

## THE MEASUREMENT OF SELECTION ON QUANTITATIVE TRAITS: BIASES DUE TO ENVIRONMENTAL COVARIANCES BETWEEN TRAITS AND FITNESS

MARK D. RAUSHER

*Department of Zoology, Duke University, Durham, NC 27706 USA*

*Abstract.*—The use of regression techniques for estimating the direction and magnitude of selection from measurements on phenotypes has become widespread in field studies. A potential problem with these techniques is that environmental correlations between fitness and the traits examined may produce biased estimates of selection gradients. This report demonstrates that the phenotypic covariance between fitness and a trait, used as an estimate of the selection differential in estimating selection gradients, has two components: a component induced by selection itself and a component due to the effect of environmental factors on fitness. The second component is shown to be responsible for biases in estimates of selection gradients. The use of regressions involving genotypic and breeding values instead of phenotypic values can yield estimates of selection gradients that are not biased by environmental covariances. Statistical methods for estimating the coefficients of such regressions, and for testing for biases in regressions involving phenotypic values, are described.

*Key words.*—Biased estimates, environmental covariance, fitness, quantitative traits, selection gradients.

Received April 15, 1991. Accepted September 16, 1991.

The recent development of a theoretical framework for the evolution of complex phenotypes (Lande, 1976, 1979, 1980; Turelli, 1984, 1985, 1988; Zeng, 1988) has engendered development of techniques for quantifying selection on suites of quantitative characters (Lande and Arnold, 1983; Arnold and Wade, 1984*a*, 1984*b*; Crespi and Bookstein, 1989; Wade and Kalisz, 1989, 1990). These techniques, which rely primarily on regression analysis using phenotypic values, have several advantages for field biologists: they can be applied relatively simply to data from natural populations, they permit the separation of direct selection on a character from indirect selection due to selection on correlated characters, and they allow overall selection within a generation to be broken down into distinct episodes corresponding to different portions of the life cycle. On the other hand, potential limitations associated with these techniques include multicollinearity of traits, biases introduced by failure to include relevant characters, deviations of data from multivariate normality assumed for hypothesis testing, and failure to consider environmental factors affecting both characters and fitness. Mitchell-Olds and Shaw (1987) discuss in detail methods of alleviating the first three difficulties and provide

brief suggestions for dealing with the fourth, while Price et al. (1988) discuss environmentally-induced biases in estimates of selection on single traits. Here I consider in more detail how problems caused by such environmentally-induced correlations between suites of characters and fitness arise, provide a method for detecting whether such correlations may be occurring, and outline a procedure that can provide estimates of selection gradients that are not biased by environmental correlations involving fitness.

### THE PROBLEM

The difficulty that arises when environmental factors influence both a character and fitness can be illustrated by considering a single trait. Consider, for example, the concentration of an hypothetical alkaloid in the foliage of an hypothetical plant species. Imagine that, for individuals of a given genotype, foliar alkaloid concentration is proportional to soil nitrate concentration. Imagine also that seed production (fitness) is enhanced by increased soil nitrate concentration. Finally, consider an experiment in which a number of individuals of each of several genotypes are planted in randomized positions in a field that is heterogeneous for soil nitrate content. Some indi-

viduals of a particular genotype, growing in patches of high-nitrate soil, will produce both foliage high in alkaloid content and large numbers of seeds. Other individuals, growing in patches of low-nitrate soil, will produce foliage with low alkaloid content and few seeds. There will thus be an environmentally induced covariance between alkaloid content and fitness for that genotype, as well as for others. However, if alkaloid content does not effect fitness causally, there may be no correlation between a genotype's mean alkaloid content and its mean fitness, despite the positive environmental covariance.

In this situation, there is genetic variation for alkaloid production and there is a non-zero selection differential for alkaloid concentration, represented by the phenotypic covariance between fitness and alkaloid concentration. Consequently, although the standard equation for response to selection,

$$R = h^2s \quad (1),$$

predicts that mean alkaloid concentration should evolve (here  $R$  is the change in the mean of a character,  $h^2$  is its heritability, and  $s$  is the selection differential), such evolution will not occur because genotypes do not differ in fitness, i.e., there is no additive genetic covariance between fitness and alkaloid concentration. The environmental covariance between alkaloid concentration and fitness causes the phenotypic covariance, which provides a distorted picture of the actual selection pressures acting on genetic variation for alkaloid concentration (see Price et al., 1988 for a similar example).

This difficulty also arises when attempting to analyze selection on more than one character using the standard Lande and Arnold (1983) approach. As shown below, it occurs because the phenotypic values of one or more measured traits are correlated with the error deviations for fitness in the Lande-Arnold regression model, which can occur if an unmeasured environmental factor affects the measured traits and also affects fitness directly or through an unmeasured, intervening phenotypic character.

Mitchell-Olds and Shaw (1987) suggest that this problem may in some cases be dealt with by including environmental covariates in a Lande-Arnold type analysis, or by using

blocked experimental designs. In many situations, however, the relevant covariates may be unknown, or the spatial scale and pattern of environmental variation may be incompatible with possible blocking designs.

An alternative approach might be the "instrumental variables" technique of econometrics (Fomby et al., 1984). In principle, breeding values are ideal instrumental variables because they are correlated with their corresponding phenotypic values but are by definition uncorrelated with error deviations for fitness as long as there is no genetic correlation between the measured traits and the unmeasured, intervening phenotypic character. However, this technique requires that both phenotypic and breeding values be measured on all individuals in an experiment, which will be difficult, if not impossible, in many investigations. Nevertheless, the instrumental variables approach suggests it may be possible to circumvent biases introduced by environmental covariances by measuring selection directly on additive genetic variation. In the next sections I describe how this may be accomplished.

#### ESTIMATING SELECTION ON GENETIC VARIATION

##### *Directional Selection*

In this section I compare estimates of directional selection resulting from the Lande-Arnold analysis and from regression analysis involving breeding values, show that these analyses yield equivalent estimates only when there are not environmentally-induced covariances between fitness and the characters, and show that such covariances bias estimates from the Lande-Arnold analysis. I assume all genetic variation for traits is additive, since with non-additive variation the validity of predicted responses of quantitative traits to selection under any model is unclear (Kempthorne, 1969 p. 509).

