Analysis of the Inheritance, Selection and Evolution of Growth Trajectories

Mark Kirkpatrick,* David Lofsvold*,[1] and Michael Bulmer†

**Department of Zoology, University of Texas, Austin, Texas 78712, and †Department of Statistics, Oxford University, Oxford OX1 3TG, England*

Manuscript received March 15, 1989
Accepted for publication December 18, 1989

ABSTRACT

We present methods for estimating the parameters of inheritance and selection that appear in a quantitative genetic model for the evolution growth trajectories and other "infinite-dimensional" traits that we recently introduced. Two methods for estimating the additive genetic covariance function are developed, a "full" model that fully fits the data and a "reduced" model that generates a smoothed estimate consistent with the sampling errors in the data. By decomposing the covariance function into its eigenvalues and eigenfunctions, it is possible to identify potential evolutionary changes in the population's mean growth trajectory for which there is (and those for which there is not) genetic variation. Algorithms for estimating these quantities, their confidence intervals, and for testing hypotheses about them are developed. These techniques are illustrated by an analysis of early growth in mice. Compatible methods for estimating the selection gradient function acting on growth trajectories in natural or domesticated populations are presented. We show how the estimates for the additive genetic covariance function and the selection gradient function can be used to predict the evolutionary change in a population's mean growth trajectory.

A predictive theory for the evolutionary response of growth trajectories to selection is an important goal of both evolutionary biologists and breeders. Evolutionary biologists are interested in growth trajectories because of their impact on morphology, size-mediated ecological interactions, and life-history characters (*e.g.*, EBENMAN and PERSSON 1988). Animal and plant breeders are concerned with growth trajectories because of the potential to increase the economic value of domesticated species by altering growth patterns through artificial selection (*e.g.*, FITZHUGH 1976). Since the sizes of individuals of the same age in a population typically vary in a quantitative (continuous) manner, it has long been recognized that quantitative genetics provides appropriate methods for the study of the inheritance and evolution of growth trajectories.

We have recently extended the classical quantitative model for the evolution of multiple characters to "infinite-dimensional" traits such as growth trajectories in which the phenotype of an individual is represented by a continuous function (KIRKPATRICK 1988; KIRKPATRICK and HECKMAN 1989). In those earlier studies, we assumed the parameters of inheritance and selection were known quantities. Our goal in this paper is to develop methods for estimating those parameters and to show how they can be used to analyze the evolution of a population's mean growth trajectory. While the example we discuss deals with body size, the methods apply to any ontogenetic process. More generally, the infinite-dimensional method can be extended to other kinds of traits in which an individual's phenotype is a continuous function, such as reaction norms and morphological shapes, and so may be of use in a variety of biological contexts. An analysis of several data sets using these methods, and a discussion of the evolutionary implications of the results, is planned for a later publication.

The infinite-dimensional model is motivated by the fact that growth trajectories do not immediately fit into the framework of conventional quantitative genetics, which treats the evolution of a finite number of traits. This is because growth trajectories are continuous functions of time, so that a trait in an individual requires an infinite rather than finite number of measurements to fully describe. The infinite-dimensional model offers several advantages over earlier attempts to adapt quantitative genetics to growth trajectories (KIRKPATRICK and HECKMAN 1989). First, it predicts the evolution of the full growth trajectory (rather than at a set of landmark ages) without making *a priori* assumptions about the family of curves that are evolutionarily possible. Second, it provides a method for analyzing patterns of genetic variation that reveal potential evolutionary changes in the growth trajectory for which there is and for which there is not substantial genetic variation. Third, the method appears to have reduced biases in the estimates of the genetic variation (and therefore of the

[1] Present address: Department of Biology, Franklin and Marshall College, P.O. Box 3003, Lancaster, Pennsylvania 17604.

The publication costs of this article were partly defrayed by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

response to selection) when compared with the alternative approaches. Two additional advantages appear from the methods presented in this paper: the spacing of the ages at which the data are collected is correctly accounted for (even when the spacing is uneven), and it allows one to project the evolution of the growth trajectory even when the data on selection and inheritance are collected at two different sets of ages.

We will begin with a brief review of the infinite-dimensional model, then turn to the problem of estimating the parameters of inheritance. To make the ideas concrete, we will illustrate the genetic methods using a subset of the data of RISKA, ATCHLEY and RUTLEDGE (1984) on the genetics of growth in ICR randombred mice. In a detailed study, these workers measured 2693 individuals at weekly intervals between ages 2 weeks and 10 weeks in a half-sib breeding design. For the sake of simplicity, we will use only their data on male body weight at ages 2, 3 and 4 weeks in the following. Next the estimation of the parameters of selection is treated. Last, we show how the estimates of the genetic and selection parameters can be used to project the evolution of the population's mean growth trajectory.

Some of the statistical methods developed in this paper can involve a substantial amount of computation. Computer programs for these operations are available from the first author.

THE INFINITE-DIMENSIONAL MODEL

The mean size of unselected individuals in a cohort through time is referred to as the cohort's *mean growth trajectory* and is denoted by the function $\bar{z}$. Thus the value of $\bar{z}(a)$ is simply the expected size of individuals at age a in the absence of selection. Selection within a given generation generally will cause the observed mean size of individuals to differ from the mean growth trajectory and also will produce an evolutionary change in the mean growth trajectory between that generation and the next.

The evolutionary change in $\bar{z}$ can be determined by extending the standard theory of quantitative genetics to infinite-dimensional characters (KIRKPATRICK and HECKMAN 1989). The growth trajectory of an individual can be thought of as the sum of two continuous functions. The first of these represents the additive genetic component of the growth trajectory inherited from the individual's parents. The second component is attributable to environmental effects, such as nutrition, and to genetic dominance. The additive and nonadditive components are defined to be independent of each other and are assumed to be multivariate normally distributed in the population. This assumption is standard in quantitative genetic models of multiple characters. The normality of genetic effects is consistent with a variety of forms of genetic variation at the individual loci involved provided the number of loci is moderate to large and linkage is loose (BULMER 1985, Chap. 8; BARTON and TURELLI 1989). When genetic effects are not normal it may be possible to transform the scale of measurement to one in which they are (for example, by taking logarithms) (WRIGHT 1968, Chap. 10; FALCONER 1981 Chap. 17). Last, we assume that the growth trajectory is autosomally inherited, that the effects of random genetic drift, mutation, epistasis, and recombination on the mean growth trajectory are negligible compared with selection, and that generations are nonoverlapping.

When selection acts on the sizes of individuals, the evolutionary dynamics of the mean growth trajectory are described by the equation

$$\Delta\bar{z}(a) = \int_0^{a_{\max}} \mathcal{G}(a, x)\beta(x)\, dx, \qquad (1)$$

where $\Delta\bar{z}(a)$ is the evolutionary change in the mean size of individuals of age a following a single generation of selection, $\mathcal{G}$ is the additive genetic covariance function, and β is the selection gradient function (KIRKPATRICK and HECKMAN 1989). Equation 1 can be modified to accommodate situations in which selection acts directly on growth rate rather than size per se; see LYNCH and ARNOLD (1988).

The additive genetic covariance function $\mathcal{G}$ plays the same role in the evolution of growth trajectories that the additive genetic covariance matrix does in the standard theory of quantitative characters (see LANDE 1979). The value of $\mathcal{G}(a_1,a_2)$ is the additive genetic covariance for size between individuals measured at age a_1 and those same individuals measured at age a_2. The selection gradient function β is a measure of the forces of directional selection acting on body size (LANDE and ARNOLD 1983). The magnitude of $\beta(a)$ reflects the strength of directional selection acting on body size at age a. A negative value of $\beta(a)$ indicates selection favors smaller size, while a positive value indicates the converse.

Equation 1 predicts the evolutionary change across only a single generation. In general, it is possible that both the strength of selection and the genetic variation will change from generation to generation. This does not present a problem for Equation 1, however, since new values can be used in each generation. This information can come either from direct estimation of the parameters or from genetic and ecological models that predict how they will change through time. We discuss methods for direct estimation below; theoretical approaches are reviewed by BARTON and TURELLI (1989) and BULMER (1989).

Predicting the evolutionary dynamics of the mean growth trajectory thus requires estimating the parameters of inheritance, described by $\mathcal{G}$, and of selection, described by β. In the next three sections, we discuss

$$\hat{\mathbf{G}} = \begin{bmatrix} 436 & 522 & 424 \\ 522 & 808 & 665 \\ 424 & 665 & 558 \end{bmatrix}$$

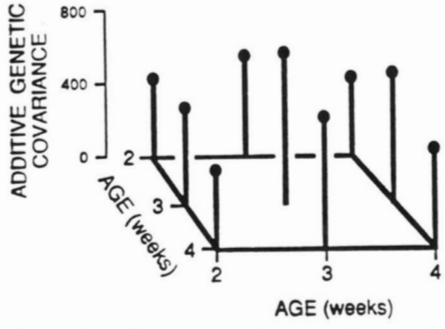

FIGURE 1.—The additive genetic covariances for log(body weight) of male mice 2 to 4 weeks of age, estimated by RISKA, ATCHLEY and RUTLEDGE (1984). *Above*: $\hat{\mathbf{G}}$, the estimated additive genetic covariance matrix for ages 2, 3 and 4 weeks of age. *Below*: A three-dimensional representation of $\hat{\mathbf{G}}$.

estimation of $\mathscr{G}$, the analysis of $\mathscr{G}$, and the estimation of β. Before proceeding, we pause here to describe the notation conventions used throughout the paper. Continuous functions, such as the mean growth trajectory $\mathscr{z}$ and the additive genetic covariance function $\mathscr{G}$, are denoted with a script font. Vectors and matrices are written in bold. We use a hat or a tilda to signify estimates of quantities; for example, the estimate of an additive genetic covariance matrix is written $\hat{\mathbf{G}}$.

