**) matrices in the unconstrained solution: three phenotypes

Number of eigenvalues >0 in matrix							
A	**C**	*N*	Mean	Range	Best df	*p*	R^2
3	3	135	0.00	0.00	6.38	0.64	0.09
3	2	5324	2.33	16.37	8.38	0.68	0.80
3	1	8410	4.11	30.32	10.13	0.04	0.80
3	0	602	6.03	22.30	12.28	0.50	0.83
2	3	410	0.91	11.38	7.17	0.37	0.85
2	2	6405	2.67	20.16	8.71	0.04	0.80
2	1	3477	4.21	23.63	10.19	0.91	0.81
1	3	51	1.90	6.53	8.00	0.30	0.93
1	2	186	3.03	11.29	8.93	0.35	0.81

N = number of simulations in that category. Best df and *p* = best fitting degrees of freedom to the χ^2 distribution from the LD model and the *p* level for this fit. R^2 = squared multiple correlation regressing the difference in χ^2 on the four eigenvalues from the unconstrained **A** and **C** matrices.

Note that when all unconstrained parameter matrices are positive definite, there are no Cholesky problems—the LD and the unconstrained models give identical answers. The ostensible χ^2 values in the LD solution do indeed follow a χ^2 distribution with the correct degrees of freedom—3 in the two phenotype simulations and 6 in the case of three phenotypes. (The best fitting df with three phenotypes, 6.38, is not significantly different from the distribution using 6 df—$\chi^2 = 1.80$, df = 1, $p = 0.18$). Also, the eigenvalues of the unconstrained matrices do not predict the difference in χ^2 values.

In all other cases, however, there is a significant difference between the average ostensible χ^2 for the LD and the χ^2 values for the unconstrained solution, with the former always being greater than the latter. The difference is also a function of the number of negative eigenvalues. If we fix the number of positive eigenvalues for, say, matrix **A**, then as the number of negative eigenvalues in **C** increases, the LD solution becomes more discrepant from the unconstrained solution. Similarly, if we fix the number of positive eigenvalues for matrix **C**, then the average ostensible χ^2 for the LD solution increases as the number of negative eigenvalues in **A** increases.

The range of the differences is, of course, sensitive to the number of simulations that fall into a category—the more simulations, the greater the chance of observing a large difference. The numbers, however, suggest that the difference between the LD and the unconstrained models is not always trivial and cannot always be safely ignored.

Note that the fit statistic for the LD in Tables III and IV often fits a χ^2 distribution, *albeit* with fractional degrees of freedom. This suggests that distribution of the unconditional fit statistic for the LD model (i.e., the fit statistic that is not conditioned on the number of positive eigenvalues in an unconstrained solution) is closely approximated by a *mixture* of χ^2 distributions. The situation is most easily seen by recalling the results on one phenotype. If we exclude the improbable situation in which both a and c hit a bound, then the fit statistic will be a mixture of a χ^2 distribution with 1 df (neither a nor c is bounded) and a χ^2 distribution with 2 df (either a or c is bounded). Such mixtures of χ^2 distributions have been reported elsewhere in genetic epidemiology (Sham *et al*., 1996; Terwilliger, 1995) as well as in other likelihood problems when the solution occurs on the boundary of the numerical space (Self and Liang, 1987).

[3]Note also that when more than one phenotype is analyzed, the best fitting degrees of freedom are not consistent with setting a parameter to a bound and adding a degree of freedom. If this were the case, then the best fitting solution when there are two positive eigenvalues in **A** but one in **C** should be the same as when there is one positive eigenvalue in **A** but two in **C**. For both two and three phenotypes, this is definitely not the case. Similarly, the addition of a degree of freedom is not constant over the number of phenotypes. With two phenotypes, when **A** is positive definite and **C** has only one negative eigenvalue, then 1.56 df need to be added in order to make the distribution equivalent to a χ^2. With three phenotypes, however, a positive definite **A** and a **C** with only one negative eigenvalue requires an addition of 2.38 df to approximate a χ^2 distribution.