The expected multivariate response to selection is given by Lande (1979) as

$$\Delta \bar{z} = \mathbf{G}\mathbf{P}^{-1}\mathbf{s} = \mathbf{G}\boldsymbol{\beta}, \quad (2)$$

where  $\bar{z}$  is the vector of means of the characters,  $\mathbf{G}$  is the matrix of additive genetic variances and covariance,  $\mathbf{P}$  is the phenotypic variance-covariance matrix,  $\mathbf{s}$  is the

vector of selection differentials, and  $\beta = \mathbf{P}^{-1}\mathbf{s}$  is the vector of selection gradients. Each element  $\beta_i$  of  $\beta$  represents the magnitude of directional selection acting directly on character  $i$ , with the effects of selection on other characters removed. Lande and Arnold (1983) demonstrated that the elements of  $\beta$  can be estimated from the coefficients of a multiple regression of relative fitness on the characters.

To derive an analogous expression for regression of breeding values, I begin with the result, first suggested by Robertson (1966, 1968), that the change in mean phenotypic value of a character from one generation to the next is equal to the additive genetic covariance between the character and fitness (for a formal derivation, see Price, 1970, 1972; Crow and Nagylaki, 1976; Uyenoyama et al., 1981; Nagalyki, 1989; see also Appendix 1). The multivariate generalization of this relationship is

$$\Delta\bar{z} = \mathcal{C} \equiv \text{cov}_a[w, \mathbf{z}], \quad (3)$$

where  $w$  is relative fitness (absolute fitness divided by mean fitness), and the subscript  $a$  indicates that the covariances are between the genotypic value of fitness and the breeding values of the characters.

Each element  $c_i$  of the vector  $\mathcal{C}$  represents the net directional selection on additive genetic variation for character  $i$  and hence is analogous to  $s$  in Equation (1). As Lande and Arnold (1983) did for  $s$ , it will often be useful to partition net selection on a character into portions due to selection acting directly on that character and to selection acting indirectly through genetically correlated characters. This may be achieved by multiplying the RHS of Equation (3) by the identity matrix in the form  $\mathbf{G}\mathbf{G}^{-1}$ :

$$\Delta\bar{z} = \mathbf{G}\mathbf{G}^{-1}\mathcal{C} = \mathbf{G}\mathbf{B}. \quad (4)$$

In this equation,  $\mathbf{B} \equiv \mathbf{G}^{-1}\mathcal{C}$  is a vector of directional selection pressures acting directly on each character, analogous to the selection gradients,  $\beta$ , in Lande and Arnold's analysis. The elements of  $\mathbf{B}$  may be interpreted as partial regression coefficients of a multiple regression of the breeding value of fitness on the breeding values of the characters (Kendall and Stuart, 1973 eq. 27.42), just as the elements of  $\beta$  are partial regression coefficients in a regression in-

volving phenotypic values (Lande and Arnold, 1983). Moreover, the net selection on trait  $i$ ,  $c_i$ , can be partitioned into direct and indirect components:

$$c_i = \sum_{j=1}^n b_j g_{ij}$$

where  $g_{ij}$  is the  $ij^{\text{th}}$  element of  $\mathbf{G}$  and  $b_j$  is the  $j^{\text{th}}$  element of  $\mathbf{B}$ . The term  $b_i g_{ii}$  represents the component of selection acting directly on additive variation for character  $i$ , while the remaining terms represent components of selection acting indirectly through genetically correlated characters (see Lande and Arnold, 1983 for a similar partitioning of  $s$ ).

Since Equations (2) and (4) both describe the same evolutionary event, it is clear that  $\mathbf{B}$  and  $\beta$  represent the same forces of selection and that therefore it is necessary that  $\mathbf{B} = \beta$ , i.e., that

$$\mathbf{G}\mathbf{P}^{-1} \text{cov}_p[w, \mathbf{z}] = \text{cov}_a[w, \mathbf{z}]. \quad (5)$$

This relationship is valid as long as all relevant characters affecting fitness are measured. In most empirical investigations, however, this will not be the case, and empirical estimates of  $\mathbf{B}$  and  $\beta$  may differ. Path analysis (Wright, 1968; Li, 1975) may be used to determine the conditions under which such differences are expected to arise. Consider the case of two genetically correlated traits portrayed in the path diagram in Figure 1. This diagram represents the possible causal influences linking the breeding values ( $x_i$ ), phenotypic values ( $z_i$ ), and environmental deviations ( $E_i$ ) of the two traits with relative fitness ( $w$ ). The phenotypic value of trait  $i$  is assumed to be determined directly by the traits' breeding values (path coefficient  $h_i$ ) and by its environmental deviation (path coefficient  $e_i$ ). Two environmental factors,  $F_1$  and  $F_2$ , are assumed to determine directly the environmental deviations of traits 1 and 2, respectively (without loss of generality, it may be assumed that the path coefficient from  $F_i$  to  $E_i$  is 1). Fitness is assumed to be influenced directly by the phenotypic values of the traits (path coefficients  $a_1$  and  $a_2$ ). Fitness may also be influenced by the environmental factors  $F_i$  (coefficients  $a'_1$  and  $a'_2$ ; note that these pathways may or may not involve intervening, unknown charac-