ESTIMATING THE ADDITIVE GENETIC COVARIANCE FUNCTION $\mathscr{G}$

To estimate the additive genetic covariance function $\mathscr{G}$, we begin with the additive genetic covariance matrix **G** familiar from standard quantitative genetics. The sizes of an individual at two ages a_i and a_j are considered to be two different characters, and the value of G_{ij} is equal to the additive genetic covariance for the sizes of an individual at those two ages. Methods for estimating genetic variances and covariances of multiple characters have been extensively developed by animal and plant breeders (FALCONER 1981; BULMER 1985), and have more recently been applied to natural populations by evolutionary biologists (*e.g.*, ARNOLD 1981; PRICE, GRANT and BOAG 1984; LOFSVOLD 1986). Given measurements of size at n ages, an $n \times n$ estimated additive genetic covariance matrix $\hat{\mathbf{G}}$ can be calculated. We refer to the vector of n ages at which these measurements have been taken as the *age vector*, denoted **a**.

The entries in the matrix $\hat{\mathbf{G}}$ provide direct estimates of the additive genetic covariance function $\mathscr{G}$ at n^2 points, since $G_{ij} = \mathscr{G}(a_i, a_j)$. The relationship between

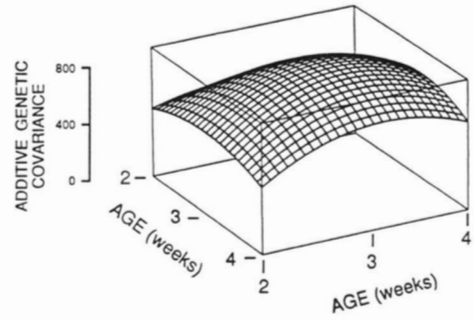

FIGURE 2.—The full estimate $\hat{\mathscr{G}}$ of the additive genetic covariance function, fit to the matrix $\hat{\mathbf{G}}$.

the covariance matrix **G** and the covariance function $\mathscr{G}$ is illustrated in Figures 1 and 2. The values of $\mathscr{G}$ between the measured ages can be estimated by interpolation using smooth curves. By using smooth curves, we make the implicit assumption that the genetic variances and covariances do not change in a discontinuous fashion. (Our method can be modified to accommodate discontinuities produced, for example, by metamorphosis by dividing the growth trajectory into pre- and post-metamorphosis periods, and determining the covariances within and between the two periods.)

A variety of techniques could be used to estimate a continuous covariance function $\mathscr{G}$ from an observed covariance matrix $\hat{\mathbf{G}}$. We have chosen to use a family of methods that involve fitting orthogonal functions to the data. The motivation for using this approach for fitting smooth functions to the data rather than some other (such as splines) is that the coefficients derived from fitting orthogonal functions are very useful for analyzing patterns of genetic variation in the growth trajectory, as we describe below.

A pair of functions ϕ_i and ϕ_j are said to be normalized and orthogonal over the interval $[u, v]$ if

$$\int_u^v \phi_i(x)\phi_j(x)\,dx = 0 \quad \text{and} \quad \int_u^v \phi_i^2(x)\,dx = 1.$$

Many families of functions that meet these criteria are available. We will analyze the mouse data using the well-studied Legendre polynomials. The choice of which family of orthogonal functions to use does not affect the estimates for the covariance function at the ages at which the data were taken (the points in $\hat{\mathbf{G}}$). The choice does, however, affect the interpolation and therefore can affect conclusions regarding ages other than those at which the data were collected. (All families of orthogonal polynomials, however, will produce the same estimate for $\mathscr{G}$ if the maximum degree of the polynomials is held constant.) We favor polynomials over series of sines and cosines (Fourier functions), for example, because on biological grounds we expect a covariance function for growth to be relatively smooth rather than oscillatory. In any event,

the element of arbitrariness introduced by the choice of orthogonal functions decreases as the number of ages at which data were sampled increases.

The jth normalized Legendre polynomial, P_j, is given by the formula

$$P_j(x) = \frac{1}{2^j}\sqrt{\frac{2j+1}{2}} \cdot \sum_{m=0}^{[j/2]} (-1)^m \binom{j}{m}\binom{2j-2m}{j} x^{j-2m}, \tag{2}$$

where $[\cdot]$ indicates that fractional values are rounded down to the nearest integer (BEYER 1976, p. 439). These polynomials are defined over the interval $[-1, 1]$, and so $u = -1$ and $v = 1$. From Equation 2, we find that the first three polynomials are:

$$\phi_0(x) = 1/\sqrt{2}, \tag{3a}$$

$$\phi_1(x) = \sqrt{3/2}\, x, \tag{3b}$$

and

$$\phi_2(x) = \sqrt{45/8}\, x^2 - \sqrt{5/8}. \tag{3c}$$

The additive genetic covariance function $\mathcal{G}$ can be approximated to any specified degree of precision using a complete set of orthogonal functions such as Legendre polynomials (COURANT and HILBERT 1953, p. 65). In this form, the covariance between body size at ages a_1 and a_2 is

$$\mathcal{G}(a_1, a_2) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} [\mathbf{C_G}]_{ij}\, \phi_i(a_1^*)\phi_j(a_2^*), \tag{4}$$

where

$$a_i^* = u + \frac{v-u}{a_{\max} - a_{\min}}(a_i - a_{\min}) \tag{5}$$

and $a_{\min}$ and $a_{\max}$ are respectively the first (smallest) and last (largest) elements of the age vector. The *adjusted age vector* $\mathbf{a}^*$, calculated from the age vector $\mathbf{a}$ using (5), rescales the ages at which the data were taken to the range of the orthogonal functions. In the case of the mouse data, the age vector is $\mathbf{a} = [2, 3, 4]^\top$. Thus $a_{\min} = 2$ and $a_{\max} = 4$, and so the adjusted age vector is $\mathbf{a}^* = [-1, 0, 1]^\top$.

The matrix $\mathbf{C_G}$ in Equation 4 is the *coefficient matrix* associated with the covariance function $\mathcal{G}$. Its elements are constants that depend both on $\mathcal{G}$ and on the family of orthogonal functions ϕ being used (Legendre polynomials, in this example). The full expansion of Equation 4 involves an infinitely large coefficient matrix which can only be estimated with an infinite amount of data. Given measurements on the sizes of individuals at n ages, however, an $n \times n$ truncated version of $\mathbf{C_G}$ can be estimated. We previously found that using the truncated estimate $\hat{\mathbf{C}}_\mathbf{G}$ consisting of relatively few dimensions often produces a good approximation (KIRKPATRICK and HECKMAN 1989), and this is our present goal.

We have developed two methods for estimating the coefficient matrix $\mathbf{C_G}$. These correspond to two different ways to estimate the additive genetic covariance function $\mathcal{G}$. The first method yields what we refer to as a "full" estimate of $\mathcal{G}$. This approach estimates the coefficient matrix in such a way that the corresponding covariance function exactly reproduces the estimated additive genetic variances and covariances at the ages that were measured (that is, $\hat{\mathbf{G}}$). Our second method produces a "reduced" estimate of $\mathcal{G}$. The motivation for this approach is the fact that any estimate of $\mathbf{G}$ includes sampling error. Fitting a function through every point in $\hat{\mathbf{G}}$ causes the sampling error to be included in the full estimate of $\mathcal{G}$. This noise makes the full estimate of $\mathcal{G}$ somewhat less smooth than the actual covariance function is. The reduced method finds a smoother and simpler estimate of $\mathcal{G}$ using information about the sampling error of $\hat{\mathbf{G}}$: the reduced estimate is the lowest-order polynomial that is statistically consistent with the data. A drawback of this method is that it excludes higher-order terms from the estimate of $\mathcal{G}$ even when they actually exist if the experiment is not sufficiently powerful to prove their presence. Because of this, we recommend investigators consider both the full and reduced estimates of $\mathcal{G}$.

The full estimate of $\mathcal{G}$: The full estimate of the additive genetic covariance function, denoted $\hat{\mathcal{G}}$, is found by calculating the coefficient matrix $\hat{\mathbf{C}}_\mathbf{G}$ whose corresponding covariance function exactly reproduces the estimated additive genetic covariance matrix $\hat{\mathbf{G}}$. We can write the observed covariance matrix in terms of the orthogonal functions using Equation 4:

$$\hat{\mathbf{G}} = \mathbf{\Phi}\, \hat{\mathbf{C}}_\mathbf{G}\, \mathbf{\Phi}^\top, \tag{6}$$

where the matrix $\mathbf{\Phi}$ is defined such that $[\mathbf{\Phi}]_{ij} = \phi_j(a_i^*)$. The matrix $\hat{\mathbf{C}}_\mathbf{G}$ is the estimate of the coefficient matrix appearing in Equation 4. It is truncated to dimensions $n \times n$ by the finite number (n) of ages represented in the data matrix $\hat{\mathbf{G}}$. Rearranging Equation 6, we find an expression that can be used to calculate the estimated coefficient matrix:

$$\hat{\mathbf{C}}_\mathbf{G} = \mathbf{\Phi}^{-1}\, \hat{\mathbf{G}}[\mathbf{\Phi}^\top]^{-1}. \tag{7}$$

The matrix $\hat{\mathbf{C}}_\mathbf{G}$ obtained from this calculation can be substituted into Equation 4 to give a continuous estimate of the covariance function $\mathcal{G}$ for all ages between the earliest and latest at which the data were taken.