Of particular interest is the extent to which the eigenvalues of the unconstrained solution predict the difference in fit functions between the LD and the unconstrained solutions. Excluding the case in which the LD gives a valid likelihood, then over 50% of the variance in these differences is predictable with two phenotypes while 80% or more is predictable with three phenotypes. In all cases, the regression coefficients were in the direction that the more negative eigenvalues for a matrix, the greater the difference between the fit statistics. This suggests that the *degree*

[3] The concept of "boundary conditions" requires comment because the term is used equivocally. Define a *simple boundary constraint* as the deliberate limitation of the subset of real numbers into which a likelihood solution must fall by specifying a lower and/or upper real bound for a parameter. Simple boundary constraints may be invoked by using boundary options in optimizing software, by explicit programming in the function to be minimized (e.g., iterating on c and computing c^2 as $c * c$), or by using explicit constraints (e.g., specifying that the sum of X_1 and X_2 must be less than 1.0). When a constraint is active at the solution (e.g., a parameter is within ε of its stated bound, or c is within ε of 0, or $X_1 + X_2$ is within ε of 1.0), then one parameter is considered fixed and a degree of freedom is gained (C.R. Rao, personal communication to John Rice). Those cases in which a likelihood ratio statistic is a mixture of χ^2 distributions appear to be simple boundary constraints. Define *complex boundary constraints* as a set of mathematical rules that may limit the subset of real numbers into which a solution may fall but cannot be explicitly stated as a series of simple boundary constraints. An example would be iterating on the elements of a covariance matrix under the constraint that the covariance matrix be positive definite. One could place simple bounds on the diagonals of this matrix but that in itself will not guarantee that the matrix be positive definite. I have not been able to uncover treatment of the properties of likelihood estimates and likelihood ratio test statistics for complex boundary conditions. I conjecture that Cholesky problems are complex boundary constraints.

of "negative definiteness" of a matrix is just as important as the *fact* of negative definiteness in assessing the fit of a LD model.

This observation suggests a reason why adjusted degrees of freedom for elements in matrix **A** differed from those in matrix **C**—when matrix **A** had (a) negative eigenvalues(s), they were not as large as the negative eigenvalues in **C**. With two phenotypes, when **A** had one negative eigenvalue, the mean of that negative eigenvalue was −0.04 (range = −0.93 to -8×10^{-6}). For **C**, however, the average was −0.18 (range = -1.67 to -3×10^{-7}). (This should not be surprising given that matrix **A** was based on moderate heritabilities while matrix **C** has a population value of 0.)

The conjecture offered here, along with the results of the multiple regressions, also suggest why parameters in the diagonals of an LD matrix often converge towards their implied mathematical bounds of 0. If an unconstrained **A** or **C** matrix is negative (semi) definite, then the LD model will move towards a solution that most closely approximates a negative (semi) definite matrix—i.e., a singular matrix, one that has one or more zeros on the diagonal of the LD matrix.

To explore this possibility, I calculated the eigenvalues for the predicted additive genetic covariance matrix and the predicted common environmental matrix from the LD solution. I arbitrarily defined a predicted matrix as singular if had an eigenvalue that was less than 10^{-10}. Table V gives the percent of singular predicted **A** matrices in the LD solution as a function of the number of positive eigenvalues of that matrix in the unconstrained solution. Table VI gives the analogous statistics for the **C** matrix.

Table V. Percent of matrices formed by $\Lambda_A\Lambda_A^t$ that are singular as a function of the number of positive eigenvalues in the unconstrained **A** matrix

m	*N*	Percent singular
Two phenotypes		
0	20	100.0
1	2408	99.7
2	1817	0.0
Three phenotypes		
0	0	
1	51	100.0
2	410	97.3
3	135	0.0

m = number of positive eigenvalues in **A**, *N* = number of simulations. Only simulations with a positive definite unconstrained **C** matrix are given.

Table VI. Percent of matrices formed by $\Lambda_C\Lambda_C^t$ that are singular as a function of the number of positive eigenvalues in the unconstrained **C** matrix

m	*N*	Percent singular
Two phenotypes		
0	4465	99.9
1	15451	99.0
2	1817	0.0
Three phenotypes		
0	602	100.0
1	8410	99.6
2	5324	97.1
3	135	0.0

m = number of positive eigenvalues in **C**, *N* = number of simulations. Only simulations with a positive definite unconstrained **A** matrix are given.