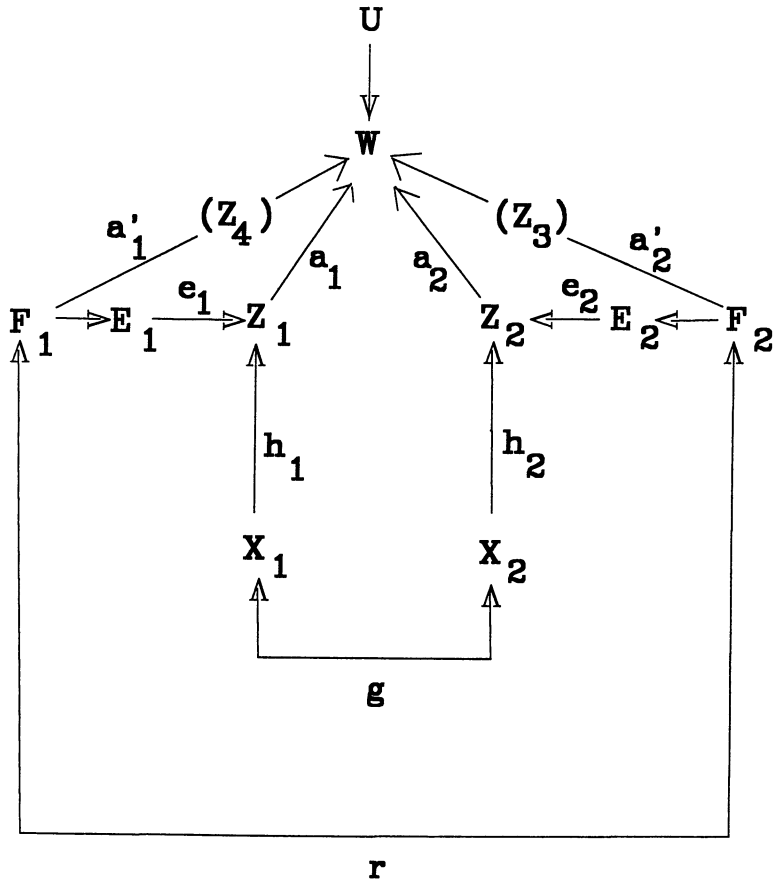


FIG. 1. Path diagram showing possible networks of causal pathways affecting fitness.  $x_i$ ,  $E_i$ , and  $z_i$  are, respectively, the breeding value, environmental deviation, and phenotypic value of trait  $i$ .  $w$  is relative fitness.  $F_i$  and  $U$  are environmental variables. The values associated with a causal arrow (e.g.,  $h$ ,  $a$ ,  $e$ ) are the path coefficients, which represent the covariance (and correlation) of the standardized variables connected by the arrow. Note that  $F_i$  determines  $E_i$  completely, so that the path coefficient between  $F_i$  and  $E_i$  is 1.  $z_3$  and  $z_4$  represent possible unmeasured, intervening traits (see text).

ters). Finally, the genetic and phenotypic correlations between the characters are assumed to be  $g$  and  $r$ , respectively.

In Appendix 2 I apply standard path-analytic techniques to Figure 1 to demonstrate that Equation (5) holds if and only if  $a'_1 = a'_2 = 0$ . In other words,  $\mathbf{B} = \beta$  only when environmental factors that affect the measured characters do not also affect fitness either directly or through unmeasured characters. When  $a'_1 \neq 0$  or  $a'_2 \neq 0$ , then  $\beta$  is a biased estimate of the forces of directional selection acting directly on the characters. This result can be derived for larger numbers of characters, but the algebra becomes too messy for presentation because of the expressions for  $\mathbf{P}^{-1}$ .

The bias introduced into  $\beta$  when some  $a'_i$  are not zero can be understood by first considering the vector of environmental covariances between the measured characters and relative fitness. By definition, the environmental covariances are the covariances between the environmental deviations of the characters,  $E_i$ , and fitness. The environmental covariance of the  $i$ th character and fitness is

$$\text{cov}_e[w, z_i] = \sigma_{E_i} \sigma_{w_p} [e_i a_i + \sum_{j \neq i} r_{ij} e_j a_j] + \sigma_{E_i} \sigma_{w_p} \left[ a'_i + \sum_{j \neq i} r_{ij} a'_j \right], \quad (6)$$

where  $\sigma^2_{E_i}$  and  $\sigma^2_{w_p}$  are the variances of the

environmental deviations of character  $i$  and the phenotypic variance of relative fitness, respectively.

The second term on the RHS of Equation (6) involves only path coefficients  $a'_i$  and thus represents the component of the total environmental covariance caused by effects of the environmental factors  $F_i$  on fitness (possibly acting through an unknown, intervening character). Consequently, I designate this term the environmental component of the environmental covariance. By contrast, the  $a_i$  represent the direct effects of the characters on fitness, i.e., selection on the phenotype. Consequently, Equation (6) says that the process of selection establishes an environmental covariance between the characters and fitness. I therefore designate the first term in Equation (6) the selection component of the environmental covariance.

The bias in  $\beta$  when  $a'_i \neq 0$  is due to the environmental component of  $\text{cov}_e[w, \mathbf{z}]$ . Specifically, it is shown in Appendix 3 that the true response to selection is given by

$$\Delta \bar{\mathbf{z}} = \mathbf{G}\mathbf{P}^{-1}\mathbf{s}' = \mathbf{G}\beta', \quad (7)$$

where the  $i^{\text{th}}$  element of  $\mathbf{s}'$  is given by

$$s'_i = s_i - \sigma_{E_i}\sigma_{w_i} \left( a'_i + \sum_{j \neq i} r_{ij}a'_j \right). \quad (8)$$

In other words,  $\beta$  is biased because the selection differential for each character is an inaccurate estimate of the total magnitude of selection on that character. From Equations (6) and (8), it can be seen that the magnitude of this inaccuracy is equal to the environmental component of the environmental covariance between the character and fitness.

#### *Stabilizing and Disruptive Selection*

I here define stabilizing (disruptive) selection as any decrease (increase) in the genetic variance of a character apart from any change caused by directional selection. This definition is similar to that adopted by Lande and Arnold (1983) for changes in phenotypic variance, though alternative definitions are possible (e.g., see Mitchell-Olds and Shaw, 1987). I limit discussion to changes in additive genetic variances and covariances within a generation caused by

selection, which allows comparison of my results to those of Lande and Arnold for phenotypic selection. I assume again that genetic variation for the characters is entirely additive.