To illustrate, the study of RISKA *et al.* produced an estimate for the additive genetic covariance matrix of

the log of male body weight at 2, 3 and 4 weeks:

$$\hat{\mathbf{G}} = \begin{bmatrix} 436.0 & 522.3 & 424.2 \\ 522.3 & 808.0 & 664.7 \\ 424.2 & 664.7 & 558.0 \end{bmatrix}.$$

The elements of $\mathbf{\Phi}$ are calculated by evaluating the first three Legendre polynomials (Equation 3a–c) at the three points of the adjusted age vector $\mathbf{a}^*$:

$$\mathbf{\Phi} = \begin{bmatrix} \phi_0(-1) & \phi_1(-1) & \phi_2(-1) \\ \phi_0(0) & \phi_1(0) & \phi_2(0) \\ \phi_0(1) & \phi_1(1) & \phi_2(1) \end{bmatrix} = \begin{bmatrix} 0.7071 & -1.2247 & 1.5811 \\ 0.7071 & 0 & -0.7906 \\ 0.7071 & 1.2247 & 1.5811 \end{bmatrix}.$$

The full estimate of the additive genetic covariance function, $\hat{\mathscr{G}}$, is found by plugging these matrices into Equation 7 to obtain $\hat{\mathbf{C}}_\mathbf{G}$:

$$\hat{\mathbf{C}}_\mathbf{G} = \begin{bmatrix} 1348 & 66.5 & -112.0 \\ 66.5 & 24.3 & -14.0 \\ -112.0 & -14.0 & 14.5 \end{bmatrix}.$$

Finally, the full estimate of $\mathscr{G}$ is obtained by substituting $\hat{\mathbf{C}}_\mathbf{G}$ into Equation 4. This gives

$$\begin{aligned} \hat{\mathscr{G}}(a_1, a_2) = {} & 808 + 71.2(a_1^* + a_2^*) \\ & + 36.4a_1^*a_2^* - 40.7(a_1^{*2}a_2^* + a_1^*a_2^{*2}) \\ & - 215.0(a_1^{*2} + a_2^{*2}) \\ & + 81.6a_1^{*2}a_2^{*2}, \end{aligned}$$

which is valid for ages between $a = 2$ and $a = 4$. The result can be verified by checking that indeed $\hat{G}_{ij} = \hat{\mathscr{G}}(a_i, a_j)$. The full estimate of the additive covariance function for the mouse data calculated in this way is shown in Figure 2.

The reduced estimate of $\mathscr{G}$: Our second approach, that of finding a reduced estimate for $\mathscr{G}$, seeks to fit a set of k orthogonal functions to $\hat{\mathbf{G}}$, where $k < n$. We denote a reduced estimate of $\mathscr{G}$ as $\tilde{\mathscr{G}}$ and the corresponding reduced estimate of the coefficient matrix as $\tilde{\mathbf{C}}_\mathbf{G}$. The method, which is described in detail in APPENDIX A, consists of two steps. First, a candidate estimate of $\mathscr{G}$ is constructed using weighted least squares to fit the simplest possible orthogonal function, that in which $\tilde{\mathscr{G}}$ is constant for all ages. Second, this candidate estimate is tested for statistical consistency with $\hat{\mathbf{G}}$. To perform this test we have developed a procedure that produces an approximate χ^2 statistic for the goodness of fit of the reduced estimate to $\hat{\mathbf{G}}$. If this test shows that $\tilde{\mathscr{G}}$ is consistent with (that is, it does not differ significantly from) $\hat{\mathbf{G}}$, then it is accepted. If $\tilde{\mathscr{G}}$ differs significantly from $\hat{\mathbf{G}}$, we then consider a more complex reduced estimate by fitting the first two orthogonal functions to the data. The fit is again tested using the χ^2 test. The procedure is iterated with successively more orthogonal functions until reduced estimates $\tilde{\mathbf{C}}_\mathbf{G}$ and $\tilde{\mathscr{G}}$ are obtained that are consistent with $\hat{\mathbf{G}}$. If no simpler combination of orthogonal functions will successfully fit the data, the full estimate consisting of n orthogonal functions will always fit the data perfectly.

Using this method on the mouse data (see APPENDIX A), we find that the least-squares estimate for $\tilde{\mathscr{G}}$ that consists of the first Legendre polynomial, ϕ_0, alone is $\tilde{\mathscr{G}}(a_1, a_2) = 324$. This estimate is rejected because the test statistic $\chi^2 = 57.3$ with 5 degrees of freedom shows the estimate is inconsistent with the data ($P \ll 0.01$). The least squares estimate of $\mathscr{G}$ produced by the first two Legendre polynomials (a constant and a linear term) is

$$\tilde{\mathscr{G}}(a_1, a_2) = 312.2 - 11.9(a_1^* + a_2^*) + 24.5\,a_1^*\,a_2^*.$$

This estimate is also inconsistent with $\hat{\mathbf{G}}$ ($\chi^2 = 38.7$, 3 d.f., $P \ll 0.01$). Consequently, it is not possible to find a reduced estimate of $\mathscr{G}$ for this data set: only the full estimate consisting of the first three Legendre polynomials, shown in Figure 2, is statistically consistent with $\hat{\mathbf{G}}$. In contrast, other data sets (particularly cases in which the number of individuals is smaller and the number of ages is larger than in this example) will often result in a reduced estimate that is consistent with the data.

Analysis of the additive genetic covariance function. The major motivation for using orthogonal functions to estimate $\mathscr{G}$ is that the coefficient matrix $\mathbf{C}_\mathbf{G}$ can be used to analyze the patterns of inheritance (KIRKPATRICK and HECKMAN 1989). In particular, the coefficient matrix can be used to calculate the *eigenfunctions* and *eigenvalues* of $\mathscr{G}$.

Eigenfunctions are analogous to the eigenvectors (principal components) familiar from the analysis of covariance matrices. Each eigenfunction is a continuous function that represents a possible evolutionary deformation of the mean growth trajectory. Any mean growth trajectory can be thought of as the sum of a population's current mean growth trajectory plus a combination of the eigenfunctions of its additive genetic covariance function. Paired with each eigenfunction is a number known as its eigenvalue. The eigenvalue is proportional to the amount of genetic variation in the population corresponding to that eigenfunction. Eigenvalues (and the eigenfunctions associated with them) are conventionally numbered in order of decreasing size, beginning with the largest.

Eigenfunctions with large eigenvalues are deformations for which the populations has substantial genetic variation. The shape of the mean growth trajectory will therefore evolve rapidly along these

deformations if they are favored by selection. Eigenfunctions with very small (or zero) eigenvalues, on the other hand, represent deformations for which there is little (or no) additive genetic variation. If selection favors a new mean growth trajectory that is obtained from the current trajectory by some combination of these deformations, there will be very slow (or no) evolutionary progress towards it. The eigenfunctions and eigenvalues therefore contain information that is of great value in understanding the evolutionary potential of growth trajectories. The ith eigenfunction and eigenvalue are denoted ψ_i and λ_i, respectively.

In principle, a covariance function has an infinite number of eigenfunctions and eigenvalues. (Many of the eigenvalues may, however, be zero.) In practice, we are able to estimate only a few of them because experiments give information about the covariance function at only a finite number of points (ages). The number of eigenfunctions and eigenvalues that can be estimated equals the dimensionality of the estimated coefficient matrix, which will be equal to the number of ages at which size was measured when dealing with a full estimate of $\mathscr{G}$ but will be smaller when using a reduced estimate.

Estimates of the eigenfunctions ψ_i and eigenvalues λ_i are calculated from the coefficient matrix $\mathbf{C_G}$. The ith eigenfunction is constructed from the relation

$$\psi_i(a) = \sum_{j=0}^{n-1} [\mathbf{c}_{\psi i}]_j \phi_j(a^*), \tag{8}$$

where $[\mathbf{c}_{\psi i}]_j$ is the jth element of the ith eigenvector of $\mathbf{C_G}$ (KIRKPATRICK and HECKMAN 1989). The ith eigenvalue of $\mathscr{G}$ is identical to the ith eigenvalue of $\mathbf{C_G}$. Eigenfunctions are adjusted to a norm of unity by convention in order to allow meaningful comparisons between the eigenvalues. This is conveniently done by requiring that the norms of the eigenvectors $\mathbf{c}_{\psi i}$ equal unity. (Virtually all software packages which compute eigenvalues and eigenvectors do this as a matter of course.) Thus to obtain estimates of the eigenfunctions and eigenvalues, we determine the eigenvectors and eigenvalues of our estimate of the coefficient matrix $\mathbf{C_G}$, then use these in Equation 8. The method can be applied using either the full coefficient matrix $\hat{\mathbf{C}}_\mathbf{G}$ or a reduced coefficient matrix $\tilde{\mathbf{C}}_\mathbf{G}$.

Sampling errors in the estimate of the genetic covariance matrix $\mathbf{G}$ produce biases in the estimates of the eigenvalues (HILL and THOMPSON 1978). Although the estimate of the arithmetic mean of the eigenvalues (*i.e.*, $1/n \sum_{i=1}^{n} \lambda_i$) is unbiased, the larger eigenvalues are consistently overestimated while the smaller eigenvalues are consistently underestimated. This problem, which is general to all multivariate quantitative genetic studies, becomes particularly obvious in data sets that produce one or more eigenvalue estimates that are negative. (Covariance matrices are by definition positive semidefinite, and so have no negative eigenvalues.) HAYES and HILL (1981) proposed transforming the estimate of $\mathbf{G}$ using a method they term "bending" in order to remedy this problem. Their method can be applied to $\hat{\mathbf{G}}$ whenever negative eigenvalues are encountered if an estimate of the phenotypic covariance matrix $\mathbf{P}$ is available.

Often one would like to know the sampling distributions of the eigenvalues estimated for the additive genetic covariance function. We have developed two methods and describe them in detail in APPENDIX C. The first method constructs separate confidence limits for each eigenvalue by numerical simulation. The approach is to generate a simulated covariance matrix whose expectation is $\hat{\mathbf{G}}$ but that includes random deviations in the elements that correspond to the sampling error. The eigenvalues for the coefficient matrix corresponding to each simulated $\hat{\mathbf{G}}$ are calculated. This procedure is iterated many times, and the distribution for each eigenvalue is constructed empirically with the results. The second method uses a chi-squared statistic to test hypotheses about one or more of the eigenvalues. Typically, the hypothesis of interest is whether or not the observed eigenvalues are statistically distinguishable from zero.

We will now illustrate the methods for analyzing genetic covariance functions with the full estimate of $\mathscr{G}$ from the mouse data. All three eigenvalues of $\hat{\mathbf{C}}_\mathbf{G}$ are positive, and so bending the data matrix is unnecessary. Using a standard computer package, we find that the first (largest) eigenvalue of $\hat{\mathbf{C}}_\mathbf{G}$ is $\lambda_1 = 1361$, and the eigenvector associated with it is

$$\hat{\mathbf{c}}_{\psi 1} = [0.995, 0.0504, -0.0831]^\top.$$

By substituting this into Equation 8, we obtain the full estimate for the first eigenfunction of $\mathscr{G}$:

$$\hat{\psi}_1(a) = 0.7693 - 0.0617a^* - 0.1971a^{*2}.$$

The second and third eigenfunctions are obtained in the same way. The three eigenfunctions are shown in Figure 3. The eigenvalues associated with the eigenfunctions are $\lambda_1 = 1361$, $\lambda_2 = 24.5$ and $\lambda_3 = 1.5$ (Figure 4).

