# A Note on the Asymptotic Distribution of Likelihood Ratio Tests to Test Variance Components

Peter M. Visscher[1,2]

[1] *Queensland Institute of Medical Research, Brisbane, Australia*
[2] *Institute of Evolutionary Biology, University of Edinburgh, United Kingdom*

When using maximum likelihood methods to estimate genetic and environmental components of (co)variance, it is common to test hypotheses using likelihood ratio tests, since such tests have desirable asymptotic properties. In particular, the standard likelihood ratio test statistic is assumed asymptotically to follow a $\chi^2$ distribution with degrees of freedom equal to the number of parameters tested. Using the relationship between least squares and maximum likelihood estimators for balanced designs, it is shown why the asymptotic distribution of the likelihood ratio test for variance components does not follow a $\chi^2$ distribution with degrees of freedom equal to the number of parameters tested when the null hypothesis is true. Instead, the distribution of the likelihood ratio test is a mixture of $\chi^2$ distributions with different degrees of freedom. Implications for testing variance components in twin designs and for quantitative trait loci mapping are discussed. The appropriate distribution of the likelihood ratio test statistic should be used in hypothesis testing and model selection.

Maximum likelihood (ML) is the main method used in practice to estimate variance components in quantitative genetics applied to human, natural and artificially selected populations. In particular, residual (or restricted) maximum likelihood (REML; Patterson & Thompson, 1971) is used in mixed linear models with large complex pedigrees in livestock and evolutionary genetics, and ML is heavily used in structural equation models (which are also mixed linear models) in human behavior genetics (Martin & Eaves, 1977; Neale & Cardon, 1992). With these methods, it has become feasible to test hypotheses about parameters using likelihood ratio tests (e.g., Mood et al., 1974). Such tests have desirable asymptotic properties (Wilks, 1938), although the distribution of the test statistic for small samples is unknown. A fairly common example of the use of likelihood ratio tests in quantitative genetics is to test the significance of variance components (e.g., Foulley et al., 1990; Martin & Eaves, 1977; Shaw, 1987). Usually, this is done by calculating the ML values under two models: one which includes the variance component to be tested, and one which excludes this component. The latter model is equivalent to the first model with the restriction of the variance component to be tested set to zero.

The standard theory regarding the asymptotic distribution of likelihood ratio test statistics is that it is distributed as $\chi^2(k)$, where $k$ is the difference in the number of parameters estimated between the full model, and the nested reduced model (Wilks, 1938). Meyer and Hill (1992) reviewed the use of likelihood ratio tests to approximate sampling variances of estimated variance components using REML, and many others have advocated their use (e.g., Foulley et al., 1990; Shaw, 1987). In behavior genetics, likelihood ratio test statistics are commonly used in model selection (e.g., Martin & Eaves, 1977; Neale & Cardon, 1992). It is commonly assumed that the asymptotic distribution of a likelihood ratio test statistic under the null hypothesis of no variance of the random (latent) effects is a (central) $\chi^2$ with one degree of freedom. However, there is ample statistical literature to show that this is not correct (e.g., Chernoff, 1954; Miller, 1977; Self & Liang, 1987; Stram & Lee, 1994). For estimation of variance components in twin research, this was recently also pointed out by Dominicus et al. (2006).

The aim of this note is to demonstrate and highlight that the asymptotic distribution of the likelihood ratio is a mixture of $\chi^2$ with different degrees of freedom when testing variance components under the null hypothesis that they are zero, and to suggest what to do in practical situations. Although the results are not new as such, they are still not well recognized in both the human behavior genetics and animal genetics communities.

*Received 31 May, 2006; accepted 6 June, 2006.*

*Address for correspondence: Peter Visscher, Genetic Epidemiology, Queensland Institute of Medical Research, PO Royal Brisbane Hospital, QLD 4029, Australia. E-mail: peter.visscher@qimr.edu.au*

## Methods

Consider a balanced one-way design, with unit residual variance, and $s$ groups each with $n$ observations within the group. Under the null hypothesis of no between-group variance, the between-group mean square ($B$) is distributed as $\chi^2(s-1)$, and the within group variance component ($W$) as $\chi^2[s(n-1)]$. For a sample of $s$ twin pairs (of the same zygosity), the distribution of the between- and within-pair mean squares are $\chi^2(s-1)$ and $\chi^2(s)$, respectively, asymptotically (large $s$) identical (e.g., Visscher, 2004).