In Appendix 1 it is shown that the within-generation change in  $\mathbf{G}$  due to selection is given by

$$\mathbf{G}^* - \mathbf{G} = \Phi - \mathcal{C}\mathcal{C}^T. \quad (9)$$

Here  $\Phi$  is a matrix whose  $ij^{\text{th}}$  element,  $\Phi_{ij}$  is given by

$$\begin{aligned} \Phi_{ij} &= \sum_{\psi} f(\psi)[w(\psi) - \bar{w}][\tilde{x}_{ij} - \bar{\tilde{x}}_{ij}] \\ &\equiv \text{cov}_q[w, \tilde{x}_{ij}], \end{aligned} \quad (10)$$

where  $\psi$  signifies a multilocus genotype,  $f(\psi)$  is the relative frequency of genotype  $\psi$ ,  $w(\psi)$  is the genotypic value of fitness for genotype  $\psi$ ,  $\tilde{x}_{ij} = [x_i(\psi) - \bar{x}_i][x_j(\psi) - \bar{x}_j]$ ,  $x_i(\psi)$  is the breeding value of trait  $i$  for an individual of genotype  $\psi$ . In other words,  $\Phi_{ij}$  is the covariance between the genotypic value of fitness and the product of deviations of trait breeding values.

The last term on the RHS of (9) represents the change in  $\mathbf{G}$  due to directional selection. The first term, by contrast, represents the net force of selection due to quadratic deviations of breeding values from their means, and thus represents the overall effect of stabilizing, disruptive, and correlational selection on the additive genetic variances and covariances. As with directional selection, it will often be of interest to decompose the net quadratic forces into those acting directly on the genetic variance (or covariance) of a particular trait and those acting indirectly through genetically correlated characters.

This decomposition can be achieved by premultiplying the RHS of (8) by  $I = \mathbf{G}\mathbf{G}^{-1}$  and postmultiplying by  $I = \mathbf{G}^{-1}\mathbf{G}$  to yield

$$\begin{aligned} \mathbf{G}^* - \mathbf{G} &= \mathbf{G}\mathbf{G}^{-1}\Phi\mathbf{G}^{-1}\mathbf{G} - \mathbf{G}\mathbf{G}^{-1}\mathcal{C}\mathcal{C}^T\mathbf{G}^{-1}\mathbf{G} \\ &= \mathbf{G}\mathbf{T}\mathbf{G} - \mathbf{G}\mathbf{B}\mathbf{B}^T\mathbf{G}, \end{aligned} \quad (11)$$

where

$$\mathbf{T} = \mathbf{G}^{-1}\Phi\mathbf{G}^{-1}. \quad (12)$$

The elements of  $\mathbf{T}$  may be interpreted as partial regression coefficients of the genotypic value of fitness on  $\tilde{x}$ , the quadratic deviations of trait breeding values. Because

they are partial regression coefficients, the diagonal elements,  $\Gamma_{ii}$ , represent the force of stabilizing or disruptive selection acting directly on the additive genetic variances of the characters, independent of the effects of stabilizing, correlational, or disruptive selection on genetically correlated characters, as well as of the effects of directional selection. Similarly, the off-diagonal elements,  $\Gamma_{ij}$ , represent correlational selection acting directly on the additive genetic covariances. The matrix  $\Gamma$  thus is analogous to the matrix of phenotypic quadratic selection gradients,  $\gamma$ , described by Lande and Arnold (1983).

Partitioning of  $\Phi$  is now achieved by rearranging (12) to produce  $\Phi = \mathbf{G}\Gamma\mathbf{G}$ , which leads directly to the expression for each element of  $\Phi$ :

$$\phi_{ij} = \sum_k \sum_l g_{ik} g_{jl} \Gamma_{kl}.$$

The term  $g_{ij}^2 \Gamma_{ij}$  represents the portion of the total change in the  $ij^{\text{th}}$  element of  $\mathbf{G}$  due to quadratic selection of magnitude  $\Gamma_{ij}$  on that element, while the remaining terms represent the portion brought about by quadratic selection acting indirectly through genetically correlated traits.

The equation analogous to (11) for the change in  $\mathbf{G}$  when selection is measured on phenotypic variation is obtained by combining Equations (12)–(14) in Lande and Arnold (1983):

$$\mathbf{G}^* - \mathbf{G} = \mathbf{G}\gamma\mathbf{G} - \mathbf{G}\beta\beta^T\mathbf{G}, \quad (13)$$

where

$$\gamma = \mathbf{P}^{-1}[\text{cov}_p[w, \tilde{\mathbf{z}}]]\mathbf{P}^{-1} \quad (14)$$

and  $\tilde{\mathbf{z}} = (\mathbf{z} - \bar{\mathbf{z}})(\mathbf{z} - \bar{\mathbf{z}})^T$ .

Because Equations (11) and (13) are equally valid when all relevant characters are included, both predict the same change in  $\mathbf{G}$  and hence, in the absence of directional selection, imply  $\Gamma = \gamma$ . By analogy with the results for directional selection, however, it might be expected that empirical estimates of  $\Gamma$  and  $\gamma$  would be equivalent only if unmeasured environmental factors do not simultaneously influence quadratic deviations of the measured traits and fitness, and that estimates of  $\Gamma$  are often likely to be less-biased. Formal proof of this conjecture has not yet been obtained, however.

### Statistical Procedures

The potential for estimates of evolutionary forces derived from examination of selection on genetic variation ( $\mathbf{B}$  and  $\Gamma$ ) to differ from those obtained from examination of selection on phenotypic variation ( $\beta$  and  $\gamma$ ) suggests that statistical procedures are needed for addressing two issues: (1) estimation of  $\mathbf{B}$  and  $\Gamma$ , and (2) testing for differences between  $\mathbf{B}$  and  $\beta$  and between  $\Gamma$  and  $\gamma$ . In this section I briefly describe procedures that may be used to accomplish these tasks.

Methods for estimating  $\mathbf{B}$  and for comparing  $\mathbf{B}$  and  $\beta$  can pose some difficult statistical problems. For example, derivation of Equation (4) assumes that  $\mathbf{G}$  is invertible, i.e., that all measured characters are genetically variable. Consequently,  $\mathbf{B}$  should not be estimated from family components of covariance unless it is shown that all traits are heritable and  $\mathbf{G}$  is positive definite (Hill and Thompson, 1978). Nevertheless, when these conditions are met, the following procedures may be used.