Any conceivable evolutionary change in a population's mean growth trajectory can be written in terms of a weighted sum of the eigenfunctions. The rate at which a population will evolve from its current mean trajectory to some new trajectory favored by selection is determined by the eigenvalues associated with the eigenfunctions responsible for that change. A large eigenvalue indicates that a change corresponding to that eigenfunction will happen rapidly, while a small (or zero) eigenvalue indicates that the change will be slow (or will not happen at all).

The first eigenfunction is a deformation involving

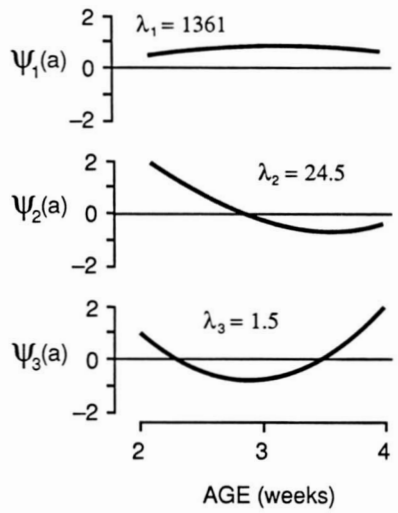

FIGURE 3.—Estimates of the three eigenfunctions and their eigenvalues for the additive genetic covariance function $\mathcal{G}$.

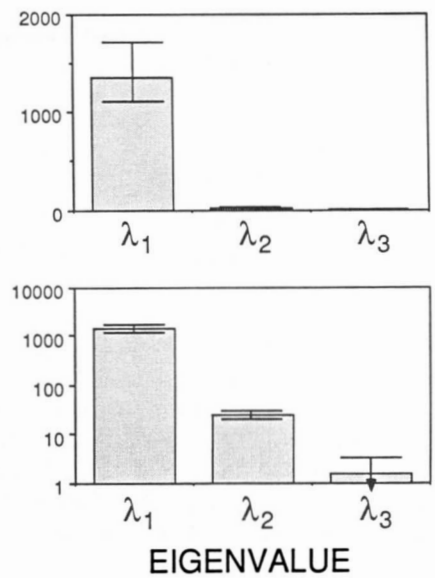

FIGURE 4.—The three eigenvalues of the additive genetic covariance function $\mathcal{G}$ and their 95% confidence limits (determined by the numerical simulation method) on linear (*above*) and logarithmic (*below*) scales. The confidence limits for λ_3 include zero.

an overall increase or decrease of size at all ages (Figure 3). The large size of the first eigenvalue indicates that selection will produce rapid changes if this kind of alteration in the mean growth trajectory is favored. The second eigenfunction corresponds to genetic changes that increase (or decrease) size between ages 2 to 3 weeks, and decrease (or increase) size after 3 weeks of age. The third eigenfunction shows a more complex pattern. The second and third eigenvalues, however, reveal that the amount of genetic variation associated with these eigenfunctions is small in comparison with the variation associated with the first eigenfunction. These eigenvalues indicate that the evolutionary response to selection would be two or more orders of magnitude slower for changes involving the second and third eigenfunctions than for those involving the first eigenfunction.

The 95% confidence regions for each of the eigenvalues constructed by the numerical simulation method (described in APPENDIX C) are [1100, 1700] for λ_1, [17, 33] for λ_2, and [−2.7, 5.1] for λ_3 (Figure 4). We are therefore quite confident that the large differences between the estimate of the first eigenvalue compared with the second and third are real. The conclusion that the estimate for λ_3 is not statistically different from zero is confirmed by the chi-squared test (also described in APPENDIX C). The hypothesis that λ_3 equals zero gives $\chi^2_{(1)} = 0.65$ which is not significant ($P > 0.1$). The hypothesis that both λ_2 and λ_3 are zero, however, is rejected ($\chi^2_{(3)} = 40.4$, $P \ll 0.01$).

A qualitatively similar picture of the pattern of genetic variation for mouse growth emerges from an analysis of the full data set for ages 2–10 weeks. This analysis and its evolutionary implications will be discussed in a later publication.

ESTIMATING THE SELECTION GRADIENT FUNCTION β

Having developed the methods for estimating the additive genetic covariance function $\mathcal{G}$, we now turn to methods for estimating the selection gradient function β. The techniques are extensions of the results of LANDE (1979) and LANDE and ARNOLD (1983). Applications and difficulties with these methods are discussed by ARNOLD and WADE (1984a, b) and MITCHELL-OLDS and SHAW (1987).

Our strategy here is the same as is used to estimate $\mathcal{G}$. The values of β at a finite number of ages are estimated, and then a continuous function is estimated by interpolating between these points. The selection gradient acting on any trait is defined as the partial regression of the phenotypic value of that character onto relative fitness, holding the phenotypic values for other traits constant (LANDE and ARNOLD 1983). In the context of growth trajectories, the partial regression coefficients of relative fitness at each of several ages onto size form an estimated selection gradient vector $\hat{\mathbf{b}}$. The continuous selection gradient function can then be estimated by fitting orthogonal functions to $\hat{\mathbf{b}}$.

A selection gradient function can be written in terms of the same orthogonal functions that were used to describe the additive genetic covariance function:

$$\beta(a) = \sum_{i=0}^{\infty} [\mathbf{c}_\beta]_i \phi_i(a^*) \qquad (9)$$

(KIRKPATRICK and HECKMAN 1989). In (9), $\mathbf{c}_\beta$ is the coefficient vector associated with the selection gradient function β. The full estimate of $\mathbf{c}_\beta$ that passes

through every point in $\hat{\mathbf{b}}$ is found using the relation

$$\hat{\mathbf{c}}_\beta = \mathbf{\Phi}^{-1}\,\hat{\mathbf{b}}. \tag{10}$$

The continuous selection gradient function is estimated by substituting the $\hat{\mathbf{c}}_\beta$ into (9). Alternatively, given information on the errors in the estimates of the elements of $\hat{\mathbf{b}}$, a reduced estimate of β can be calculated using weighted least squares as described in APPENDIX A.

Estimating the selection gradient function β thus requires an estimate of the selection gradient vector $\mathbf{b}$. The basic methodologies for estimating $\mathbf{b}$ are discussed by LANDE and ARNOLD (1983), ARNOLD and WADE (1984a, b), and MITCHELL-OLDS and SHAW (1987). The methods can be applied to growth trajectories in two ways. The first requires data on the sizes of individuals at each of several ages and a measure of their lifetime relative fitnesses. The selection gradient vector can then be estimated directly as the partial regressions of size onto relative fitness at those ages. This is the preferable approach, but is limited to cases in which there is data on the lifetime fitnesses of individuals.

In the absence of such data, an indirect method that makes use of data on the effects of size on fecundity and mortality can be used if relative fitnesses are constant in time. Under that assumption, a result from LANDE (1979) can be extended to show that

$$\beta(a)\,da = \frac{\delta}{\delta\bar{z}}\ln(\bar{W}), \tag{11}$$

where $\bar{W}$ is the population's mean fitness and $(\delta/\delta\bar{z})\ln(\bar{W})$ represents the first variation of $\ln(\bar{W})$ with respect to $\bar{z}$ (see COURANT and HILBERT 1953, p. 184; R. GOMULKIEWICZ, in preparation). Equation 11 is analogous to the equation for a finite number of quantitative characters, $\beta = \nabla\ln(\bar{W})$ (LANDE 1979; LANDE and ARNOLD 1983).

We can make use of Equation 11 if we have some understanding of how size affects fitness. If, for example, fecundity and mortality rates are functions only of size and age, then the relationship between the selection gradient and these life history attributes is

$$\beta(a) = f_1(a)\bar{m}'(a) - f_2(a)\bar{\mu}'(a) \tag{12}$$

where

$$f_1(a) = l(a)/\bar{W},\ f_2(a) = \frac{1}{\bar{W}}\int_a^{a_{\max}} l(x)\bar{m}(x)\,dx,$$

and

$$\bar{W} = \int_0^{a_{\max}} l(x)\bar{m}(x)\,dx.$$

Here, $l(a)$ is the probability a newborn survives to age a, $\bar{m}(a)$ and $\bar{\mu}(a)$ are, respectively, the average birth and mortality rates among surviving individuals at age a, and primes denote derivatives taken with respect to $\bar{z}^*(a)$, the mean size of individuals alive at age a (KIRKPATRICK 1984, 1988). Fitness, on the other hand, may be determined in part by factors other than size and age, such as growth rate. In these cases, Equation 12 can be modified to account for the way in which these other factors determine fitness (LYNCH and ARNOLD 1988).

Using the indirect method of estimating the selection gradient function β depends on evaluating the components of Equation 12 (or its analog, if fitness depends on more than size and age alone). The term $\bar{m}'(a)$ is the rate at which the average birth rate of individuals alive at age a changes per unit increase in the mean size of individuals. Given census data about a cohort of individuals at ages a_i and a_{i+1}, this term can be estimated using the regression of fecundity on body size, divided by the duration of the interval:

$$\bar{m}'(\bar{a}_i) = \frac{d}{d\bar{z}^*(\bar{a}_i)}\,\bar{m}(\bar{a}_i) \approx \frac{\mathrm{Cov}[m,\bar{z}(\bar{a}_i)]}{\sigma^{2*}(\bar{a}_i)(a_{i+1} - a_i)}, \tag{13}$$

where $\bar{a}_i = (a_i + a_{i+1})/2$ is the midpoint of the interval between a_i and a_{i+1}. Equation 13 is a linear interpolation that attributes the effects of size on birth rate to the midpoint of the interval being measured. The term $\mathrm{Cov}[m, \bar{z}(\bar{a}_i)]$ is the covariance between the number of births over the interval and body size among those individuals that survived through the entire interval. The average of an individual's size at the beginning and at the end of the interval should be used for this purpose. The term $\sigma^{2*}(\bar{a}_i)$ is the mean of the variance in size at the start of the interval and the variance in size at the end of the interval among those individuals that survived throughout the period. Only individuals that survive are used in the calculation because the fecundity of individuals that died in the interval is reduced by the reduced time they had in which to reproduce.