The probability that the ANOVA estimate of the between-group variance is negative when the population value is zero, is Prob $F_{[s-1,s(n-10)]} < 1$, a probability involving a standard central $F$-distribution. Asymptotically (for large $s$), this probability is half. When using REML, the ANOVA and REML estimates are equivalent when $B$ is greater than $W$, but when $B$ is less than $W$ the between-group variance is set to zero and an estimate of the residual variance is obtained by weighting $B$ and $W$ by their degrees of freedom (Thompson, 1962), that is, $T = [(s-1)B + s(n-1)W]/[(s-1)+s(n-1)]$ = [total *Sum of Squares*]/[$sn-1$]. In that case, the likelihood under the full model (between- and within-group variance) and reduced model (within-group variance only) are the same. When ML is used to estimate variance components, the resulting estimates are biased under both the full and reduced models. However, this bias is small if the number of fixed effects and covariates is small relative to the total number of observations. Because of the simple relationship between ANOVA and REML estimators for balanced designs, we have used REML in this study. However, the same principles apply to ML estimation.

The relationships between the $F$ ratio ($B/W$) and the likelihood ratio test statistic (LRT) for finite samples can be derived. The residual log-likelihood equation for the full model, that is, including the between component of variance, is

$$-2\log(L)|\sigma_b^2\sigma_w^2 = \frac{(s-1)B}{\left(n\sigma_b^2+\sigma_w^2\right)}+\frac{s(n-1)W}{\sigma_w^2} + (s-1)\log\left(n\sigma_b^2+\sigma_w^2\right)+s(n-1)\log\left(\sigma_w^2\right) \quad [1]$$

where $\sigma_b^2$ is the variance between groups

$\sigma_w^2$ is the variance within groups

and the likelihood equation for the reduced model, that is, the model with a residual within group variance only, is

$$-2\log(L)|\sigma_w^2 = \frac{(sn-1)T}{\sigma_w^2}+(sn-1)\log\left(\sigma_w^2\right) \quad [2]$$

When $B$ is less than $W$, the ML for the full and reduced model are equivalent, because the estimate of the between-group variance ($\sigma_b^2$) in the full model is zero. When $B$ is greater than $W$, it can be shown that the likelihood ratio test is equal to:

$$LRT = (sn-1)\log\left[\frac{s(n-1)}{(sn-1)}+\frac{(s-1)}{(sn-1)}\frac{B}{W}\right]-(s-1)\log\left(\frac{B}{W}\right) \quad [3]$$

and, since $F = B/W$,

$$LRT = (sn-1)\log\left[\frac{s(n-1)}{(sn-1)}+\frac{(s-1)}{(sn-1)}F\right]-(s-1)\log(F) \quad [4]$$

Equation 4 immediately implies that the relationship is valid only if $F > 1$, and that if $F = 1$, LRT = 0 for a one-way ANOVA model. Conditional on $F > 1$, the likelihood ratio test for $\sigma_b^2 = 0$ behaves like a standard likelihood-based test for a nested design, that is, its asymptotic distribution is a central $\chi^2$ with one degree of freedom. Hence, if the null hypothesis is true and the sample size very large, in 50% of the cases the test statistic is zero (or, follows a $\chi^2[0]$), and in 50% of the cases it follows a $\chi^2(1)$.