The results suggest that there is considerable merit to this conjecture. Whenever an unconstrained matrix is negative (semi) definite, the predicted matrix formed by the product of the LD matrix and its transpose is almost always singular. On the other hand, if the unconstrained matrix is positive definite, then the predicted matrix formed by the product of the LD matrix and its transpose is always positive definite.

MORE RESULTS

The most sobering aspect of these results is that Cholesky problems are not rare. These simulations (and many others not reported) were based on those parameter values most often reported in behavioral genetic research—heritabilities in the moderate range, genetic correlations that are the major source of phenotypic correlations, and little common environment. Yet only 135 of the 25,000 simulations reported for three phenotypes—that is, only 0.54% of all these simulations—never encountered a Cholesky problem (see Table III). And matters got worse with more phenotypes. With five phenotypes, over 99.9% of the simulations had a Cholesky problem. Cholesky problems are probably the norm, not the exception under this parameter space.

Factors such as sample size and the population covariance matrix do indeed moderate results, but not as much as one might think. Over the parameter space mentioned above, the main effect of increasing sample size was to stabilize matrix **A** and let it become negative (semi) definite less often. It had no

effect on matrix **C** when its population value was 0. To illustrate, using 100 pairs for each zygosity and two phenotypes, only 6 percent of simulations were free of Cholesky problems. Using the same population parameters but increasing sample size to 1000 pairs of each zygosity bettered that number, but only to 15%.

Cholesky problems are most likely to be encountered with the common environment matrix. Why? Empirically, common environment usually accounts for less variance than either additive genetic or unique environmental effects. Hence, sampling error is more likely to result in an unconstrained variance estimate that is less than 0 in a common environmental matrix than in a matrix with population values much larger than 0. Furthermore, the more phenotypes in the analysis, the more likely that a negative variance will be sampled in **C**. This, I suspect, is the reason why matters became worse as more phenotypes were simulated.

One can, of course, simulate data that are relatively immune from Cholesky problems. Just use a large number of twin pairs and population covariance matrices with large diagonals but very small off diagonals. The optimal situation that avoids Cholesky problems appears be one in which genetics, common environment, and unique environment contribute equally to phenotypic variance but there are no genetic and environmental correlations among traits. This scenario, however, is not particularly informative—what is the sense of doing a multivariate analysis of uncorrelated phenotypes?

Do the Cholesky problems outlined above create problems for behavioral genetic analysis? For some types of multivariate problems with some phenotypes, the answer is definitely "yes," at least for the "business as usual" approach. Here, one fits an LD matrix to, say, the common environment covariance matrix, then fits a second model in which all elements of that matrix are set to 0, and assumes that the degrees of freedom for the likelihood ratio test are $n(n+1)/2$ where n is the number of phenotypes. As the simulations have demonstrated, however, the distribution—if it is even χ^2 to begin with—is unlikely to have the hypothesized degrees of freedom. If there were five phenotypes, then the "business as usual" approach tabulates 15 free elements in the LD matrix, giving a test with ostensibly 15 df. If, in fact there were only 6.7 df because of Cholesky problems, then we lose power to detect common environmental effects. Champions of the family environment may justifiably look askance at field that claims that there is little common environment when the statistical techniques used to justify that assertion are not fully powered. Note that other model fitting criteria that are a function of the degrees of freedom—e.g., Akaike's Information Criterion—will also be influenced by Cholesky problems.

Cholesky problems also affect properties of confidence intervals. This should not be surprising—Cholesky problems perturb the fit statistic and the confidence intervals depend on this fit statistic. Here, I present an example. Table VII presents results of fitting an unconstrained model and an LD model on a simulated data set for a single phenotype. The MZ variance and covariance was, respectively, 1.0378 and 0.497 and DZ variance and covariance was 1.0074 and 0.2058 ($N = 1000$ pairs of each zygosity). Note that the estimate of c^2 is negative in the unconstrained solution while the estimate of c goes to its implied mathematical bound of 0 in the LD model. A major property of confidence limits—that they are invariant