*Estimation of  $\mathbf{B}$  and  $\Gamma$ .*—The interpretation of  $\mathbf{B}$  and  $\Gamma$  as coefficients of regressions involving genotypic and/or breeding values suggests one method for obtaining approximate estimates of these quantities is the use of family-mean multiple regressions of relative fitness on the traits. As shown in Appendix 4, the family-mean partial regression coefficients approach the appropriate genetic regression coefficients as the number of individuals per family becomes large (see also Arnold, 1981; Via, 1984). The actual statistical procedure is identical with that described by Lande and Arnold (1983) for phenotypic selection analysis, except that family means are used as observations in place of measurements on individuals. [Note that for quadratic terms, Equation (10) requires an estimate of a function of breeding values. Since in the absence of non-additive genetic variance the expected value of a family mean is proportional to the breeding value of the family's parent, regression on family means provides appropriate estimates of  $\Gamma$ , apart from proportionality constants determined by the type of families used. As with the standard phenotypic regression analysis, however, the

linear coefficients in a quadratic regression will be biased estimates of the directional selection gradients.] Cautions regarding effects of multicollinearity and limitations on hypothesis testing described for phenotypic selection by Mitchell-Olds and Shaw (1987) also apply to regressions involving family means. In addition, caution must be taken in determining the appropriate degrees of freedom for significance testing because, for example, full-sib families from diallel and factorial crosses are not statistically independent.

As long as genetic variance for the traits is entirely additive, clonal, full-sib, or half-sib families may be used to estimate  $\mathbf{B}$  and  $\mathbf{\Gamma}$ . Because non-additive deviations in one trait are uncorrelated with additive genetic values of a second trait (Kempthorne, 1969 p. 418), any non-additive genetic component of fitness variation will not contribute to the estimated regression coefficients, regardless of which type of family is used. Paternal half-sib families may be desired, however, if maternal effects are substantial. It is expected that if there is a small amount of non-additive genetic variation for the traits, the family mean regression approach will yield approximate estimates of the selection gradients, though the magnitudes of the biases introduced need exploration.

The forces of directional selection,  $\mathbf{B}$ , may also be estimated using family components of variance and covariance from a standard multivariate analysis of variance (MANOVA). In particular, estimates of each component of  $\mathcal{C}$  may be obtained from the family covariance component in a MANOVA involving fitness and a particular trait, while  $\mathbf{G}$  may be estimated from the variance-covariance matrix of family components in a MANOVA involving all the traits. The product  $\mathbf{G}^{-1} \mathcal{C}$  can then be formed to estimate  $\mathbf{B}$ . Estimates of the standard errors of elements of  $\mathbf{B}$  can be obtained by applying a jackknife or bootstrap technique to this procedure (Efron, 1982; but see caution below). Unfortunately, it is unclear how this technique can be adopted for estimating  $\mathbf{\Gamma}$ .

*Comparing  $\mathbf{B}$  and  $\mathbf{\Gamma}$  with  $\beta$  and  $\gamma$ .*—Estimates of selection obtained from phenotypic and family-mean regression analyses can be compared using the combined regression model

$$\mathbf{w} = \bar{\mathbf{x}}^T C_1 + (\bar{\mathbf{x}} - \mathbf{y})^T C_2 + \text{error},$$

where  $\bar{\mathbf{x}}$  is the vector of estimated breeding values of the traits (or their squares and products when comparing stabilizing/disruptive/correlational forces of selection),  $\mathbf{y}$  is the vector of phenotypic values of the traits (for linear terms) or of their squares and products (for quadratic terms), and  $C_1$  and  $C_2$  are regression coefficients. This formulation is a modification of the instrumental variables approach (e.g., Fomby et al., 1984) in which the family-mean value of each trait is used as the value of  $\bar{\mathbf{x}}$  for each individual, rather than that individual's own (unknown) breeding value. As appropriate, the breeding values are estimated by the family means for families of genetically identical individuals or by twice the deviation of family means from the overall population mean for paternal half-sib or full-sib families (Falconer, 1981). The hypothesis  $C_1 = C_2$ , which may be tested using standard contrasts for linear models (Searle, 1971 pp. 199–204), is equivalent to the hypothesis that  $\mathbf{B} = \beta$  (or  $\mathbf{\Gamma} = \gamma$ ) (Burdick, pers. comm.; see Rausher and Simms, 1989 for an example of this technique).

When  $\mathbf{B}$  is estimated from family components of variance and covariance, it may be compared to  $\beta$  using a bootstrap or jackknife procedure. For example, for the bootstrap procedure, each family may be considered an observation and bootstrap samples can then be generated in standard fashion by sampling families with replacement. For each sample,  $\mathbf{P}$ ,  $\mathbf{G}$  and  $\mathcal{C}$ , and hence  $\mathbf{B} - \beta$ , may be calculated, and the probability that  $\mathbf{B} - \beta = 0$  can then be ascertained from the bootstrap sampling distribution. These resampling techniques must be applied with caution, however, because the appropriate level for resampling is not known for some nested or cross-classified family structures. These techniques provide *approximate* tests of significance when sample sizes are *large*.

#### DISCUSSION

The Lande-Arnold (1983) regression technique has become widely used for estimating selection gradients in nature (e.g., Kalisz, 1986; Berenbaum et al., 1986; Rausher and Simms, 1989; Simms, 1990;

Jayne and Bennett, 1990; Weis and Gorman, 1990; Bates Smith, 1990). However, a potential limitation associated with this technique is that failure to include relevant characters in the analysis can lead to inaccurate estimates of the selection gradients corresponding to the analyzed characters (Lande and Arnold, 1983). The genetic analysis presented here also suffers from this limitation, but under a more restricted set of circumstances. In particular, the Lande-Arnold analysis produces inaccurate estimates of selection gradients whenever any trait that covaries phenotypically with fitness is not included in the analysis, regardless of whether that trait is genetically variable. By contrast, omission from the genetic analysis of a genetically non-variable trait that covaries with fitness does not yield inaccurate estimates of selection gradients. Moreover, even genetically variable traits need not be measured if they are genetically uncorrelated with the measured traits (see Appendix 5). It is only when unmeasured traits are genetically correlated with measured traits that the genetic analysis is biased.