The term $\bar{\mu}'(a)$ in Equation 12 represents the effect of a change in the mean body size on the average mortality rate at age a. This is estimated from the relation:

$$\bar{\mu}'(a) = \frac{d}{d\bar{z}^*(\bar{a}_i)}\,\bar{\mu}(\bar{a}_i) \approx \frac{\bar{z}^+(a_i) - \bar{z}(a_i)}{\sigma^{2*}(a_i)(a_{i+1} - a_i)}. \tag{14}$$

In (14), $\bar{z}^+(a_i)$ is the mean size at age a_i of individuals that survive to reach age a_{i+1}, $\bar{z}(a_i)$ is the mean size of all individuals alive at age a_i, and $\sigma^{2*}(a_i)$ is their variance in size. Equations 13 and 14 follow from Equations 11, 12, and the results of ROBERTSON (1966) and PRICE (1970).

The interpolations of Equations 13 and 14 become increasingly accurate as the amount of growth that occurs over the interval becomes small relative to the variation in size among individuals present at the start

of the interval. The remaining terms involved in Equation 12, which are the survivorships and mean birth rates at different ages, can be estimated directly from census data.

Given census data from n times during the ontogeny of a cohort, this method will estimate the selection gradient function at $n - 1$ ages. These $n - 1$ point estimates form a selection gradient vector $\hat{\mathbf{b}}$ which can then be used to estimate the continuous selection gradient function β via Equations 9 and 10.

PREDICTING THE EVOLUTIONARY DYNAMICS OF THE GROWTH TRAJECTORY

The estimates of the additive genetic covariance function $\mathscr{G}$ and the selection gradient function β can be used directly in Equation 1 to predict $\Delta\bar{z}$, the evolutionary change in the mean growth trajectory. Using Equation 1 directly is awkward, however, because the integration in (1) must be performed for each age a at which $\Delta\bar{z}(a)$ is to be evaluated. A method making use of $\mathbf{C_G}$, the coefficient matrix for the additive genetic covariance function, and $\mathbf{c}_\beta$, the coefficient vector for the selection gradient function, circumvents this difficulty. Using a result from KIRKPATRICK and HECKMAN (1989), the evolutionary change in the mean growth trajectory is

$$\Delta\bar{z}(a) = \sum_i [\mathbf{c}_{\Delta\bar{z}}]_i \phi_i(a), \tag{15}$$

where the coefficients $\mathbf{c}_{\Delta\bar{z}}$ are given by the matrix equation

$$\mathbf{c}_{\Delta\bar{z}} = \mathbf{C_G}\mathbf{c}_\beta. \tag{16}$$

The summation in (15) extends over all i for which $[\mathbf{c}_{\Delta\bar{z}}]_i$ is nonzero.

These formulas allow us to estimate the evolutionary change in the mean growth trajectory following one generation of selection. The full or reduced estimates of the coefficient matrix $\mathbf{C_G}$ and coefficient vector $\mathbf{c}_\beta$ are determined using the methods described in the last two sections. These are used in Equation 16 to estimate $\mathbf{c}_{\Delta\bar{z}}$. The result is then substituted into (15) to give an estimate of the evolutionary change.

Equation 16 can be applied regardless of whether or not the additive genetic covariance function and the selection gradient function were estimated at the same ages: transforming the measurement into loadings on orthogonal functions puts the measurements on the same basis. In the event that the number of ages used to estimate the covariance function differs from the number used to estimate the selection gradient function, $\hat{\mathbf{C}}_\mathbf{G}$ and $\hat{\mathbf{c}}_\beta$ will be of different dimensions. Equation 16 can be applied in such cases by truncating the dimensions of the larger one to match those of the smaller.

A difficulty that arises when studying natural populations is that ongoing selection makes it impossible to directly observe the unselected distribution of individuals at any given age, since mortality at earlier ages can alter the distribution that will appear at later ages. The observed mean size of individuals surviving to age a, for example, will generally differ from $\bar{z}(a)$ because of selection at earlier ages. The same problem appears when one tries to estimate the additive genetic covariance function from data on a population experiencing selection. The quantities can, however, be estimated if selection is weak by calculating what the cumulative effects of selection at earlier ages have been on the distribution of sizes among the survivors. The basic methodology has been outlined by LYNCH and ARNOLD (1988).

DISCUSSION

The infinite-dimensional method for analyzing the evolution of growth trajectories joins two alternative methods in current use. Previous workers either have treated the sizes of individuals at a set of landmark ages as a finite number of traits or have fit parametric families of curves to the growth trajectories. Our alternative offers several types of advantages over those methods, including the ability to treat the full, continuous growth trajectory without making restrictive assumptions about the form of growth curves that a population's genetic variation will allow (KIRKPATRICK and HECKMAN 1989).

Two additional benefits to the infinite-dimensional method appear from the techniques introduced in this paper. First, the method explicitly accounts for the spacing of the ages at which the data were taken. There are advantages to designing breeding experiments with unequally spaced sample intervals. Genetic variances and covariances change rapidly during certain periods of ontogeny, often corresponding to critical events such as weaning (see Figure 2; see also HERBERT, KIDWELL and CHASE 1979; CHEVERUD, RUTLEDGE and ATCHLEY 1983; ATCHLEY 1984). Periods in which the variance structure is changing rapidly should receive a greater sampling effort. (Ideally, the frequency at which data is collected should be proportional to how rapidly the variances are changing at that point in development.) The infinite-dimensional approach allows an investigator to concentrate effort on the critical periods, but also give these measurements the appropriate weights when estimating the population's response to selection.

A second additional benefit to using this approach is that the ages at which the genetic parameters are estimated and the ages at which the strength of selection is evaluated need not be the same. It may often be the case that logistical reasons make it hard or impossible to take these data at the same ages. This will immediately eliminate the possibility of using con-

ventional quantitative-genetic methods, since they require that the characters on which the genetic and selection parameters are measured are homologous.

The price paid for these advantages is that our method relies on an assumption of infinite-dimensional normality in the distribution of the additive-genetic component of the growth trajectories. The normality assumption is basic to classical quantitative genetics, and has support from both empirical studies and several kinds of models for the effect of genes at the underlying loci (FALCONER 1981; BULMER 1985; BARTON and TURELLI 1989). The genetic effects for even a single trait can, however, depart from normality (*e.g.*, ROBERTSON 1977). Thus an important question in quantitative genetics is the extent to which multiple quantitative traits (including growth trajectories) conform to multivariate normality. This is an empirical question, since at present it seems unlikely that it can be resolved by theory (TURELLI 1988). Models such as ours that are based on a normality assumption, however, may provide reasonable approximations for the evolution of the mean phenotypes even when this assumption is violated if the departures are small.

We are grateful to R. GOMULKIEWICZ and F. H. C. MARRIOTT for important suggestions regarding the mathematical analysis. We thank B. RISKA for help in analyzing his mouse data. M. LYNCH, M. TURELLI, and two anonymous reviewers made numerous helpful comments on an earlier draft. This work was supported by National Science Foundation grants BSR-8604743 and BSR-8657521 to M. Kirkpatrick.

LITERATURE CITED

ANDERSON, T. W., 1958 *An Introduction to Multivariate Analysis.* Wiley, New York.

ARNOLD, S. J., 1981 Behavioral variation in natural populations. I. Phenotypic, genetic, and environmental correlations between chemoreceptive responses to prey in the garter snake, *Thamnophis elegans.* Evolution **35:** 489–509.

ARNOLD, S. J., and M. J. WADE, 1984a On the measurement of natural and sexual selection: theory. Evolution **38:** 709–720.

ARNOLD, S. J., and M. J. WADE, 1984b On the measurement of natural and sexual selection: applications. Evolution **38:** 720–734.

ATCHLEY, W. R., 1984 Ontogeny, timing of development, and genetic variance-covariance structure. Am. Nat. **123:** 519–540.

BARTON, N. H., and M. TURELLI, 1989 Evolutionary quantitative genetics: how little do we know? Ann. Rev. Genet. **23:** 337–370.

BECKER, W., 1984 *Manual of Quantitative Genetics,* Ed. 4 Academic Enterprises, Pullman, Wash.

BEYER, W. H., 1976 *Handbook of Standard Mathematical Tables,* Ed. 25. C.R.C. Press, Boca Raton, Fla.

BOHREN, B. B., H. E. MCKEAN and Y. YAMADA, 1961 Relative efficiencies of heritability estimates based on regression of offspring on parent. Biometrics **17:** 481–491.

BULMER, M. G., 1985 *The Mathematical Theory of Quantitative Genetics.* Oxford University Press, Oxford.

BULMER, M., 1989 Maintenance of genetic variability by mutation-selection balance: a child's guide through the jungle. Genome **31:** 761–767.

CHEVERUD, J. M., J. J. RUTLEDGE and W. R. ATCHLEY, 1983 Quantitative genetics of development: genetic correlations among age-specific trait values and the evolution of ontogeny. Evolution **37:** 895–905.

COURANT, R., and D. HILBERT, 1953 *Methods of Mathematical Physics,* Vol. 1. Wiley, New York.

DRAPER, N. R., and H. SMITH, 1966 *Applied Regression Analysis.* Wiley, New York.

EBENMAN, B., and L. PERSSON (Editors), 1988 *The Dynamics of Size-Structured Populations.* Springer-Verlag, Berlin.

FALCONER, D. S., 1981 *Introduction to Quantitative Genetics,* Ed. 2 Longman, London.

FITZHUGH, H. A., 1976 Analysis of growth curves and strategies for altering their shapes. J. Anim. Sci. **33:** 1036–1051.

HAYES, J. F., and W. G. HILL, 1981 Modification of estimates of parameters in the construction of genetic selection indices ('bending'). Biometrics **37:** 483–493.

HERBERT, J. G., J. F. KIDWELL and H. B. CHASE, 1979 The inheritance of growth and form in the mouse. IV. Changes in the variance components of weight, tail length, and tail width during growth. Growth **43:** 36–46.

HILL, W. G., and R. THOMPSON, 1978 Probabilities of non-positive definite between-group or genetic covariance matrices. Biometrics **34:** 429–439.