Table VII. Effect of Cholesky constraints on 95% confidence intervals

Unconstrained				LD solution			
		Confidence limits				Confidence limits	
Parameter	Estimate	Lower	Upper	Parameter	Estimate	Lower	Upper
a^2	0.546	0.395	0.702	a	0.691	0.615	0.731
c^2	−0.062	−0.192	0.065	c	0.000	−0.280	0.280
e	0.733	0.703	0.765	e	0.738	0.710	0.768
χ^2	0.390			χ^2	1.294		
$c^2 = 0$				$c = 0$			
a^2	0.478	0.422	0.535	a	0.691	0.650	0.731
e	0.738	0.710	0.768	e	0.738	0.710	0.768
χ^2	1.294			χ^2	1.294		

Table VIII. Common environment covariance matrix (correlations above the diagonal) for the unconstrained and lower diagonal solutions for the five subtests of the National Merit Scholarship Qualifying Test

	English	Math	Social science	Natural science	Vocabulary
Unconstrained solution					
English	8.94	0.85	0.83	0.72	0.86
Math	6.32	6.20	1.00	0.91	0.90
Social science	6.40	6.41	6.59	1.04	1.02
Natural science	7.48	7.88	9.29	12.19	0.78
Vocabulary	8.42	7.31	8.54	8.83	10.60
Lower diagonal ssolution					
English	8.96	0.85	0.81	0.72	0.86
Math	6.34	6.22	0.95	0.91	0.90
Social science	6.64	6.50	7.53	0.93	0.95
Natural science	7.50	7.87	8.91	12.09	0.78
Vocabulary	8.40	7.31	8.50	8.87	10.61

over transformations of the parameters (Neale and Miller, 1997)—is violated when there is a Cholesky problem. If confidence limits were invariant, then the square of the lower confidence interval for a in the LD solution should equal the lower confidence limit for a^2 in the unconstrained solution. But it does not: $0.615^2 = 0.378$, not 0.395. The same is true of the upper confidence limit for a in the Cholesky: $0.731^2 = 0.534$, a considerable difference away from 0.702. It is intriguing that a Cholesky problem with parameter c should influence the confidence interval of a different parameter, a.

Note that the invariance property is regained once the boundary constraint on c is considered (bottom section of the Table). Here, the square of the lower interval for a in the Cholesky is now within rounding error of the lower interval of a^2 in the unconstrained solution ($0.65^2 = 0.423$ versus 0.422). The same is true of the upper limits ($0.731^2 = 0.534$ *versus* 0.535).

In summary, all of these results suggest that those who fit LD matrices to data face the burden of demonstrating the validity of the fit statistics and of the degrees of freedom for testing models. So how should this be done?

Potential Remedies

One solution that cannot be recommended is to fix diagonal elements of an LD matrix to 0 and adjust the degrees of freedom. The constraints imposed by the Cholesky model are subtle and the degrees of freedom are not always integers. Perhaps future research will develop an acceptable variant of this strategy, but for the present time it should be avoided.

A second solution is to calculate the eigenvalues from the matrix computed as the product of the LD matrix and its transpose. That is, if you iterate on the elements of the LD matrix Λ_A, then compute $\mathbf{A} = \Lambda_A \Lambda_A^t$ and take the eigenvalues of $\mathbf{A}$. (One could also use Eq. (1) to derive the Gauss factors and examine the diagonal elements of $\mathbf{D}$). If all eigenvalues are positive, then there is no problem. If one or more eigenvalues is 0, then matrix is singular and there may be a significant Cholesky problem.

There are two difficulties with this solution. First, with numerical estimation, it is not always easy to distinguish a singular matrix from one that is positive definite but happens to have an eigenvalue close to 0. The second difficulty is more serious—this solution does not inform *how much* of a problem is at hand. Recall that the difference between an LD and an unconstrained solution is not akin to falling off a cliff. As the unconstrained matrix departs more and more from positive definiteness, the LD solution becomes more and more invalid (examine the R^2 values from Tables III and IV). A singular predicted matrix, however, will occur whenever there is a trivial difference between the two solutions as well as whenever there is a large, substantive difference. Hence, the *fact* of singularity cannot be used to judge the extent to which likelihood principles are violated. Again, future research could develop fruitful avenues alone this line, but for the present singularity should not be used as the sole criterion for the diagnosis of the importance of Cholesky problems.