In a two-way design, for example twin pairs nested within schools, the asymptotic distribution of the test statistic under the null hypothesis of zero pair and zero school variance will be a mixture of $\chi^2(0)$, $\chi^2(1)$, and $\chi^2(2)$, corresponding to the ANOVA estimates of both variances less than 0, either pair or school variance less than 0, or both variance estimates greater than 0, respectively. Asymptotically, that is, for a large number of schools and a large number of pairs per school, the distribution of the likelihood ratio test statistic under the null hypothesis that both variances are zero will be $^1/_4\chi^2(0) + {}^1/_2\chi^2(1) + {}^1/_4\chi^2(2)$, because the probability of obtaining nonzero school and pair variance estimates is .25, and the probability of obtaining either a zero school component alone or a zero pair component alone is .25. For a two-way cross-classified design, the asymptotic distribution of the likelihood ratio test statistic under the null hypothesis that both variances are zero will also be $^1/_4\chi^2(0) + {}^1/_2\chi^2(1) + {}^1/_4\chi^2(2)$ (from Thompson, 1962).

A formal and rigorous way of generalizing the distribution of likelihood ratio test statistics where parameters are on the boundary of the parameter space was given by Self and Liang (1987), and explained by Stram and Lee (1994) for variance component estimation in linear mixed models.

These are not just esoteric examples. An incorrect assumption about the distribution of the likelihood ratio test statistic for variance component estimation will lead to the wrong $p$ values and may lead to incorrect inference in model selection procedures. The widely used versatile statistical package Mx (Neale, 2005) calculates $p$ values for nested models from LRT, assuming that all LRT statistics, whether for fixed or random effects, follow central $\chi^2$ distributions. As pointed out by Dominicus et al., (2006) and in this note, this leads to incorrect $p$ values for variance components.

## Discussion

It was illustrated with simple examples of balanced designs why the commonly applied likelihood ratio test for testing zero (co)variance components is



https://doi.org/10.1375/twin.9.4.490 Published online by Cambridge University Press

conservative. Many researchers in quantitative genetics (both in human behavior genetics and in livestock genetics) mistakenly assume that the distribution of the test statistic asymptotically follows a $\chi^2(p)$, with $p$ the number of (co)variance components tested. In these cases, the actual asymptotic distribution of the likelihood ratio follows a mixture of $\chi^2$ distributions, and the statistical inference may be incorrect. As stated before, these results are not new, and can be found in the statistical literature (Chernoff, 1954; Miller, 1977; Self & Liang, 1987; Stram & Lee, 1994), the human genetics gene mapping literature (e.g., Duggirala et al., 1996) and recently for twin studies (Dominicus et al., 2006). Shaw (1987) discussed this problem briefly in the application of ML approaches to quantitative genetics of natural populations. The key paper in the literature on this subject is that by Self and Liang (1987). Without reproducing the extensive mathematical proofs given by Self and Liang (1987), and partly reproduced by Stram and Lee (1994), I have attempted to explain and illustrate the behavior of the likelihood test statistic under the null hypothesis intuitively. Essentially, the mixtures of $\chi^2$ distributions occur because the true parameters are on the boundary of the parameter space, and in that sense they are properties of likelihood ratio tests under non-standard conditions (Self & Liang, 1987). Perhaps the simplest illustration is when testing the equality of two means estimated from different populations, with the null hypothesis $\mu_1 = \mu_2$ versus the alternative hypothesis $\mu_1 > \mu_2$. Under the null hypothesis of $\mu_1 = \mu_2$, the asymptotic distribution of the likelihood ratio test will be distributed as $^1/_2\chi^2(0) + {}^1/_2\chi^2(1)$. This example applies, for example, to gene mapping by allele sharing methods, when the null hypothesis is 50% sharing of affected sibling pairs, and the alternative hypothesis is more than 50% sharing. If the observed amount of sharing is less than 50% then conventionally the test statistic is set to zero, leading to the mixture of zero and $\chi^2(1)$ under the null hypothesis. More complicated examples occur in multivariate applications, as described by Carey (2005).