The hypothetical example presented in the introduction is useful in clarifying this distinction. In that example, a Lande-Arnold type analysis would indicate that alkaloid content is subject to selection because variation in an environmental factor (soil nitrate content) causes fitness and alkaloid content to covary. In such a case, it is likely that soil nitrate content does not directly affect fitness, but rather affects some intermediate variable such as plant soluble nitrogen content, which in turn determines how many seeds a plant can produce (fitness). Consider a situation in which plant soluble nitrogen content is not genetically variable. In this situation, variation in soil nitrate content would cause phenotypic (but not genetic) variation in a trait (soluble nitrogen content) affecting fitness. Failure to include this variable in the Lande-Arnold phenotypic selection analysis would then lead to inaccurate estimates of selection on alkaloid content, because the analysis would not account for the non-genetic correlation between alkaloid content and soluble nitrogen content. By contrast, failure to include plant soluble nitrogen content in the genetic

analysis of selection developed here will not lead to inaccurate estimates of selection on alkaloid content because environmentally-induced correlations play no part in the analysis. In particular, because soluble nitrogen content is not genetically variable, genetic correlations of this trait with alkaloid content do not exist and hence need not be controlled for when estimating selection on alkaloid content.

Any regression analysis of fitness should be viewed as yielding preliminary conclusions that should subsequently be confirmed experimentally (e.g., Wade and Kalisz, 1990). To this end, a combination of phenotypic and genetic analyses of selection may often be useful in helping to determine whether bias is present in estimates of selection gradients. Statistical differences in estimates obtained by the two methods would suggest either that an important character has been omitted from the analysis or that biases are being introduced into the phenotypic analysis by environmental correlations involving fitness. In either case, the discrepancy would indicate that additional information is required for estimating selection gradients accurately and for suggesting appropriate experimental manipulations.

#### ACKNOWLEDGMENTS

I thank J. Fry and E. Simms for valuable discussions during the preparation of this manuscript, and R. Brandon, M. Kirkpatrick, B. Walsh, and two anonymous reviewers for valuable comments on intermediate drafts. Supported by NSF Grant BSR 8817899.

#### LITERATURE CITED

- 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* 38:709-719.
- . 1984b. On the measurement of natural and sexual selection: Applications. *Evolution* 38:720-734.
- BATE SMITH, T. 1990. Natural selection on bill characters in the two bill morphs of the African finch *Pyrenestes ostrinus*. *Evolution* 44:832-842.
- BERENBAUM, M. R., A. R. ZANGERL, AND J. K. NITAO. 1986. Constraints on chemical coevolution: Wild

- parsnip and the parsnip webworm. *Evolution* 40: 1215–1228.
- CRESPI, B. J., AND F. L. BOOKSTEIN. 1989. A path-analytic model for the measurement of selection on morphology. *Evolution* 43:18–28.
- CROW, J. F., AND T. NAGYLAKI. 1976. The rate of change of a character correlated with fitness. *Am. Nat.* 110:207–213.
- EFRON, B. 1982. The Jackknife, the Bootstrap, and Other Resampling Plans. S.I.A.M., Philadelphia, PA USA.
- FALCONER, D. S. 1981. *Introduction to Quantitative Genetics*, 2nd ed. Longman, N.Y., USA.
- FOMBY, T. B., R. C. HILL, AND S. R. JOHNSON. 1984. *Advanced Econometric Methods*. Springer-Verlag, N.Y., USA.
- HILL, W. G., AND R. THOMPSON. 1978. Probabilities of non-positive definite between-group or genetic covariance matrices. *Biometrics* 34:429–439.
- JAYNE, B. C., AND A. F. BENNETT. 1990. Selection on locomotor performance capacity in a natural population of garter snakes. *Evolution* 44:1204–1209.
- KALISZ, S. 1986. Variable selection on the timing of germination in *Collinsia verna* (Scrophulariaceae). *Evolution* 40:479–491.
- KEMPTHORNE, O. 1969. *An Introduction to Genetic Statistics*. Iowa State Univ. Press, Ames, USA.
- KENDALL, M. G., AND A. STUART. 1973. *The Advanced Theory of Statistics*. Vol. 2. Inference and Relationship, 3rd ed. MacMillan, N.Y., USA.
- LANDE, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30: 314–334.
- . 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: Body size allometry. *Evolution* 33:402–416.
- . 1980. The genetic covariance between characters maintained by pleiotropic mutations. *Genetics* 94:203–215.
- LANDE, R., AND S. J. ARNOLD. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- LI, C. C. 1975. *Path Analysis—A Primer*. Boxwood Press, Pacific Grove, CA USA.
- MITCHELL-OLDS, T., AND R. G. SHAW. 1987. Regression analysis of natural selection: Statistical inference and biological interpretation. *Evolution* 41: 1149–1161.
- NAGYLAKI, T. 1989. Rate of evolution of a character without epistasis. *Proc. Nat. Acad. Sci.* 86:1910–1913.
- PRICE, G. R. 1970. Selection and covariance. *Nature* 227:520–521.
- . 1972. Extension of covariance selection mathematics. *Ann. Hum. Genet.* 35:485–490.
- PRICE, T., M. KIRKPATRICK, AND S. J. ARNOLD. 1988. Directional selection and the evolution of breeding date in birds. *Science* 240:798–799.
- RAUSHER, M. D., AND E. L. SIMMS. 1989. The evolution of resistance to herbivory in *Ipomoea purpurea*. I. Attempts to detect selection. *Evolution* 43: 563–572.
- ROBERTSON, A. 1966. A mathematical model of the culling process in dairy cattle. *Anim. Prod.* 8:95–108.
- . 1968. The spectrum of genetic variation, pp. 5–16. *In* R. C. Lewontin (ed.), *Population Biology and Evolution*. Syracuse University Press, Syracuse, NY USA.
- SEARLE, S. R. 1971. *Linear Models*. Wiley, N.Y., USA.
- SIMMS, E. L. 1990. Examining selection on the multivariate phenotype: Plant resistance to herbivores. *Evolution* 44:1177–1188.
- TURELLI, M. 1984. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theor. Popul. Biol.* 25:138–193.
- . 1985. Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. *Genetics* 111:165–195.
- . 1988. Population genetic models for polygenic variation and evolution, pp. 601–608. *In* E. J. Eisen, M. M. Goodman, G. Namkoong, and B. S. Weir (eds.), *The Second International Conference on Quantitative Genetics*. Sinauer, Sunderland, MA USA.
- UYENOYAMA, M. K., M. K. FELDMAN, AND L. D. MUELLER. 1981. Population genetic theory of kin selection: Multiple alleles at one locus. *Proc. Nat. Acad. Sci.* 78:5036–5040.
- VIA, S. 1984. The quantitative genetics of polyphagy in an insect herbivore. II. Genetic correlations in larval performance within and among host plants. *Evolution* 38:896–905.
- WADE, M. J., AND S. KALISZ. 1989. The additive partitioning of selection gradients. *Evolution* 43: 1567–1569.
- . 1990. The causes of natural selection. *Evolution* 44:1947–1955.
- WEIS, A. E., AND W. L. GORMAN. 1990. Measuring selection on reaction norms: An exploration of the *Eurosta-Solidago* system. *Evolution* 44:820–831.
- WRIGHT, S. 1968. *Evolution and the Genetics of Populations*. Vol. I. Genetic and Biometrical Foundations. Univ. Chicago Press, Chicago, IL USA.
- ZENG, Z-B. 1988. Long-term correlated response, interpopulation covariation, and interspecific allometry. *Evolution* 42:363–374.