Another potential solution is not to fit an LD matrix, but to use Eq. (1) and fit the Gauss factors—i.e., matrices $\mathbf{Q}$ and $\mathbf{D}$—in the model. In the general case, this parameterization will give the same answer as the unconstrained solution. There are, however, two problems with this approach, one practical, the other substantive. The practical issue is that numerical searches on this parameterization are not well conditioned, so it often takes a number of different tries to achieve convergence. The substantive problem occurs when the solution has a negative element in $\mathbf{D}$. This denotes a negative variance, and how does one interpret a negative variance? The problem goes away if one can set that element in $\mathbf{D}$—and, of course, all elements in the corresponding

column of **Q**—to 0 without a worsening in fit. But this may not always be the case[4].

A more productive approach is to eschew fitting LD matrices altogether (at least for the moment). Instead, use the unconstrained solution. Then try to extract the Cholesky factors[5] (or Gauss factors or eigenvalues) from the final parameter matrices. If Cholesky factorization succeeds, then one has the same solution that one would have observed by fitting LD matrices in the first place. If the factorization fails, then at least one has computed the baseline likelihood (or χ^2) needed to compare the fit statistic from an LD solution and has also identified the troublesome matrix (or matrices). It is also possible to reduce a general model using only the unconstrained solution. For example one could test whether the common environmental covariance matrix equals **0** using unconstrained parameters. Factorization can then be performed after model reduction.

Whenever Cholesky factorization fails but one wants to fit an LD model, then it is essential to make two comparisons. The first is the comparison of the fit statistic of the unconstrained solution with the fit statistic for the LD solution. In some cases, this difference will be trivial enough to ignore. Future research, however, is needed to provide satisfactory criteria to distinguish a "trivial" from a "substantive" difference.

The second comparison is between the unconstrained parameter matrix and the predicted value of that matrix computed as the product of an LD matrix and its transpose. One could arrive at a relatively large difference between the fit statistics of the unconstrained and the LD solutions simply because of large sample size. (Or, for that matter, a small difference between fit statistics because of modest sample size.). Only a comparison between the predicted parameter matrices can inform one of whether a comparison of fit statistics is accompanied by a substantive difference about how genes and environment influence traits.

I illustrate this remedy with an analysis of a public data set—the National Merit Twins (Loehlin and Nichols, 1976). For the subtests of the National Merit Scholarship Qualifying Test, the χ^2 goodness of fit for the unconstrained and lower diagonal models were, respectively, 19.33 and 21.26 (df = 15). Here, the difference is only 1.93 χ^2 units, and it came about because three correlations in the common environment matrix exceeded 1.0. The additive genetic covariance matrix could not be set to 0 in either the unconstrained ($\chi^2 = 167.76$, df = 15, $p < 0.0001$) or the lower diagonal solution ($\chi^2 = 165.83$, df = 15, $p < 0.0001$). Neither could the common environment matrix be set to 0 in the unconstrained ($\chi^2 = 61.28$, df = 15, $p < 0.0001$) or the lower diagonal solution ($\chi^2 = 59.35$, df = 15, $p < 0.0001$). Finally, there is no appreciable difference between the common environment covariance matrices for the two solutions (see Table VIII). Although there is a Cholesky problem here, it can safely be ignored because there is no substantive difference between the unconstrained and the lower diagonal solutions.

The situation is different if we consider the five extraversion scales of the California Psychological Inventory (Gough, 1964). The χ^2 goodness of fit statistics for the unconstrained and the lower diagonal models are, respectively 23.95 and 28.74. Both models reject the hypothesis that the additive genetic covariance matrix can be set to 0 (Unconstrained model: $\chi^2 = 69.44$, df = 15, $p < 0.0001$; lower diagonal model: $\chi^2 = 64.65$ df = 15, $p < 0.0001$). On the other hand, the two models differ in hypothesis testing about the common environmental matrix. The unconstrained model rejects the hypothesis that this matrix is 0 ($\chi^2 = 27.61$, df = 15, $p = 0.024$); the lower diagonal model fails to reject this hypothesis ($\chi^2 = 22.82$, df = 15, $p = 0.088$).