KEMPTHORNE, O., and O. B. TANDON, 1953 The estimation of heritability by regression of offspring on parent. Biometrics **9:** 90–100.

KIRKPATRICK, M., 1984 Demographic models based on size, not age, for organisms with indeterminate growth. Ecology **65:** 1874–1884.

KIRKPATRICK, M., 1988 The evolution of size in size-structured populations, pp. 13–28 in *The Dynamics of Size-Structured Populations,* edited by B. EBENMAN and L. PERSSON. Springer-Verlag, Berlin.

KIRKPATRICK, M., and N. HECKMAN, 1989 A quantitative genetic model for growth, shape, and other infinite-dimensional characters. J. Math. Biol. **27:** 429–450.

LANDE, R., 1979 Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution **33:** 402–416.

LANDE, R., and S. J. ARNOLD, 1983 The measurement of selection on correlated characters. Evolution **37:** 1210–1226.

LOFSVOLD, D., 1986 Quantitative genetics of morphological differentiation in *Peromyscus.* I. Tests of the homogeneity of genetic covariance structure among species and subspecies. Evolution **40:** 559–573.

LYNCH, M., and S. J. ARNOLD, 1988 The measurement of selection on size and growth. pp. 47–59 in *The Dynamics of Size-Structured Populations,* edited by B. EBENMAN and L. PERSSON. Springer-Verlag, Berlin.

MITCHELL-OLDS, T., and R. G. SHAW, 1987 Regression analysis of natural selection: statistical inference and biological interpretation. Evolution **41:** 1149–1161.

PRICE, G. R., 1970 Selection and covariance. Nature **227:** 520–521.

PRICE, T. D., P. R. GRANT and P. T. BOAG, 1984 Genetic changes in the morphological differentiation of Darwin's ground finches, pp. 49–66 in *Population Biology and Evolution,* edited by K. WOHRMANN and V. LOESCHKE. Springer-Verlag, Berlin.

RISKA, B., W. R. ATCHLEY and J. J. RUTLEDGE, 1984 A genetic analysis of targeted growth in mice. Genetics **107:** 79–101.

ROBERTSON, A., 1966 A mathematical model of the culling process in dairy cattle. Anim. Prod. **8:** 93–108.

ROBERTSON, A., 1977 The non-linearity of offspring-parent regression, pp. 297–304 in *Proceedings of the International Conference on Quantitative Genetics,* edited by E. POLLAK, O. KEMPTHORNE and T. B. BAILEY. Iowa State University Press, Ames.

TURELLI, M., 1988 Phenotypic evolution, constant covariances, and the maintenance of additive variance. Evolution **42:** 1342–1347.

WRIGHT, S., 1968 *Evolution and the Genetics of Populations: Vol. 1, Genetic and Biometric Foundations.* University of Chicago Press, Chicago.

Communicating editor: M. TURELLI

APPENDIX A

Here we present a method for fitting a reduced estimate of $\mathscr{G}$ and testing for its consistency with the data. We then illustrate the procedure using the data on the log of male mouse weight at ages, 2, 3 and 4 weeks from RISKA, ATCHLEY and RUTLEDGE (1984).

Finding a reduced estimate: Recall that a reduced estimate is one consisting of k orthogonal functions, where k is smaller than n (the dimensionality of $\hat{\mathbf{G}}$, which equals the number of ages at which measurements of body size were obtained). Given a set S of k orthogonal functions, we use the method of weighted least squares to fit the $k \times k$ reduced coefficient matrix $\tilde{\mathbf{C}}_\mathbf{G}$. This produces the most statistically efficient estimate of the coefficient matrix that can be obtained from a linear function of the elements of $\hat{\mathbf{G}}$ (DRAPER and SMITH 166, p. 80). To apply weighted least squares, we begin by forming the vector $\hat{\mathbf{g}}$ by stacking the successive columns of $\hat{\mathbf{G}}$:

$$\hat{\mathbf{g}} = (\hat{G}_{11}, \cdots, \hat{G}_{n1}, \hat{G}_{12}, \cdots, \hat{G}_{n2}, \cdots, \hat{G}_{nn})^\top.$$

This vector has dimension n^2. The vector $\tilde{\mathbf{c}} = (\tilde{C}_{00}, \cdots, \tilde{C}_{k0}, \tilde{C}_{01}, \cdots, \tilde{C}_{k1}, \cdots, \tilde{C}_{kk})^\top$ is formed in the same way from the (as yet undetermined) coefficient matrix $\tilde{\mathbf{C}}_\mathbf{S}$, and has dimension k^2. In this notation, the relation between the undetermined coefficients and the observed genetic covariances is given by the regression equation

$$\hat{\mathbf{g}} = \mathbf{X}_\mathbf{S}\,\tilde{\mathbf{c}} + \mathbf{e}, \qquad \text{(A1)}$$

where $\mathbf{e}$ is a vector of errors and the matrix $\mathbf{X}_\mathbf{S}$ is determined by the set S of orthogonal functions. The matrix $\mathbf{X}_\mathbf{S}$ is calculated by first forming $\mathbf{\Phi}_\mathbf{S}$, the $n \times k$ matrix obtained by deleting the columns of $\mathbf{\Phi}$ corresponding to those ϕ_j not in S, then taking the Krondecker product of $\mathbf{\Phi}_\mathbf{S}$ with itself:

$$\mathbf{X}_\mathbf{S} = \mathbf{\Phi}_\mathbf{S} \otimes \mathbf{\Phi}_\mathbf{S} = \begin{bmatrix} (\mathbf{\Phi}_\mathbf{S})_{11}\,\mathbf{\Phi}_\mathbf{S} & (\mathbf{\Phi}_\mathbf{S})_{12}\,\mathbf{\Phi}_\mathbf{S} & \cdots \\ (\mathbf{\Phi}_\mathbf{S})_{21}\,\mathbf{\Phi}_\mathbf{S} & (\mathbf{\Phi}_\mathbf{S})_{22}\,\mathbf{\Phi}_\mathbf{S} & \cdots \\ \vdots & \vdots & \end{bmatrix}. \qquad \text{(A2)}$$

This is a matrix of dimensions $n^2 \times k^2$.

Calculating $\tilde{\mathbf{c}}$ also requires the covariance matrix $\mathbf{V}$ of the errors in the estimates of $\hat{\mathbf{g}}$: $V_{ij,kl} = \mathrm{Cov}[\hat{G}_{ij}, \hat{G}_{kl}]$. $\mathbf{V}$ can be estimated given the particular design of the breeding experiment used to estimate $\mathbf{G}$. We present the general method for calculating $\hat{\mathbf{V}}$, the estimate of $\mathbf{V}$, and apply it to three widely used experimental designs (parent-offspring regression, half-sibs, and full sibs) in APPENDIX B.

In typical regression applications, a least-squares estimate of the coefficients in $\mathbf{c}$ would follow directly from the linear form of Equation A1 and the specification of $\mathbf{X}_\mathbf{S}$ and $\mathbf{V}$. The symmetry of $\mathbf{G}$, however, produces redundancies in the vector $\hat{\mathbf{g}}$ that cause $\mathbf{V}$ to be singular and so prevents us from calculating $\tilde{\mathbf{c}}$ from Equation A1 immediately. We therefore make the following modifications:

1. Delete from $\hat{\mathbf{V}}$ those columns and rows corresponding to elements of $\hat{\mathbf{g}}$ whose entry G_{ij} has $i < j$.
2. Delete from $\mathbf{X}_\mathbf{S}$ the rows corresponding to those elements of $\hat{\mathbf{g}}$ for which G_{ij} has $i < j$.
3. For each element of $\tilde{\mathbf{c}}$ for which C_{ij} has $i < j$, add the corresponding column of $\mathbf{X}_\mathbf{S}$ to the column corresponding to C_{ji}, then delete the former column.
4. Delete from $\hat{\mathbf{g}}$ the elements for which G_{ij} has $i < j$.
5. Delete from $\tilde{\mathbf{c}}$ the elements for which C_{ij} has $i < j$.

Following these operations, $\hat{\mathbf{V}}$ has dimensions $n(n+1)/2 \times n(n+1)/2$, $\mathbf{X}_\mathbf{S}$ becomes $n(n+1)/2 \times k(k+1)/2$, $\hat{\mathbf{g}}$ becomes $n(n+1)/2 \times 1$, and $\hat{\mathbf{c}}$ becomes $k(k+1)/2 \times 1$.

We now can calculate $\tilde{\mathbf{c}}$ using standard weighted least squares procedures [see, *e.g.* DRAPER and SMITH (1966, pp. 77–81) and BULMER (1985, pp. 60–61)]:

$$\tilde{\mathbf{c}} = (\mathbf{X}_\mathbf{S}^\top\,\hat{\mathbf{V}}^{-1}\,\mathbf{X}_\mathbf{S})^{-1}\,\mathbf{X}_\mathbf{S}^\top\,\hat{\mathbf{V}}^{-1}\,\hat{\mathbf{g}}. \qquad \text{(A3)}$$

The reduced coefficient matrix $\tilde{\mathbf{C}}_\mathbf{G}$ is then constructed from $\tilde{\mathbf{c}}$. First, form a matrix by restoring the elements deleted in Step 5 above, and "unstacking" the columns. Then insert a row and a column of zeroes in the positions corresponding to those orthogonal functions not included in $\mathbf{\Phi}_\mathbf{S}$ to obtain $\tilde{\mathbf{C}}_\mathbf{G}$. (For example, if the first-order orthogonal function, ϕ_1, has been omitted, a row of zeroes would be inserted into $\mathbf{C}_\mathbf{S}$ between the 0th and 2nd rows, and a column of zeroes between the 0th and 2nd columns.) The reduced estimate $\tilde{\mathscr{G}}$ of the additive genetic covariance function is then obtained by substituting $\tilde{\mathbf{C}}_\mathbf{G}$ into Equation 3.