Corresponding Editor: J. Arnold

#### APPENDIX 1

Here I derive text Equations (3) and (9). I assume all genetic variation for characters is additive.

Let  $\psi$  designate the multilocus genotype of an individual and let  $\mathbf{x}(\psi) = [x_1(\psi), \dots, x_r(\psi)]$  be the vector of additive genetic values of traits  $z = [z_1, \dots, z_n]$  for individuals of genotype  $\psi$ . Let the frequency distribution of  $\psi$  be designated by  $f(\psi)$ . Finally, let the mean relative fitness of an individual of genotype  $\psi$  be  $w(\psi)$ , where  $\bar{w} = \overline{w(\psi)} = 1$ .

From Lande (1979; Eq. 5b), the change in mean value of  $\mathbf{x}$  caused by directional selection is

$$\begin{aligned} \Delta \bar{\mathbf{x}} &= \Delta \bar{\mathbf{x}} = \left[ \sum_{\psi} f(\psi) \mathbf{x}(\psi) w(\psi) \right] - \bar{\mathbf{x}} \\ &= \bar{\mathbf{x}}^* - \bar{\mathbf{x}}, \end{aligned} \quad (\text{A.1.1})$$

where  $\bar{\mathbf{x}}^*$  is the mean vector of breeding values after selection (but before mutation, segregation, and recombination).  $\mathcal{C} = \text{cov}_g[w, z]$ , the vector of additive ge-

netic covariances between fitness and the characters, is given by

$$\mathcal{C} = \text{cov}[w(\psi), x(\psi)] = \sum_{\psi} f(\psi)[w(\psi) - \bar{w}][x - \bar{x}].$$

Expanding the RHS and noting that  $\bar{w} = 1$  yields

$$\mathcal{C} = \left[ \sum_{\psi} f(\psi)x(\psi)w(\psi) \right] - \bar{x}, \tag{A.1.2}$$

which, with (A.1.1) yields Equation (3) in the text.

Now consider the change in the additive genetic variance-covariance matrix,  $\mathbf{G}$ , due to selection within a generation. Let  $g_{ij}$  be the  $ij^{\text{th}}$  element of  $\mathbf{G}$ . Then the change in  $g_{ij}$  due to selection is

$$g^*_{ij} - g_{ij} = \left[ \sum_{\psi} f(\psi)w(\psi)[x_i(\psi) - \bar{x}_i^*][x_j(\psi) - \bar{x}_j^*] \right] - \left[ \sum_{\psi} f(\psi)[x_i(\psi) - \bar{x}_i][x_j(\psi) - \bar{x}_j] \right]. \tag{A.1.3}$$

Substituting the elements  $c_i$  and  $c_j$  from (A.1.2) into (A.1.3) and simplifying yields

$$g^*_{ij} - g_{ij} = \left[ \sum_{\psi} f(\psi)w(\psi)[x_i(\psi) - \bar{x}_i][x_j(\psi) - \bar{x}_j] \right] - \bar{x}_i \bar{x}_j - c_i c_j, \tag{A.1.4}$$

where  $\bar{x}_{ij} = [x_i(\psi) - \bar{x}_i][x_j(\psi) - \bar{x}_j]$  and  $\bar{x}_{ij}$  is the mean of  $\bar{x}_{ij}$ .

Now define  $\text{cov}_q[w, \bar{x}_{ij}]$  as the covariance between the genotypic value of fitness and  $\bar{x}_{ij}$ , i.e.,

$$\text{cov}_q[w, \bar{x}_{ij}] = \sum_{\psi} f(\psi)[w(\psi) - \bar{w}][\bar{x}_{ij} - \bar{\bar{x}}_{ij}].$$

Expanding and rearranging gives

$$\text{cov}_q[w, \bar{x}_{ij}] = \left[ \sum_{\psi} f(\psi)w(\psi)[x_i(\psi) - \bar{x}_i][x_j(\psi) - \bar{x}_j] \right] - \bar{\bar{x}}_{ij}. \tag{A.1.5}$$

Comparing (A.1.4) and (A.1.5) reveals that

$$g^*_{ij} - g_{ij} = \text{cov}_q[w, \bar{x}_{ij}] - c_i c_j,$$

which corresponds to Equation (9).

APPENDIX 2

In this analysis, standard path-analytic techniques are applied to Figure 1 to demonstrate that Equation (5) holds only if  $a'_1 = a'_2 = 0$ . From standard path analysis

$$\text{cov}_a[w, z] = \begin{bmatrix} \sigma_{g_i} \sigma_{w_i} (h_1 a_1 + g h_2 a_2) \\ \sigma_{g_i} \sigma_{w_i} (h_2 a_2 + g h_1 a_1) \end{bmatrix} \tag{A.2.1}$$

where  $\sigma^2_{g_i}$  is the additive genetic variance of trait  $i$ ,  $\sigma^2_{w_i}$  is the phenotypic variance in fitness,  $h_i$  and  $a_i$  are path coefficients (see Fig. 1), and  $g$  is the genetic correlation between traits 1 and 2.