Furthermore, there is a large substantive difference in interpretation of the common environmental matrix between the two solutions (see Table IX). This matrix is not positive definite because three diagonal elements are less than 0, making common environmental correlations incomputable. This pattern suggests nonadditive genetic variance. The lower diagonal solution on the other hand gives moderate to large common environmental correlations (with one exception). Which of the two situations is correct? Other data suggest nonadditive genetic variance for extraversion (Eaves *et al*., 1989; see also Table 2.1 of Loehlin, 1992). Furthermore, the common environmental effect for the Capacity for Status (CS) scale has been independently replicated in two other twin samples (Carey *et al*., 1978). Hence, the unconstrained solution is more consistent with the literature

[4] The observation of a negative variance should never be dismissed out of hand. If the negative variance cannot be set to 0, then it suggests that the model itself is at fault. For example, significant negative variance in **C** often implies nonadditive genetic variance.

[5] Here the distinction between the matrix-algebra and the behavioral-genetic definition of the Cholesky matrix becomes important. Software for computing Cholesky factors often applies only to positive definite matrices and will give an error if the matrix is not positive definite.

Table IX. Common environment covariance matrix (correlations above the diagonal) for the unconstrained and lower diagonal solutions for the extraversion scales of the California Psychological Inventory

	DO	CS	SO	SP	SA
Unconstrained solution					
DO	−0.31				
CS	3.22	4.88			0.48
SO	0.01	1.10	−0.07		
SP	0.51	1.78	−3.04	−2.65	
SA	0.65	0.95	−0.40	−0.66	0.80
Lower diagonal solution					
DO	2.59	0.94	0.62	0.80	0.74
CS	3.43	5.09	0.52	0.86	0.49
SO	1.49	1.75	2.20	.05	0.40
SP	1.82	2.75	0.10	2.01	0.51
SA	1.35	1.25	0.68	0.82	1.30

DO = dominance; CS = capacity for status; SO = socialization; SP = social presence; SA = self acceptance.

than the lower diagonal solution. In this example, the Cholesky problem can lead to substantive differences in inference. If a lower diagonal solution is desired, the authors should also report the unconstrained solution and allow readers to judge for themselves which solution to interpret.

These potential remedies must be tempered by the fact that there is no analytical proof about the nature of Cholesky problems. The suggested solutions are based on a *conjecture* about the source of the problem and that conjecture itself is not trouble free. Remember that the likelihood function involves the term (**A** + **C**) or, for DZ twins, (0.5***A** + **C**). Why does a Cholesky problem for the LD parameterization that computes **A** seem to influence only matrix **A** and one for **C** only matrix **C**? Could not Cholesky issues jointly affect the sum of **A** and **C**? Clearly work remains to be done on this issue.

If the conjecture offered herein has some validity, then it portends problems that extend well beyond the Cholesky. Specifically, *any* mathematical model with either explicit or implicit constraints on the maximum likelihood solution may be compromised whenever the implications of those constraints are not recognized. For example, one could iterate on a matrix of standard deviations (**S**), the off-diagonal elements of a correlation matrix (**R**), and compute the covariance matrix as **SRS**. The resulting covariance matrix may be negative definite, but it will never be negative definite because a diagonal element is less than 0. The present conjecture predicts that this parameterization will create problems whenever sampling error results in a negative variance but not when the covariance matrix is negative definite for other reasons (e.g., a genetic correlation greater than 1.0). The extent to which such allied issues present a practical problem in behavioral genetic analysis is another fertile area for future research.

One final point needs forceful emphasis—the Cholesky problems outlined above may turn out to be more of a statistical curiosity than an impediment to empirical research. The only way to discover this is to amass an empirical database that compares the LD parameterization to an unconstrained model and then views the extent to which the two models lead to a substantive difference in the interpretation of results. Because Cholesky problems are a matter of degree—not of kind—we will definitely find that many discrepancies are inconsequential. It is the frequency of the consequential discrepancies that must be empirically determined.