When testing different variance component models in linear models in quantitative genetics, we are usually not dealing with balanced nested designs. However, with unbalanced designs the asymptotic distribution of the likelihood ratio under the null hypothesis of zero (co)variances will also be a mixture of different $\chi^2$ distributions. When testing a single variance component, the asymptotic distribution under the null hypothesis is $^1/_2\chi^2(0) + {}^1/_2\chi^2(1)$ (Chernoff, 1954). In practice this means that we have been too conservative and not rejected the null hypothesis of zero variance often enough. For example, when testing the null hypothesis of common environmental variance from a standard twin design at $\alpha = .05$ (threshold from $\chi^2[1]$ is 3.84) and the test statistic is, say, 3.0, the null hypothesis will not be rejected. However, the null hypothesis should have be rejected because the asymptotic distribution of the test statistic is $^1/_2\chi^2(0) + {}^1/_2\chi^2(1)$, corresponding to a 5% significance threshold of 2.7. For testing, for example, zero additive genetic (co)variances in a multivariate linear model, the asymptotic distribution of the likelihood ratio test statistic is $^1/_2\chi^2(q\text{-}1) + {}^1/_2\chi^2(q)$, with $q$ the number of traits, that is, the order of the covariance matrix in the full model (Stram & Lee, 1994). In this case, the null hypothesis is that the covariance matrix between $(q\text{-}1)$ traits is positive-definite, and that an additional trait being tested has zero variance and zero covariances with the $(q\text{-}1)$ other traits. For example, under the null hypothesis that one trait in a trivariate analysis has zero genetic variance and zero covariances with the two other traits, the distribution of the likelihood ratio test is asymptotically $^1/_2\chi^2(2) + {}^1/_2\chi^2(3)$. When testing for a zero covariance matrix, that is, for several zero variance components simultaneously, for example for zero between school and between-twin pair variance in a two-way design, the distribution of the LRT is asymptotically a mixture of as many $\chi^2$ distributions as there are variance components, with mixing proportions factors of $^1/_2$. This case was not discussed by Stram and Lee (1994), who assumed that either only one random effect would be tested at a time, or that the test would be conditional on a positive-definite part of the complete covariance matrix.

These boundary problems are not restricted to estimating (co)variance components in mixed linear models. For example, Elsen et al. (1997) tested the null hypothesis of no quantitative trait locus (QTL) segregating at a marker locus in a balanced halfsib design using an approximate ML method, and found by simulation that about 50% of the test statistics were zero, and that the mean and variance of the test statistic were close to what would be expected if it had been distributed as $^1/_2\chi^2(0) + {}^1/_2\chi^2(1)$, that is, a mean of $^1/_2(0) + {}^1/_2(1) = 0.5$, and a variance of $^1/_2\ \mathrm{E}[\chi^2(0)]^2 + {}^1/_2\ \mathrm{E}[\chi^2(1)]^2 - \mathrm{E}^2[{}^1/_2\chi^2(0) + {}^1/_2\chi^2(1)] = 1.5 - 0.25 = 1.25$. Baret et al., (1999) explored the case of QTL mapping using either ML or linear regression in balanced halfsib designs further, and showed that there was a clear relationship between the distribution of the test statistics for the different methods, and they were able to predict the exact proportion of zero LRT for different designs.

For QTL mapping in complex or simple pedigrees using variance components, Blangero and colleagues were, to the author's knowledge, the first to point out that the distribution of *LRT* is a mixture of zero and $\chi^2(1)$ (e.g., Duggirala et al., 1996). The existence of mixture distributions of test statistics in genetics has a much longer history. A likelihood ratio test for the recombination parameter ($\theta$) in parametric linkage analysis is traditionally performed with the null hypothesis of $\theta = {}^1/_2$ and the alternative hypothesis of $\theta < {}^1/_2$, also leading to a 50:50 mixture of zero and $\chi^2(1)$ (e.g., Sham, 1998, pp. 63-64). However, there is an additional implicit assumption which is not usually



https://doi.org/10.1375/twin.9.4.490 Published online by Cambridge University Press