Having produced the reduced estimate $\tilde{\mathscr{G}}$ using the set of orthogonal functions S, we want to test the goodness of fit of $\tilde{\mathscr{G}}$ to $\hat{\mathbf{G}}$. We have adopted a procedure that approximates the distribution of errors in the estimated $\hat{G}_{ij}$ by a multivariate normal. Using this approximation, the consistency of $\tilde{\mathscr{G}}$ and $\hat{\mathbf{G}}$ can be determined using the standard test for the fit of a regression model [see DRAPER and SMITH (1966, pp. 77–81) and BULMER (1985, pp. 60–61)]. We test the chi-squared statistic

$$\chi^2_{(m-p)} = (\hat{\mathbf{g}} - \mathbf{X}_\mathbf{S}\,\tilde{\mathbf{c}})^\top\,\hat{\mathbf{V}}^{-1}\,(\hat{\mathbf{g}} - \mathbf{X}_\mathbf{S}\,\tilde{\mathbf{c}}), \qquad \text{(A4)}$$

where $m = n(n + 1)/2$ is the number of degrees of freedom in $\hat{\mathbf{G}}$ and $p = k(k + 1)/2$ is the number of parameters being fit. A significant result indicates that the model is inconsistent with the data, in which case we attempt to fit a model consisting of a larger number of orthogonal functions.

Because we are approximating the errors in the $\hat{\mathbf{G}}_{ij}$'s as multivariate normal, the chi-squares test does not produce exact probability values. We expect it, however, to be a reasonable guide that discriminates between candidate covariance functions that fit the data reasonably well and those that do not. More accurate tests could be developed with numerical simulation.

In summary, the algorithm for finding the reduced estimate of the covariance function is as follows. Estimates of the additive covariance function are obtained by fitting orthogonal functions in a stepwise manner using weighted least squares (Equation A1). Each estimate is tested against $\hat{\mathbf{G}}$ using an approximate statistical test given by Equation A4. The reduced estimate is the simplest set of orthogonal functions (*e.g.*, the polynomial of lowest degree) which when fit produces an estimate of $\mathscr{G}$ that is not statistically different from $\hat{\mathbf{G}}$.

A worked example: To illustrate the method, consider a reduced estimate for the mouse data consisting of the first two Legendre polynomials (that is, the polynomials of degrees 0 and 1). The data matrix $\hat{\mathbf{G}}$ is given in the text (following Equation 7). Following the steps outlined above, we have

$$\hat{\mathbf{g}} = [436.0,\ 522.3,\ 424.2,\ 522.3,\ 808.0,\ 664.7,\ 424.2,\ 664.7,\ 558.0]^{\mathsf{T}}.$$

The matrix $\mathbf{\Phi_S}$ is found by deleting from the matrix $\mathbf{\Phi}$ (given in the text, following Equation 7) the third column, corresponding to the missing 2nd degree polynomial. This gives:

$$\mathbf{\Phi_S} = \begin{bmatrix} 0.7071 & -1.2247 \\ 0.7071 & 0 \\ 0.7071 & 1.2247 \end{bmatrix}.$$

From Equation A2 and the steps listed above we find that

$$\mathbf{X_S} = \begin{bmatrix} 0.5 & -0.866 & -0.866 & 1.5 \\ 0.5 & 0.0 & -0.866 & 0.0 \\ 0.5 & 0.866 & -0.866 & -1.5 \\ 0.5 & -0.866 & 0.0 & 0.0 \\ 0.5 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.866 & 0.0 & 0.0 \\ 0.5 & -0.866 & 0.866 & -1.5 \\ 0.5 & 0.0 & 0.866 & 0.0 \\ 0.5 & 0.866 & 0.866 & 1.5 \end{bmatrix}.$$

Using the method described in APPENDIX B, we find that $\hat{\mathbf{V}}$, the estimated covariance matrix of errors in $\hat{\mathbf{g}}$, is

$$\hat{\mathbf{V}} = \begin{bmatrix} 2752 & 3187 & 2541 & 3187 & 3692 & 2944 & 2541 & 2944 & 2347 \\ 3187 & 4527 & 3504 & 4527 & 6210 & 4830 & 3584 & 4830 & 3754 \\ 2541 & 3504 & 3057 & 3504 & 4708 & 4058 & 3057 & 4058 & 3477 \\ 3187 & 4527 & 3504 & 4527 & 6210 & 4830 & 3584 & 4830 & 3754 \\ 3692 & 6210 & 4708 & 6210 & 10453 & 7921 & 4708 & 7921 & 6005 \\ 2944 & 4830 & 4058 & 4830 & 7921 & 6673 & 4058 & 6673 & 5562 \\ 2541 & 3584 & 3057 & 3584 & 4708 & 4058 & 3057 & 4058 & 3477 \\ 2944 & 4830 & 4058 & 4830 & 7921 & 6673 & 4058 & 6673 & 5562 \\ 2347 & 3754 & 3477 & 3754 & 6005 & 5562 & 3477 & 5562 & 5155 \end{bmatrix}.$$

We now follow the steps prescribed above. Step 1, which deletes rows and columns from $\hat{\mathbf{V}}$, produces

$$\hat{\mathbf{V}} = \begin{bmatrix} 2752 & 3187 & 2541 & 3692 & 2944 & 2347 \\ 3187 & 4527 & 3504 & 6210 & 4830 & 3754 \\ 2541 & 3504 & 3057 & 4708 & 4058 & 3477 \\ 3692 & 6210 & 4708 & 10453 & 7921 & 6005 \\ 2944 & 4830 & 4058 & 7921 & 6673 & 5562 \\ 2347 & 3754 & 3477 & 6005 & 5562 & 5155 \end{bmatrix}.$$

By deleting rows from $\mathbf{X_S}$ (Step 2 above) yields

$$\mathbf{X_S} = \begin{bmatrix} 0.5 & -0.866 & -0.866 & 1.5 \\ 0.5 & 0.0 & -0.866 & 0.0 \\ 0.5 & 0.866 & -0.866 & -1.5 \\ 0.5 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.866 & 0.0 & 0.0 \\ 0.5 & 0.866 & 0.866 & 1.5 \end{bmatrix}.$$

The vector of coefficients, $\mathbf{c} = [c_{00}, c_{10}, c_{01}, c_{11}]^{\mathsf{T}}$, contains the element c_{01} for which $i < j$. In Step 3 we therefore add the third column of $\mathbf{X_S}$ to the second, then delete the third column. This leaves the matrix in its final form:

$$\mathbf{X_S} = \begin{bmatrix} 0.5 & -1.732 & 1.5 \\ 0.5 & -0.866 & 0.0 \\ 0.5 & 0.0 & -1.5 \\ 0.5 & 0.0 & 0.0 \\ 0.5 & 0.866 & 0.0 \\ 0.5 & 1.732 & 1.5 \end{bmatrix}.$$

Removing the redundant elements in $\hat{\mathbf{g}}$ (Step 4) gives

$$\hat{\mathbf{g}} = [436.0,\ 522.3,\ 424.2,\ 808.0,\ 664.7,\ 558.0]^{\mathsf{T}},$$

and doing the same for $\tilde{\mathbf{c}}$ (Step 5) produces $\tilde{\mathbf{c}} = [c_{00}, c_{01}, c_{11}]^{\mathsf{T}}$.

The reduced coefficient vector $\tilde{\mathbf{c}}$ is calculated using Equation A3. This gives

$$\tilde{\mathbf{c}} = [624.3, -13.8, 16.3]^{\mathsf{T}},$$

and so

$$\tilde{\mathbf{C}}_{\mathbf{G}} = \begin{bmatrix} 624.3 & -13.8 & 0.0 \\ -13.8 & 16.3 & 0.0 \\ 0.0 & 0.0 & 0.0 \end{bmatrix}.$$

By using these coefficients in Equation 3, we arrive at the reduced estimate of $\mathscr{G}$ that consists of the 0 and 1st degree Legendre polynomials:

$$\tilde{\mathscr{G}}(a_1, a_2) = 312.2 - 11.9(a_1^* + a_2^*) + 24.5a_1^*a_2^*.$$

The reduced estimate $\tilde{\mathscr{G}}$ can be tested against the observed genetic covariance matrix $\hat{\mathbf{G}}$ using the chi-squared statistic of Equation A4. We find $\chi^2 = 38.68$. Since $\hat{\mathbf{G}}$ has $m = 6$ degrees of freedom and we have estimated $p = 3$ coefficients, we test the statistic with 3 degrees of freedom and find that the difference between the reduced estimate $\tilde{\mathscr{G}}$ and the observed values $\hat{\mathbf{G}}$ are highly significant. We therefore reject the reduced estimate consisting only of the 0 and 1st degree Legendre polynomials.

Following the same procedure for all other combinations of 0, 1st and 2nd degree Legendre polynomials shows that only the full estimate consisting of all three is consistent with $\hat{\mathbf{G}}$. The error variance of the G_{ij}'s in these data is therefore sufficiently small that no reduced model is acceptable, although this may often not be the case for smaller data sets. The full estimate $\hat{\mathscr{G}}$ is shown in Figure 2.

APPENDIX B

This appendix describes methods to calculate $\mathbf{V}$, the covariance matrix of errors in $\hat{\mathbf{G}}$, the estimated additive genetic covariance matrix. We use the notation $V_{ij,kl}$ to denote the covariance of $\hat{G}_{ij}$ and $\hat{G}_{kl}$. Below we present formulae for estimating $\mathbf{V}$ from three widely used breeding designs: half sibs, full sibs, and parent-offspring regression.

In the following calculations, we will often need an expression for the covariance of two mean cross products. From classical statistics theory we have the result

$$\mathrm{Cov}(M_{ij}, M_{kl}) = (M_{ik}M_{jl} + M_{il}M_{jk})/f, \qquad \text{(B1)}$$

where M_{ij} is the mean cross product of variables i and j, and f is the number of degrees of freedom (Anderson 1958, p. 161; Bulmer 1985, p. 94). Replacing each of the M's with its estimate $\hat{M}$ and dividing by $(f+2)$ rather than f yields $\hat{V}_{ij,kl}$, an unbiased estimator of the covariance.