METHODS

Software for the simulations was developed by the author as a series of programs and subroutines written in Fortran 90 and 95 that used algorithms from the NAG (Numerical Algorithms Group, http://www.nag.com), LAPACK, and BLAS libraries for random number generation, function minimization, and certain matrix operations. In each simulated data set, genetic values were generated for a twin pair based on a simple additive model with a genetic correlation of 1.0 for MZ twins and 0.5 for DZ twins. Here, pseudorandom numbers were generated from a multivariate normal distribution with a mean of 0 and a pre-specified genetic covariance matrix (and, of course, the relevant genetic correlation for the type of twin pair) giving vectors $\mathbf{g}_1$ for twin 1 and $\mathbf{g}_2$. Two uncorrelated vectors of unique environmental values, $\mathbf{e}_1$ and $\mathbf{e}_2$, were separately generated from a multivariate normal distribution with means of **0** and a pre-specified unique environmental covariance matrix. Phenotypes for a twin pair were then computed as the sum of the relevant genetic and unique environmental vectors. (In some simulations where the common environmental matrix was not **0**, then vector **c** was generated from the common environmental covariance matrix. The phenotype for a member of a twin pair was then calculated as $\mathbf{g}_i + \mathbf{c} + \mathbf{e}_i$, $i = 1, 2$).

Intraclass covariance matrices were calculated for MZ and for DZ twins based on these phenotypic

vectors by double entry of the twin pair in reversed order. The function minimized was

$$-2\,\mathrm{Log}(L) = \sum_{i=mz}^{dz} N_i(\log \Sigma_i - \log \mathbf{S}_i + \mathrm{trace}(\Sigma_i^{-1}\mathbf{S}_i) - 2p),$$

where N_i is the number of twin pairs for a zygosity, Σ_i is the predicted and $\boldsymbol{S}_i$, the observed covariance matrix for a zygosity, and p is the number of phenotypes.

The predicted phenotypic covariance matrix block was the sum of matrices **A** (additive genetic covariance matrix), **C** (common environmental covariance matrix) and **E** (unique environment covariance matrix), and the cross-twin covariance was calculated as $\gamma\mathbf{A} + \mathbf{C}$ where γ qualed 1.0 for MZ and 0.5 for DZ twins. In the unconstrained model, elements of **A** and **C** were all free parameters that were estimated without bounds. In the Cholesky model, elements of Cholesky factors Λ_A andΛ were iterated on so that $\mathbf{A} = \Lambda_A\Lambda_A^t$ and $\mathbf{C} = \Lambda\Lambda_C^t$. In both the unconstrained and Cholesky models, elements the lower diagonal matrix Λ_E were iterated on so that $\mathbf{E} = \Lambda_E\Lambda_E^t$. (Given uncorrelated error variance, matrix **E** must be positive definite and the current parameterization greatly improved the numerical search. No element in matrix Λ_E ever entered into a boundary constraint or Cholesky problem in the simulations reported here. This precept, however, should not be generalized to all models, especially those that have information on measurement error.)

Two sets of starting values were used, one based on the observed covariance matrices and the other from the genetic and unique environmental covariance matrices used to generate the simulated data. Minimization of the function was done using the NAG algorithm e04ucf. Two criteria were used to assess convergence. The first was the NAG return value indicating successful convergence along with a test that the convergence was not obtained because the parameters entered a space where a predicted twin matrix was not positive definite. If, despite using different sets of starting values, the NAG return did not indicate successful convergence but instead suggested numerical problems in the accuracy of the calculated function value or of elements in the Hessian matrix, then first derivatives were calculated. If the norm of the gradient (i.e., the vector of first derivatives) was less than 10^{-6} and the absolute value of no derivative exceeded 0.0001, then convergence was assumed. Lack of convergence was not a problem. It never occurred using one phenotype and occurred 0.05% of the time with two phenotypes.

The parameters used for a single phenotype were given in the text. The genetic covariance matrix for the two-phenotype simulations was

$$\begin{pmatrix} 0.52 & 0.48 \\ 0.48 & 0.52 \end{pmatrix}$$

and the unique environmental covariance matrix

$$\begin{pmatrix} 0.5125 & 0.39 \\ 0.39 & 0.5125 \end{pmatrix}.$$

Many other covariance matrices were also used, but I chose to present these because they were the ones used to develop and test the program and to perform the initial simulations.