stated for variance component QTL linkage analysis, and that is that the statistical power to detect (additive) genetic variance under the null model is large. That is, there is an implicit assumption that a genetic variance component for family resemblance (e.g., a sib correlation in a nuclear family design) will always be detected. If this assumption is not met then the resulting test statistic for the QTL variance has a larger probability than half of being zero. This occurs because the full model is trying to partition the observed genetic variance into a QTL component and a residual polygenic component of variance. If there is no evidence for genetic variance under the null model then there also will be no evidence for QTL variance under the full model. If there is no additive genetic variance then one would expect the LRT for the QTL variance to be zero with a probability of three quarters, and $\chi^2(1)$ with a probability of one quarter. Visscher and Hopper (2001) simulated this case and reported an expected ML test statistic of ~.2, close to the expected value of .25. A similar finding was reported by Macgregor et al. (2005) for REML analysis of longitudinal data. For complex (multivariate) models involving many (co)variance components in practice, the power to detect certain (co)variances may be compromised even under the null model of no QTL components of variance, so that the distribution of the LRT for the QTL is likely to be a complicated mixture of many $\chi^2$ distributions. I believe that the issue of power under the null model (i.e., when QTL components are not fitted) is one of the reasons why there has been controversy over the distribution of the test statistic for QTL linkage in multivariate models (Amos et al., 2001; Macgregor et al., 2005; Marlow et al., 2003; Wang, 2003).

One question is what one should do in practice. There are at least three different actions that could be taken, (i) ignore the problem, (ii) assume asymptotic properties to calculate $p$ values, and (iii) set empirical thresholds for significance testing or calculate empirical $p$ values. If we ignore the problem which has been described in this note, the statistical tests tend to be too conservative. Hence, the power of detecting a significant variance component will be reduced. For example, Shaw (1987), in a landmark paper for the evolutionary genetics community, found through simulation that the empirical power to detect a significant nonzero variance component was lower than expected when testing against a threshold obtained from a central $\chi^2$ with one degree of freedom, even when the simulated value of the variance component was nonzero (Table 1 of Shaw, 1987). This may have occurred because the probability of obtaining a zero test statistic was large for the small population sizes studied by Shaw (10 sires, three dams per sire, and two progeny per dam). Although Shaw (1987) simulated an 'unconstrained REML', that is, estimates of variances were allowed to be negative, there still was an implicit constraint in the model of Shaw, because the likelihood function can be written in a form similar to Equation 1, which implies, for a halfsib design, that $n\sigma_b^2 + n\sigma_w^2$ should be greater than zero. The author simulated the design of Shaw (1987) with the usual constraint that estimates of the variances should be positive, and powers were calculated for the null hypothesis of a zero between-sire variance component. This corresponds to Shaw's test of a zero additive genetic variance. Results for 10,000 replicates (Table 1) show that the powers were similar to those obtained by Shaw (1987) when testing against a $\chi^2(1)$ distribution, that is, against a 5% threshold of 3.84, and that the proportion of zero likelihood ratio tests varied from .12 to .39. On average, the powers that Shaw (1987) found were slightly higher, presumably because of the different constraints used. However, the results are similar enough to conclude that the reduced power when testing the LRT against a $\chi^2(1)$ distribution is because of the relatively large probability of obtaining a zero test statistic.

Depending on the purpose of the data analysis, losing power because the significance test is conservative can be unsatisfactory. For example, QTL experiments are costly and usually not very large, so that reducing the type-II error is important. For finite samples, we usually ignore the fact that we do not know the distribution of the LRT but use its asymptotic properties. The same principle could be applied when we know that the parameters to be tested are on the boundary of the parameter space. For example, when testing a single variance component, one could assume that the asymptotic distribution of the test statistic is $^1/_2\chi^2(0) + {}^1/_2\chi^2(1)$, and for testing a set of $q$ (co)variance components corresponding to an additional random effect, that the asymptotic distribution is $^1/_2\chi^2(q-1) + {}^1/_2\chi^2(q)$ (Stram & Lee, 1994). It is not clear how bad these assumptions are for finite, unbalanced, samples. If the (assumed) asymptotic distribution of the test statistic is $^1/_2\chi^2(0) + {}^1/_2\chi^2(k)$, then a practical