Also from standard path analysis,

$$s = \text{cov}_p[w, z]$$

$$= \begin{bmatrix} \sigma_{p_1} \sigma_{w_1} [a_1 + a_2 (h_1 h_2 g + e_1 e_2 r) + e_1 a'_1 + e_1 a'_2 r] \\ \sigma_{p_2} \sigma_{w_1} [a_2 + a_1 (h_1 h_2 g + e_1 e_2 r) + e_2 a'_2 + e_2 a'_1 r] \end{bmatrix}. \tag{A.2.2}$$

Finally, let

$$\mathbf{G} = \begin{bmatrix} \sigma^2_{g_1} & g_{12} \\ g_{12} & \sigma^2_{g_2} \end{bmatrix} \quad \text{and} \quad \mathbf{P} = \begin{bmatrix} \sigma^2_{p_1} & p_{12} \\ p_{12} & \sigma^2_{p_2} \end{bmatrix}.$$

Then, after simplification,

$$\mathbf{G}\mathbf{P}^{-1}\text{cov}_p[w, z] = \begin{bmatrix} \sigma_{g_i} \sigma_{w_i} (h_1 a_1 + g h_2 a_2) \\ \sigma_{g_i} \sigma_{w_i} (h_2 a_2 + g h_1 a_1) \end{bmatrix} + \begin{bmatrix} \sigma_{p_1} \sigma_{w_1} (e_1 a'_1 + r e_1 a'_2) \\ \sigma_{p_2} \sigma_{w_1} (e_2 a'_2 + r e_2 a'_1) \end{bmatrix}. \tag{A.2.3}$$

But, comparing (A.2.1) and (A.2.3) reveals that Equation (5) will be true in general only if  $a'_1 = a'_2 = 0$ .

APPENDIX 3

In this appendix I show that the predicted response to selection is given by

$$\Delta \bar{z} = \mathbf{G}\mathbf{P}^{-1}\mathbf{s}' = \mathbf{G}\mathbf{b}',$$

where  $\mathbf{s}'$  is a vector, the  $i^{\text{th}}$  element of which is given by Equation (8) in the text.

Consider the two characters portrayed in Figure 1. From Equations (5) and (A.2.1),

$$\text{cov}_a[w, z] = \begin{bmatrix} \sigma_{p_1} \sigma_{w_1} [a_1 + a_2 (h_1 h_2 g + e_1 e_2 r)] \\ \sigma_{p_2} \sigma_{w_1} [a_2 + a_1 (h_1 h_2 g + e_1 e_2 r)] \end{bmatrix}. \tag{A.3.1}$$

Let  $\mathbf{s}'$  represent the expression in brackets. Then (A.2.2) and (A.3.1) together imply

$$\Delta \bar{z} = \mathbf{G}\mathbf{P}^{-1}(\mathbf{s} - \phi) = \mathbf{G}\mathbf{P}^{-1}\mathbf{s}' = \mathbf{G}\mathbf{b}',$$

where

$$\phi = \begin{bmatrix} \sigma_{E_1} \sigma_{w_1} (a'_1 + a'_2 r) \\ \sigma_{E_2} \sigma_{w_1} (a'_2 + a'_1 r) \end{bmatrix}$$

is the vector of environmental components of the environmental covariances between the characters and fitness, as derived from standard path analysis. This argument is generalized in a straightforward manner to more than two traits.

APPENDIX 4

In this appendix I show that regressions of family means may be used to estimate  $\mathbf{B}$ . An analogous argument can be given for estimation of  $\mathbf{\Gamma}$ . The vector of coefficients,  $\mathbf{\beta}_F$ , of a multiple regression of the family means of fitness on family means for the traits is given by

$$\mathbf{\beta}_F = \mathbf{F}^{-1}\mathbf{S},$$

where  $\mathbf{F}$  is the variance-covariance matrix for family means of the traits and  $\mathbf{S}$  is the vector of covariances of the family means of fitness with the family means of the traits (Kendall and Stuart, 1973). But

$$\mathbf{F} = \delta \mathbf{G} + \frac{1}{n} \mathbf{W}_1$$

and

$$\mathbf{S} = \delta \mathcal{C} + \frac{1}{n} \mathbf{W}_2,$$

where  $\delta = \frac{1}{4}$  for half sib families,  $\delta = \frac{1}{2}$  for full-sib families, and  $\delta = 1$  for families of genetically identical individuals,  $\mathbf{W}_1$  is the matrix of within-family components of variance or covariance for the traits,  $\mathbf{W}_2$  is the vector of within-family components of covariance between fitness and the traits, and  $n$  is the average number of individuals in a family (Arnold, 1981; Via, 1984). Consequently,

$$\beta_F = [\mathbf{G} + \frac{1}{\delta n} \mathbf{W}_1]^{-1} [\mathcal{C} + \frac{1}{\delta n} \mathbf{W}_2].$$

As  $n$  gets large, the right-hand term in each bracket becomes small and

$$\beta_F \approx \mathbf{G}^{-1} \mathcal{C} = \mathbf{B}.$$

#### APPENDIX 5

Let subscripts 1 and 2 refer to the measured and unmeasured (intervening) traits, respectively. Then, for

both sets of traits together, under the assumption that intervening traits are genetically uncorrelated with measured traits,

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 & 0 \\ 0 & \mathbf{G}_2 \end{bmatrix}.$$

Now, the selection gradient calculated assuming that intervening traits are known is

$$\begin{aligned} \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} &= \mathbf{B} = \mathbf{G}^{-1} \mathcal{C} = \begin{bmatrix} \mathbf{G}_1^{-1} & 0 \\ 0 & \mathbf{G}_2^{-1} \end{bmatrix} \begin{bmatrix} \mathcal{C}_1 \\ \mathcal{C}_2 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{G}_1^{-1} \mathcal{C}_1 \\ \mathbf{G}_2^{-1} \mathcal{C}_2 \end{bmatrix}. \end{aligned}$$

In other words, the selection gradient for the measured traits,  $\mathbf{B}_1$ , involves only the genetic variance-covariance matrix for those traits and  $\mathcal{C}_1$ , the vector of additive covariances among the measured traits and fitness. In particular, values of unknown, intervening traits are not used to estimate the selection gradient on the measure traits.