For three, four and five phenotypes, heritabilities were set to 0.50, genetic correlations to 0.60, and unique environmental correlations to 0.20. A range of other values were also used—heritabilities ranging from 0.30 to 0.70, modest amounts of common environment, smaller and larger genetic correlations—in simulations that were not reported except in general terms.

For a single phenotype, 10,000 replicated were generated. I initially used 10,000 replicates for two phenotypes, but they provided too few cases of the rare combinations of parameters that encountered certain forms of Cholesky problems to permit further exploration. Hence, I used 30,000 replicates. For three, four, and five phenotypes, the program was altered so that the first 25,000 simulations that converged were accepted.

I also varied sample size in the simulations, trying 100, 500, 1000, and 5000 pairs of each zygosity. Sample size did indeed influence results, but only in the general terms described in the text. I presented the results on 100 pairs because they gave a sufficient number of Cholesky problems with the additive genetic matrix to compare to those Cholesky problems with the common environmental matrix.

To compute the distribution of the ostensible χ^2 statistic, I grouped the observed fit statistics into cells with increments of 0.02 between adjacent cells. That is, the first cell contained the number of simulations with an ostensible χ^2 between 0 and 0.02, the next cell contained the number of simulations with an ostensible χ^2 between 0.02 and 0.04, and so on. If a cell contained fewer than 5 observations, it was merged with (an) adjacent cell(s) until all cells had at least 5

observations. The goodness of fit to a χ^2 distribution was calculated as

$$\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i},$$

where k = the number of cells, O_i = the observed number of simulations in the ith cell, and E_i = the expected number of simulations in the ith cell. E_i was derived from the distribution function for χ^2 with λ degrees of freedom where λ was either a free parameter or a parameter fixed to the nominal df, depending on the problem at hand. The integral for E_i was calculated from the cell's lower real limit to its upper real limit. The degrees of freedom for this goodness of fit was $k-1$.

To compute the fit for a fixed, nominal degrees of freedom relative to a freely estimated degrees of freedom, parameter λ was fixed at the nominal value and the difference in function values for the above equation (i.e., λ fixed *versus* λ free) was computed. This was treated as χ^2 with one df.

For analysis of the National Merit Twin sample, I first removed mean differences between male and female twins, treating each member of a twin pair as a separate observation. Models were then fitted to MZ and DZ intraclass covariance matrices.

ACKNOWLEDGMENTS

This paper was supported in part by grant M01-RR00051. I thank Dorret Boosma, Conor Dolan, Lindon Eaves, and an anonymous referee for their thoughtful comments and suggestions on two earlier drafts of this paper, and Nick Martin, Mike Neale, and Andrew Heath for helpful insight into the nature of the problem.

REFERENCES

Carey, G., Goldsmith, H. H., Tellegen, A., and Gottesman, I. I. (1978). Genetics and personality inventories: the limits of replication with twin data. *Behav. Genet*. **8**:299–313.

Eaves, L. J., Eysenck, H. J., and Martin, N. G. (1989). *Genes, Culture and Personality: An Empirical Approach*. New York: Academic Press.

Fulker, D. W. (1978). Multivariate extensions of a biometrical model of twin data. *Progr. Clin. Biol. Res*. **24A**:217–236.

Gough, H. B. (1964). *CPI Manual*. Palo Alto, CA: Consulting Psychologists Press.

Loehlin, J. C. (1992). *Genes and Environment in Personality Development*. Thousand Oaks, CA: Sage.

Loehlin, J. C., and Nichols, R. C. (1976). *Heredity, Environment, and Personality*. Austin, TX: University of Texas Press.

Martin, N. G., Jardine, R., and Eaves, L. J. (1984). Is there only one set of genes for different abilities? A reanalysis of the National Merit Scholarship Qualifying Test data. *Behav. Genet*. **14**:355–370.

Neale, M. C., and Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behav. Genet*. **27**:113–120.

Self, S. G., and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc*. **82**:605–610.

Sham, P. C., Curtis, D., and MacClean, C. J. (1996). Likelihood ratio tests for linkage and linkage disequilibrium: Asymptotic distribution and power. *Am. J. Human Genet*. **58**:1093–1095.

Edited by Dorret Boomsma