Half-sib analysis: In the classic half-sib analysis, s sires are each mated to n dams, and one offspring is measured from each mating. An analysis of variance and covariance partitions the observed variation among the offspring into components among sires and a residual [see Falconer (1981, p. 140) and Becker (1984, pp. 45–54, 113–118)]. The additive genetic component of variance is estimated as

$$\hat{G}_{ij} = 4(\hat{M}_{a,ij} - \hat{M}_{e,ij})/n, \qquad \text{(B2)}$$

where $M_{a,ij}$ is the mean cross-product among sires, $M_{e,ij}$ is the residual mean crossproduct, and n is the number of offspring per sire in a balanced design. (The mean crossproducts $M_{a,ij}$ and $M_{e,ij}$ are defined so as to be independent.) The sampling covariance is therefore

$$\hat{V}_{ij,kl} = \frac{16}{n^2}[\mathrm{Cov}(\hat{M}_{a,ij}, \hat{M}_{a,kl}) + \mathrm{Cov}(\hat{M}_{e,ij}, \hat{M}_{e,kl})], \qquad \text{(B3)}$$

where the covariances of the $\hat{M}$'s are given by Equation B1.

We often want to estimate $\mathbf{V}$ from data summaries in the literature that do not include the estimated mean cross products. These quantities can, however, be back-calculated from the estimated additive genetic and phenotypic covariance matrices $\hat{\mathbf{G}}$ and $\hat{\mathbf{P}}$ that frequently are reported. In a half sib analysis, the necessary relations are

$$\hat{M}_{e,ij} = \hat{P}_{ij} - \frac{1}{4}\hat{G}_{ij} \qquad \text{(B4a)}$$

and

$$\hat{M}_{a,ij} = \frac{n-1}{4}\hat{G}_{ij} + \hat{P}_{ij}. \qquad \text{(B4b)}$$

Substituting (B4) into (B3) then gives an estimate of $V_{ij,kl}$.

Full sib analysis: In this breeding design, each of s sires is mated to d dams, and n progeny are measured per dam [see Falconer (1981, pp. 140–141) and Becker (1984, pp. 55–65, 119–127)]. The resulting nested analysis of variance and covariance yields two estimates of the genetic covariance:

$$\hat{G}_{s,ij} = 4(\hat{M}_{s,ij} - \hat{M}_{d,ij})/nd, \qquad \text{(B5a)}$$

and

$$\hat{G}_{d,ij} = 4(\hat{M}_{d,ij} + \hat{M}_{e,ij})/n, \qquad \text{(B5b)}$$

where $\hat{M}_{s,ij}$, $\hat{M}_{d,ij}$, and $\hat{M}_{e,ij}$ are respectively the estimated among sires, among dams, and residual crossproducts. The two estimates of the G's give rise to two estimates for the V's:

$$\hat{V}_{ij,kl} = \frac{16}{n^2d^2}[\mathrm{Cov}(\hat{M}_{s,ij}, \hat{M}_{s,kl}) + \mathrm{Cov}(\hat{M}_{d,ij}, \hat{M}_{d,kl})] \qquad \text{(B6a)}$$

and

$$\hat{V}_{ij,kl} = \frac{16}{n^2}[\mathrm{Cov}(\hat{M}_{e,ij}, \hat{M}_{e,kl}) + \mathrm{Cov}(\hat{M}_{d,ij}, \hat{M}_{d,kl})], \qquad \text{(B6b)}$$

where the covariances are again calculated using Equation A1. The two estimates of **V** obtained from (B6a) and (B6b) can be averaged to give a single composite estimate.

The $\hat{M}$'s that appear in (B6a,b) can be obtained from reported values of $\mathbf{G_S}$, $\mathbf{G_d}$, and **P** using

$$M_{e,ij} = P_{ij} - \frac{1}{4}(G_{s,ij} + G_{d,ij}), \tag{B7a}$$

$$\begin{aligned} M_{d,ij} &= \frac{n}{4} G_{d,ij} + M_{e,ij} \\ &= \frac{n-1}{4} G_{d,ij} - \frac{1}{4} G_{s,ij} + P_{ij}, \end{aligned} \tag{B7b}$$

and

$$\begin{aligned} M_{s,ij} &= \frac{nd}{4} G_{s,ij} + M_{d,ij} \\ &= \frac{nd}{4} G_{s,ij} + \frac{n-1}{4} G_{d,ij} + P_{ij}. \end{aligned} \tag{B7c}$$

Parent-offspring regression: When parent-offspring regression [see FALCONER (1981, pp. 136–140) and BECKER (1984, pp. 103–106, 133–134)] is used, the additive genetic covariance of trait i with trait j can be estimated using

$$G_{ij} = (M_{ij} + M_{ji})/2, \tag{B8}$$

where M_{ij} is the estimated crossproduct for trait i in the parents and trait j in the offspring. That is,

$$M_{ij} = \frac{1}{f} \sum_k \left(z^o_{ik} - \bar{z}^o_i\right)\left(z^p_{jk} - \bar{z}^p_j\right) \tag{B9}$$

where z^0_{ik} is the mean of trait i in family k, $\bar{z}^o_i$ is the overall mean of z_i in the offspring, z^p_{jk} is the midparent value of trait j in family k, $\bar{z}^p_j$ is the overall mean of trait j in the parents, and f is the degrees of freedom. Our estimate of the sample covariances of the genetic covariances are then readily obtained from Equation B1 as

$$\begin{aligned} \hat{V}_{ij,kl} &= \mathrm{Cov}\left[\frac{1}{2}(M_{ij} + M_{ji}), \frac{1}{2}(M_{kl} + M_{lk})\right] \\ &= (\hat{G}_{ik}\hat{G}_{jl} + \hat{G}_{il}\hat{G}_{jk})/(f+2). \end{aligned} \tag{B10}$$

Variation in family size can be taken into account using a form of weighted regression (KEMPTHORNE and TANDON 1953). Doing so results in each mean crossproduct, M_{ij}, being multiplied by a weight, w_i, which is the reciprocal of the variance of the offspring means about the regression line. The weight of trait z_i is

$$w_i = \left\{\left[\left(\rho_i - \frac{1}{2}\beta^2_{ij}\right) + \frac{(1-\rho_i)}{n}\right]P_{ii}\right\}^{-1}, \tag{B11}$$

where ρ_i is the intraclass correlation of trait z_i in the offspring (= $h^2/2$ for midparent regression in the absence of dominance and environmental correlations between sibs), β_{ij} is the slope of the parent-offspring regression, P_{ii} is the phenotypic variance of z_i, and n is the number of offspring per family (KEMPTHORNE and TANDON 1953; BOHREN, MCKEAN and YAMADA 1961; BULMER 1985, p. 79). If family size varies, weighted regression should be used to estimate the genetic parameters. ρ_i and β_{ij} can either be guessed, or estimated from the data and used to iteratively calculate the regression coefficients (*cf.* BULMER 1985, pp. 83–84). Note, however, the latter method yields biased estimates of the parameters (BOHREN, MCKEAN and YAMADA 1961).

APPENDIX C

This appendix describes in detail two methods for testing hypotheses about the estimated additive genetic covariance function $\mathscr{G}$. The first tests whether one or more of the eigenvalues of $\mathscr{G}$ are statistically indistinguishable from 0. The second is a numerical method for constructing the confidence limits of the eigenvalues of $\mathscr{G}$. In this appendix we make use of the notation and results of APPENDIXES A and B.

To find confidence limits on the estimates of the eigenvalues of $\mathscr{G}$, we begin by forming the $n(n+1)/2$-dimensional vector $\hat{\mathbf{g}}$ from the diagonal and subdiagonal elements of $\hat{\mathbf{G}}$ (as described in APPENDIX A) and the $n(n+1)/2 \times n(n+1)/2$ error matrix $\hat{\mathbf{V}}$ (as described in APPENDIX B). The elements of an additive genetic covariance matrix simulated with error are calculated as

$$\mathbf{g}' = \hat{\mathbf{g}} + \hat{\mathbf{V}}^{1/2}\mathbf{e}, \tag{C1}$$

where $\hat{\mathbf{V}}^{1/2}$ is the matrix square root of $\hat{\mathbf{V}}$ and **e** is a $n(n+1)/2$-dimensional vector of uncorrelated, normally distributed random variates with expectation 0 and variance 1. The simulated covariance matrix $\mathbf{G}'$ is then reconstructed from the elements of $\mathbf{g}'$. The corresponding coefficient matrix $\mathbf{C_G}'$ is determined using Equation 5, and its eigenvalues calculated. The values are recorded, and the entire procedure reiterated. We have been using 1000 iterations in our analyses.

The α-percent confidence limits for each eigenvalue can then be determined directly by the range included by $1 - \alpha$ of the values. Confidence regions for the values of the eigenfunctions at any specified points (ages) of interest can be determined at the same time.

Our second method tests the hypothesis that one or more of the estimated eigenvalues of $\mathscr{G}$ is statistically indistinguishable from 0. We can write the estimated coefficient matrix $\hat{\mathbf{C}}_\mathbf{G}$ in terms of its eigenvalues and eigenvectors:

$$\hat{\mathbf{C}}_\mathbf{G} = \mathbf{U} \wedge \mathbf{U}^\top, \tag{C2}$$

where $\wedge$ is a diagonal matrix whose elements are the eigenvalues of $\hat{\mathbf{C}}_\mathbf{G}$ and **U** is a matrix whose columns

are the corresponding eigenvectors. We then generate a coefficient matrix $\mathbf{C}_{\mathbf{G}}^{*}$ by setting one or more of the eigenvalues in Λ in Equation C_2 equal to 0. The genetic covariance matrix $\mathbf{G}^*$ is constructed using

$$\mathbf{G}^* = \mathbf{\Phi}\, \mathbf{C}_{\mathbf{G}}^{*}\, \mathbf{\Phi}^{\top}, \tag{C3}$$

from which vector $\mathbf{g}^*$ is formed from the lower diagonal elements of $\mathbf{G}^*$ in the same way that $\hat{\mathbf{g}}$ was. The hypothesis of zero eigenvalues is then tested with the chi-squared statistic

$$\chi^2 = (\hat{\mathbf{g}} - \mathbf{g}^*)^{\top}\, \hat{\mathbf{V}}^{-1}(\hat{\mathbf{g}} - \mathbf{g}^*) \tag{C4}$$

with $t(t + 1)/2$ degrees of freedom, where t is the number of eigenvalues that have been set to zero. If this reaches a significant value, then the hypothesis that those eigenvalues are zero is rejected. The same procedure can be used to test a hypothesis that one or more eigenvalues are equal to some specified values other than zero.