**Table 1**

The Power of Likelihood Ratio Test and the Proportion of LRT = 0 ($P_0$) under REML Analysis

| $V_A$:$V_D$:$V_E$ | Power | $P_0$ | Power (Shaw, 1987) |
|---|---|---|---|
| 0.2:0.1:0.7 | 5.1 | .39 | 8.2 |
| 0.2:0.5:0.3 | 5.5 | .38 | 6.2 |
| 0.2:0.7:0.1 | 5.2 | .38 | 7.6 |
| 0.5:0.1:0.4 | 16.2 | .21 | 14.8 |
| 0.5:0.25:0.25 | 15.3 | .22 | 11.6 |
| 0.5:0.4:0.1 | 15.0 | .22 | 14.8 |
| 0.8:0.1:0.1 | 28.5 | .12 | 29.0 |

Note: The power of likelihood ratio test (x100) and the proportion of LRT = 0 ($P_0$) under REML analysis of 10,000 data sets drawn from a balanced two-way design with 10 sires, three dams per sire, and two progeny per dam. Empirical power is given as the frequency of the LRT > 3.84, when testing that the additive variance is zero. All variances are constrained to be nonnegative. Parameterization of Shaw (1987) is used, in which $V_A$, $V_D$, and $V_E$ are additive, dominance, and environmental variances, respectively.



https://doi.org/10.1375/twin.9.4.490 Published online by Cambridge University Press

consequence is that the appropriate $p$ value for an observed test statistic is easily calculated by halving the corresponding $p$ value from a $\chi^2(k)$ distribution. In the simple case of testing A and C components in an ACE model using Mx (Neale, 2005), the $p$ value supplied by Mx can simply be halved (Dominicus et al., 2006).

One method which would not make any a priori assumptions about the distribution of the test statistic is the permutation test. This method, which shuffles the observations over levels of fixed and random effects, is an empirical way to set a significance threshold. In the case of the one-way design, observations within groups would be randomly assigned to groups. This method has become the method of choice to set significance thresholds for QTL mapping (Churchill & Doerge, 1994). However, this approach may be computationally too demanding when using large data sets and REML. Computations can be reduced, however, if the only parameter of interest is the mixing proportion for the assumed $\chi^2$ distributions. For each permuted data set, one could calculate the likelihood for the variance component at zero (i.e., for the reduced model, which is usually easier to calculate), and the likelihood for a small positive value (e.g., $10^{-6}$). Comparing these two likelihoods will immediately show whether the LRT is at zero for the permuted sample. Hence, implementing a permutation test like this is possible, even for large data sets. When dealing with ML analyses for simple designs, for example QTL mapping using sibling pairs, then a permutation test is easy to implement and can be used to calculate both pointwise and genome-wide empirical $p$ values. However, if the design of the experiment and/or the hypothesis to be tested are complex then permutation testing is not straightforward or even impossible to implement. For example, for QTL mapping in an arbitrary complex pedigree there is no simple permutation analysis that keeps the parameter estimates under the null hypothesis of no linkage the same.

Another resampling scheme which has been proposed to find the empirical distribution of the LRT is the parametric bootstrap (Shaw & Geyer, 1997). In particular, these authors present an asymptotic parametric bootstrap which is computationally feasible, and is recommended when inequality constraints (e.g., variance components are greater than 0) are enforced in covariance component estimation.

In conclusion, recognizing the problem that the distribution of a likelihood ratio test asymptotically is a mixture of $\chi^2$ distributions when parameters are on the boundary of the parameter space should be taken into account when testing hypotheses and performing model selection, thereby maximizing the power of the experiment and drawing the correct inference.

## Acknowledgments

I wish to thank John Blangero for making me aware of the problem at the Allerton II conference on Genetic Analysis of Economically Important Traits in Livestock in 1996, and Bill Hill, Robin Thompson, Philippe Baret, Frank Shaw, and Ruth Shaw for discussion and helpful comments in the late 1990s, and Stuart Macgregor and Mike Neale for more recent ones.

## References

Amos, C. I., De Andrade, M., & Zhu, D. K. (2001). Comparison of multivariate tests for genetic linkage. *Human Heredity, 51,* 133–144.

Baret, P. V., Knott, S. A., & Visscher, P. M. (1999). On the use of regression and maximum likelihood for QTL mapping in halfsib designs. *Genetical Research, 72,* 149–158.

Carey, G. (2005). Cholesky problems. *Behavior Genetics, 35*, 653–665.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics, 25*, 573–578.

Churchill, G. A., & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics, 138,* 963–971.

Dominicus, A., Skrondal, A., Gjessing, H. K., Pedersen, N. L., & Palmgren, J. (2006). Likelihood ratio tests in behavioral genetics: Problems and solutions. *Behavior Genetics, 36,* 331–340.

Duggirala, R., Stern, M. P., Mitchell, B. D., Reinhart, L. J., Shipman, P. A., Uresandi, O. C., Chung, W. K., Leibel, R. L., Hales, C. N., O'Connell, P., & Blangero, J. (1996). Quantitative variation in obesity-related traits and insulin precursors linked to the *OB* gene region on human chromosome 7. *American Journal of Human Genetics, 59,* 694–703.

Elsen, J. M., Knott, S. A., Le Roy, P., & Haley, C. S. (1997). Comparison between some approximate maximum likelihood methods for quantitative trait locus detection in progeny test designs. *Theoretical and Applied Genetics, 95,* 236–245.

Foulley, J. L., Gianola, D., San Cristobal, M., & Im, S. (1990). A method for assessing extent and sources of heterogeneity of residual variances in mixed linear models. *Journal of Dairy Science*, *73,* 1612–1624.

Macgregor, S., Knott, S. A., White, I., & Visscher P. M. (2005). Analysis of longitudinal quantitative trait data in complex pedigrees. *Genetics, 171,* 1365–1376.

Marlow, A. J., Fisher, S. E., Francks, C., Macphie, I. L., Cherny, S. S. Richardson A. J., Talcott J. B., Stein J. F., Monaco A. P., & Cardon L. R. (2003). Use of multivariate linkage analysis for dissection of a complex cognitive trait. *American Journal of Human Genetics, 72,* 561–570.

Martin, N. G., & Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity, 38,* 79–95.

Meyer, K., & Hill, W. G. (1992). Approximation of sampling variances and confidence intervals for maximum likelihood estimates of variance components. *Journal of Animal Breeding and Genetics, 109*, 264–280.

Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Annals of Statistics, 5,* 746–762.

Mood A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). Tokyo: McGraw-Hill.

Neale, M. C. (2005). Mx Software [Compute software]. Richmond, VA: Department of Psychiatry, Virginia Commonwealth University.

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht, the Netherlands: Kluwer Academic.

Patterson H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika, 58,* 545–554.

Self, S. G., & Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association, 82,* 605–610.

Sham, P. (1998). *Statistics in human genetics*. London: Arnold.

Shaw, F. H., & Geyer, C. J. (1997). Estimation and testing in constrained covariance component models. *Biometrika, 84,* 95–102.

Shaw, R. G. (1987). Maximum likelihood approaches applied to quantitative genetics of natural populations. *Evolution, 41*, 812–826.

Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed model. *Biometrics, 50,* 1171–1177.

Thompson, W. A. (1962). The problem of negative estimates of variance components. *Annals of Mathematical Statistics, 33,* 273–289.

Visscher, P. M. (2004). Power of the classical twin design revisited. *Twin Research, 7,* 505–512.

Visscher, P. M., & Hopper, J. L. (2001). Power of regression and maximum likelihood methods to map QTL from sib-pair and DZ twin data. *Annals of Human Genetics, 65,* 583–601.

Wang, K. (2003). Mapping quantitative trait loci using multiple phenotypes in general pedigrees. *Human Heredity*, *55,* 1–15.

Wilks, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics, 9*, 60–70.



https://doi.org/10.1375/twin.9.4.490 Published online by Cambridge University Press